Identification and functional characterization of adhesins involved in attachment of methanogens to rumen protozoa

A thesis presented in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Microbiology
at Massey University, Palmerston North, New Zealand

Filomena Ng

2016

Abstract

Symbiotic interactions are frequently observed amongst members of the complex microbial community inhabiting the fermentative forestomach (rumen) of ruminant animals. In this ecosystem, hydrogen (H_2)-using methanogens can be found as ecto- and endo-symbionts of H_2-producing protozoa, and this interaction contributes to ruminant methane emissions. Rumen symbionts must have the ability to attach to protozoal hosts, presumably *via* protozoa-binding cell surface proteins, however the identity and specificity of these proteins are not known.

A protein of the methanogenic archaeon *Methanobrevibacter ruminantium* M1 that binds to rumen protozoa was identified using phage display technology. A large shot-gun phage display library was constructed from M1 DNA, and affinity screened by biopanning using rumen protozoa as bait. After two rounds of biopanning, a recombinant clone encoding part of a previously annotated putative adhesin, Mru_1499, was identified as a protozoa-binding protein. The protozoal binding region of the affinity selected protein was mapped, and a "reverse panning" procedure was developed to identify protozoal species that bind to the affinity selected protein.

Next, the protozoa-associated methanogen and bacterial communities were characterized, and several taxa of archaea and bacteria were found to be over-represented in the protozoa-associated community relative to their abundance in the rumen contents. Adhesins from this protozoa-associated community were identified by affinity screening of a community-scale phage display library using rumen protozoa as bait, combined with high-throughput single molecule amplicon sequencing. The comparison between pre- and post-panning sequence datasets showed seven highly enriched candidate adhesin-encoding ORFs after affinity-panning of the library on protozoa as bait.

In conclusion, several adhesins mediating interactions between methanogenic archaea, bacteria and protozoa were identified using phage display at both single-organism and metagenome scales. Further assays are required to verify the function of these candidate adhesins as "molecular bridges" in interactions involving rumen protozoa. This is the first report for characterization of the protozoa-associated symbiont community by next generation sequencing of the 16S rRNA gene.

Acknowledgments

I am grateful to my primary supervisor, Dr. Dragana Gagic, for introducing me to the world of rumen microbial symbioses, the autonomy that she trusted me to work with, and her kindness and support during the time we have known each other. Her enthusiasm for research and openness to exploring new ideas have been motivating influences for me in the last four years.

I feel very fortunate for the supportive community that I was surrounded by during the course of my studies. Many thanks to my co-supervisors, Drs. Jasna Rakonjac, Graeme Attwood, and Mark Patchett, who have contributed valuable time and unique perspectives to the project. I would also like to thank Dr. Sandra Kittelmann for her insight into rumen protozoa and her guidance with community analyses, Drs. Peter Janssen and Gemma Henderson for helpful discussions about the rumen microbial community, Michelle Kirk and Sarah Lewis for technical assistance with sample processing for pyrosequencing, Greg Skelton for assistance with rumen sampling, Dr. Ruy Jauregui for writing scripts to process datasets acquired from PacBio sequencing, and Dr. Eric Altermann for running sequences through GAMOLA. It has been a privilege to work with the technicians and scientists in the Rumen Micro lab. It truly feels like a team.

To fellow inhabitants of the FeedTech student office, it has been a pleasure to share an office with thirteen lovely people. I will miss the lively lunchtime discussions.

Thank you to my family, and friends who have become family, for their love and support.

Table of contents

List of figures

List of tables

Abbreviations

ALP	Adhesin-like protein
BSA	Bovine serum albumin
cfu	Colony forming units
DGGE	Denaturing gradient gel electrophoresis
EDS	Energy dispersive X-ray spectroscopy
FISH	Fluorescence *in situ* hybridization
GAPDH	Glycerladehyde-3-phosphate dehydrogenase
GlcNac	*N*-acetylglucosamine
Ig-like	Immunoglobulin-like
IQR	Interquartile range
m.o.i.	Multiplicity of infection
nr	Non-redundant
OD	Optical density
ORF	Open reading frame
OTU	Operational taxonomic unit
PAM1	Shot-gun phage display primary library created from metagenomic DNA
PAM2	Shot-gun phage display primary library after affinity selection against c-myc
PAM_FINAL	PAM2 library after two rounds of biopanning against protozoal bait, followed by affinity selection against c-myc
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
pfu	Plaque forming units
PMB	Pseudomurein binding
pMru_1499^{A}	Phagemid vector encoding Mru_1499^{A}
PP	Phagemid particle
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
TBS-T	Tris buffered saline containing Tween 20

Chapter 1. Introduction

1.1 Symbiosis

Most of the Earth has been colonized with diverse organisms that rarely live in isolation. In population-dense environments, interspecies interactions occur frequently, as organisms compete or collaborate in their individual pursuits for resources. Symbiosis is generally defined as the close relationships between organisms sharing a habitat (Leung & Poulin, 2008). The outcomes of these interactions range from co-operative (mutualistic) to antagonistic (parasitic). The focus of this thesis is on mutualistic interspecies interactions.

1.1.1 Ecological basis for symbiosis

Mutualistic symbioses are based on the exchange of functions or nutrients between partners (Kouzuma *et al.*, 2015). Although they are often considered to be stable long term relationships, the nature of these relationships is also dependent on the external environment. Mechanisms for symbiosis maintenance are also necessary to ensure that symbiont abundance remains within the carrying capacity of the host, and to continuously select for symbionts that contribute to host fitness (Dunn, 2012). These features will be described in greater detail using the symbiosis between *Vibrio fischeri* and *Euprymna scolopes* as an example.

1.1.2 Simple systems (host/single symbiont)

Interactions between bacteria and their eukaryotic host have been extensively studied in the model system of symbiosis between *Vibrio fischeri*/*Euprymna scolopes* (Hawaiian bobtail squid) (Dunn, 2012). The squid host feeds from the water column at night; therefore, it is vulnerable to nocturnal predators that can observe its silhouette from below, against moonlight and starlight. *V. fischeri* is a facultative anaerobic bacterium that protects the squid host from predators by emitting light to camouflage its host. In

exchange for the luminescent function, the host provides symbionts with nutrients and a place to proliferate (Dunn, 2012). The squid-*Vibrio* relationship is regulated externally by the Diel cycle (McFall-Ngai, 2014). Bacterial symbionts begin to colonize the host and accumulate in the oxygen-limited light organ of the squid during daylight hours. Once the bacterial symbionts have reached the threshold population density at night, genes involved in luminescence production are activated by quorum sensing, and the bacteria emit light to conceal their host from predators. At dawn, the host vents most of the symbiotic *V. fischeri cells* (90%) into the environment, and then the cycle of bacterial colonization and venting starts again. In an experiment where the squid host was co-colonized with wild-type *V. fischeri* and a *Δlux* strain (a mutant *V. fischeri* strain that lacks the suite of genes required for producing luminescence) at a ratio of 1:1, wild-type bacteria outcompeted the *Δlux* mutant strain at a ratio of 10^6:10 (wild type: *Δlux* mutant) per squid (Koch *et al.*, 2014). This evidence supports the hypothesis for a symbiosis maintenance mechanism within the squid host that eliminates errant non-luminescent symbionts. The squid/*V. fischeri* model system of symbiosis has been well-characterized because it is experimentally tractable. Both partners can be grown in the laboratory setting, the bacterial symbiont can be genetically modified, and the symbiosis only involves two species, which decreases the complexity of analyses.

1.1.3 Multi-species symbioses (host/multiple symbionts)

Co-operative symbioses between multiple species are ubiquitous in nature. Many metazoans are hosts to complex microbial ecosystems. Due to the complexity of the system, it can be difficult to elucidate the factors that drive these associations; however, there is heightening interest in understanding these relationships due to their impact on health and the environment. Recent research has shown that several idiopathic autoimmune conditions (e.g. Crohn's disease and irritable bowel disease) are related to

the depletion of bacterial species in the order Clostridiales from the human gut microbiota (Velasquez-Manoff, 2015). In the murine model, individuals that are less densely colonized by clostridial species are more susceptible to developing colitis; however, the disease state can be prevented by the introduction of a cocktail of native clostridial species to neonatal mice (Atarashi *et al.*, 2011).

In ruminants, there is a core set of microbiota that resides in the rumen (forestomach; Figure 1.1) of the animals (Henderson *et al.*, 2015). The rumen microbial community is affected by diet (Henderson *et al.*, 2015); moreover, certain microbial profiles are correlated with animal productivity (Zhou *et al.*, 2009, Jami *et al.*, 2014) and methane emissions (Kittelmann *et al.*, 2014). In New Zealand, methane and nitrous oxide emissions which mainly resulted from animal production systems accounted for 54.8% of total greenhouse gas emissions in 2013 (New Zealand Ministry for the Environment, 2015); therefore, there is a strong incentive for better understanding microbial relationships that contribute to high or low methane emitting phenotypes.

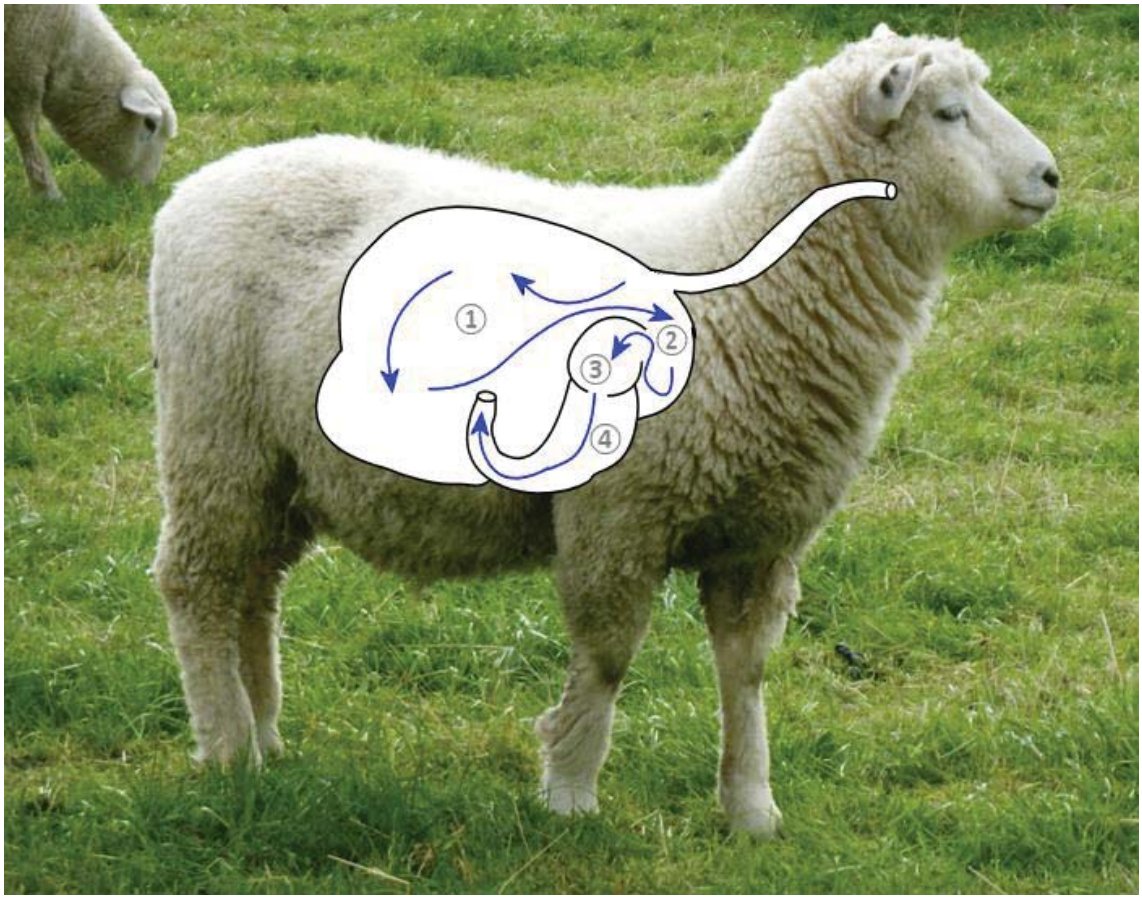


Figure 1.1. Anatomy of the gastrointestinal tract in ruminants.

In ruminants, such as sheep, ingested feed passes through the esophagus into the rumen ①, where fermentation takes place. The feed then travels into the reticulum ②, the omasum ③, and then the abomasum ④ for further processing before exiting *via* the small and large intestines (Hobson & Stewart, 1997).

1.2 Rumen as an example of a co-operative ecosystem

Ruminants, such as cows, sheep, goats, and deer, harbour a fermentative forestomach called the "rumen" (Figure 1.1), which is an anaerobic, pH-controlled habitat that houses a complex microbial community comprised of bacteria, archaea, ciliated protozoa, fungi, and phage (Table 1.1) (Sirohi *et al.*, 2012). The rumen microbial community of domesticated ruminants important in agriculture have been extensively studied (Henderson *et al.*, 2015). The rumen microbiomes of ruminants that consume unusual feed material, and wild ruminants, have also been studied. Alginate-degrading bacteria

have been isolated from seaweed-consuming North Ronaldsay sheep on the coast of Scotland, and novel microbial species have been identified from the rumen microbiome of wild ruminants (for example, North American moose, Norwegian reindeer, and African wildebeests) (Williams *et al.*, 2013, Ishaq & Wright, 2015).

The metabolic activities executed by the rumen microbial community are essential for feed digestion, and the community composition is dependent on the ruminant host's diet (Mao *et al.*, 2014, Henderson *et al.*, 2015). Many examples of symbiotic relationships occurring amongst members of the rumen microbial ecosystem have been reported (Hobson & Stewart, 1997, Russell & Rychlik, 2001). Bacteria, protozoa and fungi act together to digest forage by breaking down the structural polysaccharides, and fermenting the liberated sugars to volatile fatty acids (mainly acetic, propionic and butyric acids) that are absorbed across the rumen epithelium and oxidised by the host animal for energy. Other fermentation end products, such as H_2, carbon dioxide (CO_2), formate, short-chain alcohols, and methylamines, are not used by the host and are further reduced to methane (CH_4) by the action of methanogenic archaea or converted into acetate by acetogenic bacteria (Buddle *et al.*, 2011). The cooperation between these H_2-producing, polysaccharide-degrading organisms and their H_2-consuming methanogen partners is known as "interspecies H_2 transfer".

Table 1.1. Microbial groups present in the rumen environment.

Microbial group	Density (cells or particles per mL of rument contents)
Bacteria	10^{10}-10^{11}
Archaea	10^{7}-10^{9}
Protozoa (ciliates)	10^{5}-10^{7}
Fungi	10^{3}-10^{6}
Phage	10^{7}-10^{10}

Adapted from Sirohi *et al.* (2012).

1.2.1 Protozoa and their partner organisms

Rumen ciliated protozoa account for up to 50% of the rumen microbial biomass (Newbold *et al.*, 2015). Protozoal communities in the rumen are categorized into types A, B, O, and K (Table 1.2) based on the key species present (Eadie, 1967, Kittelmann & Janssen, 2011). The type A community is characterized by the presence of *Polyplastron multivesiculatum*, which has been observed to prey upon other protozoa (with the exception of small *Entodinium*) and cannibalize members of its own species (Eadie, 1967). The type B community is characterized by the presence of *Epidinium ecaudatum* or *Eudiplodinium maggii*. In type O communities, only *Entodinium*, *Dasytricha*, and *Isotricha* species are found. The type K community is found solely in cattle, and *Elytroplastron bubali* is unique to this community type. Ciliated protozoa vary greatly in size. Rumen ciliate species of the genera *Polyplastron*, *Metadinium* and *Ophryoscolex* can be up to approximately 200 µm in length and 150 µm in width. In contrast, many *Entodinium* species as well as *Charonina ventriculi* are much smaller (length ~35 µm, width ~12 µm) (Ogimoto & Imai, 1981). *C. ventriculi* has been frequently observed in different ruminant hosts. Recently, it was identified by the full cycle rRNA approach, where *C. ventriculi* cells were picked from a bovine rumen sample harbouring the type A protozoal community, and the 18S rRNA gene sequence of this protozoal species was analyzed and verified by fluorescence *in situ* hybridization (Kittelmann *et al.*, 2015). *C. ventriculi* was not found in the ovine sample examined in the same study; however, a multinational survey noted the presence of this species in sheep rumen samples from the former Yugoslavia, Iceland, China, and Japan (Booyse & Dehority, 2011).

Table 1.2. Protozoa community types identified in the rumen.

Protozoa community type	Species present
A	*Polyplastron multivesiculatum* *Diploplastron affine* *Ophryoscolex tricoronatus* *Entodinium* spp. *Dasytricha* spp. *Isotricha* spp.
B	*Eudiplodinium maggii* *Epidinium* spp. *Eremoplastron* sp. *Ostracodinium* sp. *Metadinium medium* *Entodinium* spp. *Dasytricha* spp. *Isotricha* spp.
O	*Entodinium* spp. *Dasytricha* spp. *Isotricha* spp.
K (cattle-specific)	*Elytroplastron bubali* *Entodinium* spp. *Dasytricha* spp. *Isotricha* spp.

Most ciliated protozoal species in the sheep rumen can be classified into two families: Ophryoscolecidae and Isotrichidae (Kittelmann *et al.*, 2015). Species in the two families are morphologically distinct. Members of Ophryoscolecidae have one or two localized tufts of cilia, and caudal spines are sometimes observed (Wright, 2015) (Figure 1.2). In contrast, members of Isotrichidae are covered with cilia throughout the entire cell surface, and caudal spines are not found in this taxon. Classification by morphological features of rumen ciliates correlates well with 18S rRNA gene sequence analysis at the genus level; however, gene sequence information helps to corroborate taxonomic assignments, especially at the species level. In the case of *Entodinium*, species in this genus can be difficult to differentiate without expert knowledge as their morphologies are similar (Tymensen *et al.*, 2012a). Moreover, not all morphological features are useful for taxonomic classification. For example, *Entodinium, Eudiplodinium, Diplodinium, Epidinium* species were previously classified by the number of caudal spines observed; however, researchers later found that caudal spination in these species can vary depending on environmental conditions (Dehority, 2006).

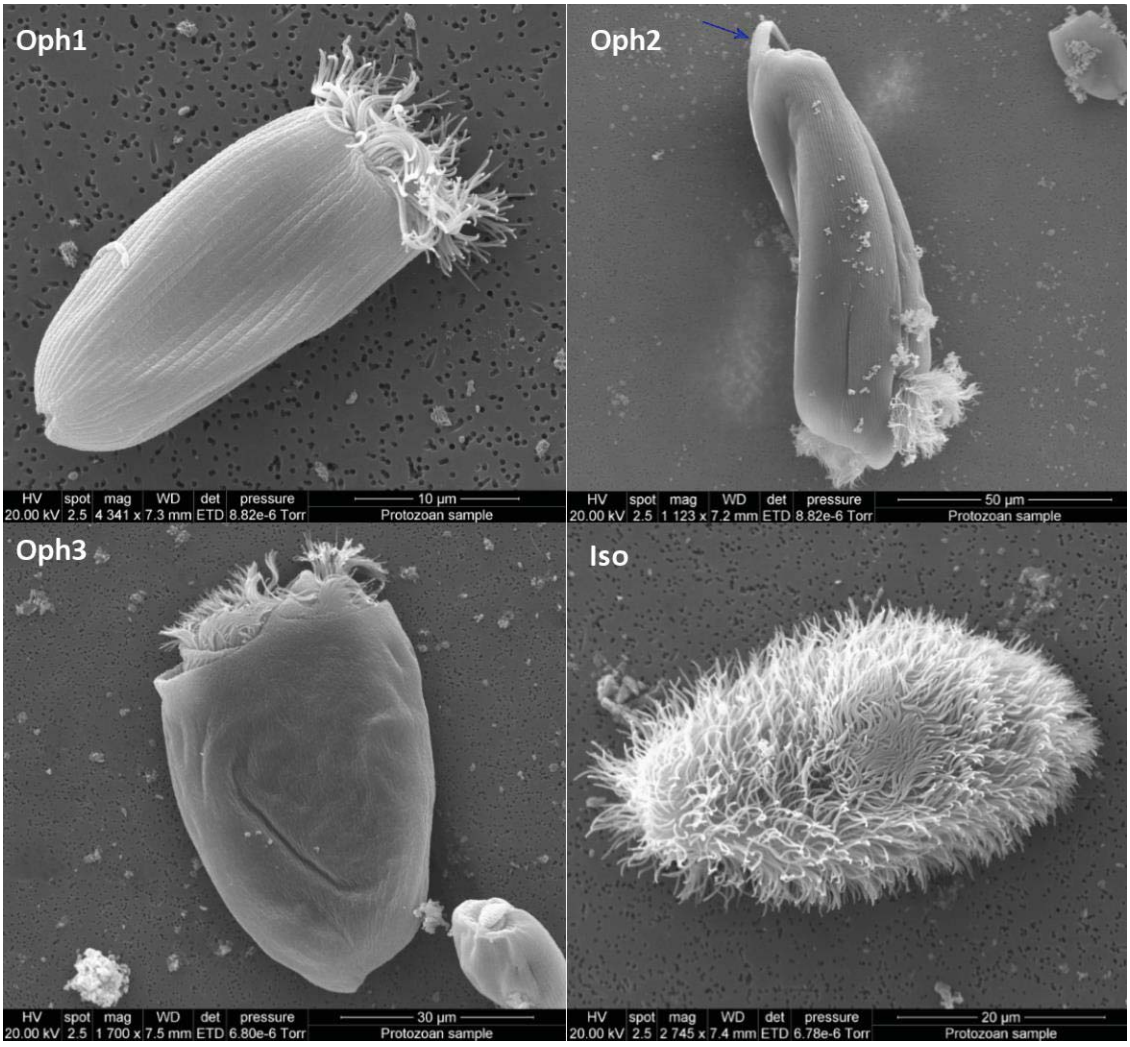


Figure 1.2. Rumen protozoa from the families Ophryoscolecidae and Isotrichidae have different morphological features.

Scanning electron microscopy (SEM) images show that members of Isotrichidae (bottom right image labelled Iso) are covered with cilia over the whole cell surface, whereas members of Ophryoscolecidae (Oph1, Oph2, Oph3) have one or two tufts of cilia in localized regions of the cell. Caudal spines (blue arrow in Oph2) can sometimes be observed in the genera *Epidinium* and *Entodinium* (members of Ophryoscolecidae).

(SEM images produced by F. Ng and the Manawatu Microscopy and Imaging Centre, Massey University, New Zealand)

The species composition of the rumen ciliate community varies depending on diet, genetics of the individual host animals, and animal host species. A study on cattle showed that high grain diet supported the presence of the type A community, whereas high forage diet supported the presence of the type B community (Tymensen *et al.*,

2012a). Regardless of diet, *Entodinium* was the most abundant protozoal genus in the rumen as determined by microscopy (94.2% for high grain diet, 57.4% for high forage diet) (Tymensen *et al.*, 2012a). In cattle harbouring type O ciliated protozoa, the abundance of *Entodinium* and *Isotricha* decreased under high forage diet, while the *Dasytricha* population increased (Monteils et al., 2012). However, even when host animals are fed the same diet, differences can still be found in the rumen protozoal community. In a study where rumen samples were collected from several goat species that were grazed on the same plant material, community profiling by denaturing gradient gel electrophoresis (DGGE) of 18S rRNA protozoal gene sequences revealed both interspecies and intraspecies variations amongst the animals (Shi *et al.*, 2008). Comparative analysis of protozoal communities found in deer, sheep, and cattle showed that the sheep rumen harboured large intraspecies variations which can sometimes obscure diet effects (Kittelmann & Janssen, 2011). Together, these studies indicate that host species and genetics have an effect on the protozoal population present in the rumen. In addition, planktonic protozoa species found in the rumen fluid differ from the species adhered to the feed particulates (Shin *et al.*, 2004). No protozoa were found to attach to the rumen epithelium (Shin et al., 2004).

Rumen protozoa have overlapping functions in fiber degradation, but some taxa also occupy unique ecological niches. For example, *Eudiplodinium maggii* and *Epidinium* sp. exhibit cellulase activity, whereas *Entodinium* spp. have a lower level of cellulase activity but greater amylase activity, and members of the Isotrichidae family are principally involved in the degradation of soluble carbohydrates (Williams & Coleman, 1992). After the ruminant host has consumed its feed, there is a burst of plant material available for digestion in the rumen. Members of the Isotrichidae family are able to accumulate reserves of starch granules for storage during this time, and they display sensitivity

towards glucose as chemoattractant and peptides as chemorepellent. Perhaps bacterial degradation of plant material resulting in glucose liberation facilitates protozoal motility towards plant particles. In contrast, members of the Ophryoscolecidae family exhibit a lower level of chemotactic behaviour, but they are chemoattracted to both glucose and peptides.

Complete genome sequences for rumen ciliates have not been published, but protozoal proteins that contribute to fiber-degrading activity have been identified by functional screening of rumen metagenomic cDNA libraries against cellulose and xylan (Findley *et al.*, 2011). A candidate cellulase containing glycoside hydrolase family 5 domain and three xylanases containing glycoside hydrolase family 10 or family 11 domains were identified from this screen, and the closest relatives to these proteins are encoded in an expressed sequence tag library derived from sheep mono-faunated with *Epidinium ecaudatum*. Centrins (cytoskeletal proteins that are components of ciliary structures) from *Entodinium caudatum* have also been identified using the cDNA library screening approach (Eschenlauer *et al.*, 1998). This approach can be adopted to gain insight into the protein components of the protozoal cell surface as well.

Electron microscopy has been used to characterize components of the cell surface in a range of protozoal species. Members of the Ophryoscolecidae family have a glycocalyx layer on their cell surfaces. For *Entodinium* spp. and *Epidinium caudatum*, this layer is ~50 nm thick (Furness & Butler, 1983, Furness & Butler, 1985b). In *Eudiplodinium maggii*, it is ~200 nm (Furness & Butler, 1985a). The glycocalyx rests upon two or three membrane layers. The epiplasm underlies these membrane layers (45 nm-400 nm, depending on the species) and is tightly associated with the cytoskeleton underneath. On the other hand, members of the Isotrichidae family are covered with cilia, and they have not been reported to possess a glycocalyx layer. Instead, basal bodies and fibres

which form the ciliature have been observed (Vigues & Groliere, 1985). In terms of cell surface association with substrates, *Isotricha intestinalis* has been observed to attach to plant particles *via* its dorsilateral ridge, but it was not possible to capture this interaction by electron microscopy as the protozoal cells detached from the substrate upon fixation with formaldehyde during sample preparation (Orpin & Hall, 1983).

1.2.1.1 Bacteria

A diverse range of interactions occur between rumen protozoa and other members of the rumen microbial community. Predatory behaviour is common, as protozoa ingest bacteria for nutrients and therefore contribute to nitrogen turnover in the rumen by releasing bacterial proteins (Jouany, 1996). Protozoa also engulf plant material and use endogenous fibrolytic enzymes to break it down. Both protozoa and bacteria contribute to the breakdown of plant material (Henderson *et al.*, 2015, Newbold *et al.*, 2015); therefore, ciliated protozoa and bacteria have been proposed to work in synergy to digest feed ingested by the host (Zhang et al., 2007). This hypothesis is supported by experimental evidence where the removal of either partner resulted in a decrease in fiber degradation activities (Shin *et al.*, 2004, Shi *et al.*, 2008). For example, significant decrease in cellulase and xylanase activities was observed in the rumen after defaunation treatment (where protozoa were removed by chemical treatment) (Eugene *et al.*, 2004). Likewise, overall decreased efficiency in the rumen fluid-based *in vitro* degradation of fiber (corn stover, alfalfa hay, and wheat straw as substrates) was reported when bacteria were eliminated by antibiotics (Zhang *et al.*, 2007, Foroozandeh *et al.*, 2009). In addition to the syntrophic relationships observed, the bacterial community associated with protozoa also contribute to host tolerance against toxic substances, such as mercury (Kišidayová *et al.*, 2010).

1.2.1.2 Protozoa-associated bacterial community and their effect on the ruminant host

Species composition of the protozoa-associated bacterial community has been extensively explored over the past two decades, however studies prior to development of next-generation sequencing technology had limited resolution in taxonomic determination of detected bacteria. Protozoa-associated bacteria belonging to the phyla Firmicutes (streptococci, *Butyrivibrio* spp., and Gram positive rods which may be clostridial species), Bacteroidetes (*Bacteroides* spp.), and Actinobacteria (micrococci, corynebacteria) have been identified by Gram staining followed by microscopy (Bonhomme, 1990). Clone library analyses have also been performed on DNA samples isolated from single protozoal cells. In a previous study, protozoa-associated bacterial species isolated from single cells of *Polyplastron multivesiculatum* were identified by sequencing of clone libraries derived from PCR-amplified partial 16S rRNA gene sequences (Irbis & Ushida, 2004). *P. multivesiculatum* are known to predate on bacteria; therefore, in this study, the protozoa were incubated with an antibiotic cocktail for 48 h under anaerobic conditions to remove unassociated bacteria and to allow for protozoa to completely digest ingested bacteria, in an attempt to elucidate the identities of bacteria that have a mutualistic relationship with the protozoan. Sequences corresponding to *Ruminococcus albus* and *Streptococcus bovis* were found in protozoal cells prior to treatment with an antibiotic cocktail. After the protozoa were incubated in the presence of antibiotics for 48 h, only sequences corresponding to the phylum Proteobacteria were detected.

Some well-studied examples of bacteria-eukaryote symbioses are highly specific. For instance, in the symbiosis between *Vibrio fischeri* and *Euprymna*, the squid light organ selects specifically for the species *V. fischeri* from the seawater (McFall-Ngai, 2014). Associations between protozoa and bacteria in the rumen appear to be less specific, as

bacteria from several phyla have been observed (Bonhomme, 1990, Irbis & Ushida, 2004). In the termite hindgut, which is a diverse anaerobic environment similar to the rumen, a diverse community of bacterial symbionts are associated with flagellated protozoa. Species from spirochaetes and Bacteroidales have been identified, and they are speculated to be involved in syntrophic interactions (production of acetate and nitrogen-containing nutrients, respectively) (Gast *et al.*, 2009). Specific cognate pairing of symbionts and hosts may be attributed to the co-evolution of the interacting partners, whereas phylogenetically distinct symbionts found in the host could suggest functional redundancy (i.e. multiple organisms can fulfil a given function in the ecosystem, and can therefore satisfy the functional basis behind the symbiosis) (Dziallas *et al.*, 2012).

The faunated rumen harbours a more complex bacterial community than the defaunated rumen (Ozutsumi *et al.*, 2005), suggesting that the faunated rumen may harbour bacterial species with specialized functions that would ultimately confer survival advantages to the ruminant host. However, this advantage may only be observable when the host is exposed to environmental challenges. Experimental data indicate that some protozoa-associated bacteria can protect their protozoal host from mercury toxicity (Kišidayová *et al.*, 2010). The mechanism of detoxification was tentatively attributed to the conversion of $HgCl_2$ into insoluble complexes by bacterial mercury reductase enzymes (Kišidayová *et al.*, 2010). In addition to protecting the protozoal host, it is plausible that this functionality may protect the ruminant host from mercury toxicity as well.

1.2.1.3 Methanogenic archaea

Methane produced in the rumen is attributed to methanogenic archaea, commonly known as methanogens. Rumen methanogens harbour a unique pathway that enables energy for metabolism to be generated by using H_2 or formate to reduce CO_2 into CH_4

(Buddle *et al.*, 2011). Rumen methanogens mainly belong to the genera *Methanobrevibacter*, *Methanosphaera* and *Methanomicrobium*, and the order *Methanomassiliicoccales* (previously referred to as *Methanoplasmatales*, *Thermoplasmatales*-affiliated lineage C, or rumen cluster C) *(*Table 1.3). Members of the genus *Methanomicrobium* were found to represent >5% of total rumen archaea in ruminants of a number of geographical locations, but were rarely observed in New Zealand cattle and sheep (Henderson *et al.*, 2015, Seedorf *et al.*, 2015). In contrast, the *Methanobrevibacter ruminantium* clade and *Methanobrevibacter gottschalkii* clade were shown to be universally prevalent, and accounted for an average of 74% of rumen archaea in all samples collected in a global census of rumen microbiomes (Henderson *et al.*, 2015). In New Zealand cattle and sheep, these two clades are predominant groups observed in the rumen, representing 75.3% of the total rumen methanogen population (Seedorf *et al.*, 2015). At the global level, *Methanosphaera* sp. ISO3-F5, *Methanosphaera* sp. Group 5 and *Methanomassiliicoccales*-affiliated species together account for 17.7% of the total rumen methanogen population (Henderson *et al.*, 2015). In New Zealand cattle and sheep, these taxa represent 17.9% of the rumen methanogen population. Although *Methanomassiliicoccales* sp. Group 11 only has a mean abundance of 1% in New Zealand pasture-fed cattle and sheep, it is important to note that this number is much higher (average: 15.6%; maximum observed: 68.4%) in grain-fed animals (Seedorf *et al.*, 2015). Interestingly, *Methanomassiliicoccales*-affiliated species, in contrast to other methanogens, do not have blue-green autofluorescence (excitation: 420 nm, emission: 470 nm) exhibited by many methanogen species due to the presence of co-enzyme F_{420}, which is part of the methanogenesis pathway (Valle *et al.*, 2015). Given that *Methanomassiliicoccales*-affiliated species produce methane, their methanogenesis pathway excludes enzymatic steps that depend on co-enzyme F_{420} (Paul *et al.*, 2012).

Table 1.3. Abundance and prevalence of rumen methanogen taxa reported in global census of ruminants and study on New Zealand cattle and sheep.

Taxon	% abundance of total archaea in rumen contents (% prevalence of taxon in samples)	
	Global census of ruminants (Henderson *et al.*, 2015)	NZ cattle and sheep (Seedorf *et al.*, 2015)
Methanobrevibacter (Mbb)		
Mbb. ruminantium clade	27.1 (99.1)	32.9 (100)
Mbb. gottschalkii clade	46.9 (100)	42.4 (99.1)
Methanosphaera (Msp)		
Msp. sp. ISO3-F5	5.7 (97.4)	8.2 (94.8)
Msp. sp. Group 5	2.1 (71.3)	5.5 (91.3)
Methanomassiliicoccales (Mmc)-affiliated species		
Mmc. Group 10 sp.	3.0 (84.8)	2.2 (71.0)
Mmc. Group 11 spp.	0.4 (1.7-43.5*)	1.0 (5.5)
Mmc. Group 12 sp. ISO4-H5	6.5 (87.1)	1.0 (53.2)
Methanomicrobium		
M. mobile	0.7 (27.6)	Not observed

Note: % abundance refers to the mean abundance of a given taxon in the samples analyzed. Prevalence indicates the percentage of samples in which a given taxon is observed. *Depending on species.

Methanobrevibacter ruminantium M1 was the first rumen methanogen to have a complete genome sequence determined (Leahy *et al.*, 2010). The size of the M1 genome is 2.93 Mb, larger by about 1 Mb in comparison to the genome of *Methanobrevibacter smithii* PS, a closely related isolate. Comparative *in silico* analysis suggested that the difference in genome size is due to the presence of a greater number of putative adhesin-encoding genes in the M1 genome (105 *versus* 48 in *M. smithii* PS) (Leahy *et al.*, 2010).

When *M. ruminantium* M1 was co-cultured with the H_2-producing bacterium *Butyrivibrio proteoclasticus* B316, the two species were observed to aggregate (Figure 1.3), and six

adhesin-encoding genes in *M. ruminantium* M1 were up-regulated in the co-culture compared to the monoculture (Leahy *et al.*, 2010). These findings indicate that some adhesin-like proteins may be involved in interspecies interactions *via* cell-cell attachment. Besides putative adhesins, genes encoding proteins involved in methanogenesis were also up-regulated in co-cultures (Leahy *et al.*, 2010). The latter phenomenon had also been observed in the syntrophic interaction [defined as the mutualistic exchange of metabolic products between co-operating partners (Morris *et al.*, 2013)] between hydrogenotrophic *Methanothermobacter thermautotrophicus* and propionate-oxidizing bacterium *Pelotomaculum thermopropionicum*, where genes from the methanogenesis pathway were activated upon attachment of the methanogen to the bacterial flagellum (Shimoyama *et al.*, 2009).

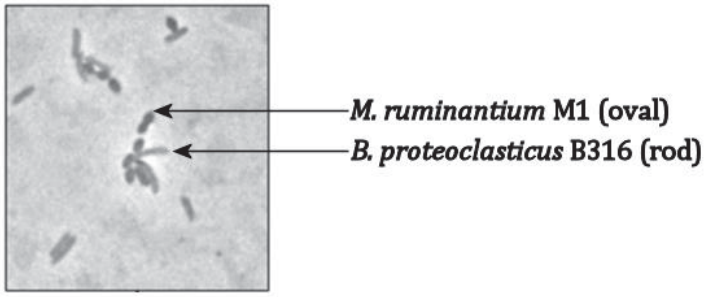


Figure 1.3. Microscopy image showing physical associations between *Methanobrevibacter ruminantium* M1 and *Butyrivibrio proteclasticus* B316 in co-culture.

Microscopy image reproduced with permission from Leahy *et al.* (2010).

Although methanogens are responsible for CH_4 production in the rumen, no correlation has been shown between methanogen abundance in the rumen and the level of CH_4 emitted by an animal (Machmuller *et al.*, 2003, Zhou *et al.*, 2011, Morgavi *et al.*, 2012). Methane emissions are correlated with transcript levels of genes in methanogenesis pathway (Shi *et al.*, 2014). The CH_4 output may be determined by access of methanogens

to H_2-producing microbial partners rather than methanogen abundance. This hypothesis is consistent with evidence of methanogenesis activation in archaea upon detection of H_2-producing partners in two independent studies (Shimoyama *et al.*, 2009, Leahy *et al.*, 2010). A similar mechanism could exist in rumen protozoa-methanogen symbiosis, whereby genes encoding proteins in the methanogenesis pathway are up-regulated upon attachment to H_2-producing ciliate hosts.

1.2.1.4 Protozoa-associated methanogen community

Methanogens can reside as intracellular or extracellular symbionts of protozoa (Finlay *et al.*, 1994) (Figure 1.4). Protozoal metabolism results in the formation of H_2 and the attached methanogens benefit by using this H_2 in their energy-generating methanogenesis pathway. In return, the action of the methanogens is thought to lower the H_2 partial pressure inside the protozoan, allowing protozoal fermentation to proceed to energetically more favourable end products (Muller, 1993, Finlay *et al.*, 1994, Akhmanova *et al.*, 1998, Vogel, 2008). The rate of interspecies H_2 transfer in anaerobic methanogenic environments is significantly enhanced when H_2-forming bacteria and H_2-consuming methanogens aggregate (Schink & Thauer, 1988, Schink, 1997, Ishii *et al.*, 2005). By extension, it is speculated that physical association between methanogens and eukaryotic H_2 producers, such as rumen protozoa, would similarly increase the efficiency of interspecies H_2 transfer, leading to increased methane production.

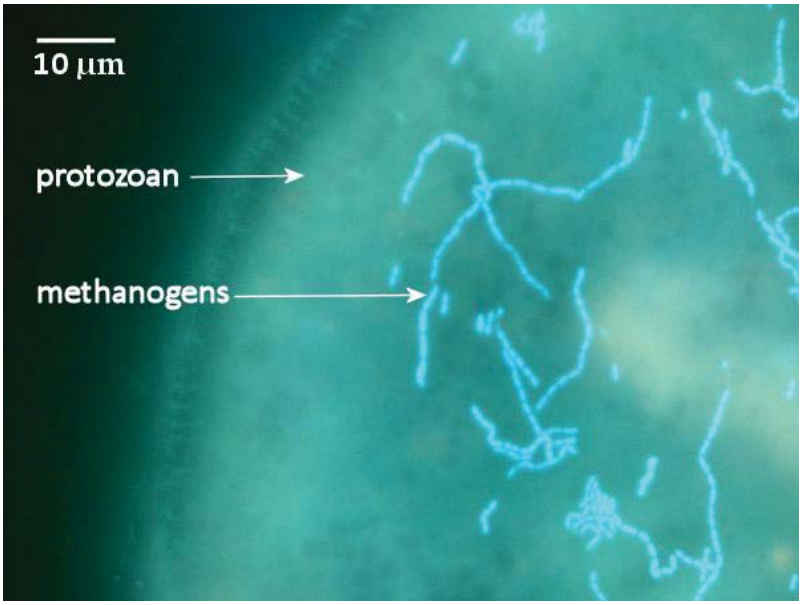


Figure 1.4. Autofluorescent methanogens attached to the surface of a rumen protozoal cell.

This fluorescence microscopy image shows the presence of methanogen cells with autofluorescence on the surface of a protozoan. (Microscopy image produced by F. Ng).

Several studies have aimed to characterize the protozoa-associated methanogen community by examining methanogen species associated with single protozoal cells, as well as those from entire protozoal communities. In the former approach, single protozoal cells were isolated manually, and the resident methanogen symbionts were identified by PCR amplification and sequence analysis of 16S rRNA genes (Irbis & Ushida, 2004, Regensbogenova *et al.*, 2004). In the latter method, entire protozoa communities were isolated from rumen samples, followed by characterization of methanogenic symbionts based on their 16S rRNA gene, or an alternative methanogen-specific marker gene used for taxonomic analysis, *mcrA* (Chagan *et al.*, 1999, Tokura *et al.*, 1999, Tymensen *et al.*, 2012b, Xia *et al.*, 2014). The *mcrA* gene is highly conserved amongst methanogens (Snelling *et al.*, 2014). It encodes the alpha subunit of methyl coenzyme M reductase, an enzyme that catalyzes the final step of the methanogenesis pathway (Friedrich, 2005). In both types of studies, *Methanobrevibacter* species were

reported to dominate this ecological niche, accompanied by minor contributions of species from the order *Methanomassiliicoccales*, and the genera *Methanosphaera* and *Methanomicrobium*. The abundance of symbiotic methanogens varies amongst protozoal species, suggesting that methanogens may prefer particular protozoal hosts (Finlay *et al.*, 1994, Lloyd *et al.*, 1996).

Not all ciliated protozoa appeared to host the same number of methanogen symbionts. For example, few endosymbiotic archaea were observed in *Epidinium* species, while, in the same study, methanogenic symbionts were found in almost all examined species of *Entodinium* and in large entodinomorphs (Lloyd *et al.*, 1996). Contradictory evidence was presented with respect to symbiont colonization of *Dasytricha*. Although DNA hybridization with archaea-specific DNA probes were used in both studies, Lloyd *et al* (1996) did not observe any endosymbiotic methanogens within *Dasytricha*, while another group did observe endosymbionts in this ciliate host (Finlay *et al.*, 1994, Lloyd *et al.*, 1996). It is possible that there are unknown factors governing protozoa-methanogen symbiosis that determine whether methanogen symbionts are present or absent. Although the protozoa-associated methanogen community has been explored in previous studies, host specificity has not been defined for a specific methanogen species, and the molecular mechanism(s) of methanogen attachment to symbiotic hosts have yet to be revealed.

1.2.1.5 Effects of protozoa-methanogen symbiosis on host ruminant physiology

Current research focuses on manipulating protozoa in the rumen for two reasons: to increase animal productivity by improving feed efficiency and to reduce enteric CH_4 emissions. Notably, animals that are more feed efficient also tend to produce less CH_4 (Zhou et al., 2009).

Although protozoa do not produce CH_4, calculations based on rumen volume suggested that protozoa-methanogen symbiosis can contribute up to 37% of methane produced by ruminants (Finlay *et al.*, 1994). As protozoa are not essential for host survival (despite their ability to facilitate digestion in the rumen) defaunation was viewed as a potential measure for CH_4 mitigation (Hegarty, 1999). A meta-analysis study showed that short term defaunation trended towards reduced CH_4 emissions at the expense of feed digestibility (Newbold *et al.*, 2015); however, this effect was not consistent amongst all defaunation studies and methanogenesis was not suppressed upon long term defaunation (Morgavi *et al.*, 2012, Guyader *et al.*, 2014).

Protozoa are not required for feed fermentation in the rumen, but they may play a role in stabilizing rumen pH by sequestering mono- and poly-saccharides away from bacteria, thus preventing rumen acidosis (Denton *et al.*, 2015). They have also been shown to prevent toxicity from metals and mycotoxins in the ruminant host (Kiessling *et al.*, 1984, Ivan *et al.*, 1986). Considering that protozoa contribute to the ruminant's resilience against environmental challenges and that defaunation does not have prolonged effect on reduction of enteric CH_4 production (Machmuller *et al.*, 2003, Morgavi *et al.*, 2012), protozoa may not be a good target for methane mitigation. Instead, methods that target methanogens directly, through vaccine or small molecule inhibitors, are under investigation. Alternatively, novel strategies that indirectly target methanogenesis by blocking the rumen interspecies associations between H_2 producers and methanogens for the mitigation of enteric methane emissions may be a solution that does not compromise the wellbeing of animals.

1.3 Role of adhesins in establishing symbiosis

Cell attachment is an important step in symbiosis, and host-symbiont cell adhesion is generally mediated by cell-surface anchored and surface-associated proteins (Kline *et al.*,

2009, Lebeer *et al.*, 2010). Proteins that mediate binding to other microbes or host tissues and extracellular matrix or other surfaces are termed "adhesins", and they encompass a structurally diverse group of membrane-associated proteins. In both Gram positive and Gram negative bacteria, cell appendages and afimbrial adhesins have been found to mediate adhesion (Kline *et al.*, 2009). Among various protein folds found in adhesion domains in pili and afimbrial adhesins, immunoglobulin-like (Ig-like) folds appear to be dominant in mediating host cell binding (Halaby & Mornon, 1998).

Sequence homology-based annotation of genes is the standard approach for gene function prediction. It is a powerful method for recognizing well-conserved features such as type IV pili and the LPxTG motif for sortase-mediated surface localization. However, the gene homology-based approach fails to capture the full repertoire of proteins involved in adhesive functions in the absence of well conserved sequences. Within the immunoglobulin fold family, amino acid sequence similarity is low (identity as low as 10%, similarity as low as 20%); therefore, in the absence of functional data, it is not possible to predict the function and, in particular, specificity of binding for putative adhesins in this family (Halaby & Mornon, 1998, Wang *et al.*, 2013, Mei *et al.*, 2015). Furthermore, proteins that are not annotated as adhesins may harbour cell binding function. A number of proteins with "moonlighting" adhesive function have been reported for proteins of central metabolism (Kainulainen & Korhonen, 2014), challenging the one gene-one protein-one function paradigm. As current gene function prediction tools were designed based on this paradigm and are based on existing functional data that are available only for a small subset of proteins within the ever growing sequence databases, the full complement of functions a protein harbours cannot be captured by the current approach to genome sequence annotation (Khan *et al.*, 2014).

This is a particular problem with archaea, which are hard to cultivate, making the genomic studies disproportionally dominant over functional studies.

1.3.1 Archaeal adhesins involved in forming interspecies cell-to-cell associations

Although only a small number of adhesins have been functionally and morphologically characterised in archaea, some interesting examples can be found in the literature (Figure 1.5). Adhesins and protein structures containing adhesins (e.g. fimbriae) enable archaeal cells to attach to abiotic surfaces and/or other cells. For example, the methanogenic archaeon species *Methanothermobacter thermoautotrophicus* was observed to produce large amounts of fimbriae when grown on abiotic surfaces (Figure 1.5C) (Thoma *et al.*, 2008). Mass spectrometry analysis of the isolated fimbriae revealed the protein Mth60, a 15 kDa gene product encoded by the gene *mth60*, as the major component of this structure. In the presence of Mth60 antibodies, *M. thermoautotrophicus* cells detached from the surface that they were grown on. In addition, physical connections between the fimbriae and the abiotic surface were observed by scanning electron microscopy (SEM); therefore, it was concluded that Mth60 was a fimbrial adhesin that mediated attachment to abiotic surfaces. Mth60 has no known protein homologs in Bacteria.

A second example of archaeal adhesive surface appendages were found in uncultured archaeal species *Candidatus Altiarchaeum hamiconexum* (also known as SM1 Euryarchaeon). Electron microscopy showed that the filaments, named "hami", resembled barbed wire decorated with a grappling hook at the tip (end distal from the cell surface) and were up to 3.7 μm in length (Figure 1.5A,B). The protein composition and sequence analysis of the hami revealed that they were composed of 97 kDa subunits containing predicted glycosylation sites (Perras *et al.*, 2014, Perras *et al.*, 2015). As physical associations *via* hami between this archaeal species and filamentous bacteria

have been observed, these appendages were speculated to play an important role in interspecies cell adhesion (Perras *et al.*, 2014).

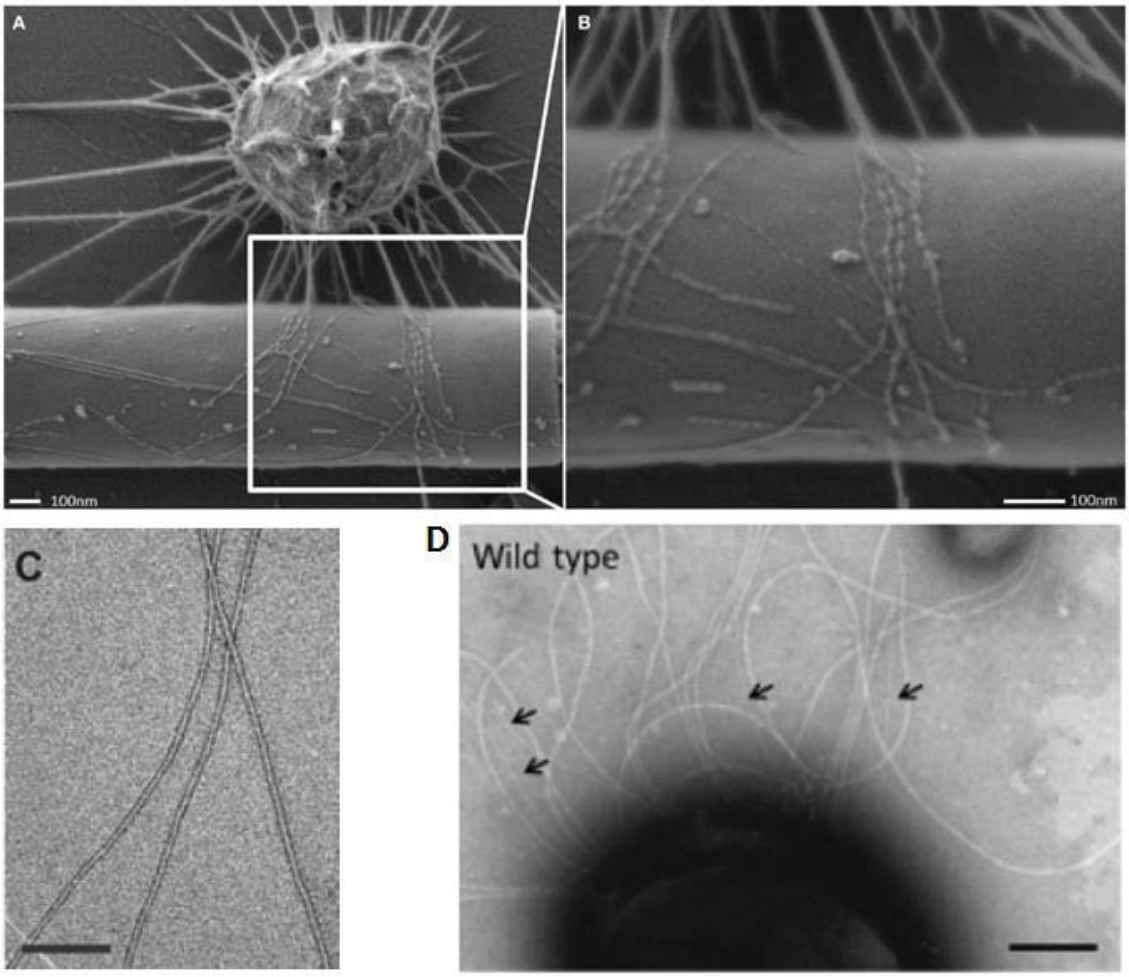


Figure 1.5. Electron micrographs of adhesive structures found in archaea.

Electron micrographs are reproduced from Perras *et al.* (2014) (Panels A and B), Thoma *et al.* (2008) (Panel C), and Jarrell *et al.* (2011) (Panel D) with permission.

Panels A and B show the SEM images of hami structures from *Candidatus Altiarchaeum hamiconexum* in physical association with a bacterium (Perras *et al.*, 2014). Panel C shows a transmission electron micrograph of *Methanothermobacter thermoautotrophicus* Mth60 fimbriae (Thoma *et al.*, 2008). Panel D depicts a wild-type *Methanococcus maripaludis* that produces both flagella and pili for adhesive function (Jarrell *et al.*, 2011).

Scale bars indicate 100 nm for Panels A, B, C; 200 nm for Panel D.

In contrast to hami and Mth60 fimbriae which do not have bacterial counterparts, cell appendages analogous to the bacterial type IV pili and flagella have been observed in methanogenic archaea *Methanococcus maripaludis* (Jarrell *et al.*, 2011). Gene knockout experiments showed that both appendages were required for attachment to abiotic surfaces, and SEM images showed that flagella bridge the neighbouring cells.

Besides complex adhesive appendages or fibrils, interspecies associations mediated by afimbrial adhesins have been noted in the bi-species symbiosis between *Pyrococcus furiosus* and *Methanopyrus kandleri* (Schopf *et al.*, 2008). Cell-to-cell contacts have been observed by SEM, but the molecular basis for adhesion remains to be elucidated in this system.

1.3.2 Annotated adhesins in rumen methanogenic archaea

The genome sequences of rumen methanogen *Methanobrevibacter ruminantium* M1 and human gut methanogen *Methanobrevibacter smithii* harbour a large repertoire of adhesin-encoding genes (Table 1.4) (Leahy *et al.*, 2010, Hansen *et al.*, 2011). Domains commonly identified in methanogen adhesin-like proteins (ALPs) include immunoglobulin-like domains, pseudomurein-binding repeat regions, carboxypeptidase and pectate lyase domains (Table 1.5). During co-culture of rumen methanogen M1 and bacterium *B. proteoclasticus* B316, transcription of genes encoding ALPs that contain pectate lyase domains and immunoglobulin-like domains was up-regulated (Leahy *et al.*, 2010). Hansen *et al.* (2011) proposed that the abundance and diversity of ALPs may be important for methanogens to establish symbiotic relationships with multiple diverse bacterial partners present in a given ecological niche (Hansen *et al.*, 2011). Cell binding activity has not been demonstrated in functional assays for any methanogen ALP as yet.

Table 1.4. Abundance of genes encoding putative adhesin-like proteins in methanogen genomes.

Taxon	Environment	Number of putative adhesin-like proteins (ALPs)	Number of ALPs with predicted surface localization
Methanobrevibacter ruminantium M1 (Leahy *et al.*, 2010) (ruminantium clade)	Rumen	105	62
Methanobrevibacter olleyae YLM1 (Kelly *et al.*, 2016b) (ruminantium clade)	Rumen	59	**48
Methanobrevibacter smithii PS (Samuel *et al.*, 2007)	Sewage digester	48	41
Methanobrevibacter millerae SM9 (Kelly *et al.*, 2016a) (gottschalkii clade)	Rumen	88	**67
Methanobrevibacter sp. AbM4 (Leahy *et al.*, 2013) (wolinii clade)	Rumen	29	26
Methanosphaera stadtmanae MCB-3 (Fricke *et al.*, 2006)	Human gut	37	37
Methanothermobacter thermoautotrophicus ΔH (Smith *et al.*, 1997)	Sewage digester	*10	8

*Putative ALPs for *M. thermoautotrophicus* ΔH were identified by BLASTP search of *M. smithii* PS putative ALPs against the genome of *M. thermoautotrophicus* ΔH, published in the supplementary materials of Samuel *et al.* (2007)

**Based on SignalP prediction of signal sequences in sequences annotated as adhesin-like proteins

Table 1.5. Domains predicted for adhesin-like proteins annotated in methanogen genomes.

	Pseudomurein-binding repeat region	Immunoglobulin-like	Polymorphic outer membrane protein repeat	Fibronectin type III	Bacterial surface protein 26-residue repeat	Filamin repeat-like	Fimbrial isopeptide formation D2 domain	Transglutaminase	Carboxypeptidase	Cysteine peptidase; Peptidase C1A	Collagen-binding protein B-type domain	Periplasmic copper-binding protein NosD	Carbohydrate binding-like fold	Chitinase/polycystic kidney disease domain	Chitobiase	Glycoside hydrolase	Pectin lyase-like	DUF11	DUF1565	DUF285	DUF3344	DUF4062	DUF4399
Mbb. ruminantium M1	X	X	X	X	X	X		X	X	X	X	X			X		X	X		X		X	X
Mbb. olleyae YLM1	X	X	X					X	X	X		X					X	X			X		
Mbb. smithii PS		X	X			X	X		X	X			X				X	X	X		X		X
Mbb. millerae SM9		X	X					X	X	X						X	X	X			X		
Mbb. sp. AbM4	X	X	X			X	X	X	X							X	X	X					
Msp. stadtmanae MCB-3		X				X			X	X							X						
Mth. thermoautotrophicus ΔH	X								X			X	X	X			X	X			X		

1.3.3 Bacterial adhesins

Bacterial adhesins have been studied for many decades. They encompass a diverse group of proteins, including pilins and afimbrial adhesins (Kline *et al.*, 2009). Their involvement in the attachment of pathogens to eukaryotic host cells has been well studied. Similar mechanisms may also be used by mutualistic symbionts to attach to their hosts (Lebeer *et al.*, 2010).

Cell appendages, such as different types of pili/fimbriae, are generally composed of multiple repeating subunits, thus they are termed polymeric adhesins (Kline *et al.*, 2009, Chahales & Thanassi, 2015). These structures enable pathogenic bacteria to adhere to sugar modifications present in glycoproteins and/or glycolipids present on host cell surfaces (Kline *et al.*, 2009). The pilus is composed of thousands of identical subunits that form the shaft of a long fiber, and is capped by an adhesive "tip" structure (made up of one or more minor pilin subunits) at the apical end of the fiber (Chahales & Thanassi, 2015) (Figure 1.6a). Monoadhesive pili interact with host receptors primarily through the pilus tip, whereas polyadhesive pili form physical associations with the host *via* the major subunits that make up the fibril as well as the minor tip subunit (Chahales & Thanassi, 2015) (Figure 1.6b). GafD, PapG and FimH are tip pilins located at the apical end of monoadhesive *E. coli* G fimbriae, P pilus, and type 1 pilus, respectively (Westerlund-Wikstrom & Korhonen, 2005). Despite the weak homology in their primary amino acid sequences, these proteins are adhesins that share similarities in structure [each contains a beta barrel jelly roll fold (Merckel *et al.*, 2003); PapG and FimH each contains an Ig-like fold (Westerlund-Wikstrom & Korhonen, 2005)] (Figure 1.7). Each of these adhesins bind to different ligands (Table 1.6), which likely contribute to tissue/host tropism.

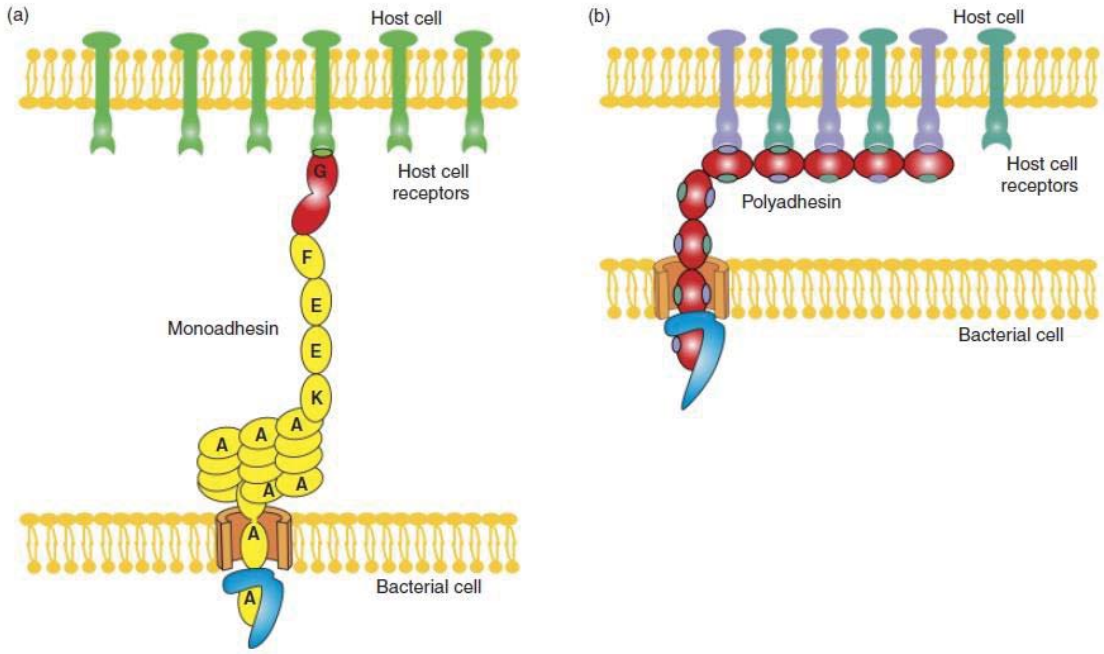


Figure 1.6. Monoadhesive pili *vs*. polyadhesive pili.

Schematic figure reproduced from Zavialov *et al*. (2007) with permission.

The shaft component of monoadhesive pili is made up of protein subunit A, and the tip component is composed of subunits K, E, F, G. Interactions with host cell receptors are mediated primarily through the tip component. Polyadhesive pili interact with host cell receptors through the major pilin subunit as well as minor tip subunit.

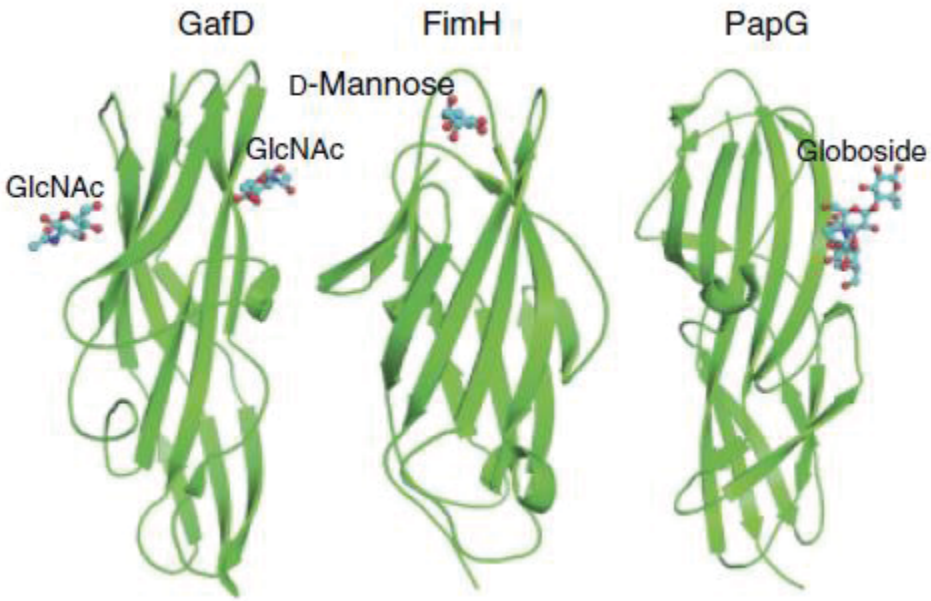


Figure 1.7. Tip subunit pilins GafD, FimH, and PapG share similarities in structure.

Figure reproduced from Zavialov *et al*. (2007) with permission.

An example of polyadhesive pili in *E. coli* is curli. Curli is composed of CsgA as a major subunit protein along the length of the fiber, and capped with the minor subunit CsgB (Chahales & Thanassi, 2015). CsgA and CsgB both exhibit adhesive properties, and share similarities at the amino acid sequence level (30% sequence identity and shared sequence motifs) (Chahales & Thanassi, 2015). Monoadhesive and polyadhesive pili/fimbriae have also been identified in other Gram negative bacterial pathogens, such as *Salmonella enterica* and *Klebsiella pneumoniae* (Kisiela *et al.*, 2006, Wagner & Hensel, 2011); however, many of these host adhesion protein structures have yet to be determined. Pili have been identified in Gram positive bacteria as well. Protein structure data for several pilus tip proteins in pathogenic *Streptococcus* species are available, and revealed the presence of CnaA and CnaB domains (variations of Ig-like folds), domains similar to von Willebrand factor type A domain, and thioester containing adhesin pilin domain (Krishnan, 2015). Therefore, Ig-like domains and variants appear to be a shared feature in Gram positive and Gram negative tip pilins.

Table 1.6. Examples of monoadhesive and polyadhesive pili in Gram negative and Gram positive pathogens.

Adhesin	Organism	Host ligands	Protein structure features	Reference
Gram negative				
Monoadhesive pili				
P pilus tip protein PapG	*E. coli*	Galabiose modified glycolipids	beta barrel jellyroll fold, Ig-like fold	Westerlund-Wikstrom & Korhonen, 2005
Type 1 pilus tip protein FimH	*E. coli*	Mannose modified glycoproteins	beta barrel jellyroll fold, Ig-like fold	Westerlund-Wikstrom & Korhonen, 2005
G fimbriae GafD tip protein	*E. coli*	N-acetylglucosamine	beta barrel jellyroll fold	Westerlund-Wikstrom & Korhonen, 2005
Type 1 fimbriae FimH	*S. enterica*	Mannose modified glycoproteins		Kisiela *et al.*, 2006
Type 3 pilus MrkD	*K. pneumoniae*	Collagen	beta barrel jellyroll fold	Rego *et al.*, 2012
Polyadhesive pili				
Dr fimbriae major subunit protein DraE	*E. coli*	Decay-accelerating factor and carcinoembryonic antigen family proteins; type IV collagen		Kline *et al.*, 2009
Dr fimbriae minor subunit protein DraD	*E. coli*	Integrins	Incomplete Ig-like fold	Bodelon *et al.*, 2013
CFA/I fimbriae major subunit pilin CfaB	*E. coli*		Ig-like fold	Li *et al.*, 2009
CFA/I fimbriae minor subunit pilin CfaE	*E. coli*		Ig-like fold	Li *et al.*, 2007
Curli CsgA and CsgB	*E. coi*	Fibronectin, laminin, fibrinogen		Chahales & Thanassi, 2015
Thin aggregative fimbriae AgfA and AgfB	*E. coli*	Fibronectin, laminin, fibrinogen		Wagner & Hensel, 2011
Gram positive				
Sortase-assembled pilus tip protein SpaC	*C. diptheriae*			Rogers *et al.*, 2011
Sortase-assembled pilus tip protein RrgA	*S. pneumonia*	Fibronectin, collagen, laminin	CnaB-CnaA-vAPD-CnaB	Krishnan, 2015
Sortase-assembled pilus tip protein GBS104	*S. agalactiae*	Fibronectin, collagen	CnaB-CnaA-vAPD-CnaB	Krishnan, 2015
Sortase-assembled pilus tip protein Cpa	*S. pyogenes*		TAPD-CnaB-TAPD-CnaB	Krishnan, 2015

Abbreviations: vAPD = von Willebrand factor type A domain-like adhesin pilin domain, TAPD = thioester-containing adhesin pilin domain

Table 1.7. Examples of mono- and oligo-meric adhesins in Gram negative and Gram positive pathogens.

Adhesin	Organism	Host ligands	Sequence motifs/domains	Protein structure features	References
Gram negative					
Autotransporters					
Ag43	*E. coli*	Collagen, laminin	19 aa repeated motif containing 5 glycine residues		Chahales & Thanassi, 2015
YadA	*Y. enterocolitica*	ECM binding	Repeated motif SVAIGXXS		Girard & Mourez, 2006, Kline *et al.*, 2009
ShdA	*S. enterica*	Fibronectin	Repeated motif containing 63 aa Imperfect repeated motif with 102 aa		Barlag & Hensel, 2015
Other mono- or oligo-meric adhesins					
Intimin	*E. coli*	Integrins		Ig-like folds	Leo & Skurnik, 2011
Invasin	*Y. enterocolitica*, *Y. pseudotuberculosis*	Integrins		Ig-like folds	Leo & Skurnik, 2011
LigA, LigB, LigC	*Leptospira* sp.	Fibronectin, fibrinogen, laminin, elastin, tropoelastin, collagen	13 tandem Ig-like domains	Ig-like folds	Lin *et al.*, 2010
SiiE	*S. enterica*	GlcNAc, sialic acid	53 tandem Ig-like domains		Barlag & Hensel, 2015
Gram positive ***Mono- or oligo-meric adhesins***					
CNA	*S. aureus*	Collagen		Ig-like folds	Berisio & Vitagliano, 2012
FbsC	*S. agalactiae*	Fibrinogen	Ig-like domains		Buscetta *et al.*, 2014
FnBPA and FnBPB	*S. aureus*	Integrin, fibrinogen, elastin, fibronectin	30 aa motif repeated 2x (no known function), 37 aa motif repeated 3x (Fn binding)		Massey *et al.*, 2001
AgI/II (SpaP)	*S. mutans*	Glycoprotein 340, fibronectin, collagen	3 repeats: TELARVQKANADAKAAY 4 repeats: TYEAALKAQYEADL	Alpha and polyproline II helices	Brady *et al.*, 2010
Serine rich protein Fap1	*S. parasanguis* FW213	Unknown	(E/V/I)S		Lizcano *et al.*, 2012
Serine rich protein GspB	*S. gordonii* M99	sialyl-T antigen	(A/E/V/T)S		Lizcano *et al.*, 2012
Serine rich protein PsrP	*S. pneumoniase* TIGR4	unglycosylated keratin 10	SAS (A/E/V)SAST		Lizcano *et al.*, 2012
Serine rich protein SraP	*S. aureus*	unknown glyco modified component	STSLSD		Lizcano *et al.*, 2012

In addition to cell appendages, afimbrial adhesins that exist as monomers or oligomers also play a role in host adhesion. These include autotransporters, cell wall-anchored proteins with adhesion domains [e.g. immunoglobulin-like folds (Bodelon *et al.*, 2013)] or characteristic patterns in the amino acid sequence [e.g. serine-rich repeats (Lizcano *et al.*, 2012)], as well as moonlighting adhesins (the latter group will be discussed in Section 1.3.4).

Trimeric autotransporters in Gram negative bacteria have been implicated in adhesion to extracellular matrix proteins (Kline *et al.*, 2009). Each monomer follows a head-stalk-anchor architecture (Chahales & Thanassi, 2015). The anchor region is well-conserved within the family, and the beta-barrel structure is integrated into the outer membrane where it acts as a pore that allows the rest of the protein (the stalk and head regions) to translocate through the membrane (Lyskowski *et al.*, 2011). The long stalk region extends outwards from the bacterial cell surface, whereas the head region at the tip of the stalk is responsible for interacting with host cells *via* extracellular matrix components (Hartmann *et al.*, 2012, Bassler *et al.*, 2015). Motif repeats have been identified in both the stalk and head structures (Hartmann *et al.*, 2012, Bassler *et al.*, 2015).

Repetitive motifs have also been observed in other afimbrial adhesins from both Gram negative and Gram positive bacteria. Repeats of Ig-like domains have been noted in the invasin/intimin family of proteins and some fibronectin-binding proteins (Table 1.7). In several Gram positive pathogenic bacteria, adhesins containing either serine-rich repeats, or both alanine rich and proline-rich motifs have been observed (Table 1.7), but adhesive function was attributed to the non-repetitive region of the proteins in both cases (Brady *et al.*, 2010, Lizcano *et al.*, 2012).

There are similarities in host cell adhesion between pathogenic and mutualistic bacteria. For example, adhesins containing serine-rich repeats and Ig-like domains have been

reported to be involved in host cell adhesion. In the probiotic bacterium *Lactobacillus reuteri* 100-23, the putative cell surface adhesin containing serine-rich repeats, Lr70902, is thought to be crucial for host attachment (Bokhari *et al.*, 2012, Frese *et al.*, 2013). The gene encoding Lr70902 was up-regulated 18-fold during the first six hours of host colonization; moreover, a gene knockout study demonstrated that colonization of the mouse gastrointestinal tract was unsuccessful in the absence of this gene product (Frese *et al.*, 2011, Frese *et al.*, 2013). Another example of cell surface adhesins involved in cell-binding is the protein pair SpcA and SpcB in probiotic bacterium *Lactobacillus rhamnosus* HN001 (Gagic *et al.*, 2013). SpcA contains two bacterial immunoglobulin-like class 3 domains, whereas SpcB is an alanine-rich polypeptide with several ASKD repeats close to the C-terminal end.

1.3.4 Moonlighting proteins and other multi-functional proteins with cell adhesion function

Moonlighting proteins are multi-functional proteins whose function is determined by the polypeptide's cellular localization, the ligand bound, and/or its oligomerization state (Jeffery, 1999). Their occurrence has been noted in many branches of life, including prokaryotes (Huberts & van der Klei, 2010, Jia *et al.*, 2013). In the Group A streptococcal species, *Streptococcus pyogenes*, about 30 cytosolic proteins were discovered to localize to the cell surface during an effort to identify cell surface proteins as potential vaccine candidates. Some of these cytosolic proteins were already known to exhibit moonlighting adhesive properties (Cole *et al.*, 2005).

Many moonlighting proteins are housekeeping proteins with intracellular functions, but act as adhesins that bind to other cells when localized to the cell surface (Table 1.8). As an example, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) catalyzes the

phosphorylation of glyceraldehyde-3-phosphate to 1,3-biphosphoglycerate in the glycolytic pathway. In pathogenic strains of *Escherichia coli*, it has been shown that GAPDH can be exported to the cell surface, and binding to plasminogen, fibrinogen, and human intestinal epithelial Caco-2 cells has been observed (Cole *et al.*, 2005, Egea *et al.*, 2007). It was speculated that multi-functional moonlighting proteins may be an evolutionary vestige from a world with less protein diversity, where the functional complexity required for life must be fulfilled by proteins that can execute multiple molecular functions (Henderson, 2014). Alternatively, it has also been suggested that a moonlighting protein may be shared amongst multiple pathways to co-ordinate disparate cellular activities (Khan *et al.*, 2014, Jeffery, 2015).

Table 1.8. Moonlighting proteins with adhesive functions that can bind to other cell types.

Intracellular function (cytoplasmic or periplasmic localization)	Species	Adhesive function (when localized to cell surface) §						Reference
		Plg	Fbg	Fn	C	Lam	Cell binding	
Glycolysis								
Glucose-6-phosphate isomerase	*Lactobacillus crispatus*				X		*Lactobacillus rhamnosus* GG	a
Triosephosphate isomerase	*Lactobacillus plantarum*						Human intestinal epithelial Caco-2 cells	a
GAPDH*	*Escherichia coli*	X	X				Human intestinal epithelial Caco-2 cells	a
GAPDH	*Lactobacillus plantarum*	X	X				Human intestinal epithelial Caco-2 cells	a
GAPDH	Oral streptococci						*Porphyromonas gingivalis*	b
GAPDH	*Streptococcus agalactiae*						*Streptococcus pyogenes, Lactobacillus lactis, Staphylococcus aureus, Escherichia coli*	a
GAPDH	*Streptococcus suis*						Porcine trachaeal rings	b
Enolase	*Lactobacillus crispatus*	X			X	X	*Lactobacillus rhamnosus* GG	a
Enolase	*Lactobacillus plantarum*	X		X			Human intestinal epithelial Caco-2 cells	a
Enolase	*Streptococcus suis*			X			Human laryngeal epithelial cell line Hep-2	a, b
Pentose phosphate pathway								
6-phosphogluconate dehydrogenase	*Streptococcus pneumoniae*						Human lung alveolar epithelial A549 cells	b
Protease								
Dipeptidyl peptidase	*Streptococcus suis*			X			Human laryngeal epithelial cell line Hep-2	c
Endopeptidase O	*Streptococcus pneumoniae*	X		X			Human lung alveolar epithelial A549 cells, human umbilical vein endothelial cells	a, c

Note: Data collated from Kainulainen & Korhonen (2014)[a], Henderson (2014)[b], Jarocki *et al.* (2015)[c], Gallotta *et al.* (2014)[d].

§Plasminogen (Plg), fibrinogen (Fbg), fibronectin (Fn), collagen (C), laminin (Lam); *glyceraldehyde-3-phosphate dehydrogenase (GAPDH)

Table 1.8. (continued)

Intracellular function (cytoplasmic or periplasmic localization)	Species	Adhesive function (when localized to cell surface)						Reference
		Plg	Fbg	Fn	C	Lam	Cell binding	
Protein folding chaperone								
DnaK	*Lactobacillus lactis*						*Saccharomyces cerevisiae*	[a]
GroEL	*Lactobacillus johnsonii*						Human adenocarcinoma HT29 cells	[a]
Peptidylprolyl isomerase	*Streptococcus pneumoniae*						Human lung alveolar epithelial A549 cells	[b]
Translation								
EF-Tu	*Lactobacillus johnsonii*						Human intestinal epithelial Caco-2 cells	[a]
EF-Tu	*Lactobacillus plantarum*						Human intestinal epithelial Caco-2 cells	[a]
Others								
Alcohol acetaldehyde dehydrogenase	*Listeria monocytogenes*						Human intestinal epithelial Caco-2 cells	[a]
ATP-binding cassette transporter	*Streptococcus suis*						Human laryngeal epithelial cell line Hep-2	[b]
Cell division protein SpyAD (homologous to negative regulator of FtsZ)	*Streptococcus pyogenes*						Human lung alveolar epithelial A549 cells	[d]
Choline binding protein	*Streptococcus pneumoniae*						Human lung alveolar epithelial A549 cells, human umbilical vein endothelial cells	[b]
Glutamine synthase	*Lactobacillus crispatus*	X			X	X	*Lactobacillus rhamnosus* GG	[a]
Pyruvate oxidase	*Streptococcus pneumoniae*						Deletion mutant failed to attach to human endothelial and lung cells	[b]

Note: Data collated from Kainulainen & Korhonen (2014)[a], Henderson (2014)[b], Jarocki *et al.* (2015)[c], Gallotta *et al.* (2014)[d].

[§]Plasminogen (Plg), fibrinogen (Fbg), fibronectin (Fn), collagen (C), laminin (Lam)

Aside from multifunctional moonlighting proteins, proteins with promiscuous binding activities can execute multiple functions as well. Cellulose processing enzymes in the cellulose binding module CBM37 family are encoded by the rumen bacterium *Ruminococcus albus* strain 8, two of which are known to be important for cellulose degradation as mutants lacking the corresponding genes failed to bind cellulose (Devillard *et al.*, 2004). Functional assays with recombinant Cel48A showed that it can also bind to *R. albus* cells (Ezer *et al.*, 2008). Chitinases and chitin-binding proteins have been described as virulence factors in several bacterial pathogens (including *P. aeruginosa*, *Enterococcus faecalis*, and *Listeria monocytogenes*) (Frederiksen *et al.*, 2013). The transcription of genes encoding these proteins were up-regulated during infection. At the protein level, proteomics analysis of *Francisella tularensis* demonstrated that a chitinase was overexpressed during infection of the mouse spleen (Frederiksen *et al.*, 2013). As *N*-acetylglucosamine (GlcNac) is a chemical constituent that is in common between chitin and mammalian cells, it was speculated that these proteins bind to host cells *via* GlcNac modified glycans present on the cell surface.

1.3.5 *In silico* adhesin identification based on the primary amino acid sequence

Traditionally, *in silico* assignment of protein function is based on sequence homology to proteins with known functions; however, the results may not be sufficiently informative. The emerging prominence of multi-functional proteins can complicate sequence analysis by homology (Jeffery, 2015). In addition, much of the functionality encoded in metagenomes currently remains unknown [>25% and 65% of identified open reading frames (ORFs) for human gut and bovine rumen microbiome, respectively; >85% for human lung and bovine viromes] (Prakash & Taylor, 2012, Oulas *et al.*, 2015), as the query sequence may not be similar to any sequence in the database, or the BLAST hits may correspond to proteins with unknown functions.

To more fully capture the repertoire of adhesins harboured by an organism, a computational approach for adhesin prediction, SPAAN (a **s**oftware program for **p**redicting **a**dhesins and **a**dhesin-like proteins by **n**eural networks) has been developed. This is a non-homology based method for determining the probability that a given primary amino acid sequence may be an adhesin (Sachdeva *et al.*, 2005). Parameters that are considered include amino acid frequencies, presence of homopolymers, dipeptide frequencies, charge composition, and hydrophobic composition of the input sequence. The output is a score, P_{ad}, between 0 and 1, indicating the probability that the input sequence encodes an adhesin. The software is trained on a data set containing known adhesins and non-adhesins, and then tested on a small data set of 37 experimentally verified adhesins and non-adhesins. At a threshold value of 0.51 for P_{ad}, 89% of adhesins in the test data set were identified correctly without false positives. When the test data set was expanded to 194 adhesins from enteropathogenic bacteria, 97% of these adhesins were identified ($P_{ad} > 0.51$). The SPAAN algorithm was also used to identify adhesins from the eukaryote *Aspergillus fumigatus* and from the bacterium *Mycobacterium tuberculosis* (Upadhyay *et al.*, 2009, Kumar *et al.*, 2013). Twenty and 25 putative adhesins were identified *in silico*, respectively. *In vitro* assays have been performed to verify binding for a small subset of these putative adhesins (one for *A. fumigatus* and three for *M. tuberculosis*) (Upadhyay *et al.*, 2009, Kumar *et al.*, 2013).

1.4 Phage display and its application in identification of adhesins

Cell adhesion proteins involved in host-symbiont attachment have been identified in many bacterial pathogens using phage display technology (Mullen *et al.*, 2006, Lima *et al.*, 2013, Evangelista *et al.*, 2014, Gagic *et al.*, 2016). More recently, this method was also used to identify adhesins in probiotic lactobacilli species (Jankovic *et al.*, 2007, Rosander *et al.*, 2011, Gagic *et al.*, 2013). Phage display can be used to study functional activities in

systems where genetic manipulation methods have not yet been established or the organism of interest is not readily cultivable. Many anaerobic rumen microbes cannot be genetically manipulated or cultured; therefore, the phage display approach provides a culture-independent approach to screen for polypeptides with adhesive function at both the whole genome and metagenomic scales.

1.4.1 Principles of phage display

Filamentous phage display is a technique that enables a library of DNA fragments to be displayed as peptides or proteins on the surface of recombinant phagemid particles (PPs) through translational fusion to any of the phage coat proteins (pIII, pVI, pVII, pVIII, or pIX) (Gagic *et al.*, 2016) (Figure 1.8). The encoded population of variant DNA sequences can subsequently be screened for binding affinity against substrate(s) of interest in a process called "biopanning". Aside from filamentous phage, "tailed" phages (such as T7) can also be used for phage display (Gagic *et al.*, 2016). Recombinant fusion proteins are folded in the cytoplasm in tailed phage display systems, whereas in filamentous phage display, they are directed to the host's inner membrane and folded in the periplasm (a reducing environment that supports the formation of disulphide bridges between cysteine residues) prior to export to the cell surface (Gagic *et al.*, 2016). Therefore, tailed phage are better suited to the display of cytoplasmic proteins, and filamentous phages are more suited for displaying periplasmic and cell surface proteins.

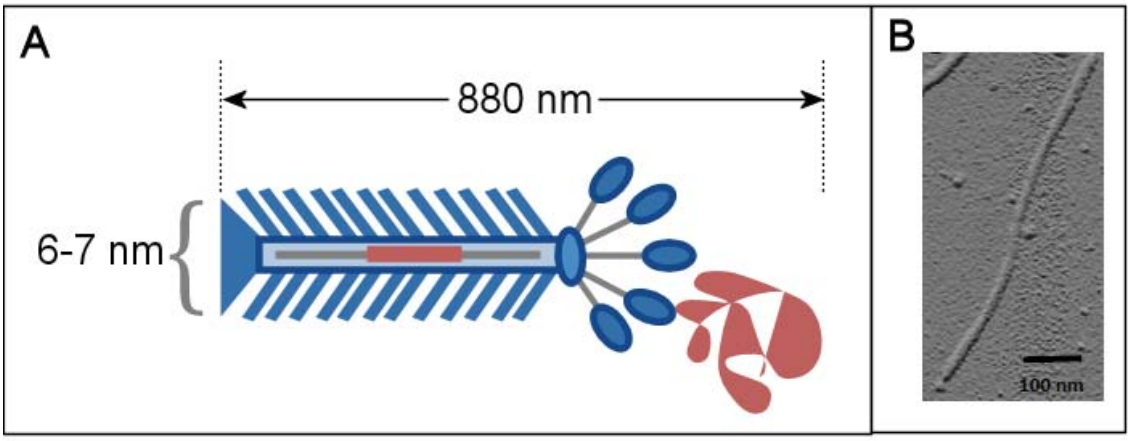


Figure 1.8. Structure of filamentous phage particle displaying recombinant pIII fusion protein.

pIII (blue oval) is present in 3-5 copies (Rakonjac *et al.*, 2011). In Panel A, pIII fusion protein (red shape), encoded by the recombinant DNA (red rectangle) packaged inside the phagemid particle (PP), is displayed on the surface of a virion. Together, minor coat proteins pVI and pIII cap one end of the PP, while pVII and pIX cap the other end. Around 2700 copies of major coat protein pVIII enclose the single stranded genome. Panel B shows an atomic force microscopy image of a filamentous phage (reproduced with permission of J. Rakonjac).

Within filamentous phage display systems, there are two most commonly used platforms: one based on phage vector and another that uses a phagemid vector and helper phage. Many variants of both have been used in a wide range of applications that are beyond the scope of this thesis (Qi *et al.*, 2012). This overview will focus on filamentous phage display through recombinant fusion to phage protein pIII in a phagemid system, as it was the platform used in the experimental method in this thesis and in all published reports of phage display used for identification of bacterial adhesins.

Phagemid vectors are an essential component of phagemid-based phage display systems. They are plasmids containing a filamentous phage (f1) origin of replication and a packaging signal, in addition to standard plasmid vector components (plasmid origin

of replication and antibiotic resistance marker). The f1 origin of replication and packaging signal allow replication and packaging of phagemid DNA into PPs upon infection with a helper phage. In addition, the phagemid vectors used in phage display have an expression or "display" cassette, composed of a promoter, ribosome binding site and a sequence encoding phage protein pIII that contains a multiple cloning site, which facilitates the construction of recombinant translational fusion pIII proteins. The protein pIII is composed of three domains, N1, N2, and C (Figure 1.9). The domains N1 and N2 are required for host infection functions, through interaction with primary and secondary host receptors, F-pilus (N2) and periplasmic TolA protein (N1). The C domain of pIII is an essential structural component of PPs. It is required for PP stabilization and PP release from the host during the last step of virion assembly (Rakonjac *et al.*, 1999). Several variants of phagemid vectors can be used for the display of recombinant pIII fusion protein. Inserts may be cloned upstream of the gene encoding pIII (Figure 1.9; Variant 1) or sandwiched between the coding sequence that correspond to N2 and C domains of pIII (Figure 1.9; Variant 2) (Smith, 1985, Gupta *et al.*, 2013). Small inserts can likely be cloned in the region between N1 and N2 of pIII in the phagemid system, as Tjhung *et al. (*2015) has demonstrated that intra-domain fusion proteins were successfully generated in a closely-related phage vector platform (Tjhung *et al.*, 2015) (Figure 1.9; Variant 3). In addition to translation fusions with full length pIII, phagemid vectors that only contain the C domain have also been widely used (Figure 1.9; Variant 4), because the presence of N1 and N2 domains may cause changes to the membrane of the *E. coli* host that interfere with helper phage infection, which would ultimately have a detrimental effect on PP production (Boeke *et al.*, 1982, Scott & Barbas, 2001, Gagic *et al.*, 2013).

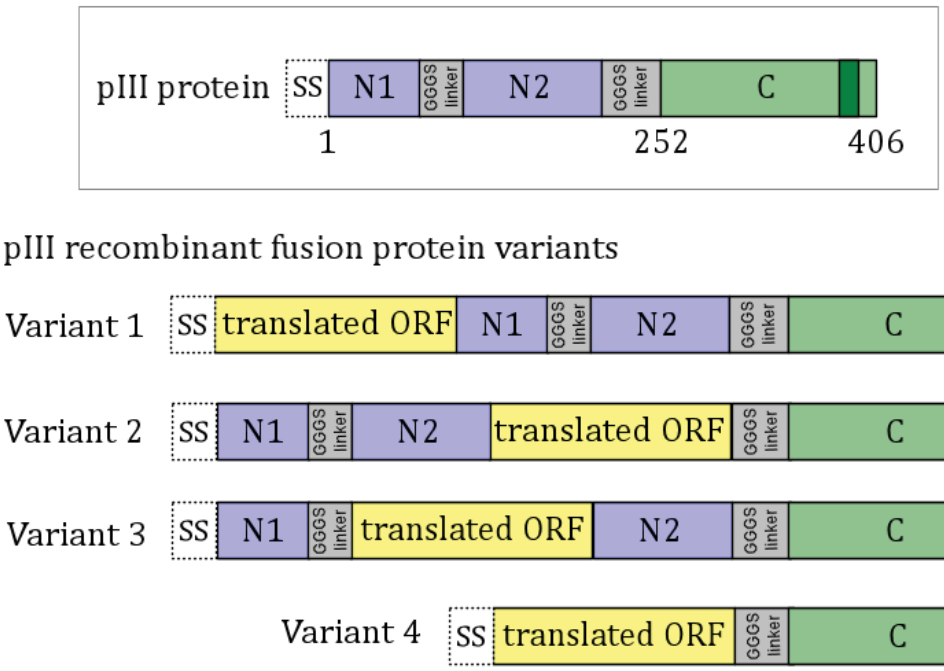


Figure 1.9. pIII protein domain structure and fusion protein variants used in phage display.

Phage pIII protein is composed of signal sequence (SS), that is absent from the final (mature) protein, N1 and N2 domains (yellow) that are involved in interactions with host receptors, and C region (light green) that is required for PP stabilization and release from the host. A membrane anchor (dark green) is present close to the C-terminus of the protein. Glycine-rich linkers (repeats of GGGS) are found in between N1 and N2, and N2 and C regions (Rakonjac *et al.*, 1999).

Monovalent and polyvalent display of recombinant pIII proteins can be achieved by using a wild-type helper phage or a helper phage deficient in pIII, respectively. Given that the phagemid vectors encode a recombinant fusion of a translated inserted coding sequence with phage protein pIII, phagemid-derived recombinant pIII and helper phage-derived wild-type pIII are commonly co-assembled into the same PP. Infection with wild-type helper phage results in monovalent display of recombinant pIII fusion proteins (Figure 1.10), as availability of helper phagemid-derived pIII tends to be lower than the availability of phage-derived pIII (Gagic *et al.*, 2016). To achieve polyvalent display, several helper phage variants including VCSM13d3 (a helper phage deficient in

pIII) and hyperphage (a helper phage harbouring truncated pIII) have been constructed (Rakonjac *et al.*, 1997, Rondot *et al.*, 2001). As pIII is not encoded by the VCSM13d3 genome and this protein is essential in PP assembly, all copies of pIII found on the PP are phagemid-derived. Monovalent display enables the selection of proteins with higher affinity for the bait/target during affinity screening. In contrast, polyvalent display enables the selection of proteins with lower affinity to the bait/target, as the displayed protein may be associated with the bait more strongly through increased avidity.

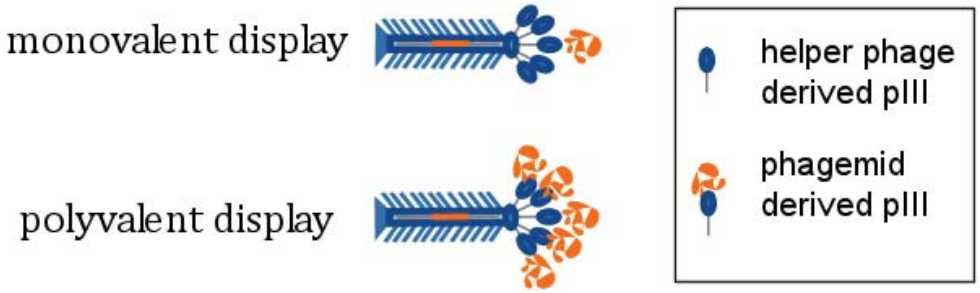


Figure 1.10. Monovalent and polyvalent phage display.

1.4.2 Biopanning

Once a library of PPs has been established, this library can be screened to select and enrich for displayed polypeptides which have binding affinity to the substrate of interest in a process known as "biopanning" (Rakonjac *et al.*, 2011). Polypeptides of interest are typically rare in the phage display library; therefore, successive rounds of enrichment by biopanning may be required for their identification. During the selection procedure, PPs are exposed to the substrate of interest ("bait") (Figure 1.11). PPs displaying polypeptides with affinity for the substrate of interest will remain bound, whereas non-binding PPs are removed during washing steps. The substrate-bound PPs are eluted, and subsequently used to infect appropriate host cells for amplification. Through several rounds of selection and enrichment, PPs displaying polypeptides with the desired binding property are enriched.

Conditions for affinity binding and elution can be modified according to the needs of a given experiment. For example, selection rounds can be performed with increasing wash stringencies to select for PPs with high affinity to the substrate (Lunder *et al.*, 2008). After the non-binding PPs have been removed, PPs can be eluted by competitive or non-specific elution (Lunder *et al.*, 2008). In competitive elution, a high affinity ligand is added to the PP-bound bait to release the PPs. For instance, appropriate *E. coli* host strain culture can be used to elute PPs displaying recombinant fusion pIII proteins, as pIII has a high affinity to *E. coli* receptors. Non-specific elution, on the other hand, encompass physical methods (e.g. sonication) and pH changes (e.g. adjusting the PP-bait mixture to a pH of 2.2) to release PPs from bait. The optimal method varies depending on the nature of bait-ligand interactions.

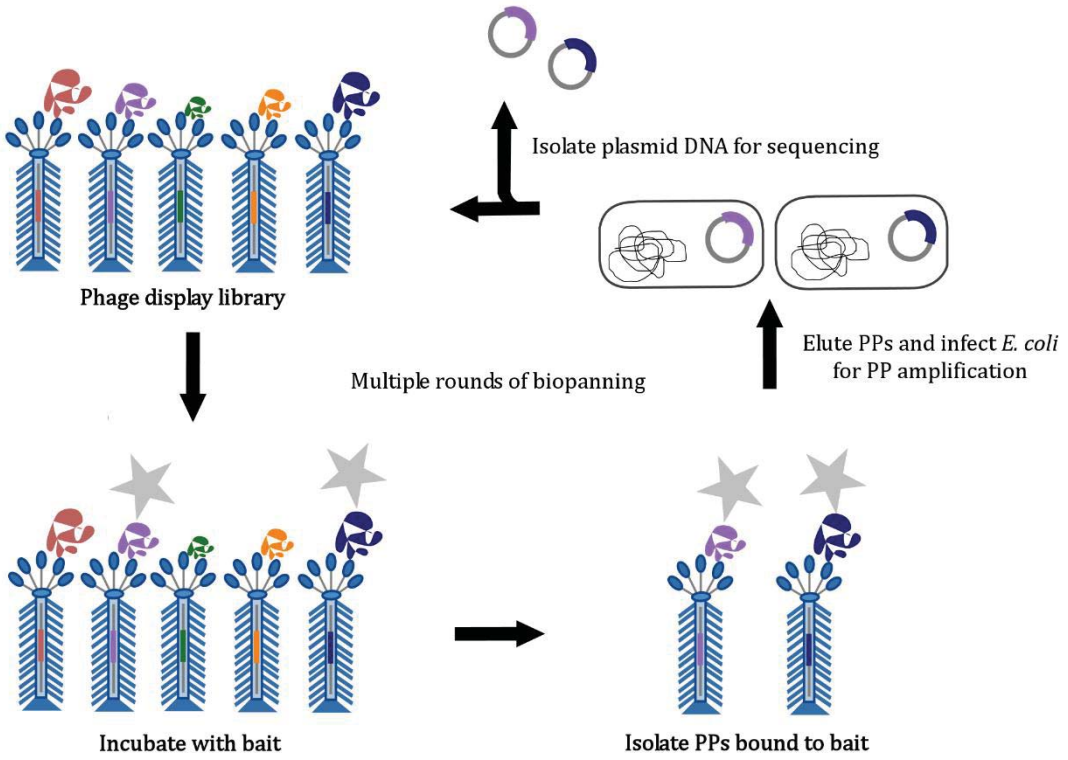


Figure 1.11. Biopanning procedure for affinity selection of phagemid particles with binding affinity to substrate of interest.

The substrate of interest is represented by grey stars. Recombinant proteins displayed as fusions to phage protein pIII are represented by multi-coloured shapes.

1.4.3 Phage display at a microbial community scale

Although phage display has been proposed as a culture-independent method to explore metagenomic functional potential more than 10 years ago (Cowan *et al.*, 2005), there has only been one study reporting the successful application of the technology in metagenomic functional screening using the T7 phage system. In a study by Zhang *et al.* (2009), proteins involved in natural product synthesis were identified from a metagenomic DNA sample isolated from soil-dwelling microbes (Zhang *et al.*, 2009). The T7 phage display platform was used to identify acyl carrier proteins and peptidyl carrier proteins that were involved in shuttling the reaction intermediates during natural product biosynthesis, by screening a metagenomic phage display library constructed

using DNA isolated from soil. The library of PPs was subjected to enzymatic modification by phosphopantetheinyl transferase, which catalyzes the specific covalent linkage of the conserved serine residue in acyl and peptidyl carrier proteins with biotin-labelled CoA, thereby biotinylating the phage of interest. The biotinylated phage were then selected by incubation with immobilized streptavidin. After six rounds of biopanning, Sanger sequencing was performed on selected clones, and 11 acyl or peptidyl carrier proteins were identified from a primary library of 10^8 clones.

1.5 Next generation sequencing

Next generation sequencing technology is frequently used in taxonomic analyses of microbial communities and elucidation of their functional potential (Franzosa *et al.*, 2015). As a large amount of data can be generated from a next generation sequencing run, it offers significant advantages over previous methods used for microbial community analysis. More recently, researchers have also applied this technology to identify enriched genes contained in phage display libraries after affinity selection ('t Hoen *et al.*, 2012, Matochko & Derda, 2015).

Currently, 454 pyrosequencing, Illumina, SOLID, and Ion Torrent sequencing are technologies commonly used for next generation sequencing. Each strategy has its strengths and shortcomings (Table 1.9). PacBio single molecule real time sequencing is becoming more frequently used as well, due to the long reads that can be generated using this technology (Rhoads & Au, 2015). 454 pyrosequencing and PacBio single molecule real time sequencing will be described in greater detail in the following sections, as these technologies were used to generate data in this thesis.

Table 1.9. Comparison of next generation sequencing technologies commonly in use.

Technology	454 pyrosequencing	Illumina HiSeq	SOLiD	Ion Torrent	PacBio RS II
Number of reads	1×10^6	up to 8×10^9	8×10^8	8×10^7	up to 7×10^4
Read length (# bases)	700	up to 250	60	200	10,000
Sequencing approach	Sequencing by synthesis (DNA polymerase)	Sequencing by synthesis (DNA polymerase)	Sequencing by ligation	Sequencing by synthesis (DNA polymerase)	Sequencing by synthesis (Φ29 polymerase)
Amplification procedure used to generate detectable signal	Emulsion PCR	Bridge PCR	Emulsion PCR	Emulsion PCR	None required; single molecule technology used
Detection method	Optical detection of pyrophosphate released by enzymatic coupling to luciferin oxidation	Optical detection of reversible terminator nucleotides labelled with fluorophores	Optical detection of fluorophore labels on 8-mer oligonucleotides	Semiconductor detection of pH changes caused by H^+ released during polymerase reaction	Optical detection of phospho-linked fluorophore on nucleotide added by polymerase
Error rate for single pass read (%)	1	0.1	5	1	13
Run time	23 h	7 h to 6 days	6 days	2-4 h	0.5-4 h
Cost (per million bases)	$8.57	$0.04	$0.11	$0.10	$0.40-0.80
Advantages/ disadvatanges	Poor detection of homopolymer and expensive Long read lengths compared to other 2nd generation sequencing technologies	Good for detection homopolymers, but read lengths limited due to off-phasing effects	No unique advantage over other systems	Fast run-time Short reads due to phasing effects during homopolymer runs	No need for clonal amplification of DNA Long reads, but fewer number of reads

Note: Data compiled from Mardis (2008), Rothberg *et al.* (2011), Myllykangas *et al.* (2012), Rhoads & Au (2015).

1.5.1 454 pyrosequencing

Pyrosequencing is the next generation sequencing technique that was most often used to characterize microbial communities in rumen microbiology and medical microbiology in the period between 2007 and 2015 (Siqueira Jé *et al.*, 2012, McCann *et al.*, 2014). It is now being replaced by other methods, such as Illumina sequencing (lower cost) and PacBio sequencing (longer read) (Table 1.9). The average read length varies depending on the pyrosequencing platform used. Of the pyrosequencing platforms that are available, GS FLX+ produces the maximum read length, at an average of 700 bases (Liu *et al.*, 2012).

The DNA sample to be sequenced must contain fragments that are of suitable length. When the samples are (meta)genomic DNA or long PCR amplicons, they are sheared to produce fragments of the required size before sequencing adapters are ligated to the ends of the fragments. Alternatively, for PCR amplicons that are within the size limit, adapters can be introduced during PCR. Adapter-ligated fragments are then hybridized to beads *via* a complementary primer. Each bead binds to one fragment, and they are emulsified in water-oil droplets that contain PCR reagents in order to amplify each fragment. The amplification step is performed to ensure that the signals generated during sequencing reach a detectable intensity. The suspension is transferred to a picotitre plate with wells that can only accommodate a single bead (Myllykangas *et al.*, 2012).

During sequencing, the DNA is denatured and the single-stranded fragments are sequenced by synthesis. Each dNTP is applied sequentially. If a nucleotide complementary to the sequence is applied, the polymerase can extend the PCR product, and this reaction would result in the production of a pyrophosphate molecule. A detectable signal can be generated from pyrophosphate production by enzymatic

coupling of ATP sulfurylase and luciferase. ATP sulfurylase converts pyrophosphate into ATP, and then the ATP is used by the enzyme luciferase to generate a light signal which is captured by a camera. Unused nucleotides are removed by the enzyme apyrase, and then the cycle begins again. The template sequence is complementary to the nucleotides added which allows the identity of the template sequence to be deciphered easily (Ronaghi, 2001).

1.5.2 PacBio single molecule sequencing

Similar to other next generation sequencing methods, the DNA used in single molecule sequencing must be within a suitable size range prior to sequencing; however, this technology can accommodate longer fragment sizes as it can produce longer reads than other next generation sequencing methods (average of 2000 bases). Adapters are ligated to the ends of fragments to form circularized molecules for sequencing (Travers *et al.*, 2010). In single molecule sequencing technology, the generation of clonal populations of DNA by PCR amplification is not required as detectors are sensitive enough to sense signals emitted from single molecules. Sequencing is executed by synthesis. Single DNA polymerase molecules are immobilized to zero mode wave guides (a sequencing unit that holds a very small volume for light detection), which enables the detection of signals within zeptolitre (10^{-21}) volumes (Eid *et al.*, 2009). Each of the four dNTPs is labelled with a different fluorophore; therefore, the newly incorporated nucleotide would emit a characteristic signal that can be detected by the laser and sensors situated below the zero mode waveguide. As the template is a circularized molecule, it can be read by the polymerase multiple times, thus enhancing the accuracy of reads.

The performance of the PacBio single molecule sequencing platform has been compared against the 454 pyrosequencing platform in 16S rRNA gene sequencing (Mosher *et al.*,

2014). A test run on DNA extracted from a pure culture of *Shewanella oneidensis* revealed that a greater number of reads were assigned to the genus *Shewanella* (using a 97% cut-off against reference sequences on the NCBI database) using the PacBio RS II system with P4/C2 chemistry (99.7%) than pyrosequencing (98.1%) as well as an older version of the PacBio system with XL/C2 chemistry (~82%). The composition of PacBio sequencing reagents is proprietary; therefore, different generations are denoted by P*x*/C*y*, where *x* denotes the polymerase version released and *y* indicates the chemistry version release (XL/C2 is an older version that existed prior to the current nomenclature). On the other hand, when soil metagenome samples were tested on each system, it was apparent that pyrosequencing provided the greatest coverage (~74%) compared to PacBio sequencing with either the P4/C2 or XL/C2 chemistry (approximately 50% and 35%, respectively). In contrast with 454 pyrosequencing, errors introduced during sequencing are randomly distributed in PacBio single molecule sequencing.

1.5.3 Application of next generation sequencing in phylogenetic analyses

Prior to the advent of next generation sequencing technology, methods used for microbial community characterization included: cultivation and phenotypic characterization, microscopic identification, denaturing gradient gel electrophoresis (DGGE), and Sanger sequencing of 16S rRNA gene clone libraries. Cultivation, phenotypic characterization and microscopic identification are time-consuming, and it can be difficult to resolve different prokaryotic species by microscopy unless fluorescence *in situ* hybridization (FISH) is used to probe the 16S rRNA present in the cell. FISH is also time consuming. Furthermore, sequence information must already exist for the species of interest, and optimization steps are often necessary [for example, cell

wall permeabilization is a challenge for identification of rumen methanogens by FISH (Valle *et al.*, 2015)]. DGGE involves PCR amplification of the marker gene present in a microbial community, followed by the electrophoresis of the amplicons in an increasing gradient of denaturants (formamide and urea) to separate the amplicons based on their melting characteristics. DGGE acts as a fingerprinting method and is used to obtain the profile of a community's species composition and diversity (Prosdocimi *et al.*, 2015). This method is useful for comparing overall differences in communities, but individual gel bands must be excised for DNA extraction and sequencing if species identification is desired. Species identification and discovery is also achievable with Sanger sequencing of clone libraries created from PCR amplicons of marker genes. However, individual clones must be handpicked prior to sequencing, and the scale of the data set acquired is much smaller than that of next generation sequencing technology. Moreover, the large-scale data sets yielded by next generation sequencing methods enable low abundance species in complex communities to be identified (Lynch & Neufeld, 2015).

In metagenome sequencing, the proportion of sequence reads corresponding to genomes of rare species would be low, and a large volume of sequencing data must be generated to access these rare genomes. This issue persists even when Illumina HiSeq technology is used for sequencing (which generates the largest number of reads amongst current available technologies) (Gagic, 2015). DNA subtraction methods can, therefore, be applied to enrich for DNA derived from rare species in an ecosystem. For example, novel bacterial species that can degrade polycyclic aromatic hydrocarbons that are rare in the seawater environment were identified by DNA stable isotope labelling followed by pyrosequencing (Sauret *et al.*, 2014). Seawater microcosms that contained a low abundance of polycyclic aromatic hydrocarbon degraders were supplemented with ^{13}C-labelled phenanthrene, which results in the incorporation of ^{13}C isotopes into the

genomic DNA of phenanthrene-using bacteria. Metagenomic DNA was isolated from the bacterial community, and then ^{12}C- and ^{13}C-containing DNA were fractionated by ultracentrifugation in a CsCl gradient. DNA fractions containing ^{13}C radiolabel were used as template for PCR amplification of 16S rRNA gene sequences and then pyrosequencing was performed for species identification. When dealing with microorganisms that cannot be maintained in culture, a different strategy would be required. Gagic *et al* (2015) have demonstrated that rare sequences in a metagenomic DNA sample can be enriched using a method based on the likelihood that abundant sequences are more quickly reannealed than rare sequences after thermal denaturation (Gagic *et al.*, 2015). The reannealed double-stranded DNA corresponding to abundant microorganisms is then removed using a duplex-specific nuclease, whereas single-stranded DNA is used for subsequent rounds of enrichment or for metagenome sequencing.

As costs for deep sequencing decrease and the technology improves, it is becoming more feasible to perform whole metagenome sequencing for the purpose of community profiling (Franzosa *et al.*, 2015). PCR amplification of marker genes is not required in shot-gun metagenome sequencing; therefore, compared to methods where PCR amplicons of marker genes are sequenced, shot-gun sequencing without PCR amplification reduces amplification bias in the dataset. Furthermore, rather than simply observing changes in the species composition of the community, the metagenome dataset can simultaneously provide taxonomy and gene function information that could be useful for elucidating the functional relationships between members of the community (Franzosa *et al.*, 2015). As gene function is assigned by homology search, it is important to note that the quality of gene function assignment in metagenome

sequencing depends heavily on the representation of the species of interest in available databases (Oulas *et al.,* 2015).

1.5.4 Application of next generation sequencing in phage display

The power of next generation sequencing has also been coupled to metagenomic phage display, as there are significant advantages to combining these technologies. In standard phage display technology, the PPs displaying polypeptides that can bind to the substrate of interest may be rare in the starting material; therefore, multiple rounds of biopanning are typically carried out to enrich for binders, and then the identity of selected clones is determined by Sanger sequencing. In contrast, the large amount of data generated by next generation sequencing methods could make it possible to identify rare sequences of interest in a population and to quantify their relative abundances before and after one or two rounds of affinity screening.

A study by 't Hoen et al. (2012) demonstrated that osteoblast binding peptides could be identified from a random peptide library after just one round of biopanning ('t Hoen et al., 2012). This improvement is important to note, as enrichment of fast growing clones have been noted in phage display experiments in the past (Derda et al., 2011, Vodnik et al., 2011); therefore, by executing fewer biopanning rounds, the chances of selecting for fast growing clones in the population from the phage amplification step that precedes biopanning would decrease, and the amount of time required to perform the affinity selection experiment would decrease as well. In addition, using a large data set, it would be possible to identify more candidate binders.

Another example that demonstrates the utility of combining phage display and next generation sequencing technologies is a study where secretome phage display was performed to characterize the secreted gene products within the fibre-adherent microbial community from the bovine rumen, which is expected to contain diverse

secreted fibrolytic enzymes (Ciric *et al.*, 2014). In this variant of standard phage display technology, phagemids harbouring inserts that carry secretion signals are enriched to generate a secretome phage display library. Biopanning is not part of the experimental pipeline, but the enriched library can be screened for specific binding proteins if required. As the secretome phage display library is composed of a high diversity of insert DNA, Ciric *et al.* analyzed the library by pyrosequencing, rather than Sanger sequencing of individual clones, in order to explore the gene functions represented in the secretome of the fibre-adherent rumen microbial community to a greater depth (Ciric *et al.*, 2014). Components of cellulosomes, which are secreted multi-enzyme subunits that enable microbes to degrade cellulosic and hemi-cellulosic substrates found in plant material, were reported to be enriched in the secretome (Ciric *et al.*, 2014). Specifically, cohesin and dockerin modules, which provide a physical linkage between catalytic components in the cellulosome and the cell surface, were more abundant in the metasecretome dataset.

1.6 Project aims

The relationships between rumen protozoa and methanogenic archaea have been studied previously, however the molecular mediators involved in initiating and facilitating these interactions have not yet been discovered prior to the commencement of this thesis work. I hypothesize that the methanogen genomes encode protozoa-binding adhesins, and that phage display technology can be used to identify these archaeal adhesins.

Objectives:

(1) The first aim of this project is to demonstrate that phage display can be used to identify archaeal adhesins. I created a shotgun phage display library from a single methanogen species, *Methanobrevibacter ruminantium* M1, and screened this library for protozoa-binding adhesins. Domain mapping was performed on the identified protein to determine regions of the protein that are important for binding to protozoal hosts. Furthermore, a "reverse panning" method was established to examine the host tropism of this protein.

(2) Rumen protozoa harbour various methanogen and bacterial species as symbionts. These other protozoa-associated symbionts may also harbour ALPs that bind to protozoal hosts; therefore, I created a metagenomic shotgun phage display library derived from a protozoa-associated symbiont community to identify protozoa-binding adhesins from the multi-species environmental sample. This library was also screened on rumen-derived protozoa as bait. Several novel putative adhesins were identified using this approach.

In the course of this study, a method was established to enrich for protozoa-associated methanogens. I assessed the symbiont species present in the metagenomic DNA isolated from the enriched sample by next generation sequencing of 16S rRNA gene sequences,

and then compared the acquired data with the microbial species composition of rumen contents (without enrichment) to provide a snapshot of the rumen protozoa-associated symbiont community.

Chapter 2. Material and Methods

2.1 Chemicals and enzymes

Restriction endonucleases and ligase were obtained from Roche Molecular Biochemicals (Germany) or New England Biolabs Inc. (MA, USA). Unless otherwise stated, Platinum Taq polymerase (Life Technologies, CA, USA) was used to generate PCR amplicons. Difco™ 2×YT was purchased from Fort Richards Laboratories (NZ).

2.2 Oligonucleotides

Oligonucleotide primers used for cloning, sequencing, and PCR reactions are listed in Table 2.1. They were synthesized by Integrated DNA Technologies (Singapore).

Table 2.1. Oligonucleotide primers used in this thesis.

Primer name	Sequence (5´→3´)	Application	Reference
Ar915aF	AGG AAT TGG CGG GGG AGC AC	qPCR, sequencing	Watanabe *et al.*, 2004
Ar1386R	GCG GTG TGT GCA AGG AGC	qPCR, sequencing	Skillman *et al.*, 2006
Ba519F	CAGCMGCCGCGGTAANWC	qPCR	Kittelmann & Janssen, 2011
Ba907R	CCG TCA ATT CMT TTR AGT TT	qPCR	Kittelmann & Janssen, 2011
Ba9F	GAG TTT GAT CMT GGC TCA G	Sequencing	Kittelmann *et al.*, 2013
Ba515Rmod1	CCG CGG CKG CTG GCA C	Sequencing	Kittelmann *et al.*, 2013
RP841F	(AA) GAC TAG GGA TTG GAG TGG	Clone library, sequencing	Kittelmann *et al.*, 2013
Reg1302R	(TC) AAT TGC AAA GAT CTA TCC C	Clone library, sequencing	Kittelmann *et al.*, 2013
Syl316F	GCT TTC GWT GGT AGT GTA TT	qPCR, DGGE	Sylvester *et al.*, 2005
Syl539R	CTT GCC CTC YAA TCG TWC T	qPCR	Sylvester *et al.*, 2005
Syl539R-GC	CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGGG ACT TGC CCT CYA ATC GTW CT	DGGE	Sylvester *et al.*, 2005
pspF03	ATG TTG CTG TTG ATT CTT CA	Sequencing	Gagic *et al.*, 2013
pspR03	TGC CTT TAG CGT CAG ACT GTA GC	Sequencing	Gagic *et al.*, 2013
PelBF1	TGAAATACCTGCTGCCGACC	Sequencing	This study
mru1499_SphI-F40	ccgccggcatgcTTAGTAGCAAACTCTGGAGATGATTC	Cloning	This study
mru1499_EcoRI-R197	ccgccggaattcCAAAGCTACAGATAGCTTTGATGAAGC	Cloning	This study
mru1499_SphI-F195	cggccggcatgcGTAGCTTTGACAGTTAATTTAATGTCC AC	Cloning	This study
mru1499_EcoRI-R283	ccgccggaattcCTTGGAAGAGAGATAATTTGCAGAGC	Cloning	This study
mru1499_SphI-F286	ccgccggcatgcGCAACTGTTACTGTTCAAAAAGGAG	Cloning	This study
mru1499_EcoRI-R541	ccgcgggaattcCTTGAATACAGAACTGTCAAAGCTGC	Cloning	This study

2.3 Preparation of anaerobic solutions and media

2.3.1 Anaerobic salts solution

Anaerobic salts solution (500 mL) contained 85 mL of Salt Solution A [6 g/L NaCl, 3 g/L KH_2PO_4, 1.5 g/L $(NH_4)_2SO_4$, 0.79 g/L $CaCl_2 \cdot 2H_2O$, 1.2 g/L $MgSO_4 \cdot 7H_2O$], 85 mL of Salt Solution B (6 g/L K_2HPO_4), 330 mL of distilled water, and five drops of resazurin (0.1% w/v). The solution was heated to boiling in a microwave, then gassed with CO_2 for 20 min. $NaHCO_3$ (2.5 g) and cysteine-HCl (0.25 g) were added to the solution after cooling. The Schott bottle was sealed with a rubber stopper and autoclaved.

2.3.2 RM02 base

RM02 base (1 L) contained 1.4 g KH_2PO_4, 0.6 g $(NH_4)_2SO_4$, 1.5 g KCl, 1 mL of trace element solution SL10, 1 mL of selenite/tungstate solution, four drops of 0.1% (w/v) resazurin solution, and 950 mL distilled water. The solution was heated to boiling in a microwave, then gassed with CO_2 for 20 min. $NaHCO_3$ (4.2 g) and cysteine-HCl (0.5 g) were added to the solution after cooling. The medium was dispensed into the appropriate vessels under a stream of CO_2, sealed with rubber stoppers, then autoclaved. Clarified rumen fluid and supplement solution were added before use.

2.3.3 Clarified rumen fluid

Rumen contents collected from a fistulated cow was filtered through a layer of cheesecloth, and then centrifuged at 10,000 × *g* for 20 min. The supernatant centrifuged again under the same conditions, and stored at -20 °C until required. After thawing, the rumen liquor was centrifuged at 5,500 × *g* for 20 min, bubbled with N_2 for 20 min, sealed with rubber stopper, and autoclaved to de-activate bacteria and phage. Magnesium chloride and calcium chloride were added to the autoclaved rumen liquor at 1.63 g/100 mL and 1.18 g/100 mL, respectively. The mixture was centrifuged at 5,500 × *g* for 20 min at 4 °C to sediment precipitates. Yeast extract was added to the

supernatant at 2% (w/v) final concentration, bubbled with N_2 for 20 min, and then filter sterilized into N_2-filled bottles. Vitamin 10 concentrate [4-aminobenzoate (40 mg/L), D-(+)-biotin (10 mg/L), nicotinic acid (100 mg/L), hemicalcium D-(+)-pantothenate (50 mg/L), pyridoxamine hydrochloride (150 mg/L), thiamine chloride hydrochloride (100 mg/L), cyanocobalamin (50 mg/L), D,L-6,8-thioctic acid (30 mg/L), riboflavin (30 mg/L), folic acid (10 mg/L)] was added at a concentration of 2 mL/100 mL of yeast-supplemented rumen fluid.

2.3.4 Formate, acetate and methanol solution

A solution containing 3 M sodium formate, 1 M sodium acetate, and 1 M methanol was bubbled with N_2 gas for 20 min, then filter sterilized into sterile N_2-filled serum bottles through 0.22 µm cellulose acetate filters (Millipore, Ireland).

2.3.5 RM02 nosubRFV medium

RM02 nosubRFV medium contained 9 mL of RM02 base, 0.5 mL of clarified rumen fluid, and 0.2 mL of formate-acetate-methanol solution. When required, larger volumes were scaled up accordingly.

2.4 Bacteria, methanogen, and helper phage strains

Escherichia coli TG1 [(F' *tra*D36 *pro*AB *lac*I^{q}Z ΔM15) *sup*E *thi-1* Δ *(lac-proAB)* Δ *(mcrB-hsdSM)5 (r_{K^-} m_{K^-})*] (Lucigen, WI, USA) was used to construct the phage display libraries, and for propagation of helper phage and phagemid particles. Cells were cultured in yeast extract tryptone broth (2×YT) at 37 °C with aeration. Chloramphenicol (20 µg/mL) was added to media as required. Wild-type helper phage VCSM13 (Stratagene, CA, USA) was propagated in *E. coli* TG1. Standard methods were used for phage propagation as described in Section 2.6.1.

Escherichia coli One Shot TOP 10 [F- *mcr*A Δ *(mrr-hsd*RMS-*mcr*BC) Φ80*lac*Z ΔM15 Δl*acX*74 *rec*A1 *ara*D139] (Life Technologies, Carlsbad, USA) was used to construct clone libraries for 18S rRNA gene sequences. Cells were cultured in Luria Bertani (LB) broth at 37 °C with aeration. Ampicillin (60 µg/mL) was added to media as required.

Methanobrevibacter ruminantium M1 was cultivated in RM02 medium with clarified rumen fluid, formate, and acetate under anaerobic conditions, as described previously (Leahy *et al.*, 2010). Starter cultures (10 mL) were incubated for 5 days at 39 °C, and then 5 mL of the culture were transferred into 50 mL of fresh medium in serum bottles. After incubation for 5 days at 39 °C, 50 mL of this culture were transferred into 500 mL of fresh medium in pressure-resistant Schott bottles (Duran, Germany), in order to cultivate this organism on large scale. To obtain a minimum of 100 µg genomic DNA, 1.5 L of M1 culture was harvested by centrifugation (5000 × *g*, 20 min, 4 °C).

2.5 Molecular biology methods

2.5.1 DNA isolation

Cell pellets (*M. ruminantium* M1 or protozoa-associated methanogen community) were frozen in liquid nitrogen and ground into a fine powder with a sterile mortar and pestle. For the pellet containing protozoa-associated methanogens, the pellet was heated to 70 °C for 10 min to de-activate any remaining deoxyribonucleases before freeze-grinding. DNA was then extracted using QIAGEN genomic tip 500/G according to the manufacturer's recommendations (Qiagen, Germany).

2.5.2 Plasmid DNA isolation

Cell pellets obtained from *E. coli* cultures (5 mL) were obtained from centrifugation at 12,000 × *g* for 5 min at room temperature. Plasmid DNA was extracted using QIAGEN plasmid DNA isolation kit according to the manufacturer's recommendations. For larger

scale plasmid preparations (from 20-50 mL of *E. coli* culture), PureLink® HiPure Plasmid Midiprep Kit (Life Technologies, CA, USA) was used according to the manufacturer's instructions.

2.5.3 Colony PCR

A PCR master mix was prepared for the appropriate primer pair, and 25 µL of the master mix was dispensed into 0.2 mL PCR tubes. For amplification of 18S rRNA gene sequences, the reaction mix contained 1× PCR buffer, primer pair RP841F/Reg1302R (0.5 µM for each primer), 1.5 mM $MgCl_2$, 0.02% bovine serum albumin (BSA) (w/v), and 0.15 U/µL of Taq polymerase. PCR cycling parameters used for this reaction are described in Table 2.2. For amplification of phage display library inserts, the reaction mix contained 1× PCR buffer, 2.5 mM $MgCl_2$, primer pair pspF03/pspR03 (0.2 µM for each primer), and 0.02 U/µL of Taq polymerase. PCR cycling parameters used for this reaction are described in Table 2.3. For each PCR reaction, a random colony was picked up using a sterile pipette tip, spot inoculated onto a Petri plate containing the appropriate medium and antibiotics, and then the remaining bacterial cells were transferred into the PCR reaction mix by pipetting up and down. PCR amplicons were generated using a Mastercycle Pro thermocycler (Eppendorf, Germany).

Table 2.2. Thermal profile for PCR amplification of partial 18S rRNA gene sequences.

Initial denaturation:	94 °C for 3 min
Amplification (35 cycles):	94 °C for 30 s 54 °C for 45 s 72 °C for 1 min
Final extension:	72 °C for 7 min

Table 2.3. Thermal profile for PCR amplification of phage display library inserts.

Initial denaturation:	95 °C for 2 min
Amplification (30 cycles):	95 °C for 15 s 59 °C for 30 s 72 °C for 1 min
Final extension:	72 °C for 7 min

2.5.4 Preparation of PCR amplicons for sequencing

2.5.4.1 Sanger sequencing of the 16S archaeal gene from *M. ruminantium* M1

Archaeal 16S rRNA gene sequences were amplified by PCR using the primer pair Ar915F/Ar1386R (1 µM for each primer) in a reaction mix containing 1.5 mM $MgCl_2$, 0.02% BSA (w/v), and 0.05 U/µL of Taq polymerase. PCR cycling parameters used for this reaction is described in Table 2.4. Sequence analysis was performed by Massey Genome Service (Palmerston North, New Zealand).

Table 2.4. Thermal profile for PCR amplification of archaeal 16S rRNA gene sequences.

Initial denaturation:	95 °C for 2 min
Amplification (30 cycles):	95 °C for 15 s 59 °C for 30 s 72 °C for 1 min
Final extension:	72 °C for 7 min

2.5.4.2 Sanger sequencing of the partial 18S rRNA gene sequence from rumen protozoa

Protozoal 18S rRNA gene sequences were amplified by PCR using the primer pair RP841F/Reg1302R (0.5 µM for each primer) in a reaction mix containing 1.5 mM $MgCl_2$, 0.02% BSA (w/v), and 0.15 U/µL of Taq polymerase. PCR cycling parameters used for this reaction is described in Table 2.2. Sequence analysis was performed by Macrogen Inc. (Seoul, Korea).

2.5.4.3 Sanger sequencing of phage display inserts

Insert sequences present in the phage display vector pYW01 were amplified by PCR using the primer pair pspF03/pspR03 (0.2 µM for each primer) in a reaction mix containing 2.5 mM $MgCl_2$ and 0.02 U/µL of Taq polymerase. PCR cycling parameters used for this reaction are described in Table 2.3. Sequence analysis was performed either by Massey Genome Service (Palmerston North, NZ) or Macrogen (Seoul, Korea).

2.5.4.4 16S and 18S rRNA gene sequencing by pyrosequencing

Archaeal 16S rRNA gene sequences were amplified by PCR using the primer pair Ar915F/Ar1386R (1 µM for each primer) in a reaction mix containing 1.5 mM $MgCl_2$, 0.02% BSA (w/v), and 0.05 U/µL of Taq polymerase. Bacterial 16S rRNA gene sequences were amplified by PCR using primer pair Ba9F/Ba515Rmod1 (0.1 µM for each primer) in a reaction mix containing 2.5 mM $MgCl_2$ and 0.02 U/µL of Taq polymerase. Protozoal 18S rRNA gene sequences were amplified by PCR using the primer pair RP841F/Reg1302R (0.5 µM for each primer; final concentration) in a reaction mix containing 1.5 mM $MgCl_2$, 0.02% BSA (w/v), 50 pg/µL of DNA, and 0.15 U/µL of Taq polymerase. Barcoded PCR reactions were performed as described previously (Kittelmann *et al.*, 2013).

2.5.4.5 PacBio sequencing of phage display inserts

Insert sequences present in the phage display vector pYW01 were amplified by PCR using the primer pair PelBF1/pspR03 (0.5 μM for each primer; final concentration) in a reaction mix containing 2.5 mM $MgCl_2$ and 0.02 U/μL of Taq polymerase. PCR cycling parameters used for this reaction is described in Table 2.3. For the libraries prior to biopanning against protozoal bait, PAM1 and PAM2, the PCR amplicons were electrophoresed on 1% (w/v) agarose gels. Fragments between 0.4-1.5 kb were excised and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega; Wisconsin, USA). For the PAM_FINAL library (after two rounds of biopanning against protozoal cells as bait), PCR amplicons were purified and size selected by using NucleoMag NGS Clean-up and Size Select magnetic beads (medi'Ray; Auckland, NZ). DNA samples were submitted to the Leibniz-Institut DSMZ (Germany) for sequencing.

2.5.5 Denaturing gradient gel electrophoresis (DGGE)

Protozoal 18S rRNA gene sequences were amplified by PCR using the primer pair Syl316F and Syl539-GC (1 μM for each primer) in a reaction mix containing 2 mM $MgCl_2$, 0.05% BSA (w/v), and 1 U/μL of Taq polymerase, as described previously (Sylvester *et al.*, 2005). PCR cycling parameters used for the reaction are detailed in Table 2.5. Denaturing gradient gel electrophoresis (DGGE) was performed on the amplicons in an 8% agarose gel with a gradient from 25% to 45% denaturants (where 100% denaturants represents 7 M urea and 40% formamide) at 60 V for 16 h at 60 °C in 0.5× TAE buffer (Sylvester *et al.*, 2005). The gel was stained with SYBRgold (Life Technologies) according to manufacturer's instructions.

Table 2.5. Thermal profile for PCR amplification of protozoal 18S rRNA gene sequences in preparation for DGGE.

First cycle:	94 °C for 4 min 56 °C for 30 s 72 °C for 1 min
35 cycles:	94 °C for 1 min 56 °C for 30 s 72 °C for 1 min
Final cycle:	94 °C for 4 min 56 °C for 30 s 72 °C for 30 min

2.6 Phage protocols

2.6.1 Phage propagation

An *E. coli* culture was inoculated into fresh 2×YT containing 20 µg/mL chloramphenicol (1 mL culture into 100 mL medium), and incubated at 37 °C with shaking at 200 rpm until reaching exponential phase [optical density at wavelength of 600 nm (OD_{600}) ~ 0.2]. The culture was then infected with helper phage VCSM13 [multiplicity of infection (m.o.i.) of 50, i.e. 50 phage per bacterium], and incubated at 37 °C for 1 h without shaking. *E. coli* cells were pelleted by centrifugation (3200 × *g*, 10 min, room temperature), and resuspended in fresh 2×YT containing 20 µg/mL chloramphenicol. These cultures were incubated for 6 to 8 h at 37 °C, and *E. coli* cells were pelleted by

centrifugation at 10,000 × *g* for 20 min at 4 °C. The culture supernatant was filtered through 0.2 µm PVDF membranes (Millipore), and phagemid particles were precipitated in a solution containing 5% (w/v) PEG and 0.5 M NaCl (final concentrations) on ice for 1-2 h. The precipitated phagemid particles were pelleted by centrifugation at 10,000 × *g* for 30 min at 4 °C, and resuspended in 10 mM Tris-HCl pH 7.5.

2.6.2 Quantification

Helper phage preparations were quantified by titration with *E. coli* TG1 cells. Spot titration assays were performed to estimate phage titres (Gagic *et al.*, 2013). Ten microliters of phage at various dilutions were spotted on 2×YT agar (20 mL) overlaid with *E. coli* TG1 culture [3 mL of 2×YT containing 0.6% Bacto-agar (Difco) and 100 µL of overnight culture] (Gagic *et al.*, 2013). After overnight incubation at 37 °C, the numbers of plaque forming units (pfu) were counted for each spot. For a more accurate measurement, the amount of phage estimated to give 100 to 300 pfu was used for a whole-plate titration assay. Plates containing 2×YT agar (20 mL) were overlaid with *E. coli* TG1 culture and helper phage (3 mL of 2×YT containing 0.6% Bacto-agar, 100 µL of overnight culture, and the appropriate amount of phage). After overnight incubation at 37 °C, plaque forming units were enumerated. The assay was performed in triplicate.

Phagemid particles were titrated in a similar way, but media containing antibiotics were used to select for infected bacterial cells. Plates containing 21 mL of 2×YT agar supplemented with 25 µg/mL of chloramphenicol were overlaid with 9 mL of 2×YT agar (without antibiotics) (Gagic *et al.*, 2013). For spot titrations, *E. coli* TG1 culture (3 mL of 2×YT containing 0.6% Bacto-agar and 100 µL of overnight culture) were plated, and then 10 µL of phagemid particles at various dilutions were spotted on the plate. For whole-plate titrations, 3 mL of 2×YT containing 0.6% Bacto-agar, 100 µL of overnight *E. coli* TG1 culture, and the appropriate volume of phage were mixed and plated. After

overnight incubation at 37 °C, colony forming units (cfu) were enumerated. Whole-plate titration assays were performed in triplicate, with 100 to 300 cfu counted per plate.

2.6.3 Western blotting

PPs (10^9 cfu) were incubated for 5 min in 1× sample buffer for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) at 100°C, and then cooled on ice for 5 min. The disassembled phagemid particles were electrophoresed by SDS-PAGE on a Tris-glycine gel system using 12% Mini-PROTEAN® TGX™ precast protein gels (Bio-Rad; CA, USA), alongside the SeeBlue Plus2 pre-stained standard (Life Technologies). Proteins were transferred to PVDF membranes using an iBlot dry blotting system (Life Technologies) at 20 V for 7 min. The composition of reagents used for western blotting are detailed in Table 2.6, and all incubations and washes of the membrane were performed on a shaker. The membrane was incubated in blocking buffer at 4°C overnight, washed three times for 10 min in Tris-buffered saline supplemented with 0.05% Tween-20 (TBS-T) at room temperature, and then hybridized in primary antibody for 1 h at room temperature. The membrane was washed three times for 10 min in TBS-T to remove excess primary antibody, incubated in secondary antibody for 1 h at room temperature, and then washed three times for 10 min in TBS-T. Clarity™ western ECL substrate (Bio-Rad) was applied to the membrane, according to the manufacturer's instructions. The ChemiDoc™ MP system (Bio-Rad) was used for detection of luminescent signals on the immunoblot and image collection.

Table 2.6. Western blot reagents used for detection of recombinant pIII fusion proteins.

Reagent	Composition
1×SDS-PAGE sample buffer	
TBS; Tris buffered saline	30 mM Tris, 150 mM NaCl, pH 7.6
TBS-T; Tris buffered saline with Tween 20	TBS containing 0.05% Tween 20
Blocking buffer	TBS-T containing 5% (w/v) skim milk powder
Primary antibody	c-myc antibody produced in rabbit (Sigma, MO, USA) was diluted in blocking buffer at a ratio of 1:2000
Secondary antibody	anti-rabbit IgG conjugated to horseradish peroxidase (Sigma) was diluted in blocking buffer at a ratio of 1:2000

2.7 Isolation of protozoal epibionts and enrichment for methanogens

Rumen contents were collected from fistulated pasture-fed (ryegrass and clover) sheep 2 h after feeding. The method for isolation of protozoa-associated methanogens was adapted from a previous study (Tymensen *et al.*, 2012b). Procedures were carried out at 39 °C, and exposure to oxygen was minimized as much as possible. Rumen contents were filtered twice through two layers of PETEX mesh (335 µm pore size; Sefar Inc., Switzerland). The filtrate was diluted in 0.5 volumes of anaerobic salts solution, transferred to a separation funnel, and then incubated for 1 h. During this time, protozoa

settled to the bottom, while plant particulates floated to the top. The protozoa-containing fraction was collected, and passed through NITEX mesh (11 μm pore size; Sefar Inc.) to retain protozoa.

To remove free-living bacteria and archaea, the retentate was transferred to a bag made of NITEX mesh. This bag was submerged in 0.5 L of anaerobic salts solution with gentle stirring for 15 min to allow free-living bacteria and archaea to filter out of the bag. This procedure was repeated four times. Finally, 40 mL of protozoa-containing retentate was collected and treated with lysozyme and mutanolysin (1 mg/mL and 15 U/mL, respectively) for 1 h at 39 °C to lyse protozoa-associated bacterial ectosymbionts. Intact protozoa with attached archaeal ectosymbionts were sedimented by centrifugation at 100 × *g* for 5 min. Free-living bacteria and archaea that remain would be present in the supernatant, which was discarded.

To separate protozoa from symbionts, four methods were tested. For proteinase treatment, a mixture of pronase (0.1 mg/mL), trypsin (0.17 mg/mL) and proteinase K (0.17 mg/mL) was added to the protozoa suspension. For NaCl treatment, the protozoa sample was incubated in anaerobic salts solution containing 1 M NaCl. For H_2 treatment, the protozoa sample was pressurized with H_2 to 180 kPa. All samples had a final volume of 10 mL and were incubated at 39 °C for 1 h in 15 mL Hungate tubes pre-flushed with CO_2, then samples was withdrawn for fixation in 2% paraformaldehyde (PFA) and microscopic examination (Leica DM2500). For freeze-thaw treatment, protozoa were lysed by freeze-thaw (freeze at -20 °C for at least 18 hours, thaw at 39 °C for 15 min next day). Any protozoa that remained intact were removed by centrifugation at 100 × *g* for 5 min.

In the freeze-thaw method for detaching ectosymbiotic methanogens from protozoa, supernatant containing symbiotic archaea and bacteria was treated with lysozyme

(10 mg/mL) and mutanolysin (15 U/mL) for 1 h at 39 °C to lyse bacterial endosymbionts released during protozoa lysis. Bacterial DNA released was then removed by DNase I (Roche, Germany) treatment (40 U/mL) for 30 min at 39 °C. Co-factors $MgCl_2$ (5 mM) and $CaCl_2$ (0.5 mM) were added to facilitate DNase I treatment. DNase I was deactivated by EDTA (final concentration of 20 mM). The suspension was centrifuged again at 100 × *g* for 5 min to remove any remaining intact protozoa. Finally, cells remaining in the methanogen-enriched fraction was pelleted by centrifugation at 6,000 × *g* for 15 min at 4 °C, and the pellet was stored at -80 °C for DNA extraction (Section 2.8.2).

2.7.1 Quantification of relative amounts of archaeal, bacterial, and protozoal DNA

Relative abundances of protozoal, bacterial, and archaeal genomic DNA present in metagenomic DNA samples isolated from protozoal epibionts were measured by real time quantitative PCR. Levels of protozoal and bacterial genomic DNA were quantified as adapted from Kittelmann and Janssen using a Rotor-Gene 6000 real-time rotary analyzer by SYBR Green fluorescence (Kittelmann & Janssen, 2011). Reaction volume was 10 µL, containing 1 µL of LightCycler FastStart pre-mix (Roche), 1 µL of template DNA, 0.02% (w/v) BSA, 2 mM $MgCl_2$ for protozoa or 4 mM $MgCl_2$ for bacteria, and corresponding primer pairs (Table 2.1) at 1 µM each. After initial denaturation for 10 min at 95 °C, 40 cycles of amplification followed (95 °C for 10 s, 59 °C for 5 s, 72 °C for 10 s). A melt curve from 72 °C to 95 °C was run to check for unspecific products. Archaeal genomic DNA was quantified by SYBR Green I fluorescence (LightCycler 480 SYBR Green I Master; Roche). Reaction volume was 10 µL, containing 5 µL of LightCycler 480 pre-mix, 1 µL of template DNA, and primer pair Ar915aF/Ar1386R (Table 2.1) at 0.5 µM each. After initial denaturation for 10 min at 95 °C, 40 cycles of amplification followed (95 °C for 10 s, 59 °C for 5 s, 72 °C for 10 s). A melt curve from 72 °C to 95 °C was run to check for non-specific products.

2.8 Construction of phage display libraries

2.8.1 M1 phage display library

Genomic DNA from M1 was isolated as previously described (Leahy *et al.*, 2010). Briefly, the cell pellet from 1.5 L of culture was frozen in liquid N_2 and ground into a fine powder with a mortar and pestle. Genomic DNA was then extracted using a QIAGEN genomic tip (500/G) according to manufacturer's instructions (Qiagen). To generate randomly sheared fragments (1 to 4 kb), 10 µg genomic DNA was resuspended in shearing buffer (55 mM Tris pH 8.0, 15 mM $MgCl_2$, 25% glycerol) and mechanically sheared by nebulization at 10 psi (nitrogen gas) for 30 s in disposable medical nebulizers (Unomedical Inc., Texas, USA). The samples were concentrated in centrifugal concentrators with molecular weight cut-off of 125 kDa (Sartorius, Germany), and washed with one volume of sterile distilled water. The sheared DNA was treated with end repair enzymes (DNATerminator End Repair Kit, Lucigen) according to manufacturer's instructions, to generate blunt-ended products for ligation. The end-repaired products were purified using QIAGEN PCR purification kit (Qiagen). pYW01 phagemid vector was digested with SmaI (Roche) and treated with alkaline phosphatase (Roche) to prevent vector re-circularization. Digested phagemid DNA was purified using a QIAGEN gel extraction kit (Qiagen). Ten micrograms of end-repaired inserts were ligated with 5 µg of vector, using T4 ligase (Roche) at 16 °C overnight. Ligation products were purified using a QIAGEN PCR purification kit and eluted in 100 µL of elution buffer (10 mM Tris-HCl pH 8.0). *E. coli* TG1 electrocompetent cells (Lucigen) were transformed by electroporation (1.8 kV, 200 Ω, 25 µF) in 1 mm gap electroporation cuvettes (Bio-Rad) with the purified ligation mixture. A total of 17 transformations were carried out. After electroporation, 950 µL of SOC medium was added, and the cells were

allowed to recover for 1 h at 37 °C. Aliquots of transformed cells were taken from each culture, pooled, and plated on antibiotic-selective media to estimate library size. To each culture, 9 mL of 2×YT medium containing 20 μg/mL chloramphenicol was added and incubated at 37 °C for 8 h. Aliquots were taken from each transformed culture and pooled to inoculate 1 L of 2×YT containing 20 μg/mL chloramphenicol at a 1:100 ratio. This culture represented the primary library. To generate M1 phage display library, early log phase (OD_{600} 0.2) culture was infected with VCSM13 helper phage at 50 m.o.i. and incubated at 37 °C for 30 min without shaking. The infected cells were pelleted by centrifugation at 5,000 × *g* for 10 min, resuspended in 1 L of fresh medium supplemented with 20 μg/mL chloramphenicol, and incubated at 37 °C for 6 h with aeration. PPs were purified and quantified as detailed in Section 2.6.

2.8.2 Metagenomic phage display library of protozoa-associated methanogens

Protozoa-associated symbionts were collected as described in Section 2.7, and then the cell pellet was frozen in liquid N_2 and ground into a fine powder with a mortar and pestle. Metagenomic DNA was extracted using a QIAGEN genomic tip (500/G) according to manufacturer's instructions. Random fragmentation of DNA was performed by resuspending 10 μg of DNA in shearing buffer (55 mM Tris pH 8.0, 15 mM $MgCl_2$, 25% glycerol) and mechanically shearing by nebulization at 10 psi (nitrogen gas) for 10 s in disposable medical nebulizers (Unomedical Inc.). The sheared DNA was washed, end repaired, and purified as described in the above section (Section 2.8.1), and then 4.7 μg of the sheared and purified DNA was ligated with 5.5 μg of SmaI-digested vector using T4 ligase (Roche). The ligation reaction was purified using a QIAGEN PCR purification column and eluted in 40 μL of elution buffer. *E. coli* TG1 electrocompetent cells (Lucigen) were transformed by electroporation (1.8 kV, 200 Ω, 25 μF) in 1 mm gap electroporation cuvettes (Bio-Rad) with the purified ligation mixture. A total of 10

transformations were carried out. The procedure for recovery of transformed cells, estimation of the library size, and production of recombinant phagemid particles was as described in Section 2.8.1.

2.9 Affinity selection of protozoa-binding phagemid particles

Protozoal bait was isolated from fistulated pasture-fed sheep, as described in the first paragraph of Section 2.7. The retentate was washed five times with 100 mL of anaerobic salts solution to remove free-living bacteria and archaea. Approximately 10 mL of retentate was collected with a serological pipette, fixed in 2% (w/v) paraformaldehyde (PFA) at 4 °C overnight, then washed on NITEX mesh and resuspended in an equal volume of 1× phosphate buffered saline (PBS). Protozoa were quantified by counting under the microscope. Samples were enumerated in triplicate.

To enrich for protozoa-binding clones, affinity screening was performed with the M1 phage display library against protozoa bait. PPs (10^{11} particles) were blocked with 1% (w/v) BSA for 1 h at room temperature with rotary action, and then incubated with 10^6 protozoa in a final volume of 2 mL of 1×PBS containing 1% (w/v) BSA and 1 mM $CaCl_2$ for 3 h with rotary action. After incubation, the mixture was washed 10 times with 20 mL of 1×PBS on NITEX mesh (11 µm) to remove unbound phagemid particles. The retentate was collected, transferred to a microfuge tube, and protozoa were sedimented by centrifugation at 1,000 × *g* for 5 min. The supernatant (containing unbound phagemid particles) was removed, and bound phagemid particles were released from protozoa cells by elution under acidic conditions (2 mL of elution buffer containing 100 mM glycine-HCl pH 2.2 and 1 mg/mL BSA was added). The mixture was incubated at room temperature for 30 min, neutralized with 120 µL of 1 M Tris (unbuffered), and then incubated at room temperature for 10 min. Protozoa and cell debris were removed by centrifugation at 10,000 × *g* for 5 min, and the supernatant containing eluted PPs was

quantified and then propagated by infecting *E. coli* TG1 cells in exponential phase for 30 min at 37 °C. Uninfected cells were minimized by antibiotic selection, as the medium contains chloramphenicol. An aliquot of infected cells was inoculated into 200 mL 2×YT containing chloramphenicol. At early log phase, VCSM13 helper phage was added at 50 m.o.i. to generate PPs from the enriched library. PPs were produced and purified as detailed above. The protozoa-binding PPs were used in the next round of affinity selection. Two rounds of affinity selection were carried out.

2.9.1 Screening of M1 phage display library

Plasmid DNA was extracted after each round of selection, and electrophoresed to monitor enrichment. After the second round of selection, distinct plasmid bands were excised from agarose gel, and transformed into *E. coli*. Clones were randomly selected for sequencing.

2.9.2 Screening of phage display library derived from protozoa-associated methanogens

Biopanning was performed on the primary phage display library against c-myc antibody (Sigma) to select for PPs displaying recombinant pIII fusion proteins. These PPs were amplified in *E. coli* and then used for two rounds biopanning against protozoal cells, as detailed in Section 2.9. Affinity screening against c-myc was performed on the PPs eluted from the second round of biopanning against protozoa to eliminate fast-growers that do not display recombinant pIII fusion proteins. For PacBio sequencing, phagemid DNA extracted after each round of selection was PCR amplified using primer pair PelBF1/pspR03 and primer dimers removed by gel extraction or size-selection beads, as detailed in Section 2.5.4.5. Five micrograms of purified amplicons from each library was submitted for PacBio sequencing (P6/C4 chemistry) at DSMZ (Braunschweig, Germany). One SMRT cell was used for each library.

2.10 Construction of Mru_1499^{A} fragments for domain mapping

Sequences corresponding to fragments of the Mru_1499^{A} gene that contains each Big_1 domain (Figure 3.9) were amplified by PCR with the corresponding primers in Table 2.1. The primer pairs mru1499_SphI-F40/mru1499_EcoRI-R197, mru1499_SphI-F/ mru1499_EcoRI-R283, and mru1499_SphI-F286/mru1499_EcoRI-R541 were used to amplify fragments containing Domain 1, Domain 2, and Domain 3, respectively. A linker region (containing six nucleotides), SphI and EcoRI restriction sites (underlined in Table 2.1) were appended to the primers to facilitate cloning into pYW01. The amplified fragments were ligated into pYW01 digested with SphI and EcoRI downstream of the PelB secretion signal. Plasmid DNA was extracted from transformants (Qiagen plasmid extraction kit; QIAGEN) and sequenced to verify that the fragment was amplified and inserted into the vector correctly.

2.11 Reverse panning of protozoa on immobilized phagemid particles

To identify the protozoa species that Mru_1499^{A} can bind, recombinant PPs displaying Mru_1499^{A} were immobilized to magnetic beads and then incubated with rumen protozoa (Figure 3.10). Unbound protozoa were washed away, and the captured protozoa were identified by their 18S rRNA gene sequences. Mru_1499^{A}-displaying PPs were blocked in 1% (w/v) BSA for 30 min at room temperature with rotary action, incubated with protozoa bait (10^{11} PPs:10^{6} protozoa) in a final volume of 1 mL of 1×PBS containing 1% (w/v) BSA and 1 mM $CaCl_2$ for 1 h with rotary action. PPs were immobilized to streptavidin-coated magnetic beads (New England Biolabs, MA, USA) *via* biotinylated phage coat protein antibody (PROGEN Biotechnik GmbH, Germany). Biotinylated phage coat protein antibody was added to the sample containing PPs and formalin-fixed protozoa at 1 μg/mL (final concentration), incubated for 20 min with rotary action, then 0.5 mg of streptavidin-coated magnetic beads (pre-blocked in 1% BSA) was added and the mixture was incubated for 20 min with rotary action. PPs

immobilized to magnetic beads were captured on a magnet (Invitrogen) and gently washed with 1 mL of 1×PBS ten times. Reverse panning was also performed on negative control (PPs generated from pYW01 without insert) concurrently. A 50 µL aliquot of the sample was taken for protozoa counts by microscopy. Three replicate counts were conducted, and where possible, a total of at least 100 protozoa cells were counted for each replicate. Genomic DNA was extracted from the remaining sample and from the protozoa sample before reverse panning, using NucleoSpin Tissue XS kit (Macherey-Nagel GmbH & Co., Germany) after three freeze-thaw cycles.

2.12 18S rRNA clone library construction and analysis

Clone libraries for partial protozoa 18S rRNA gene sequences (spanning V5-V7 regions) were constructed as described previously (Kittelmann & Janssen, 2011). DNA was extracted from protozoa samples using Macherey-Nagel NucleoSpin® Tissue XS columns (Norrie Biotech, NZ) after three rounds of freeze-thaw (1 min in liquid N_2, 2 min at 45 °C). PCR amplification of the 18S rRNA gene sequences present in the protozoa within the samples was performed as described in Section 2.5.4.2. Cloning of the PCR amplicons into the vector pCR2.1 was performed by using the TOPO TA cloning kit (Invitrogen). The TOPO ligation reaction was set up and then transformed into chemically competent *E. coli* TOP10 cells, as directed by the manufacturer's instructions. Colony PCR was performed on 53 clones for each sample using the primer pair RP841F/Reg1302, and the amplicons were submitted for sequencing at Macrogen Inc. Taxonomic assignment of 18S rRNA gene sequences was performed by BLAST assignment against a custom reference database and taxonomic framework of rumen protozoa sequences (Kittelmann *et al.*, 2015).

2.13 Next generation sequencing of 18S rRNA gene sequences

Partial 18S rRNA gene sequences (spanning V5-V7 regions) were determined by Roche 454 GS FLX Titanium amplicon pyrosequencing (Eurofins Genomics, Ebersberg, Germany) using barcoded primers RP841F and Reg1302R (Kittelmann *et al.*, 2013), and the data was processed and analysed using QIIME (Caporaso *et al.*, 2010) as described previously (Kittelmann *et al.*, 2015). Taxonomic assignment of 18S rRNA gene sequences was performed by BLAST assignment against a custom reference database and taxonomic framework of rumen protozoa sequences (Kittelmann *et al.*, 2015). Rarefaction analysis was performed using the multiple_rarefactions.py script in QIIME and then the number of species subsampled collated in Microsoft Excel to ensure that the number of reads obtained was adequate to cover the species diversity present in the samples. Titanium 454 sequence data obtained in this study were deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP062572.

2.14 Affinity binding assays

Affinity binding assays were performed with protozoa as bait to measure binding affinity of recombinant PPs displaying Mru_1499^{A}. Mru_1499^{A}-displaying PPs were generated by infecting *E. coli* harbouring pYW01 that contains the insert encoding for Mru_1499^{A} with VCSM13 helper phage. PPs generated from pYW01 without insert were used as negative control. PPs were purified and quantified as detailed in the phage methods section. Mru_1499^{A}-displaying PPs were blocked in 1% (w/v) BSA, incubated with protozoa bait (10^{10} PPs:10^{5} protozoa), and eluted from the bait under the same conditions as the affinity selection procedure. PPs eluted from protozoa bait were quantified by standard methods (Section 2.6.2) and compared against the control.

2.15 Immunogold labelling of protozoa-associated phagemid particles for SEM

PPs were blocked in 1% (w/v) BSA and incubated with protozoa bait (10^{10} PPs:10^5 protozoa) for 3 h at room temperature in PBS supplemented with 2 mM $CaCl_2$ in a volume of 1 mL. At the end of the incubation, protozoa and attached PPs were fixed in PFA (2%; final concentration) for 15 min at room temperature, and then sodium borohydride was added to a final concentration of 10 mM and incubated for 5 min at room temperature to terminate the fixation process. The suspension was centrifuged at 1,000 × *g* for 5 min to sediment the protozoa cells and attached PPs, and the pellet was washed with 1 mL of PBS. The cells were resuspended in 1 mL of 1:100 diluted filamentous phage coat proteins (pVIII, pIII, pVI, pIX) polyclonal antibody produced in rabbit (Sapphire Bioscience, NZ) for 1 h at room temperature, and then centrifuged at 1,000 × *g* for 5 min and resuspended in anti-rabbit IgG conjugated to 20 nm gold produced in goat (Sapphire Bioscience). The sample was centrifuged at 1,000 × *g* for 5 min, and resuspended in 200 µL of MilliQ water. Imaging and detection of gold signal by electron dispersive spectroscopy was performed by the Manawatu Microscopy and Imaging Centre (Massey University, Palmerston North, NZ).

2.16 Bioinformatics analyses

2.16.1 M1 phage display library

BLAST searches were performed to identify Mru_1499 and homologues. The genome context of genes encoding potential Mru_1499 homologs were examined by manually searching methanogen genomes available on the Joint Genome Institute's Integrated Microbial Genomes database (Markowitz *et al.*, 2012). InterPro and SMART were used for *in silico* predictions of domain architecture of Mru_1499 and related proteins (Hunter

et al., 2012, Letunic *et al.*, 2012). Clustal Omega webserver was used for multiple sequence alignment of protein sequences (Sievers *et al.*, 2011).

2.16.2 Sequence analysis for PacBio sequencing data generated for phage display library derived from protozoa-associated methanogens

The circular consensus approach was used for increased sequencing accuracy (Travers *et al.*, 2010). Raw data collected from PacBio sequencing was filtered to eliminate circular consensus reads with less than two passes (i.e. the sequence was only read once) and accuracy less than 90% at the ends of amplicons. As multiple amplicons can be ligated between the sequencing adapters, these concatemers were separated so that each sequence corresponds to a single amplicon (Figure 4.9). Amplicon sequences were trimmed to remove phagemid vector-derived sequences. Trimmed sequences with a length of 59 nucleotides or less were eliminated from further analysis. For the primary library (PAM1), these sequences were submitted for automated annotation of function and domains in the Integrated Microbial Genome (IMG) platform hosted by the Joint Genomic Institute (Markowitz *et al.*, 2015) and for automated BLASTX search and species assignment by GAMOLA (Altermann & Klaenhammer, 2003). Cluster analysis was performed on the trimmed sequences using the software Cd-hit-est (Li & Godzik, 2006) with identity threshold of 90% and by *de novo* assembly of the sequences in Geneious R9. Enriched sequences in the PAM_FINAL library (clusters representing >1% of total sequence reads) were assembled, and ORFs upstream of the c-myc sequence present in the phage display vector were identified in Geneious R9.

BLASTP search was performed on the translations of the ORFs against the non-redundant (nr) database and a database composed of rumen methanogen strains [genome sequences that are publicly available on NCBI, acquired as part of the Hungate1000 initiative, and the following genomes which were sequenced through

projects funded by the New Zealand Agricultural Greenhouse Gas Research Centre and the Pastoral Greenhouse Gas Research Consortium: *Methanobrevibacter olleyae* YLM1, *Methanobacterium formicicum* BRM9, *Methanobrevibacter millerae* SM9, *Methanobrevibacter* sp. D5, *Methanosarcina* CM1, *Methanosphaera* sp. 3F5, Methanogenic Archaeon isolate ISO4-G1, Methanogenic Archaeon isolate ISO4-G11, and Methanogenic Archaeon isolate ISO4-H5 (Creevey *et al.*, 2014)]. Conserved domains were identified by searching the databases InterPro (Hunter *et al.*, 2012) and SMART (Letunic *et al.*, 2012). As many of the sequences do not match to proteins or domains with known functions, the probability that identified ORFs encode adhesins was also determined by SPAAN (computation basis of this tool was discussed in Section 1.3.5) (Sachdeva *et al.*, 2005). Other amino acid sequence features, including repeated motifs and over-representation of amino acid residues, were identified using MEME (Bailey *et al.*, 2009) and the MotifScan tool available through the ExPASy portal (Artimo *et al.*, 2012).

2.16.3 Sequence analysis for protozoa-associated symbiont community and bovine epimural bacterial community

Pyrosequencing data collected for the protozoa-associated methanogen and bacterial communities were analysed using the QIIME pipeline as described previously (Kittelmann *et al.*, 2013). Sequence data for the bacterial community that is attached to bovine rumen epithelial tissues has been reported by Li et al (2012) and is publicly available on the NCBI database (accession numbers: GU302522-GU304593 and HQ399694-HQ400406) (Li *et al.*, 2012a). In this study, the sequence data was re-analyzed by species assignment of these sequences using QIIME to ensure that an updated reference database was used for the assignments.

Chapter 3. Identification and Characterization of a Protozoa-Binding Adhesin from *Methanobrevibacter ruminantium* M1 by Phage Display

3.1 *Methanobrevibacter ruminantium* M1 phage display library construction and affinity screening

A *M. ruminantium* M1 phage display library was constructed in the phage display vector pYW01, which enables recombinant proteins to be produced as fusions to the C domain of phage pIII protein (Gagic *et al.*, 2013) (Figure 3.1A). The primary library contained 3.2×10^8 independent clones containing insert sizes ranging from 1 to 4 kb. Theoretically, complete coverage of the M1 genome (2.93 Mb) (Leahy *et al.*, 2010) requires a minimum of 6.7×10^3 clones containing an average insert length of 2 kb (Jacobsson *et al.*, 2003) (Equation 3.1)]; therefore, it is estimated that the genome coverage of this library is 4.8×10^4. This library was infected with helper phage VCSM13 to generate a master library of PPs, displaying the M1 proteome. The master library was used for affinity screening to identify adhesins that bind protozoa.

Equation 3.1. Minimum number of clones required for coverage of a given genome of size *b*, from Jacobsson *et al.* (2003).

N denotes the number of clones required, *P* denotes the probability that a given fragment is present in the library, *a* denotes the average fragment size, and *b* denotes the genome size. To estimate the number of clones required to represent the M1 genome, $P = 0.99$, $a = 2000$, $b = 2{,}930{,}000$)

$$N = \frac{\ln\,(1-P)}{\ln\,(1-\frac{a}{b})}$$

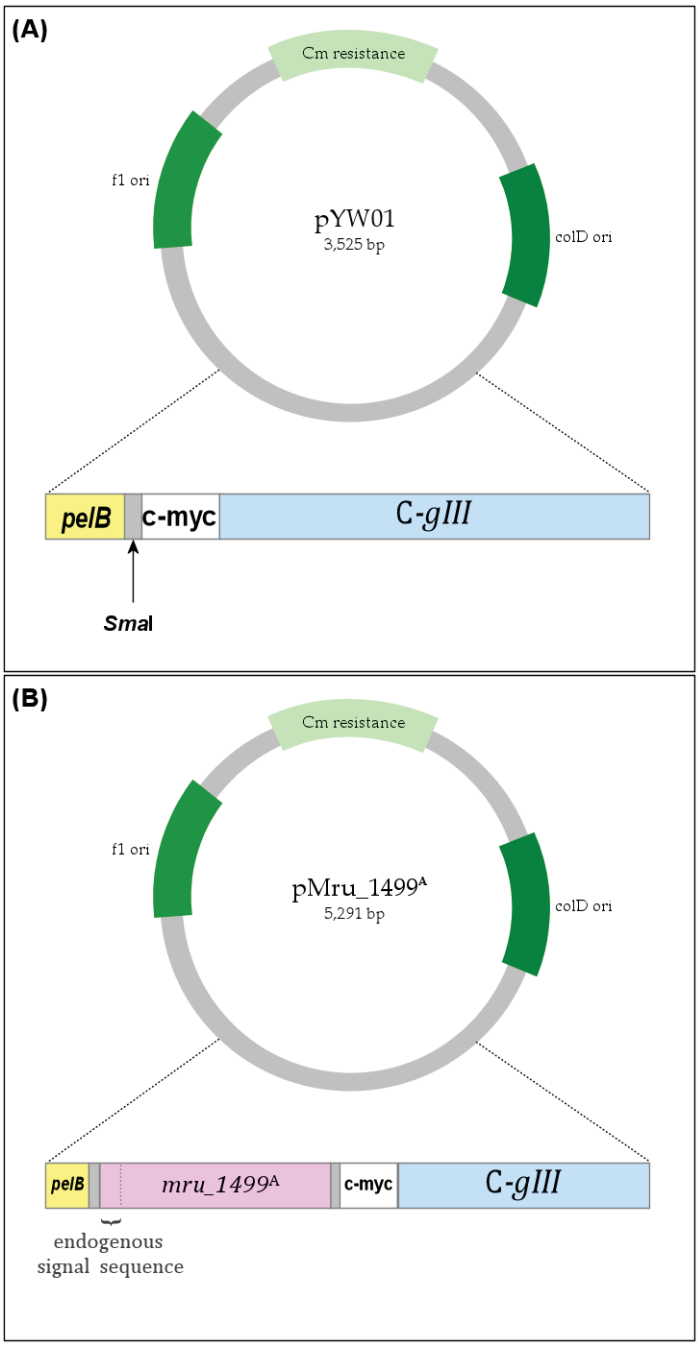


Figure 3.1. Plasmid map of phagemid vector pYW01 and pMru_1499^{A}.

Panel A shows the phagemid vector pYW01 used to create the M1 phage display library. A chloramphenicol (Cm) resistance cassette, phage origin of replication, and plasmid origin of replication (shaded in green) can be found in the phagemid. Inserts were cloned into the *Sma*I restriction site. The gene sequences encoding a c-myc tag and the C domain of phage protein pIII are downstream of the restriction site, which enables recombinant fusion proteins to be generated. Panel B shows the phagemid vector pMru_1499^{A} which encodes the affinity-selected protein Mru_1499^{A}. The insert sequence encodes an endogenous signal sequence. Also, the insert is in frame with the genes encoding c-myc and phage protein pIII (C domain).

As rumen protozoa are fragile, sensitive to oxygen, and difficult to maintain in culture, cells harvested from the sheep rumen contents were formalin-fixed prior to affinity screening to prevent cell lysis and to preserve the integrity of their cell surfaces. It is known that formalin fixation of live cells causes cross-linking of cell-surface proteins and can reduce the diversity of binders recovered (Qiao *et al.*, 2012); however this step was necessary to prevent disintegration of cells. This preparation was used as "bait" or ligand to bind PPs expressing protein domains to protozoa. During affinity screening, PPs were incubated with protozoa, followed by a series of washes to remove non-specific or background binders. The protozoa-attached PPs were subsequently eluted from bait, collected, and then amplified through infection of the *E. coli* host.

Plasmid profiles of the library pools corresponding to protozoa-attached PPs were monitored by agarose gel electrophoresis after each round of affinity screening. Before affinity screening, a smear was observed upon agarose gel electrophoresis of purified plasmids due to the random size distribution of inserts in the phage display library. After two rounds of affinity screening, four discrete plasmid bands were observed, indicating enrichment for plasmids of specific sizes (Figure 3.2). The plasmid bands were excised and purified, and then transformed into *E. coli* host strain TG1. After sequence analysis of 13 clones, a recombinant clone encoding the partial gene sequence of *mru_1499* in frame with the phagemid vector c-myc tag and *gIII* was selected for further analysis (Figure 3.1B). This clone originated from plasmid band 2 in Figure 3.2. The remaining 12 clones were not pursued as the encoded protein was not in frame with the gene encoding phage protein pIII. The gene *mru_1499* was also up-regulated when M1 was co-cultured with rumen bacterium *Butyrivibrio proteoclasticus* B316 (Leahy *et al.*, 2010). Before the start of this project, the export of an archaeal protein harbouring an archaeal signal sequence was tested in *E. coli* to confirm that archaeal signal sequences

are recognized in the bacterial host (data not shown). Here, Mru_1499^{A} contains an archaeal-derived signal sequence, which also shows that the *E. coli* host can recognize some archaeal signal sequences. Other recombinant clones were "background" clones that contained DNA inserts encoding peptides that were out of the frame with pIII, and therefore not expected to display functional fusion peptides on the surface of the virion. Background clones are frequently observed in affinity-selected phage display libraries (Vodnik *et al.*, 2011). The protein encoded by the 5´ moiety of gene sequence *mru_1499* that corresponded to the insert in the recombinant phagemid was designated as Mru_1499^{A} (the 'A' stands for affinity selected portion of the gene), and the phagemid vector encoding Mru_1499^{A} is denoted pMru_1499^{A}.

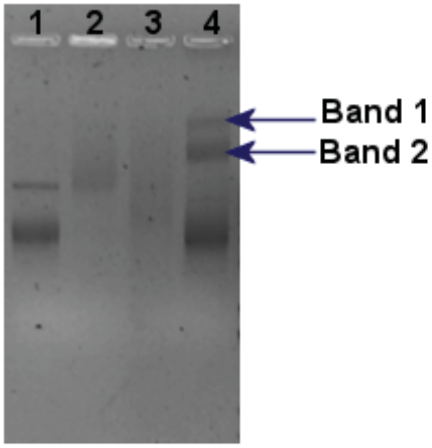


Figure 3.2. Plasmid profile of M1 library before and after panning against protozoal bait.

Lane 1 contains pYW01 plasmid DNA (no insert). Lane 2 contains plasmid DNA extracted from the primary library. Lanes 3 and 4 contain plasmid DNA extracted from *E. coli* cells after the first and second round of panning against protozoal bait, respectively. After two rounds of biopanning, enriched plasmid DNA bands indicated by arrows (Band 1 and Band 2) were extracted and transformed into *E. coli* for further analysis. The lower molecular weight bands were not further pursued, as they likely correspond to empty vector or short inserts that have been enriched due to growth advantages.

3.2 Confirmation of Mru_1499 as protozoa binding adhesin

3.2.1 The Mru_1499^{A} fusion protein is displayed on the surface of the virion

Prior to functional characterisation of Mru_1499^{A}, confirmation that this polypeptide is displayed on the surface of the recombinant PPs was required. As the DNA sequence *mru_1499*^{A} was cloned in frame with the gene sequences encoding c-myc and C domain of pIII, it is expected that the encoded protein would be translated as a recombinant pIII fusion protein containing a c-myc tag; therefore, an antibody against c-myc was used to detect its presence by immunoblotting. The expected size of the recombinant fusion protein is 73 kDa, which approximately corresponds to the band present in the immunoblot (Figure 3.3).

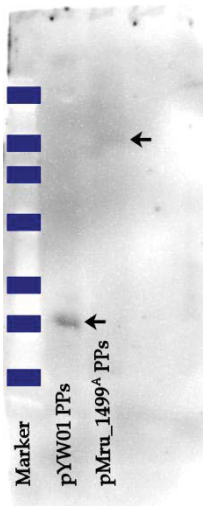


Figure 3.3. Confirmation of Mru_1499^{A}-c-myc-pIII fusion protein incorporation into phagemid particles by western blot.

The protein standard was electrophoresed in the first lane. Each rectangle represents the following protein marker sizes, from top to bottom: 98 kDa, 62 kDa, 49 kDa, 33 kDa, 28 kDa, 17 kDa, 14 kDa. Phagemid particles produced from the phagemid vectors pYW01 and pMru_1499^{A} were electrophoresed in the second and third lanes, with expected protein fusion sizes of 17 kDa and 73 kDa, respectively.

3.2.2 Validation of Mru_1499^{A} adhesion using affinity binding assays

The affinity binding assays using protozoa as bait (Section 2.14) showed that PPs displaying Mru1499^{A} could bind protozoa more efficiently, with a 558±18 fold-enrichment compared to PPs generated from the pYW01 phage display vector (no insert) (Table 3.1). This increase in binding of pMru_1499^{A} PPs over vector PPs confirmed that the affinity-selected polypeptide binds to protozoal cell surfaces.

Table 3.1. Total number of eluted phagemid particles from protozoa bait (data from two affinity binding assay experiments).

Phagemid particles	Input titre (cfu)	Output titre (cfu)	Output/Input titre ratio	Binding above control[b]	Mean fold difference (± standard deviation)
Affinity binding assay I[a]					558±18
pMru_1499^{A}	1.00×10^{9}	2.10×10^{7}	2.10×10^{-2}	545	
pYW01 (control)	1.00×10^{9}	3.85×10^{4}	3.85×10^{-5}		
Affinity binding assay II[a]					
pMru_1499^{A}	1.00×10^{9}	2.23×10^{6}	2.23×10^{-3}	571	
pYW01 (control)	1.00×10^{9}	3.90×10^{3}	3.90×10^{-6}		

[a]Data from two separate assays are presented.
[b]Fold difference is obtained by dividing output/input titre ratios for pMru_1499^{A} PPs relative to that for the vector (pYW01) PPs.

3.2.3 Attachment of pMru_1499^{A} PPs to protozoa - scanning electron microscopy

To visualize PPs bound to protozoa and thereby demonstrate that Mru_1499^{A} binds directly to protozoa, immunogold staining using PP-binding antibodies was performed. A polyclonal phage coat antibody was used as primary antibody to hybridize to PPs, then a secondary antibody conjugated to 20 nm gold particles was applied to generate a detectable signal. Detection of the heavy metal on the surface of protozoa was performed by scanning electron microscopy (SEM) with energy-dispersive X-ray spectroscopy (EDS). When excited by a high energy source, each element emits characteristic peaks in the X-ray spectrum, which allows specific elements (such as gold) to be identified. In this experiment, gold particles were detected when PPs displaying Mru_1499^{A} were used, whereas no gold signal was detected when negative control PPs derived from the vector pYW01 (no insert) were used (Figure 3.4).

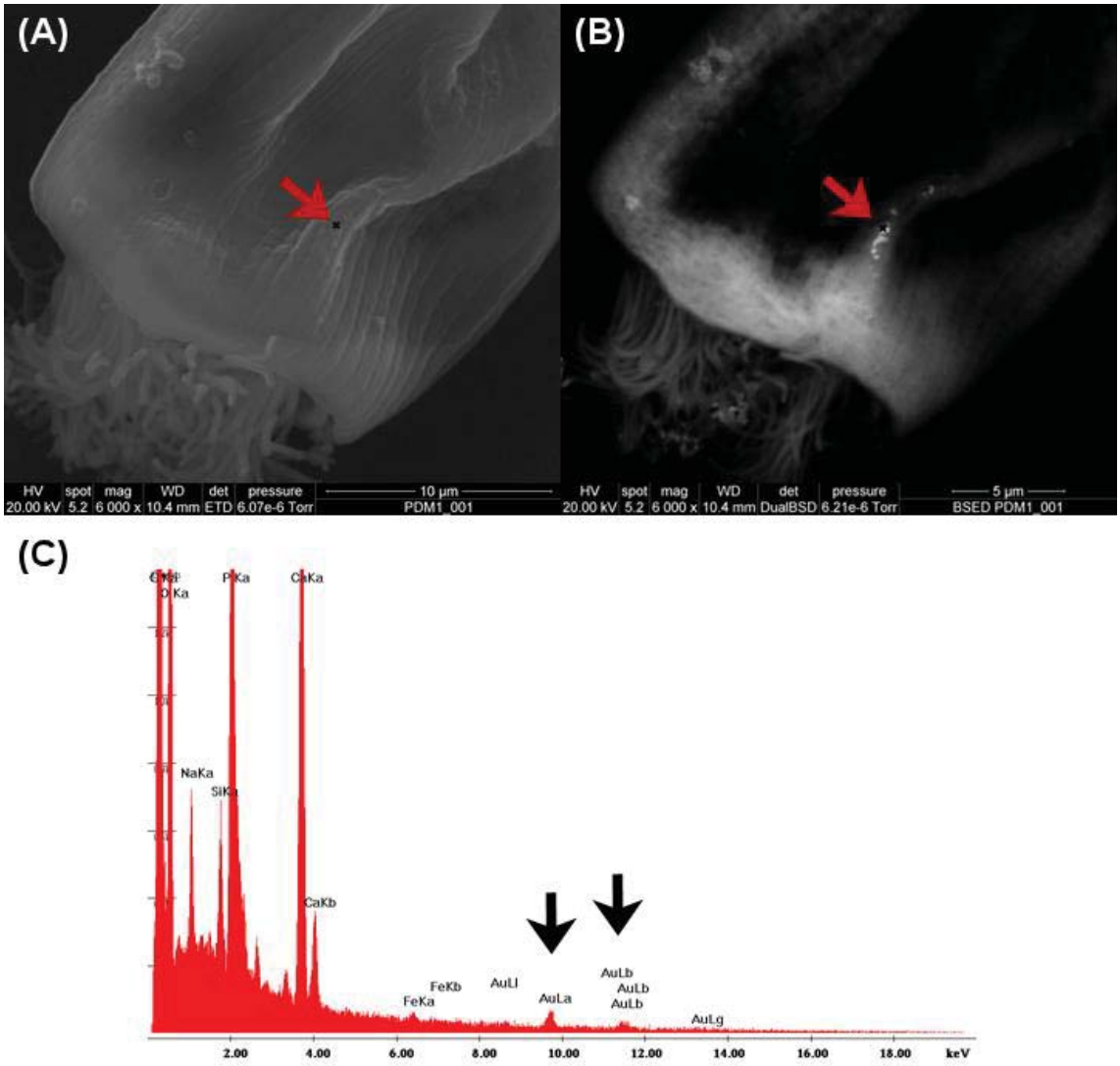


Figure 3.4. Scanning electron microscopy images of phagemid particles displaying Mru_1499A on protozoal cell surface.

(A) SEM image for a protozoan cell. (B) SEM backscattered image for the same cell, where a cluster of immunogold particles could be observed. (C) Energy dispersive X-ray spectroscopy (EDS) scan of the region indicated by an arrow in panels A and B. The presence of peaks corresponding to gold confirmed the presence of immunogold particles.

3.3 *In silico* analyses of the Mru_1499 DNA and amino acid sequences

3.3.1 Protein domains predicted from the amino acid sequence

Mru1499^{A} is a polypeptide that encodes the N-terminal 541 amino acids of the predicted *mru_1499* gene product. It contains a signal sequence and three tandem bacterial immunoglobulin-like class 1 (Big_1) domains predicted by both SMART (Letunic *et al.*, 2012) and InterPro (Hunter *et al.*, 2012) database searches. A region with low similarity to pseudomurein binding (PMB) repeats (amino acids 465 to 489; E-value 1E-3) was predicted by SMART database search and may act as an anchor to the methanogen cell surface (Leahy *et al.*, 2010). The full length gene product of *mru_1499* consists of an additional Big_1 domain and a C-terminal transglutaminase-like (TG-like) domain (Figure 3.5A). The two Big_1 domains closest to the N-terminus are most similar to each other (61% amino acid identity), whereas the second Big_1 domain (in "Domain 2") and the C-terminal Big_1 domain are the least similar (47% amino acid identity) (Figure 3.5B).

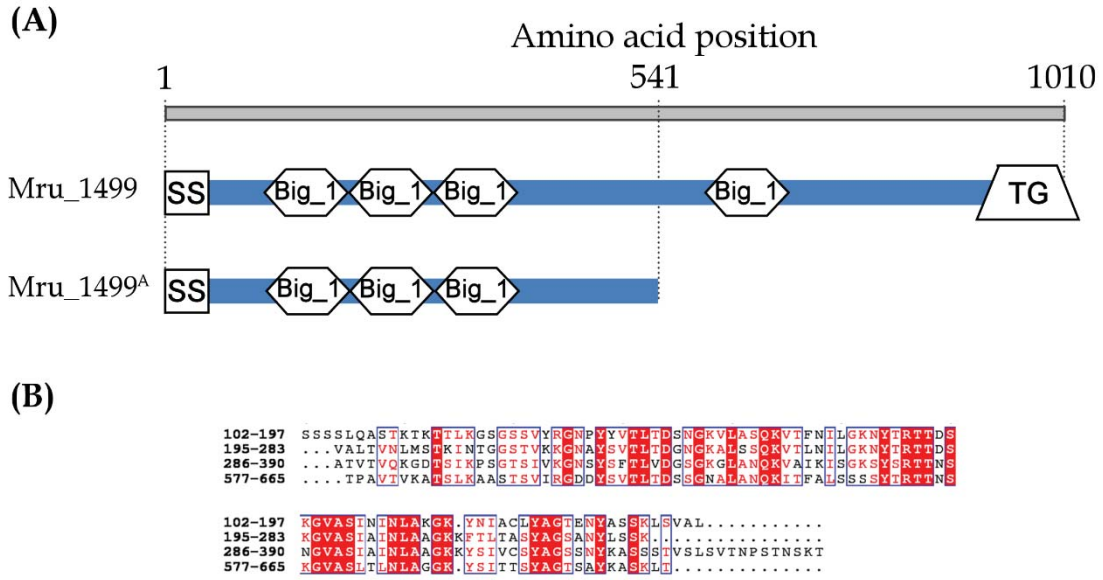


Figure 3.5. Mru_1499 domain organization.

(A) Predicted domain architecture of Mru_1499 (full length) and Mru_1499^{A} (affinity selected clone). The peptide encoded by amino acids 1-18 is a signal sequence (SS). Three tandem domains were recognized as bacterial immunoglobulin class 1 (Big_1) domains at residues 102-197 (E-value 1.4×10^{-3}), 195-283 (E-value 9.5×10^{-4}), and 286-390 (E-value 2.8×10^{-3}) by a SMART database search (Letunic *et al.*, 2012) . A fourth Big_1 domain (residues 577-665; E-value 6.6×10^{-5}) and C-terminal transglutaminase-like domain (residues 878 to 981; E-value 9.8×10^{-23}) were present in the full length Mru_1499 protein. (B) Multiple sequence alignment of the predicted Big_1 domains in Mru_1499 by Clustal Omega webserver (Sievers *et al.*, 2011).

3.3.2 Mru_1499^{A} homologs were found in other methanogens

In the M1 genome, *mru_1499* is not located within an operon, as shown by OperonDB predictions in the UCSC Genome Browser (Chan *et al.*, 2012); however, it is found in close proximity to two other ALP-encoding genes: *mru_1497* and *mru_1500*. Gene neighbourhood analysis was performed on ORFs encoding proteins with at least 25% identity to the Mru_1499 protein sequence to identify potential homologs. Genes in several rumen methanogen species within the genus *Methanobrevibacter* met this criterion and were found in a similar genome context to the gene encoding Mru_1499. These species included two members of the *M. ruminantium* clade (Seedorf *et al.*, 2014) [*M. olleyae* KM1H5-1P (sequenced as part of the Hungate1000 initative (Creevey *et al.*,

2014) and *Methanobrevibacter* sp. YLM1 (Kelly *et al.*, 2016b)], members of the *M. wolinii* clade [*M. millerae* ZA-10 (sequenced as part of the Hungate1000 initative), SM9 (Kelly *et al.*, 2016a), and *Methanobrevibacter* sp. D5 (Y. Li *et al.*, unpublished)], *M. smithii* PS (Samuel *et al.*, 2007), and *Methanobrevibacter* sp. AbM4 (Leahy *et al.*, 2013). Within the *Methanobrevibacter ruminantium* clade, homologs identified in *M. olleyae* KM1H5-1P and *Methanobrevibacter* sp. YLM1 (the most closely related species to M1) are adjacent to only one ALP-encoding gene (rather than two ALP-encoding genes, as observed in M1) (Figure 3.6). In genomes of *M. smithii* PS as well as members of the *Methanobrevibacter gottschalkii* and *Methanobrevibacter wolinii* clades, no other ALP-encoding ORFs were adjacent to the corresponding *mru_1499* homolog.

At the protein level, Mru_1499 homologs from methanogen species *M. olleyae* and *Methanobrevibacter* sp. YLM1 are the most similar to Mru_1499 (56% amino acid identity), as expected for species that are most closely related to M1. Big_1 domains were not predicted for the N-terminal half of these proteins even though regions corresponding to Mru_1499 Big_1 domains 1 and 3 aligned well with the homologs (Figure 3.7). However, given that the amino acid sequences have a high level of similarity, Mru_1499 homologs in *M. olleyae* KM1H5-1P and *Methanobrevibacter* sp. YLM1 may also have protozoa-binding activity and correspond to Big_1 domains despite being below the threshold used in the prediction software. The amino acid sequences of putative protein homologs from *Methanobrevibacter* sp. AbM4, *M. smithii* PS, *Methanobrevibacter* sp. D5, *Methanobrevibacter* sp. SM9, and *M. millerae* ZA-10 were found to be 28-35% identical to Mru_1499 (Figure 3.8).

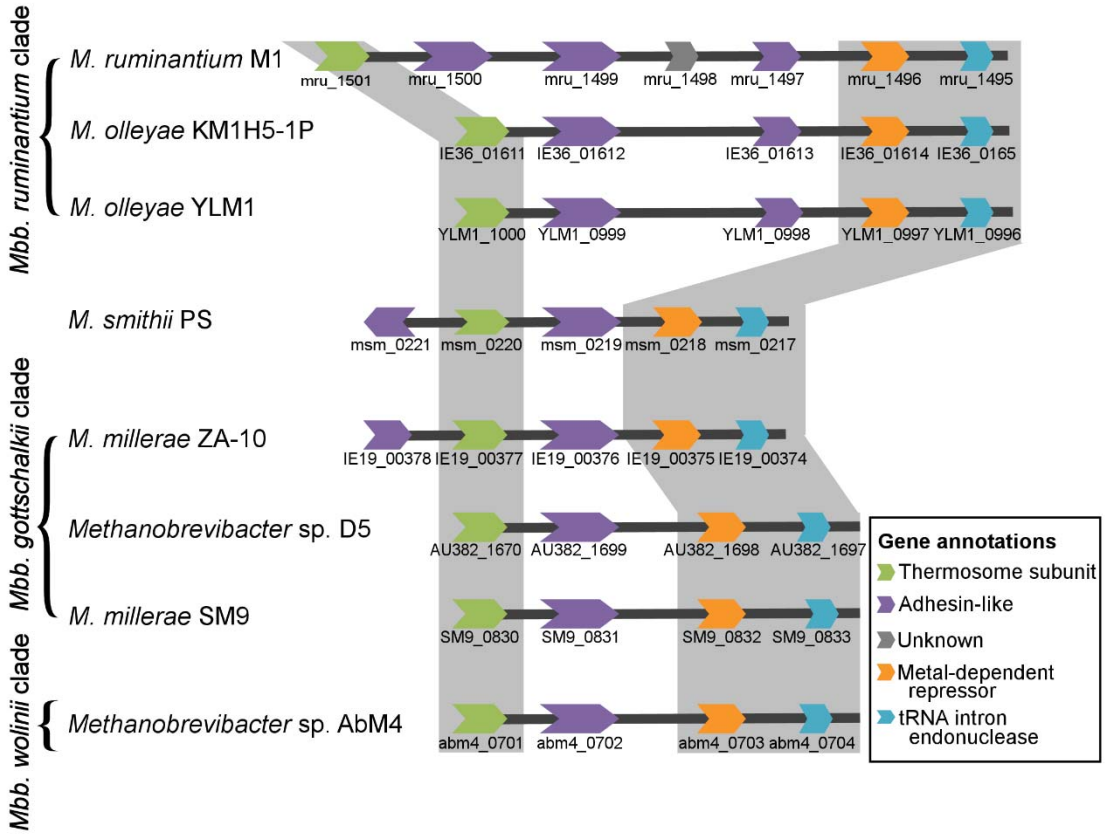


Figure 3.6. Gene neighbourhood of Mru_1499 homologs in *Methanobrevibacter* species.

Genes (not to scale) encoding the Mru_1499 homologs can be found in the same gene neighbourhood within genomes of other *Methanobrevibacter (Mbb.)* species. Green, purple, grey, orange, and blue arrows represent predicted open reading frames.

A similar protein (>25% amino acid sequence identity) containing a Big_1 domain was identified from a methanogen outside of the genus *Methanobrevibacter*. A BLAST search of the NCBI non-redundant database, using the full length Mru_1499 sequence as a query, revealed a distantly-related putative protein coded in the genome of *Methanosphaera stadtmanae* MCB-3, a human gut methanogen (Fricke *et al.*, 2006) (Figure 3.8). No other Mru_1499^{A} homologs with Big_1 domains were identified by BLASTP query of available methanogen genomes.

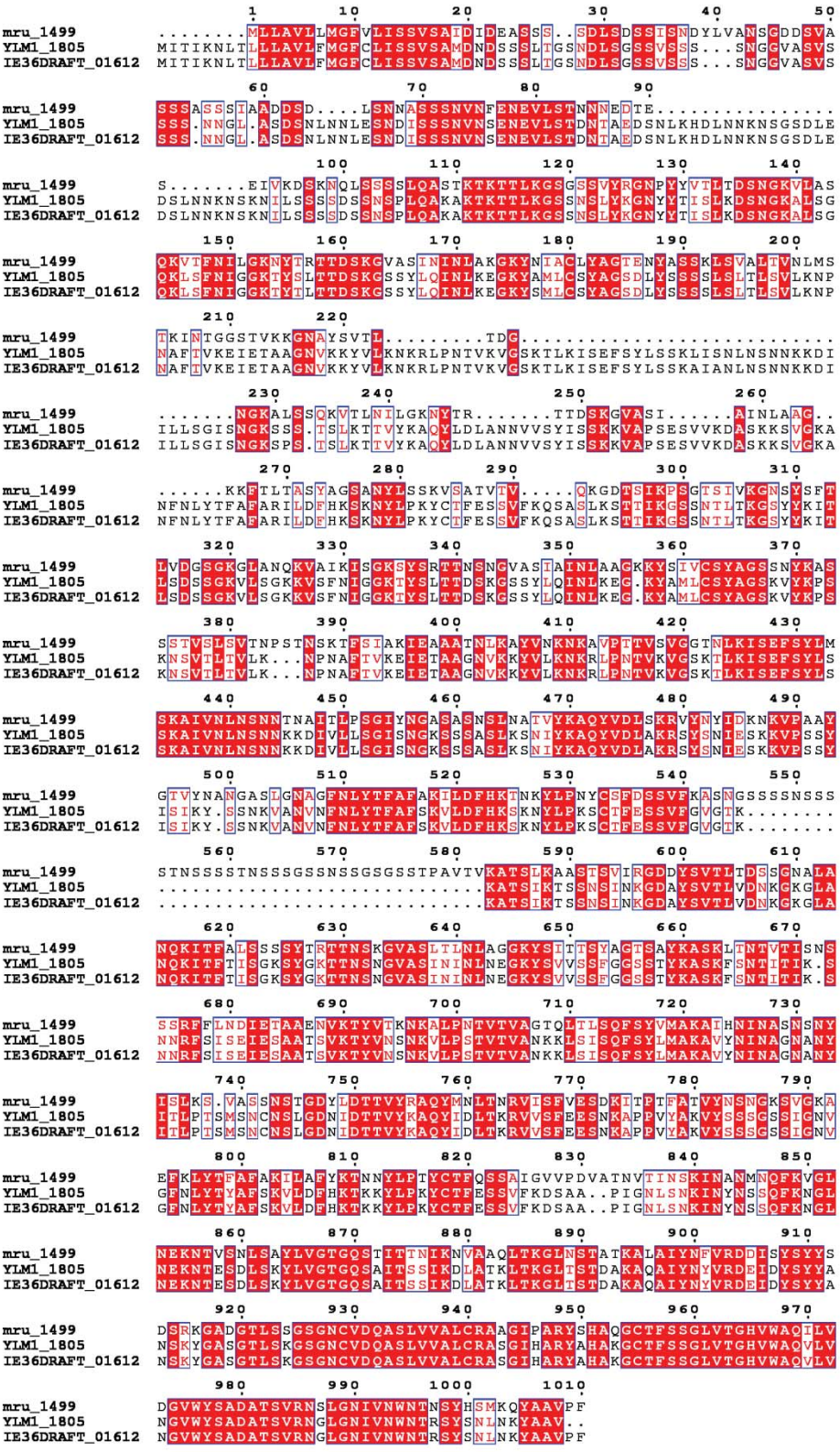


Figure 3.7. Clustal Omega protein sequence alignment for Mru_1499 homologs from the *M. ruminantium* clade.

Alignment visualized by ESPRIPT 3.0 (http://espript.ibcp.fr/ESPript/ESPript/) (Robert & Gouet, 2014). YLM1_1805 is from *Methanobrevibacter* sp. YLM1, and IE36DRAFT_01612 is from *M. olleyae* KM1H5-1P.

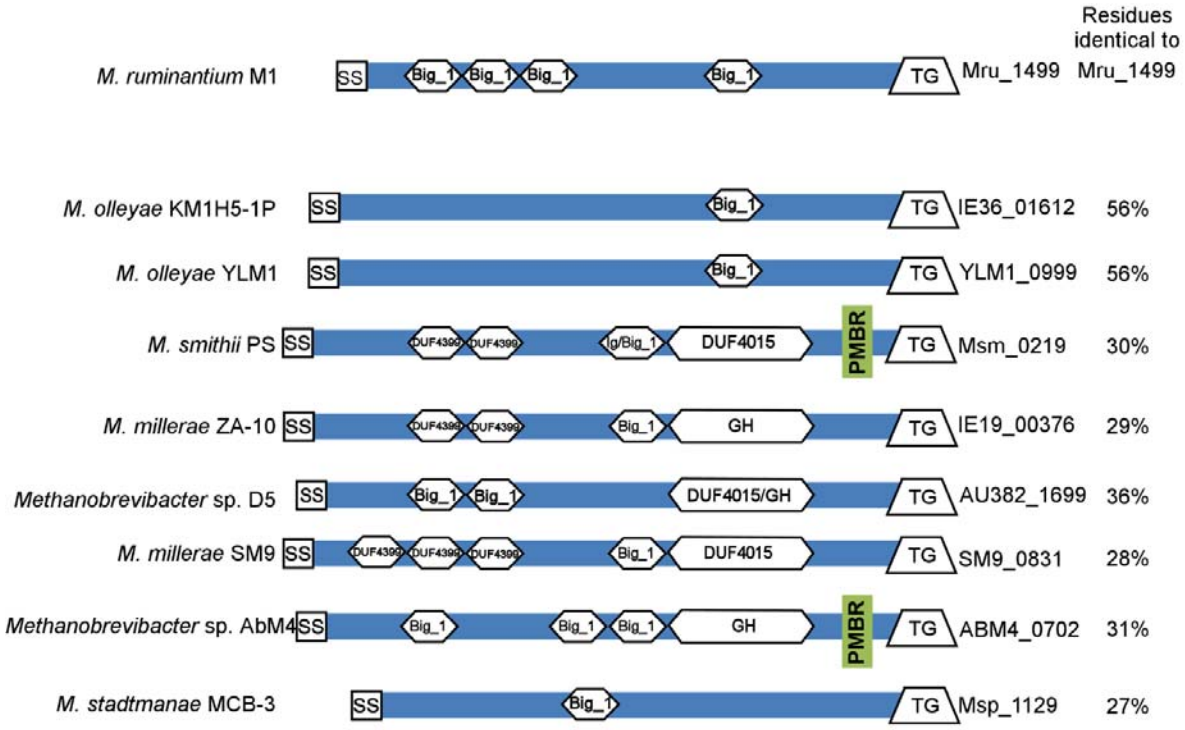


Figure 3.8. Mru_1499 protein homologs present in other rumen methanogen species.

Protein domain architecture of Mru_1499 homologs. Annotated features were predicted by both SMART and InterPro database searches, with the exception of the shaded domains. Pseudomurein binding repeats (PMBR) were predicted for Msm_0219 and AbM4_0702. The Big_1 domains in *M. olleyae* and *Methanobrevibacter* sp. YLM1 were predicted only by SMART, with E-values of 3.3×10^{-5}, and glycoside hydrolase (GH) domain in *M. millerae* was identified by an InterPro database search. Genome information for *M. ruminantium* M1, *M. smithii* PS, *Methanobrevibacter* sp. AbM4, and *Methanosphaera stadtmanae* MCB-3 is publicly available (Fricke *et al*., 2006, Samuel *et al*., 2007, Leahy *et al*., 2010, Leahy *et al*., 2013). Draft genome sequences for *M. olleyae* KM1H5-1P and *M. millerae* ZA-10 were obtained as part of the Hungate1000 initiative and are available through the Joint Genome Institute's Integrated Microbial Genomes database (Markowitz *et al*., 2012, Creevey *et al*., 2014). Other information is from *Methanobrevibacter* sp. YLM1 (W. J. Kelly et al., unpublished), *Methanobrevibacter* sp. D5 (Y. Li et al., unpublished), and *Methanobrevibacter* sp. SM9 (Kelly *et al*., 2016a).

3.4 Mapping the Mru_1499 protozoa-binding domain

Phage display technology has been used for decades in epitope mapping; therefore, to map the protein domains of Mru_1499^{A} required for protozoa binding, three recombinant phage display constructs containing each of the three Big_1 domains were created (Figure 3.9A). Each protein fragment was displayed on PPs, and assessed for binding affinity to protozoa bait as described in Section 2.14. The results of affinity binding assays showed that the number of eluted PPs corresponding to the phagemid termed "Domain 1" (which displayed amino acids 40 to 197, encompassing the N-terminus proximal Big_1 domain) was greater than the negative control, and within the same order of magnitude as the positive control (PPs displaying Mru_1499^{A}, originally affinity selected from the phage display library). The numbers of eluted PPs for constructs "Domain 2" and "Domain 3" were in the same magnitude as the negative control; therefore, the polypeptide encoded by the Domain 1 construct facilitates binding of Mru_1499^{A} to protozoal cell surfaces, but the polypeptides encoded by the Domain 2 and Domain 3 constructs do not play a role in binding (Figure 3.9B).

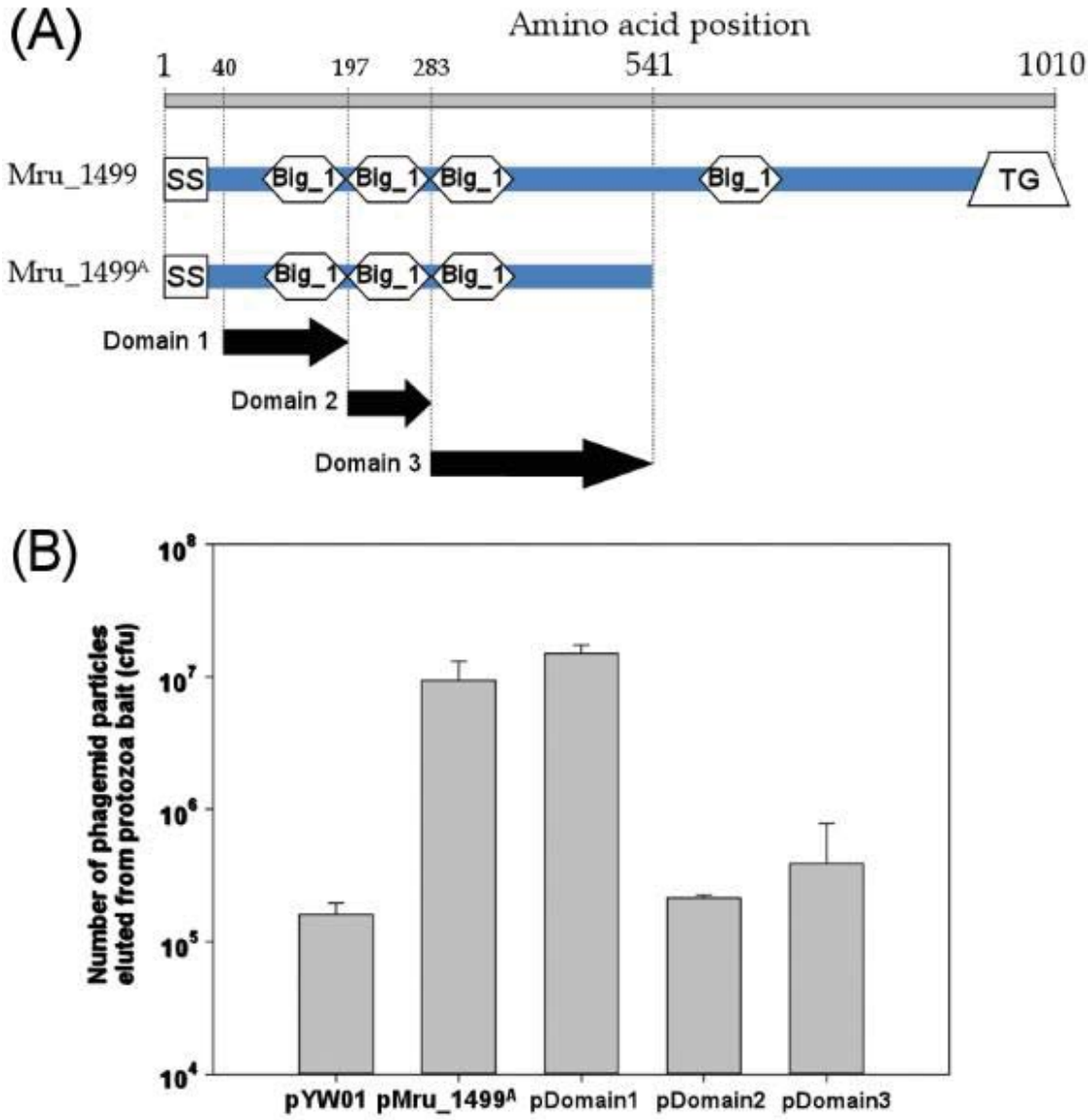


Figure 3.9. Mru_1499^{A} domain mapping.

(A) Clones encoding regions encompassing predicted Big_1 domains (Domain 1, Domain 2, Domain 3) were constructed to map the domains involved in host adhesion. (B) Results of affinity binding assay for fragments encompassing each Big_1 domain with protozoa as bait. Binding of phagemid particles (PPs) displaying Mru_1499^{A} (pMru_1499^{A}) to protozoa cells was measured as the number of bound PPs eluted from protozoa bait. PPs produced from pYW01 (vector only) were used as a negative control.

3.5 Determination of Mru_1499^{A} protozoa tropism by reverse panning

To determine whether Mru_1499^{A} binds to a broad range of rumen protozoa, or to specific proozoal species, PPs displaying Mru_1499^{A} were used as a ligand to capture protozoa that can bind to this polypeptide. In this "reverse panning" procedure, PPs displaying Mru_1499^{A} were immobilized on paramagnetic beads captured *via* streptavidin-biotin interactions (detailed in Section 2.11). Vector only pYW01 PPs were also tested by reverse panning to ensure that protozoa were bound specifically to the displayed Mru_1499^{A} protein, and not to phage coat proteins. The number of protozoa captured by PPs was enumerated by microscopy (Tymensen *et al.*, 2012a). Only a background level of binding was observed in pYW01 negative control samples (57±18 protozoa per 10^{11} PPs), whereas approximately 400-fold more protozoa (22,200±2,580 protozoa per 10^{11} PPs) were bound to Mru_1499^{A} PPs, demonstrating that physical association between protozoa and Mru_1499^{A} PPs were specific to the displayed protein.

Due to the low number of protozoa present in the pYW01 negative control samples, DNA extraction and further analysis was not performed for these samples. For protozoa bound to Mru_1499^{A} PPs, species compositions of the protozoal communities before and after reverse panning were compared to determine whether Mru_1499^{A} exhibits selectivity towards certain protozoa species.

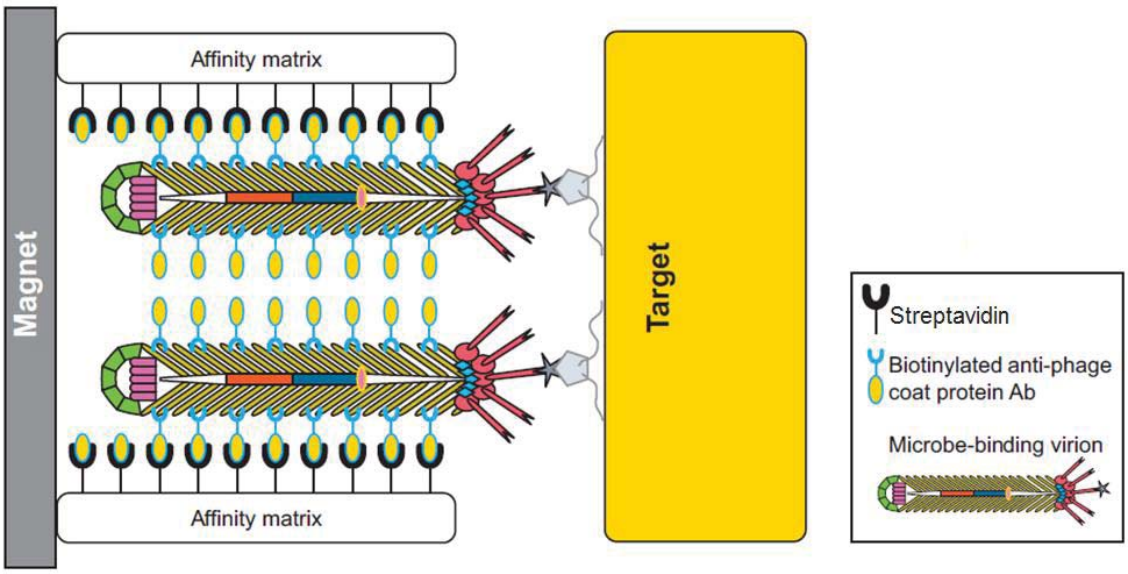


Figure 3.10. Schematic representation of reverse panning procedure used to isolate Mru_1499-binding protozoa.

Phagemid particles displaying Mru_1499^{A} were immobilized to the affinity matrix (streptavidin-coated paramagnetic beads) *via* biotinylated anti-phage coat protein antibodies. PPs and the adherent protozoa ("target") complexed with paramagnetic beads were separated from unbound protozoa using a magnet. Targets (captured protozoa) were lysed for DNA isolation, and an 18S rRNA gene region was amplified by PCR to identify the species present in the sample and estimate their relative abundances. Biotinylated antibody was raised against pVIII phage coat protein. Protozoal receptors are depicted as grey pentagons.

3.5.1 Analysis of affinity-selected protozoans by 18S rRNA amplicon DGGE

DGGE was performed as an initial assessment of the protozoal community diversity before and after reverse panning (Figure 3.11). Upon electrophoresis of the partial 18S rRNA gene sequencing amplicons, similar amplicon bands were observed in the samples before and after reverse panning; however, two bands were enriched in the sample after reverse panning. As rumen protozoal species share a high level of identity in their 18S rRNA gene sequences (Kittelmann & Janssen, 2011), it can be difficult to resolve partial gene amplicons into distinct bands that correspond to individual protozoal species. Further analysis was required to elucidate the species present in the protozoa communities.

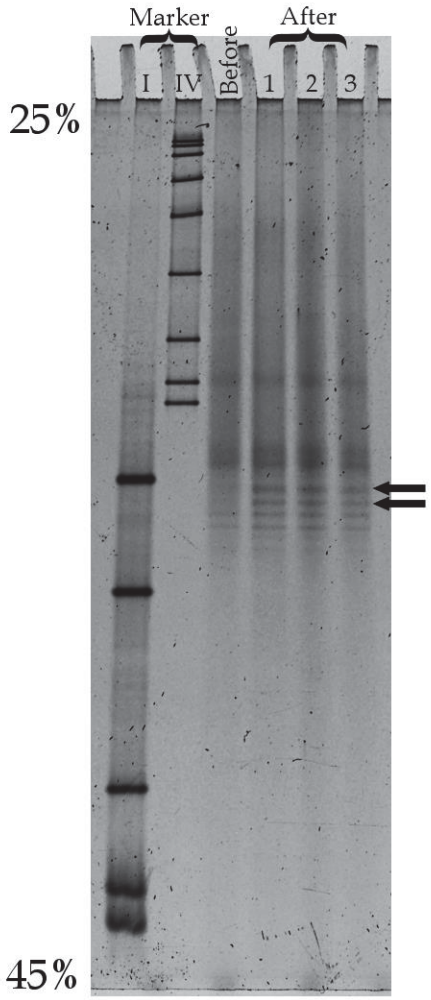


Figure 3.11. Denatured gradient gel electrophoresis analysis of 18S rRNA gene amplicons derived from protozoa samples before and after reverse panning.

Two bands (indicated by arrows) were enriched in the protozoa community after reverse panning.

3.5.2 18S rRNA amplicon derived clone library

Clone libraries were constructed from PCR amplicon products of protozoa samples before and after reverse panning to identify the protozoa species present (as described in Section 2.12). Colony PCR was performed on 53 randomly selected clones from each library, and 50 samples were submitted for sequencing analysis for each library. Species assignment of the sequences by BLAST revealed that sequences corresponding to *Epidinium sp.* and *Entodinium longinucleatum* increased by 26% and 24% after reverse panning, respectively (Table 3.2). Moreover, phylogenetic analysis of the sequences showed the sequences from the same clone library tended to cluster together (Figure 3.12). Protozoa in the family Isotrichidae were observed in the sample before reverse panning, but their 18S rRNA gene sequences were not represented in these data, suggesting that sequence coverage was inadequate.

Table 3.2. Composition of protozoa communities derived from samples before and after reverse panning determined by clone library sequence analysis.

	Number of sequences		Percentage of sequences	
	Before	**After**	**Before**	**After**
Diplodinium dentatum	1	0	2%	0%
Entodinium, other	11	4	23%	8%
Entodinium bursa	3	1	6%	2%
Entodinium dubardi	7	2	15%	4%
Entodinium furca dilobum	11	1	23%	2%
Entodinium furca monolobum	3	8	6%	16%
*Entodinium longinucleatum**	3	15	6%	30%
Entodinium nanellum	1	0	2%	0%
*Epidinium caudatum**	1	14	2%	28%
Eudiplodinium maggii	7	5	15%	10%
Total number of sequences	48	50		

* Denotes species abundance that increased by more than 15%.

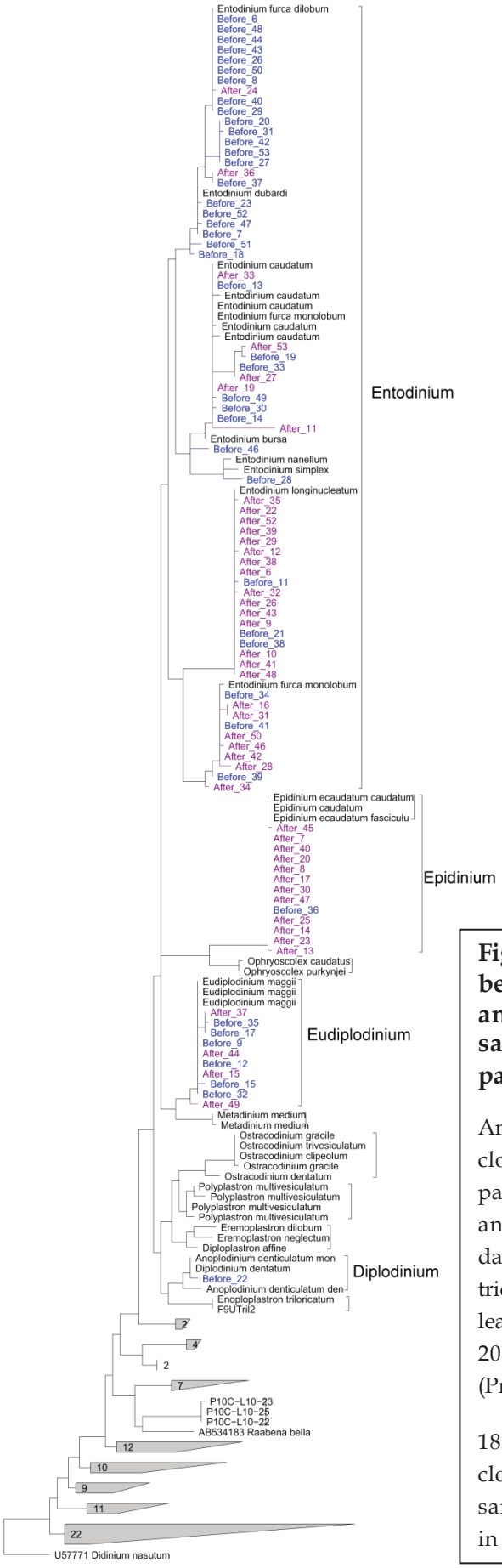


Figure 3.12. Phylogenetic relationship between known rumen protozoa species and protozoa species in clone library samples from before and after reverse panning.

Amplicon sequences (~460 nucleotides) from clone libraries before and after reverse panning were aligned using SINA aligner and then imported into a ciliate protozoa database previously constructed from trichostome ciliate reference sequences (at least 1,500 nucleotides) (Kittelmann *et al.*, 2015) by the parsimony tool in ARB software (Pruesse *et al.*, 2012).

18S rRNA gene sequences identified in the clone library representing the protozoa samples before and after reverse panning are in blue and purple, respectively.

3.5.3 18S rRNA next generation sequencing

Next generation sequencing of the amplicon pools was performed to determine species compositions of the protozoal communities before and after reverse panning with greater depth. The numbers of sequencing reads obtained for protozoa samples before and after reverse panning were 335 and 290, respectively. The data revealed that reverse panning enriched for members of the genera *Epidinium* (from 7.5% to 23% of total sequences obtained) and *Entodinium (*from 25% to 57% of total sequences obtained) (Table 3.3, Figure 3.13). *Isotricha* and *Dasytricha* sequences accounted for 20% and 0.3% of the protozoa sample before reverse panning, respectively, but no sequences were detected in the sample after reverse panning. When identities of members of the genus *Entodinium* were resolved to the species level, it was apparent that the proportion of *Entodinium longinucleatum* and *Entodinium furca monolobum* sequences increased after reverse panning. These data suggest that Mru_1499^{A} is a broad spectrum protozoa binder with a preference for binding to cells of genera from the family Ophryoscolecidae. Specifically, it exhibits strong affinity for members of the genera *Epidinium* and *Entodinium*, but it also appears to discriminate between species within the *Entodinium* genus. No binding was observed for protozoa in the Isotrichidae family (*Isotricha* spp. and *Dasytricha* spp.). The trends observed for *Entodinium* (other), *Entodinium bursa*, *Entododinium dubardi*, *Entodinium furca dilobum*, and *Entodinium nanellum* differed between Sanger sequencing of clone libraries and pyrosequencing (Table 3.2, Table 3.3), and the difference can be attributed to inadequate sampling of clones in Sanger sequencing of clone libraries. Rarefaction analyses of pyrosequencing data showed that the sequence coverage was sufficient to represent the protozoa species present in the sample (Figure 3.14).

Table 3.3. Composition of protozoa communities derived from samples before and after reverse panning determined by pyrosequencing analysis.

	Number of sequences		Percentage of sequences	
	Before	**After**	**Before**	**After**
Anoplodinium-Diplodinium	4	0	1.2%	0%
Diplodinium dentatum	8	13	2.4%	4.5%
Entodinium, other	33	31	9.9%	10.7%
Entodinium bursa	11	9	3.3%	3.1%
Entodinium dubardi	16	9	4.8%	3.1%
Entodinium furca dilobum	15	21	4.5%	7.2%
Entodinium furca monolobum	7	26	2.1%	9%
*Entodinium longinucleatum**	3	61	0.30%	21%
Entodinium nanellum	1	7	0.30%	2.4%
*Epidinium caudatum**	25	67	7.5%	23%
Eudiplodinium maggii†	145	46	43.3%	16%
Dasytricha, other	1	0	0.30%	0%
Isotricha, other	15	0	4.5%	0%
Isotricha intestinalis	2	0	0.60%	0%
Isotricha prostoma	49	0	15%	0%
Total number of sequences	335	290		

* Denotes species abundance that increased by more than 15%.

† Denotes species abundance that decreased by more than 15%.

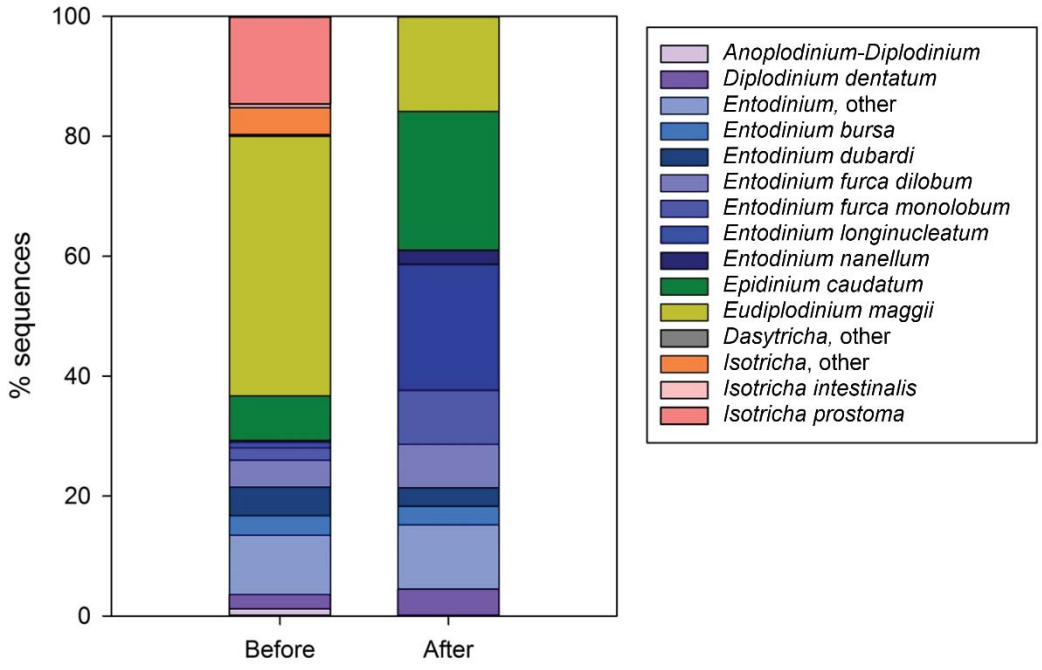


Figure 3.13. Relative abundances of ciliate protozoal species present in samples before and after reverse panning.

Specific protozoal taxa were enriched after subjected to reverse panning with Mru_1499^{A} phagemid particles.

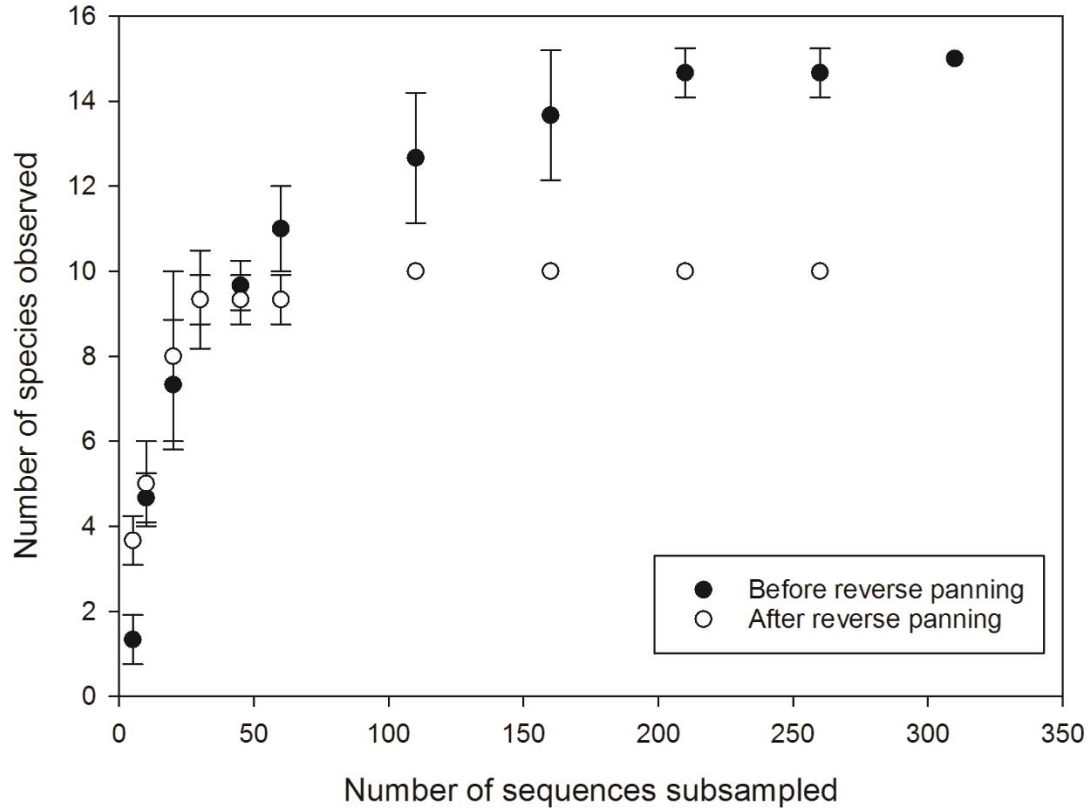


Figure 3.14. Rarefaction curves showing protozoa species coverage in pyrosequencing data for samples before and after reverse panning.

Each data point represents the mean of three replicates. The error bars represent one standard deviation on either side of the mean.

3.6 Summary

A large phage display library (composed of 10^8 independent clones) was constructed for rumen methanogenic archaeon *M. ruminantium* M1, and an adhesin that can bind to protozoa was identified from this library by affinity screening. The clone identified from affinity screening, pMru_1499^A, encodes amino acids 1 to 541 of the *mru_1499* gene product, previously annotated as adhesin-like protein. *In silico* analyses revealed that Mru_1499^A contains three tandem Big_1 domains which are typically present in proteins involved in adhesion. Domain mapping showed that the region proximal to the N-

terminus (amino acids 40 to 197), containing a single Big_1 domain, is sufficient for protozoa binding. A newly developed "reverse panning" strategy was used to capture protozoa that bind to Mru_1499^{A}. Preliminary analysis by DGGE showed that there were differences between the initial protozoa sample before reverse panning, and the protozoa community captured by Mru_1499^{A} after reverse panning. Analyses of clone library sequencing and pyrosequencing data for each sample demonstrated that the number of sequences corresponding to *Entodinium longinucleatum* and *Epidinium* sp. increased, whereas sequences corresponding to *Isotricha* spp. and *Dasytricha* spp. were not present after reverse panning.

Chapter 4. Identification of Protozoa-Binding Adhesins from Protozoa-Associated Symbionts by Phage Display

Associations between methanogenic archaea and protozoa can have an effect on ruminant methane production (Section 1.2.1.5); therefore, adhesion proteins that mediate their attachment are potential targets for disrupting protozoa-methanogen interactions as a strategy for methane mitigation (e.g. as targets for vaccine against methanogens). Results from Chapter 3 demonstrated that phage display technology can be used to discover adhesins derived from a single methanogenic archaeal species. Here, the phage display approach was used to attempt to identify adhesins from a metagenomic phage display library derived from the protozoa-associated symbiont community. In addition to methanogenic archaea, bacterial species can also be found as protozoal endosymbionts and ectosymbionts. To increase the likelihood of identifying adhesins of methanogen (rather than bacterial) origin, the protozoal-associated symbiont community was enriched for methanogens prior to extracting DNA for library construction. The protozoa-associated symbiont community has been characterized previously by Sanger sequencing of clone libraries, but not by high throughput next generation sequencing methods; therefore, methanogen and bacterial species present in the methanogen-enriched metagenomic sample were characterized by pyrosequencing of 16S rRNA gene sequences. Lastly, a phage display library was constructed from the metagenomic DNA, and affinity screened against protozoa for identification of novel adhesion peptides.

4.1 Development of a procedure for enriching protozoa-associated methanogens

A method was developed to isolate protozoa-associated symbionts and enrich for methanogens in the population. Rumen content samples were collected and pooled from

five fistulated sheep, and then filtered through a 335 µm mesh to remove the plant material present in the digesta. The sample was visually inspected for the presence of auto-fluorescent methanogens attached to protozoa by fluorescence microscopy. Observations from five samples collected on different days indicated that 9 to 23% of rumen protozoa had more than 20 associated methanogens cells [a criterion used by Vogels *et al.* (1980)]. At least 50 protozoa were examined in each sample (Table 4.1).

To isolate protozoa along with their attached symbionts, the sample was diluted in anaerobic salts solution and incubated at 39 °C in a separation funnel [as detailed in Section 2.7 and Tymensen *et al.* (2012b)]. During the incubation, protozoa settled to the bottom of the funnel, and particulate plant matter (and plant-adherent protozoa) rose to the surface (Figure 4.1). The bottom fraction was collected and retained on a mesh with 11 µm pores. The retentate was transferred to a bag made from the 11 µm mesh and then the bag was submerged in anaerobic salts solution to allow free-living bacteria and archaea to move out of the bag. Table 4.2 shows typical protozoa yield from this process. After this process, free-living bacteria were still observed in the sample by microscopy; therefore, the sample was treated with lysozyme and mutanolysin to lyse free-living and protozoa-associated bacteria that remained in the sample. Figure 4.2 summarizes the isolation procedure.

Table 4.1. Enumeration of methanogen symbionts associated with rumen protozoa.

		Protozoa cells with greater than *n* autofluorescent methanogen symbionts (%)*	
Sample	**# protozoa counted**	***n* > 5**	***n* > 20**
1	50	82%	10%
2	98	56%	15%
3	101	65%	9%
4	99	64%	15%
5	100	68%	23%

* Autofluorescent methanogens were counted as proxy for the methanogen population.

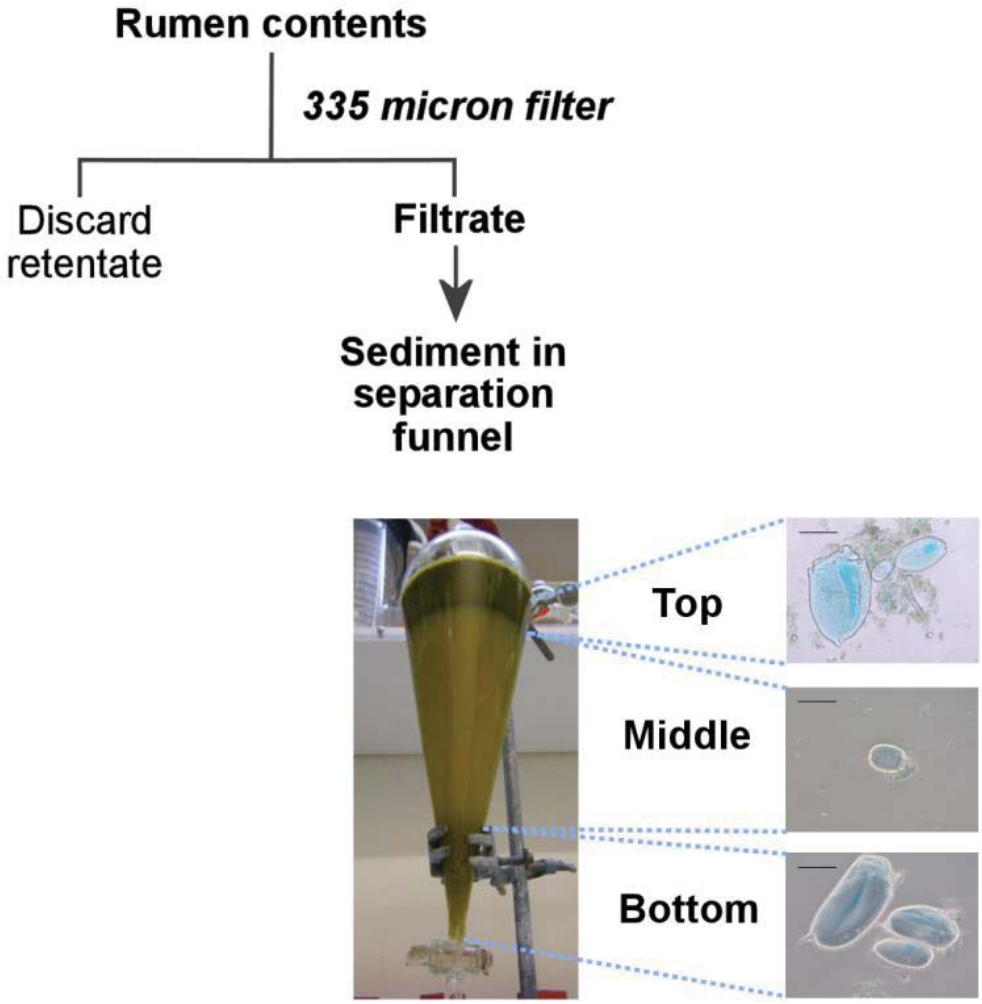


Figure 4.1. Rumen contents fractionation by sedimentation in separation funnel.

Samples from each fraction was examined by microscopy. Larger protozoa tended to settle to the bottom layer, whereas small protozoa were observed in the middle layer. Scale bars indicate 50 μm, 20 μm, and 50 μm in the microscopy images corresponding to the top, middle, and bottom layers, respectively.

Table 4.2. Example of data obtained from a single protozoa isolation experiment.

Separation step	Number of protozoa per mL of sample	Volume of sample (mL)	Total number of protozoa in sample	% protozoa retained from previous step
Filtrate from 335 μm mesh	3.25×10^5	1125	3.7×10^8	---
Sedimentation	1.50×10^5	800	1.2×10^8	33
Retentate from 11 μm mesh	3.08×10^6	22	6.8×10^7	57
Retentate from 11 μm mesh after additional filtering	2.06×10^6	24.5	5.1×10^7	75

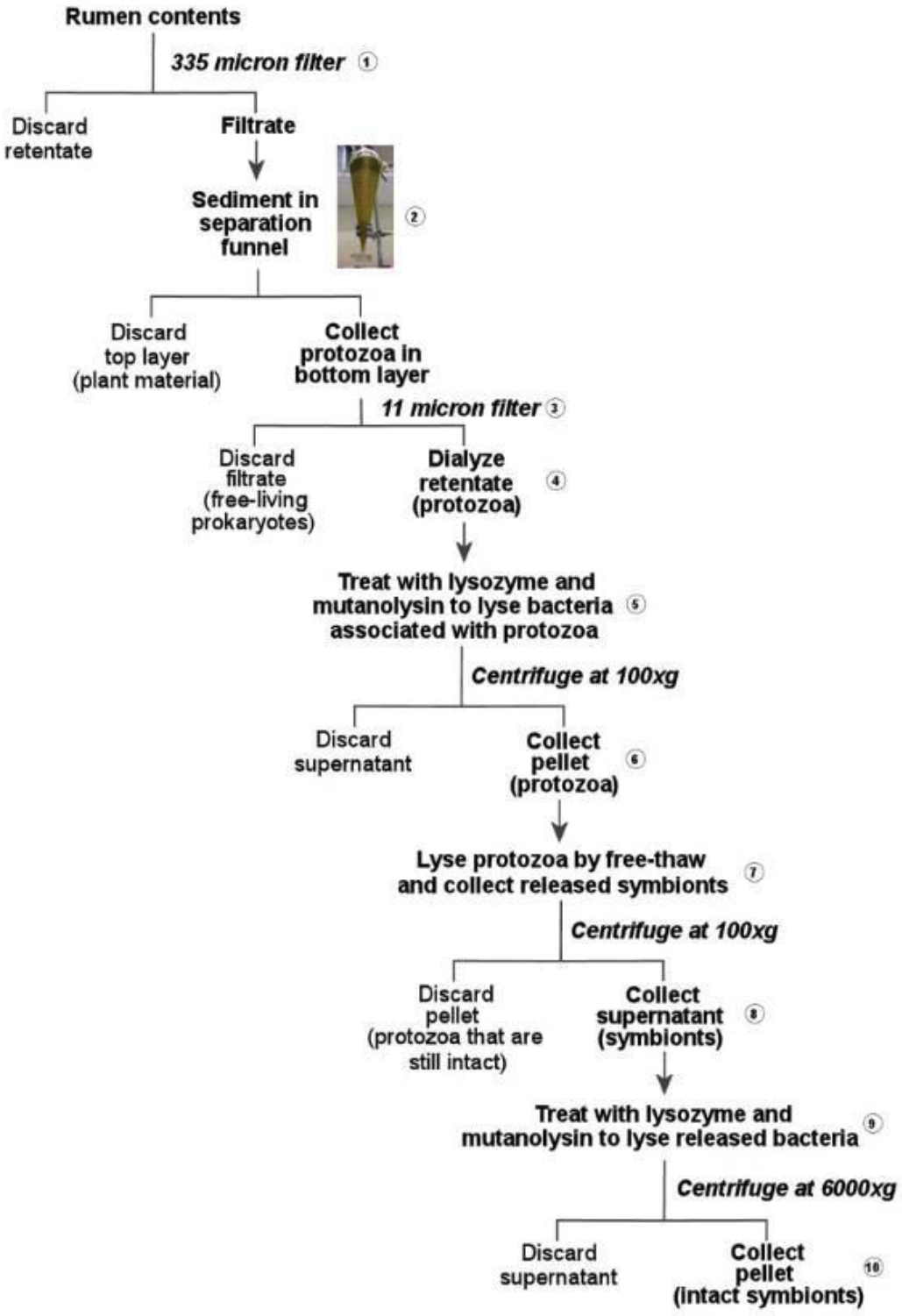


Figure 4.2. Workflow for harvesting protozoa-associated symbionts.

Instead of releasing both ecto- and endo-symbionts by freeze-thaw (Step 7 in Figure 4.2), several conditions were tested in an attempt to only harvest methanogenic ectosymbionts by detachment from protozoa using: (1) proteinase treatment, (2) sodium chloride treatment, and (3) application of H_2 gas. For Treatment 1, we expected that host-symbiont associations are mediated through membrane-anchored adhesins displayed on the methanogen cell surface; therefore, the application of proteinase should degrade these proteinaceous connections, thereby releasing ectosymbionts from the host. For Treatment 2, if the interacting proteins are associated through ionic interactions, an increase in NaCl concentration could also facilitate detachment, as it would become more energetically favourable for the proteins to associate with the Na^+ or Cl^- ions than the protein partner. Lastly, in Treatment 3, application of H_2 gas to the rumen protozoa sample was tested, as the frequency of methanogen attachment to protozoa was observed to decrease when the rumen was flushed with H_2 gas, suggesting that methanogens no longer associate with the ciliate host when there is an alternate H_2 source available (Stumm et al., 1982).

Proteinase treatment (Treatment 1) resulted in complete methanogen and protozoal cell lysis; therefore, this method was not pursued further. Sodium chloride treatment (Treatment 2) resulted in cell lysis of ~40% protozoa. The protozoa that belong to the Isotrichidae family represented 4% of the total ciliated protozoa population by microscopic counts before the treatment, but none was detected by microscopy after sodium chloride treatment. For the protozoa that remained intact, only 5 out of 100 ciliates examined had more than 5 methanogen symbionts, whereas 48 out of 100 ciliates in the control sample had more than 5 methanogen symbionts, as observed by fluorescence microscopy. A mix of fluorescent methanogens and non-fluorescent cells were observed in the sample containing detached symbionts. The non-fluorescent cells

likely represented protozoa-associated bacteria and methanogens that were no longer metabolically active (some methanogens are resistant to high salt concentration, but not all methanogen species can survive these conditions (Tokura et al., 1997). Together, these data suggest that during sodium chloride treatment, ecto-symbiotic methanogens were released by disrupting ionic interactions between cell surface proteins; however, protozoal lysis also released ecto- and endo-symbionts. To avoid introducing a bias towards halophilic methanogens and certain genera of protozoa, this method for detachment was not used. For H_2 treatment (Treatment 3), after incubation in the presence of excess H_2, 40 out of 100 ciliates in the sample had more than 5 methanogen symbionts, compared to 48 out of 100 ciliates in the control, indicating that this method is not sufficiently effective for detachment. It may be possible to optimize the tested conditions to selectively release ectosymbionts from protozoa (for example, experimenting with varying concentrations of proteinase K or NaCl); however, due to the time limitations of this project, freeze-thaw was used to release both endo- and ectosymbionts in a non-selective manner instead.

4.1.1 Relative amounts of archaea, bacterial, protozoal DNA by real time qPCR

The relative abundance of archaeal, bacterial, and protozoal DNA in the metagenomic sample was assessed by real time qPCR of the partial gene sequences of small subunit ribosomes. Bacterial DNA was the most abundant in this sample out of the three groups, despite application of lysozyme and mutanolysin to disrupt bacteria cells prior to harvesting protozoa-associated symbionts (Table 4.3). Protozoal DNA was also found in the sample. However, compared to the relative abundance of archaea to bacteria 16S rRNA gene copies present in the total rumen sample from a sheep on the same diet (Henderson *et al.*, 2012), archaea were enriched by 54-fold.

Table 4.3. Relative amounts of archaeal, bacterial, and protozoal genomic DNA present in the metagenomic DNA sample.

	Small subunit rRNA gene copies	
	Metagenomic DNA from protozoa-associated symbionts (gene copies/ng total DNA)	**Rumen contents (Henderson *et al.*, 2013)** (gene copies/g rumen contents)
Archaea 16S	7.30×10^3	1.6×10^6
Bacteria 16S	5.38×10^4	6.3×10^8
Protozoa 18S	1.33×10^4	nd
Archaea marker gene relative to bacteria (%)	13.5%	0.25%
Archaea marker gene relative to protozoa (%)	55.6%	nd
DNA Yield	83 µg	

nd = not determined

4.1.2 Analysis of the protozoa-associated symbiont community

To facilitate downstream taxonomic assignment of sequences in the metagenomic phage display library, 16S rRNA gene sequences present in the metagenomic DNA sample were analysed by pyrosequencing to identify the microbial species represented in the library. In total, 1453 sequencing reads were obtained from 16S rRNA gene sequencing of methanogenic archaea present in the metagenomic DNA sample, and species in the genus *Methanobrevibacter* accounted for more than 88% of the archaeal sequences. The genus *Methanosphaera* represented ~8% and species in the family Methanomassiliicoccaceae represented less than 3% of the methanogen gene sequences (Table 4.4).

Table 4.4. Distribution of archaeal species present in the DNA sample derived from protozoa-associated symbionts.

Family	Species	Sequencing reads (%)
	Unassigned	0.34
Methanobacteriaceae	*Methanobrevibacter gottshalkii* clade	41.57
	Methanobrevibacter ruminantium clade	45.98
	Methanobrevibacter smithii	1.03
	Methanosphaera sp. Group 5	3.58
	Methanosphaera sp. ISO3-F5	4.68
Methanomassiliicoccaceae	Group 8 sp. WGK1	1.58
	Group 9 sp. ISO4-G1	0.34
	Group 10 sp.	0.28
	Group 11 sp. BRNA1	0.07
	Group 12 sp. ISO4-H5	0.55

Note: Taxonomic assignment of archaeal 16S rRNA gene sequences were based on Greengenes database, as described in Henderson *et al.* (2015).

To identify methanogen taxa that may be protozoa-associated, the protozoa-associated symbiont community would ideally be compared with rumen contents collected in the same experiment prior to protozoa enrichment. In the absence of this data set, sequence data from a study by Henderson *et al.* (2013), was used as a reference. Both samples were extracted from the rumen of wether sheep on ryegrass pasture that was housed at the same facility in the same season [September 2009 for Henderson *et al.* (2013), August 2012 for this study]. The cell lysis procedure for DNA extraction differed (heating to 95 °C in Henderson *et al.* (2013), freeze-grinding in this study); however, QIAGEN columns were used for DNA purification in both studies. These data sets were also compared with rumen content samples collected from sheep on pasture diet in other geographic locations (Henderson *et al.*, 2015).

Compared to the reference rumen content sample from an NZ sheep (Henderson *et al.*, 2013), a notable difference is the increased abundance of 16S rRNA gene sequences from the *Methanobrevibacter gottschalkii* clade in the protozoa-associated sample (Figure 4.3, top panel). Sequences that belong to *Methanomassiliicoccales* spp., *Methanosphaera* spp., and *Methanobrevibacter ruminantium* clade were more abundant in the rumen content sample.

When compared to rumen contents of sheep on the same diet (pasture diet) from the global dataset, the proportion of sequences representing the *Methanobrevibacter gottschalkii* clade in the protozoa-associated sample clustered with rumen content samples containing a low abundance of *M. gottschalkii* clade members (found in the first quartile of the global data set). In contrast, the protozoa-associated sample clustered with rumen content samples containing a high proportion of sequences from the *Methanobrevibacter ruminantium* clade (found in the third quartile of the global data set). However, the proportion of *Methanobrevibacter ruminantium* clade sequences in the reference NZ rumen contents sample is still higher than in the protozoa-associated sample, as the reference rumen contents sample clustered with data points above the third quartile (Figure 4.3, bottom panel), which suggests that high abundance of *M. ruminantium* clade sequences may not be a characteristic that distinguishes the protozoa-associated methanogen community from the methanogen community found in rumen contents.

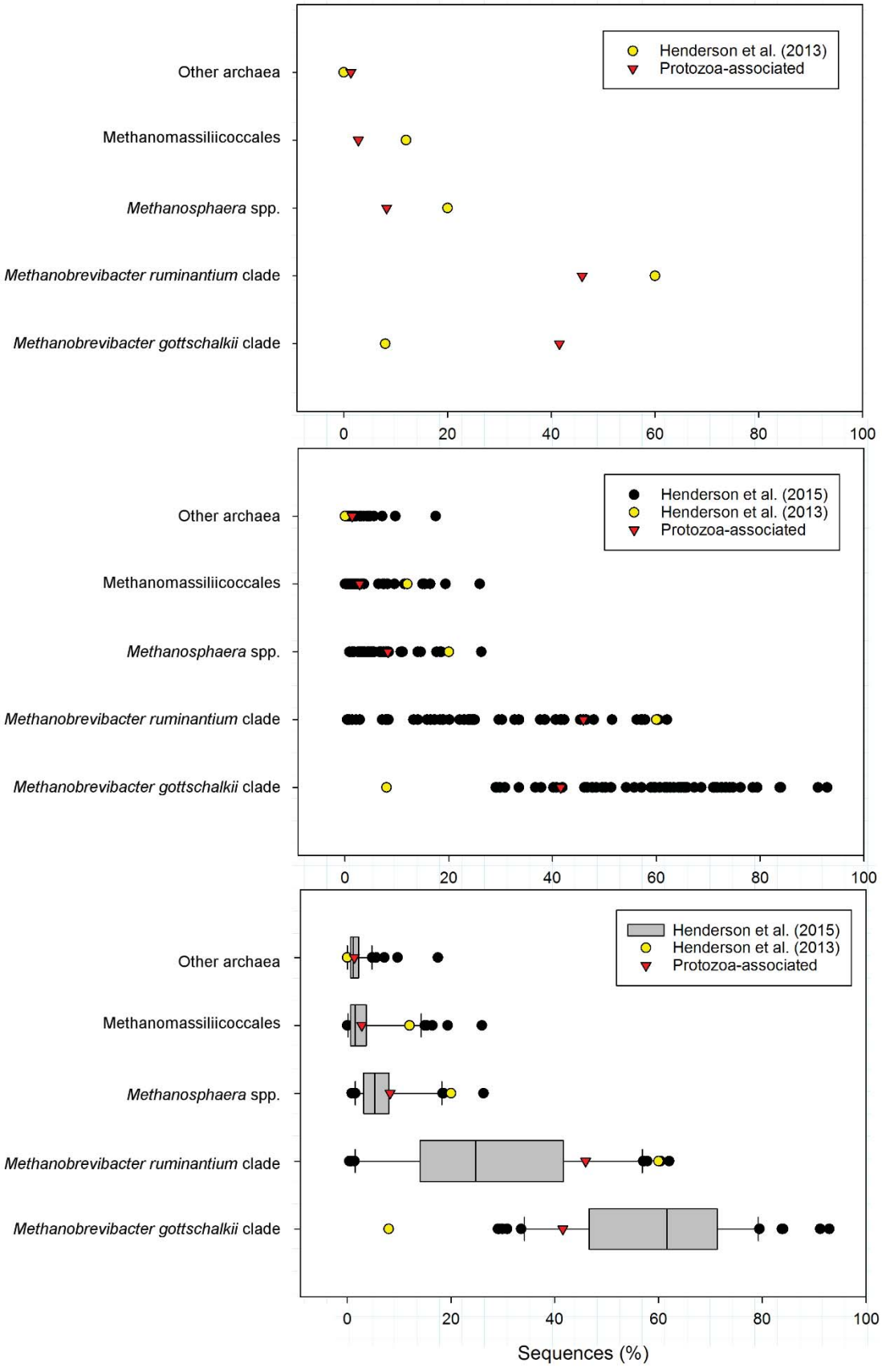

Henderson et al. (2013)
Protozoa-associated
Other archaea
Methanomassiliicoccales
Methanosphaera spp.
Methanobrevibacter ruminantium clade
Methanobrevibacter gottschalkii clade
0
20
40
60
80
100
Henderson et al. (2015)
Henderson et al. (2013)
Protozoa-associated
Other archaea
Methanomassiliicoccales
Methanosphaera spp.
Methanobrevibacter ruminantium clade
Methanobrevibacter gottschalkii clade
0
20
40
60
80
100
Henderson et al. (2015)
Henderson et al. (2013)
Protozoa-associated
Other archaea
Methanomassiliicoccales
Methanosphaera spp.
Methanobrevibacter ruminantium clade
Methanobrevibacter gottschalkii clade
0
20
40
60
80
100
Sequences (%)

Figure 4.3. Methanogenic archaeal taxa represented in the protozoa-associated sample *vs.* rumen contents.

The top panel show the abundance of each methanogenic archaeal taxon found in the sheep rumen contents by Henderson *et al.* (2013) (yellow circles) and the protozoa-associated community (red triangles). These data were overlaid with sheep rumen contents data from a global census (solid dots) in the middle panel. In the bottom panel, the global census data were represented by box plots, where the median and the interquartile range (IQR, all data points residing within the second and third quartiles) were represented by the shaded box, the bars indicated 1.5× IQR, and solid dots indicated data points that are outside of 1.5× IQR.

With respect to bacteria present in the protozoa-associated sample, a total of 18,788 sequencing reads were obtained, and sequences belonging to the Firmicutes phylum dominated the sequence reads (~75% of sequences), whereas Bacteroidetes and Actinobacteria represented 12% and 8.3% of the sequences, respectively (Figure 4.4). When the Firmicutes sequences were further resolved, 73% of sequences were found in the order Clostridiales. The distribution of phyla in the protozoa-associated sample was also compared with bacterial sequences present in the rumen contents (Henderson *et al.*, 2013). In the rumen contents, Bacteroidetes dominated the sequence reads (71.6% of sequences), and Firmicutes represented 22% the sequences. Therefore, sequences assigned to Bacteroidetes and Fibrobacteres were more prevalent in rumen contents (six-fold and 150-fold difference, respectively), whereas bacteria that belong to the phyla Firmicutes, Actinobacteria, and Proteobacteria were more prevalent in the protozoa-associated sample (three-fold, 300-fold, and 90-fold differences were observed, respectively) (Figure 4.4 and Appendix Table 7.1).

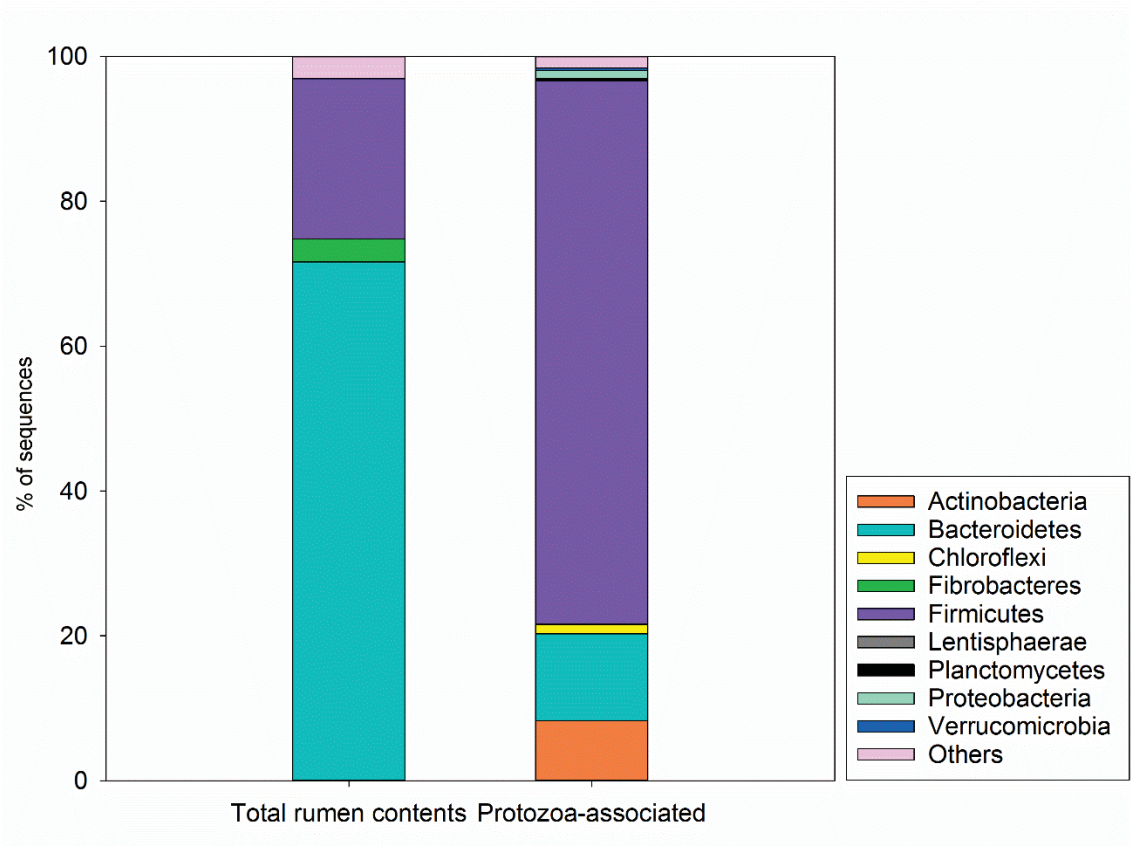


Figure 4.4. Distribution of bacterial phyla in the rumen *versus* protozoa-associated fraction.

Taxa that represented at least 1% of sequences in either the rumen content sample (Henderson *et al.*, 2013) or protozoa-associated sample are listed in Table 4.5. Their distributions were compared to determine whether specific bacterial taxa may be associated with protozoa. The listed taxa (Table 4.5) accounted for 94% and 92% of the total sequence reads in the rumen content and the protozoa-associated samples, respectively. The genus *Prevotella* constituted the majority of bacterial sequences in the rumen content sample (61% of total sequence reads; 20-fold greater than the number of reads in the protozoa-associated sample); whereas unclassified sequences that belong to the class *Clostridiales* were predominant in the protozoa-associated sample (27% of total sequence reads; 10-fold greater than the number of reads in the rumen content sample). Differences were also observed between these datasets at the lower taxonomic levels. In

the protozoa-associated sample, sequences belonging to unclassified members of the *Ruminococcaceae* family, *Butyrivibrio* genus, and *Ruminococcus* genus were more abundant by 9-fold, 4.5-fold, and 3.6-fold, respectively. Furthermore, sequences corresponding to unclassified members of the families *Mogibacteriaceae*, *Coriobacteriaceae*, *Christensenellaceae* and BS11, the genera *Mogibacterium*, SHD-231, and *Blautia* were only observed in the protozoa-associated sample. In the rumen content sample, unclassified members of the family *Veillonellaceae*, and the genera *Bacteroides* and *Pseudobutyrivibrio* each occupied 1-3% of the total gene sequences; however, these sequences accounted for less than 0.1% of the protozoa-associated sample. The genera *Fibrobacter*, *Parabacteroides*, and *Sharpea* were unique to the rumen content sample.

Table 4.5. List of bacteria genera that represent more than 1% of sequence reads in either the protozoa-associated bacteria or total rumen content sample.

Taxon*	Sequence reads (%)	
	Rumen contents (Henderson *et al.*, 2013)	Protozoa-associated
Clostridiales, unknown family and genus affiliations	2.5	27
Ruminococcaceae, unknown genus affiliations	1.0	9.7
Lachnospiraceae, unknown genus affiliations	7.4	9.3
Mogibacterium	0	8.8
Coriobacteriaceae, unknown genus affiliations	0.01	7.9
Bacteroidales, unknown family and genus affiliations	6.6	6.7
Butyrivibrio	1.4	6.4
Mogibacteriaceae, unknown genus affiliations	0	4.2
Prevotella	61	2.9
Ruminococcus	0.6	2.1
Bulleidia	0.3	1.4
SHD-231	0	1.3
Christensenellaceae, unknown genus affiliations	0	1.2
Blautia	0	1.2
BS11, unknown genus affiliations	0	1.1
Coprococcus	1.8	0.92
Veillonellaceae, unknown genus affiliations	1.7	0.043
Bacteroides	1.2	0.011
Pseudobutyrivibrio	2.9	0.005
Fibrobacter	3.2	0
Parabacteroides	1.6	0
Sharpea	1.2	0
Other taxa	5.6	7.8

*Taxonomic assignment of bacterial 16S rRNA gene sequences were based on Greengenes database, as described in Henderson *et al.* (2015).

These data sets were also analysed against rumen content samples collected from sheep on pasture diet in other geographic locations (Henderson *et al.*, 2015) (Figure 4.5). Comparisons against the rumen global census data set revealed that sequences corresponding to the genera *Mogibacterium*, *Blautia*, *Bulleidia*, and unclassified members of the family *Coriobacteriaceae* were over-represented in the protozoa-associated sample, as the abundances of these taxa are greater than the highest outlier in the global data set. For sequences corresponding to unclassified members of the family *Mogibacteriaceae* and the genus *Butyrivibrio*, their abundances in the protozoa-associated sample clustered with data points above the upper quartile (high abundance) in the global data set, whereas their abundances in the rumen content sample were found within the lower quartile samples (low abundance) in the global data set. This observation suggests that these taxa are more abundant in the protozoa-associated symbiont community.

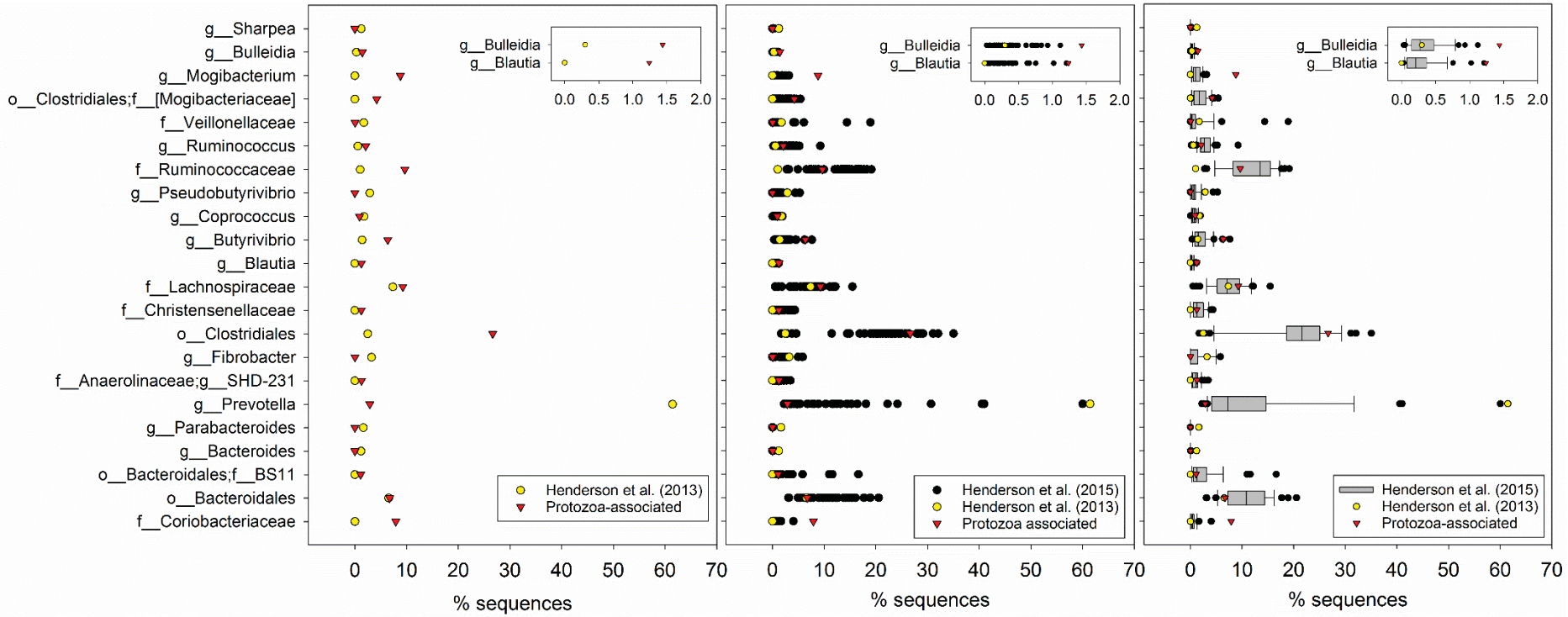


Figure 4.5. Abundance of dominant bacterial taxa (>1% of sequence reads in sample) represented in protozoa-associated sample *versus* rumen contents collected from sheep on pasture diet.

The left panel show the abundance of each bacterial taxon found in the sheep rumen contents by Henderson *et al.* (2013) (yellow circles) and the protozoa-associated community (red triangles). These data were overlaid with sheep rumen contents data from a global census (solid dots) in the middle panel. In the rightmost panel, the global census data were represented by box plots, where the median and the interquartile range (IQR, containing all data points in the second and third quartiles) were represented by the shaded box, the bars indicated 1.5× IQR, and solid dots indicated data points that are outside of 1.5× IQR. The inset panels show two genera that have protozoa-associated abundances outside of the IQR but could not be visualized in the chart with a larger scale.

The aforementioned taxa were not assigned to particular genera in the Greengenes database; therefore, representative sequences from each operational taxonomic unit (OTU) were queried against the NCBI non-redundant database to determine their closest relatives (Table 4.6). Assuming a threshold of 93% identity (Kenters *et al.*, 2011), sequences in five out of 14 OTUs without family assignments in the order *Clostridiales* belong to the same genus as rumen bacterium R-7. Sequences that belong to these OTUs account for 44% of the unclassified *Clostridiales* sequences. Relatives of rumen bacterium R-7 are generally abundant in rumen contents (Henderson *et al.*, 2015), and R-7 has been isolated as a plant/cellulose-adherent species (Noel, 2013) (personal communication, Shinkai T).

Members of OTUs in *Coriobacteriaceae* without genus affiliations were found to have at least 95% sequence identity to *Olsenella* spp. The role of this genus in the rumen has not been well-studied. They were more abundant in cattle consuming a high-starch diet, and this difference was even more pronounced when the high-starch diet was supplemented with 5% sunflower oil (Zened *et al.*, 2013). This is consistent with the observation that lipase activity was detected in many members of the *Atopobium* cluster within *Coriobacteriaceae*, including several *Olsenella umbonata* and *Olsenella profusa* strains (in particular, *Olsenella umbonata* A2, which was isolated from sheep rumen), as well as *Eggerthellaceae* spp. and *Colinsella* spp. isolates (Kraatz *et al.*, 2011, Thorasin *et al.*, 2015). Another interesting observation is that *Olsenella* spp. are mucin-utilizing and peptidolytic bacteria (Kraatz *et al.*, 2011), and they have also been identified as one of the predominant plant-adherent bacterial species during an *in sacco* study of perennial ryegrass degradation (Huws *et al.*, 2016).

With the exception of one OTU, sequences in the family *Mogibacteriaceae* which were not assigned to a particular genus are distantly related to *Eubacterium* spp. or *Anaerovorax*

odorimutans. The remaining OTU has 99% sequence identity with *Eubacterium* sp. C2. The high level of nucleotide sequence identity suggests that they are the same species. The relative *Anaerovorax odorimutans* uses putrescine as a carbon source, and the tested isolates were found to produce H_2 in pure culture (Schink, 2009). Little is known about species in the *Eubacterium* genus. *Eubacterium nodatum, Eubacterium brachy*, and *Eubacterium infirmum* have been isolated from human periodontal patients, and they were shown to be slow-growing asaccharolytic species (Hill *et al.*, 1987, Cheeseman *et al.*, 1996). Several species from *Eubacterium* have been noted to produce H_2, including *E. brachy* (Schink, 2009). A related isolate, denoted AD3011, has been cultured from the plant-adherent fraction of rumen contents previously, but it has not been characterized (Noel, 2013).

Lastly, OTUs in the unclassified sequences of family *Ruminococcaceae* were also analysed, as its abundance is 9-fold greater than the rumen sample derived from an NZ sheep housed in the same location consuming the same diet. Sequences related to strain NK4A214 of family *Ruminococcaceae* to varying degrees were identified (nucleotide sequence identities ranged from 92-95%). One OTU matched strain NK3A39 [closely related to *Ruminococcus albus*, a H_2 producer in the rumen (Kenters *et al.*, 2011, Zheng *et al.*, 2014)] with 100% identity. Only distant relatives could be identified for the remaining two OTUs.

Table 4.6. BLASTN results for sequences representing OTUs without genus assignments that were more prevalent in the protozoa-associated bacterial community.

OTU	Number of sequences	Closest relative	% identity
Clostridiales, unknown family and genus affiliations			
denovo8127	38	Rumen bacterium R-7	88
denovo6369	40	Rumen bacterium R-7	87
denovo5416	23	Rumen bacterium R-7	87
denovo2147	33	Rumen bacterium R-7	86
denovo2385	21	Rumen bacterium R-7	90
denovo3795	51	Rumen bacterium R-7	97
denovo4226	31	Rumen bacterium R-7	89
denovo71	32	Bacterium YE57	87
denovo2276	162	Rumen bacterium R-7	96
denovo1194	37	Rumen bacterium R-7	96
denovo8136	26	Rumen bacterium R-7	95
denovo848	110	*Ruminococcus gnavus* strain A2	92
denovo5536	22	Rumen bacterium R-7	95
denovo8143	22	*Syntrophococcus sucromutans* strain DSM 3224	91
denovo7764	29	*Syntrophococcus sucromutans* strain S195	92
Coriobacteriaceae, unknown genus affiliations			
denovo775	331	*Olsenella umbonata* strain lac15	95
denovo5738	66	*Olsenella* sp. F0004	95
denovo2731	33	*Olsenella umbonata* strain lac15	99
denovo4275	22	*Olsenella* sp. F0004	96
denovo7526	238	*Olsenella* sp. F0004	96
denovo7825	57	*Olsenella umbonata* strain lac15	95
Mogibacteriaceae, unknown genus affiliations			
denovo1130	21	*Eubacterium* sp. C2	99
denovo5056	57	*Eubacterium* sp. WAL 18692	88
denovo7373	41	*Anaerovorax odorimutans* strain NorPut	88
denovo7964	31	*Anaerovorax odorimutans* strain NorPut	88
denovo550	24	Eubacterium sp. WAL 17363	89
Ruminococcaceae, unknown genus affiliations			
denovo4212	21	*Saccharofermentans* sp. CA24	83
denovo2497	30	*Ruminococcaceae* bacterium GD5	87
denovo4712	32	Rumen bacterium NK4A214	95
denovo1885	41	Rumen bacterium NK3A39	100
denovo3318	39	Rumen bacterium NK4A214	94
denovo2983	28	Rumen bacterium NK4A214	92

4.1.2.1 Comparison of protozoa-adherent bacterial community with rumen epithelium-adherent bacterial community

The bovine rumen epithelium-adherent bacterial community has been described previously, and it was shown that the rumen epithelium-adherent community differs significantly from the bacterial community found in rumen contents (Li *et al.*, 2012a). Rumen protozoa and ruminant hosts are both eukaryotic in origin; therefore, we hypothesized that there may be common features (protein receptors or sugar modifications) shared between the surfaces of protozoal cells and rumen epithelial cells that allow for interactions with bacteria. If a set of common features exists, we would expect to find bacterial species that are predominantly associated with eukaryotic cell surfaces (protozoal and rumen epithelial cells) in comparison to their presence in rumen contents.

Comparative analysis of bacterial species found in rumen contents versus rumen epithelium from three steers by 16S rRNA gene sequencing published by Li et al (2012) revealed that the phylum Bacteroidetes is dominant in rumen contents, whereas Firmicutes is dominant in the rumen epithelium-adherent community (Li *et al.*, 2012a). The genus *Mogibacterium* was over-represented in the rumen epithelium for two of three animals. The genus *Prevotella* was over-represented in rumen contents for all three animals, and this difference was statistically significant when the sample size was increased to 22 animals. We observed similar trends in comparative analysis between rumen contents and protozoa-associated bacterial communities (Figure 4.5). To verify whether these observations remain consistent when both sets of sequence data are analysed using the same method and against the same reference database, the sequence data published by Li et al (2012) were extracted from the NCBI database and analysed by QIIME. Indeed, 16S rRNA gene analysis using QIIME also showed that sequences

corresponding to the *Mogibacterium* genus were more abundant in the rumen epithelium, whereas *Prevotella* were more abundant in rumen contents (Figure 4.6).

Differences between the rumen epithelial-adherent bacterial community and the protozoa-associated community were also observed (Figure 4.6). Unclassified sequences in the order Clostridiales accounted for 27% of all sequences found in the protozoa-associated community, but they contributed to less than 5% of sequences in the rumen epithelium-adherent community. For unclassified sequences in the order Bacteroidales, similar proportions of these sequences were found in the bacterial community present in rumen contents and the community associated with rumen protozoa (6.6% and 6.7%, respectively). In contrast, this taxon was under-represented in the rumen epithelium-adherent community compared to the corresponding rumen contents samples. Lastly, unclassified sequences in the family *Coriobacteriaceae* represented less than 1% of the rumen epithelium-adherent community, in contrast with 7% of the protozoa-associated community. Again, it is important to note that these trends should be further verified by comparing the protozoa-associated symbiont community with a control derived from the same rumen contents sample.

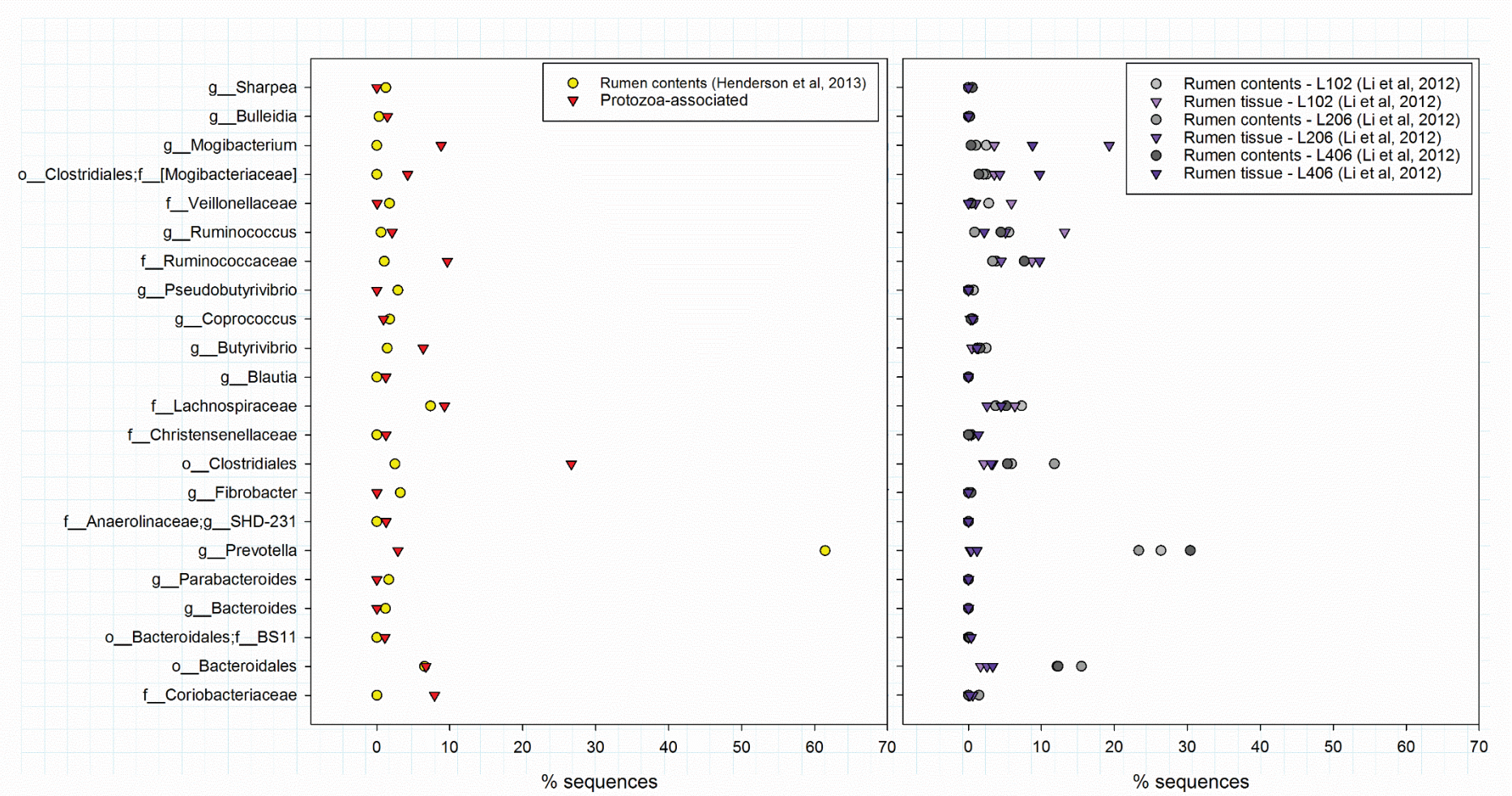


Figure 4.6. Similarities between the protozoa-associated and rumen epithelium-adherent bacterial communities.

4.2 Metagenomic phage display library construction and affinity screening

In addition to characterizing the protozoa-associated microbial community, we also aimed to use phage display technology to identify adhesion proteins encoded by these organisms that allow them to attach to their protozoal hosts. The metagenomic DNA sample isolated from protozoa-associated symbionts was mechanically sheared to generate random inserts (size range from 1 kb to 3.5 kb) and then cloned into the shot-gun phage display vector pYW01 (Figure 4.7). The primary *E. coli* TG1 recombinant library contained 1.2×10^8 independent transformants. PCR amplification of the inserts in the primary library generated amplicons with a distinct band at around 0.3 kb, and a smear corresponding to 0.45 to 1.5 kb of the size marker (Figure 4.8). The amplicon band around 0.3 kb resulted from clones containing empty vector and the smear resulted from clones containing inserts of varying sizes.

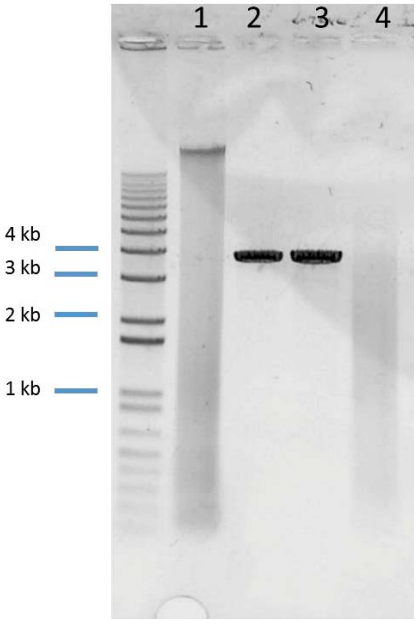


Figure 4.7. Phage display vector and metagenomic DNA samples used in metagenomic library preparation.

Lane 1: Metagenomic DNA isolated from protozoa-associated microbes; lane 2: pYW01 phage display vector digested with restriction endonuclease *Sma*I; lane 3: pYW01 phage display vector digested after treatment with restriction endonuclease *Sma*I and alkaline phosphatase; lane 4: metagenomic DNA after shearing by nebulization for ten seconds.

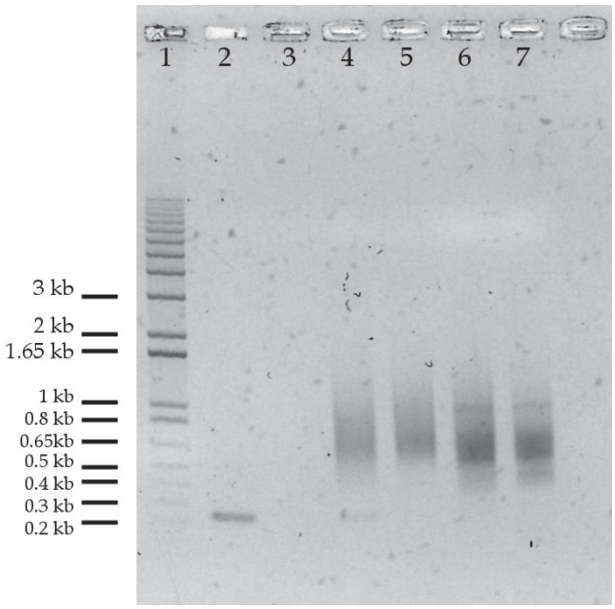


Figure 4.8. PCR amplicons of metagenomic phage display library after each screening step.

PCR amplification with primer pair PelBF1/pspR03 was performed on the positive control pYW01 plasmid DNA (lane 2), negative control pDJ01 plasmid DNA (a phage display vector that does not harbour the PelB secretion signal, and therefore the sequence corresponding to the primer PelBF1 is absent from this vector (Jankovic *et al*., 2007)) (lane 3), plasmid DNA isolated from the primary library (lane 4), the library after screening against c-myc antibody (lane 5), after the first round of screening against protozoa (lane 6), and after the second round of screening against protozoa (lane 7). Lane 1 contains the DNA ladder.

Prior to panning against protozoa, PPs derived from the primary phage display library were screened against c-myc antibody to enrich for PPs that display recombinant c-myc/pIII fusion proteins. PPs eluted after panning against c-myc antibody were amplified in the host *E. coli* TG1 strain, and the enriched PPs were then affinity selected against the rumen protozoa used as bait. Two rounds of affinity selection against protozoa were performed. In the biopanning rounds where protozoa were used as bait, the number of PPs eluted decreased from the first to the second round of biopanning (Table 4.7). The output:input ratio is expected to increase from one biopanning round to the next when adhesin-encoding PPs have been enriched. After the first biopanning

against protozoa, the number of recovered PPs was 3×10^6 and in the second round was 1×10^4, indicating that non-binders were removed from the system, but the input population in both biopanning rounds have not been enriched for protozoa-binding adhesins. The accumulation of background clones that do not encode for recombinant c-myc/pIII fusion proteins has been noted in the past (Easton, 2009, Gupta *et al.*, 2013, Ciric, 2014); therefore, affinity screening against c-myc antibody was performed on PPs eluted from Round 2 biopanning prior to sample preparation for PacBio sequencing.

Table 4.7. Number of phagemid particles that have bound to bait during biopanning.

Ligand/bait	Number of PPs used as input for screen (cfu)	Number of PPs eluted as output (cfu)	Output:input ratio
c-myc	6×10^{11}	3×10^9	5×10^{-3}
Protozoa (round 1)	5×10^{11}	3×10^6	6×10^{-6}
Protozoa (round 2)	2×10^{12}	1×10^4	5×10^{-9}

4.2.1 PacBio sequencing

Upon initial analysis of the data generated by PacBio sequencing, it was apparent that some reads contained concatemers of two or more amplicons, which were likely introduced during the adapter ligation stage of sequencing library preparation; therefore, an additional data processing step was introduced to split reads that contain multiple amplicons such that each sequence corresponds to a single amplicon (Figure 4.9). Amplicons that encode inserts less than 60 nucleotides and amplicons that correspond to non-specific PCR amplification products (6.7 to 13.2% of amplicons in each library) were also eliminated from further analyses. Data on the number of sequences recovered after each data processing step are detailed in Table 4.8.

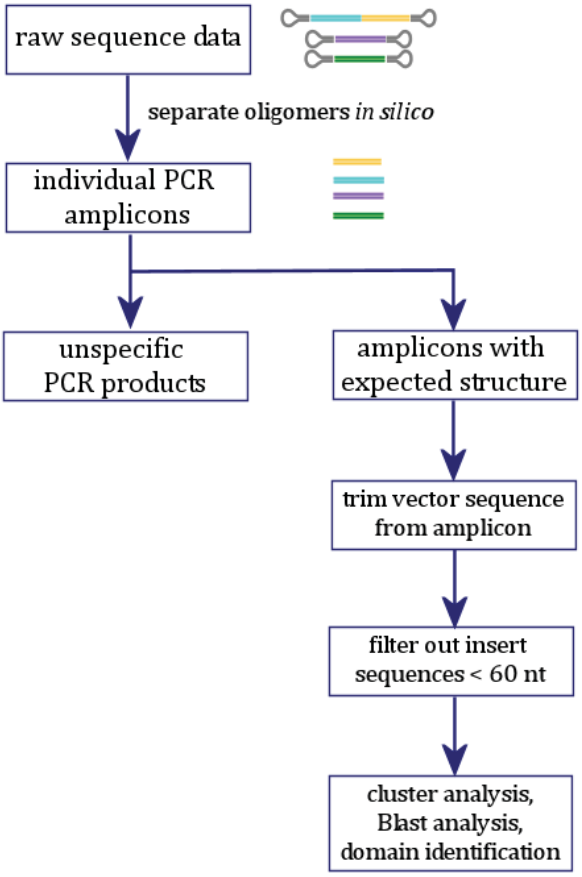


Figure 4.9. Workflow for processing raw data from PacBio sequencing.

Schematic drawings represent PCR amplicons (coloured lines) and sequencing adapters (grey hairpin loops).

Table 4.8. Summary of number of sequences recovered after each data processing step.

	Primary library (PAM1)	After c-myc screening (PAM2)	After panning against protozoa bait (two rounds) (PAM_FINAL)
Number of bases read	2.2 Mb	2.7 Mb	15.6 Mb
Number of reads	4112	5015	22,527
Mean # passes	37	38	28
Mean read quality	99.0%	99.2%	99.4%
Number of amplicons analyzed*	4160	5259	22,239
Unspecific PCR products	213 (5.1%)	59 (1.1%)	2353 (11%)
Amplicons <60bp	338 (8.1%)	295 (5.6%)	14 (0.063%)
Number of amplicons after filtering out unspecific PCR products and short sequences	3609 (86.8%)	4905 (93.3%)	19,873 (89.4%)
Average length of insert within amplicon	280 nt	274 nt	482 nt
Number of unique reads	3597 (99.7% of filtered amplicons)	4889 (99.7% of filtered amplicons)	1910 (9.6% of filtered amplicons)
Number of clusters identified in library (2 or more sequences in cluster; sequence identity > 90%)	6	8	989

*For libraries PAM1 and PAM2, sequence reads containing concatemers of PCR amplicons were separated to maximize the amount of sequence information available for analysis. For library PAM_FINAL, sequence reads containing concatemers of PCR amplicons were discarded (as this library contained five-fold more reads than PAM1 and PAM2 without processing).

4.2.1.1 Primary library sequence analysis

GAMOLA was used to perform BLASTX analysis of metagenomic insert sequences (Altermann & Klaenhammer, 2003). Preliminary taxonomic assignment was based on BLASTX hits to reference organisms with an E-value threshold of 1e-08, and this analysis suggested that the proportion of methanogen sequences relative to bacterial sequences in the primary library was 4%. This proportion is lower than expected, as qPCR analysis of small subunit ribosomal gene sequences present in the metagenomic DNA sample had indicated that the proportion of archaeal to bacterial 16S rRNA gene sequences was 13% (Table 4.3). The difference observed between qPCR analysis and *in silico* analysis of metagenomic insert sequences can be attributed to the insufficient representation of sequenced methanogen genomes (a lack of sequence information would lead to difficulties in assigning metagenomic sequences to methanogen species by computational techniques).

We examined the primary library to determine whether specific functional domains are over-represented by the protozoa-associated symbiont community. ORFs identified in the primary library were annotated against the Pfam database using the IMG annotation pipeline (Markowitz *et al.*, 2015), and 884 domains were identified (27.9% of filtered amplicon sequences). The ten most abundant Pfam domains identified are presented in Figure 4.10. Out of these 10 domains, the two most abundant domains corresponded to protein kinase (2.9%) and Ras superfamily of small GTPase (2.5%), which relate to cell signalling and signal transduction functions (Kyriakis, 2014, Wuichet & Sogaard-Andersen, 2015). Two Pfam transporter domains, pfam00005 and pfam01554, represented 1.2% and 0.57% of the identified domains, respectively. The ABC transporter domain, pfam00005, has been noted as an abundant domain in rumen metagenomic samples in previous studies (Li *et al.*, 2012b, Li *et al.*, 2014). Protein domains relating to structural components of the cell, actin and multiple membrane occupation

and recognition nexus (MORN) protein, accounted respectively for 0.57% and 0.68% of the identified domains. Interestingly, three domains (DUF285, leucine-rich repeats, and WD domain) found in proteins implicated in cell adhesion together accounted for 3.6% of the Pfam domains identified in the dataset. DUF285 is a conserved domain found in predicted surface proteins of several bacterial species, as described by the Interpro database (Hunter *et al.*, 2012). This domain with unknown function was also found in predicted ALPs of *M. ruminantium* M1 (Table 1.5). Leucine-rich repeats are widely distributed amongst proteins with diverse functions in eukaryotes, bacteria, and archaea (Bober *et al.*, 2011). Bacterial surface proteins with leucine-rich repeats have been implicated in their interactions with eukaryotic hosts (Loimaranta *et al.*, 2009, Bober *et al.*, 2011, Ishida *et al.*, 2014). Similar to leucine-rich repeats, the WD repeat domain has also been found in proteins with diverse functions, including proteins with roles in cell adhesion and protein-protein interactions (Smith *et al.*, 1999).

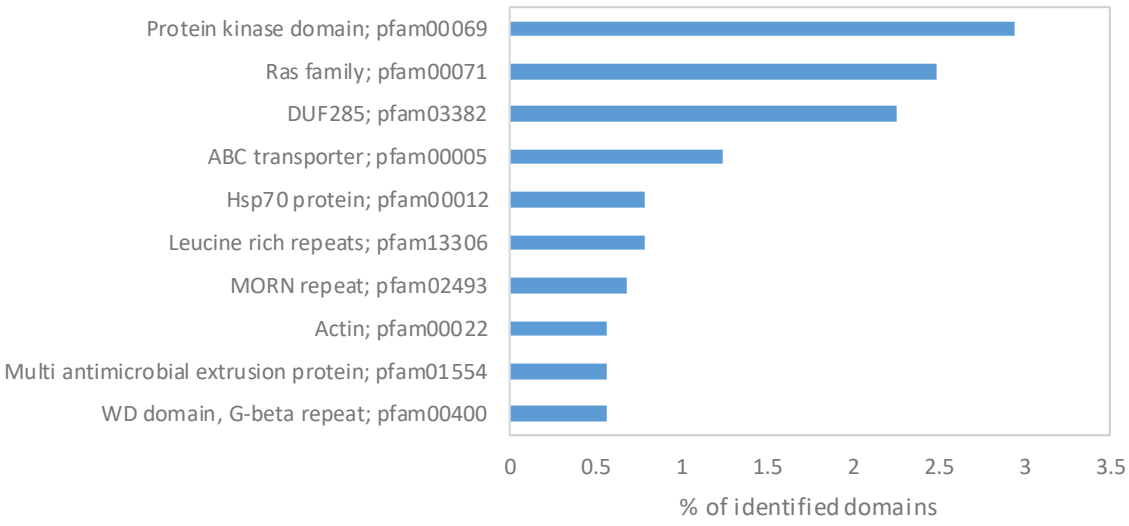


Figure 4.10. Ten most abundant Pfam domains annotated for sequence reads in shotgun metagenomic library derived from protozoa-associated symbionts.

4.2.1.2 Comparisons between libraries

The libraries will henceforth be denoted as PAM1 for the primary library, PAM2 for the primary library after it has undergone affinity screening against c-myc, and PAM_FINAL for the library after two rounds of biopanning against protozoal bait. The average insert size was 280 nucleotides in PAM1, 274 nucleotides in PAM2, and increased to 482 nucleotides in PAM_FINAL (after two rounds of biopanning against protozoal bait). Cluster analysis was performed to identify redundant reads within each library, and thereby determine whether particular sequences were over-represented within the libraries or enriched after affinity selection. Analysis of each library individually revealed that most of the reads (99.7%) were unique for the libraries PAM1 and PAM2, which also indicates that the sequencing depth in this experiment was insufficient to capture the complete diversity of sequences in these libraries. As the size of the primary library was 1.2×10^8 and the estimated size of the PAM2 library is 7.2×10^6 (based on the probability that one out of 18 blunt ended insert sequences generated from random shearing would be cloned in frame with the phagemid vector PelB signal sequence sequence at the 5´ end and the fusion proteins c-myc and pIII at the 3´ end), and the number of sequencing reads acquired were around 4500, it was not surprising that few redundant sequences were detected. In PAM_FINAL, 989 clusters containing at least two sequences were identified, i.e. this library contained the largest number of redundant amplicons, which also suggests that the phage display library may be enriched for adhesins. A greater number of sequencing reads was acquired for this library than PAM1 and PAM2 (library loading was optimized for PAM_FINAL, but not for PAM1 and PAM2 due to constraints in time and resources); therefore, cluster analysis was performed on a randomly generated subset of the PAM_FINAL data in order to demonstrate that more redundant reads are present in this library even when the number of reads in PAM_FINAL is normalized to the number of reads acquired in

PAM1 and PAM2, and the most prominent clusters (encompassing reads that represent at least 1% of the PAM_FINAL library) could still be identified. Table 4.9 shows that all seven of these clusters were detected in both the full data set and the subset of PAM_FINAL with the same ranking, and less than 20% of the amplicon sequences were unique even for the PAM_FINAL data subset. Furthermore, the largest cluster in PAM_FINAL represented almost 40% of the library, whereas clusters identified in PAM1 and PAM2 contained a maximum of two sequences (<0.025%).

Table 4.9. PAM_FINAL library clusters that represent at least 1% of library amplicon sequences.

	PAM_FINAL full data set	PAM_FINAL subset
Number of amplicon sequences analyzed	19,873	4,469
Number of unique reads found	1910 (9.6% of filtered amplicons)	732 (16% of filtered amplicons)
Number of clusters identified in library (2 or more sequences in cluster)	989	223
# reads in Cluster 1 (% of total sequences analyzed)	7309 (37.8%)	1782 (39.9%)
# reads in Cluster 2 (% of total sequences analyzed)	1675 (8.4%)	386 (8.6%)
# reads in Cluster 3 (% of total sequences analyzed)	1548 (7.8%)	356 (8.0%)
# reads in Cluster 4 (% of total sequences analyzed)	483 (2.4%)	108 (2.4%)
# reads in Cluster 5 (% of total sequences analyzed)	458 (2.3%)	87 (1.9%)
# reads in Cluster 6 (% of total sequences analyzed)	336 (1.7%)	67 (1.5%)
# reads in Cluster 7 (% of total sequences analyzed)	261 (1.3%)	66 (1.5%)

4.2.1.3 Enriched sequences after biopanning against protozoa as bait

Seven ORFs encoded in clusters representing at least 1% of the total number of sequence reads are listed in Table 4.10. Some methanogenic archaea use the *E. coli* stop codons TGA and TAG for translation to selenocysteine and pyrrolysine, respectively (Zhang *et al.*, 2005). Alternate codon usage in archaea is irrelevant for the protein translations presented in Table 4.10, as these insert sequences do not contain stop codons. The seven ORFs were over-represented in the PAM_FINAL library, and they were not found in the libraries PAM1 and PAM2. Assembly of sequences in individual clusters revealed that all seven ORFs were in frame with both the PelB signal sequence and the recombinant fusion c-myc/pIII fusion proteins. To determine whether common epitopes may exist, *in silico* analysis was performed on the seven protein sequences using the motif detection tool MEME to identify common motifs (Bailey *et al.*, 2009) (Table 4.11). The most abundant sequences belong to Cluster 1. The recombinant protein has weak similarity to a protein kinase in bacterium P3 and the P_{ad} score assigned by SPAAN is not strongly indicative of adhesion function. Further analysis showed that this protein contains a lysine-rich region (Table 4.11). Clusters 2, 3, and 4 contain ORFs that encode proteins with internal repeats and low complexity regions (Table 4.11). Clusters 2 and 3 have P_{ad} scores > 0.5 and are similar to known surface proteins, which suggests that they likely encode adhesins; furthermore, the encoded proteins contain regions that are rich in asparagine and glutamine residues, respectively. The protein sequence encoded by Cluster 4 has the lowest P_{ad} score of all seven proteins, and contains regions rich in lysine and proline. However, this sequence is also similar to a bacterial cell wall anchor protein. No domains or repeated motifs were noted in Cluster 5, nor does the encoded protein have any similarity to known proteins. It is the only protein sequence that has a BLASTP hit with a stronger match in the methanogen database than the non-redundant database, but the E-value of the methanogen match is high (E-value = 2.9). A repeated motif and

asparagine-rich regions (E-value 5.5e-04) were identified in the ORF encoded by Cluster 6. This protein also has similarity to a putative cell surface protein from *Clostridium* sp. CAG:62. Cluster 7 encodes a protein with 82% amino acid identity (matched to 132 amino acids) to an annotated periplasmic sugar-binding ABC transporter in *Oscillibacter* sp. CAG:155 (366 amino acids). This protein likely functions as an adhesin, as Pfam domain PF13407 (periplasmic binding protein-like) was identified and it was assigned a P_{ad} score > 0.5. To verify that these sequences truly encode protozoa-binding adhesins, they can be cloned into the phage display vector for PP production, and the PPs displaying the corresponding recombinant fusion proteins can be tested by affinity binding assays against protozoa as bait.

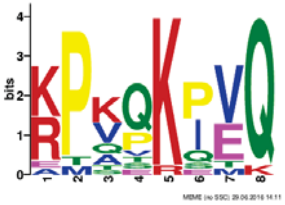
Table 4.10. ***In silico*** **homology search results for enriched amplicons in PAM_FINAL library.**

Cluster number	Encoded ORF	# residues	Top BLASTP hit *(Informative BLASTP hit)*	E value	Database for top BLASTP hit	Domains identified by Interpro or SMART database searches
1	LAKLGLQEEGNKWMNKEMSVFPASTTYVKALKTKY AGAYASDCSSLPSSRKTDEKTLASFMFNDADEELE PVFPDVSDEFSDDPASIVVTVDQNPEMSEKEVQAL ERSKYAKTDEEKALSAEERAKAKKQAAKEKEKAKK QAAKEKEKAKKQAAKEKAKAKKQAEKEKAIA	171	CDC-like kinase [bacterium P3]	0.27	nr	Coil predicted for amino acids 114..170 DUF4604 (23..170; E-value 1.9E-3)
2	TYNNSSINKPSNGSINKPNNGSINKPNNGSVNKPN NGSVNKPNNGSINKPNNGSINKPNNGSVNKPNNGS INKPSNGSINKPINGSINKPSNGSINKPSNGSINK PNNSSNNISSINKHNNSSINKPNNGSINKPNNGSI NKP	143	PREDICTED: hyphally regulated cell wall protein 3-like [*Coturnix japonica*] *(Surface antigen ariel1 [Entamoeba histolytica HM-1:1MSS]*	1E-22 (6E-19)	nr	Low complexity region (96..143) Similarity to cornifin, eukaryotic proline rich protein with role in cell differentiation (3..121; E-value 9.4e-5)
3	IARNNPSAGAVNPKSAPTQRSTQTQGQRPVQKPVQ RTAQSVPQRTAQKPVQGNNQRPVQKPVQRPVQGSS QRPQTKPVQGQRPVQKQMQKPQQGQKPAVKPVQRP ASKPQPKQVQRPVSKPVQKKPEI	128	hypothetical protein NOR_02291 [*Metarhizium rileyi* RCEF 4871] *(Collagen-binding protein [Bacillus thuringiensis], LPxTG-motif cell wall anchor domain protein [Bacillus thuringiensis IBL 200])*	6E-11 (5E-09)	nr	Low complexity region 81..97
4	SESICPENVDFLNISHAYSIPINSFNDLESDRIEM FIQSQPTIEKPKPKQKIEKPKPKPKIEQPKPKQKI EQPKPKPKIEQPKPKQKIEQPKPKPKIEKPKPKPK IEK	108	hypothetical protein [*Crocosphaera watsonii*] *(Cell wall anchor [Bacillus cereus])*	1E-08 *(2E-08)*	nr	Low complexity region 43..108

Table 4.10 (continued)

Cluster number	Encoded ORF	# residues	Top BLASTP hit	E value	Database for top BLASTP hit	Domains identified by Interpro or SMART database searches
5	HDQDQEPQGVVQDPRLPHRNGLPGPHDLSEPGS DHRQADHLRHPQAPGLHRGRGPRPRPGAHAQGS CATKCRCTKSNGMLKALPSTAQMTSSSPLTIRN TPYTARSSSSTPILSGGCRSAYNGWLKQACNKQ RTSQNGSSLSHDTKPY	148	TALC_00648 Arylsulfatase regulator (Fe-S oxidoreductase)	2.9	Methanogens	Low complexity region 47..60
6	NTNTNPRGCTNTNPRGNSNTNPRGCTNTSPRGN TNTYSRGCTNTSPRGNSNTNPRGCTNTYSRGCT NSYSRGCTNSNPRRNSNSYTRGNSNSYTRGNSN	99	Putative uncharacterized protein FLJ44967 [*Cricetulus griseus*] *(Putative cell surface protein [Clostridium sp. CAG:62]; serine repeat antigen 5 [Plasmodium falciparum])*	1E-16 *(3E-04; 5E-04)*	nr	None detected
7	RFENILSSNYADGTKLDAVLASNDSTALGVENA LNSSYTGEWPVITGQDCDIAIMKNLVEGKQSMS VFKDTRTLASKVVEMVDALMQGKEPPVNDTETY DNGTGVIPSFLCEPVACTVENYKELLIDSGYYT MDQIGVDAAAAEEAAPAEEA	152	sugar ABC transporter periplasmic sugar-binding protein [*Oscillibacter* sp. CAG:155]	2E-71	nr	Periplasmic binding protein PF13407 (1..90; E-value 3.6e-13)

Table 4.11. ***In silico*** **analysis of amino acid sequence features encoded by enriched amplicons in PAM_FINAL library.**

Cluster number	Encoded ORF	Estimated pI[a]	Motifs identified by MEME[b]	*p*- value	Sequence characteristics identified by MotifScan[a]	E value	P_{ad} score[c]
1	LAKLGLQEEGNKWMNKEMSVFPASTTYVKAL KTKYAGAYASDCSSLPSSRKTDEKTLASFMF NDADEELEPVFPDVSDEFSDDPASIVVTVDQ NPEMSEKEVQALERSKYAKTDEEKALSAEER AKAKKQAAKEKEKAKKQAAKEKEKAKKQAAK EKAKAKKQAEKEKAIA	8.4		6.0e-2	Lysine-rich region (109..168)	9.1e-06	0.446
2	TYNNSSINKP SNGSINKP NNGSINKP NNGSVNKP NNGSVNKP NNGSINKP NNGSINKP NNGSVNKP NNGSINKP SNGSINKP INGSINKP SNGSINKP SNGSINKP NNSSN NISSINKH NNSSINKP NNGSINKP NNGSINKP	10.9		5.6e-5	Asparagine-rich region (3..141)	6.8e-23	0.676
3	IARNNPSAGAVNPKS APTQRSTQ TQGQ RPVQKPVQ RTAQSVPQ RTAQKPVQ GNNQ RPVQKPVQ RPVQGSSQ RPQTKPVQ GQ RPVQKQMQ KPQQGQ KPAVKPVQ RPAS KPQPKQVQ RPVSKPVQ KKPEI	12.7		4.0e-6	Glutamine-rich region (19..115)	1.1e-07	0.688
4	SESICPENVDFLNISHAYSIPINSFNDLESD RIEMFIQSQPTIEKP KPKQKIEK PKPKPKIEQ PKPKQKIEQ PKPKPKIEQ PKPKQKIEQ PKPKPKIEKPKPKPKIEK	9.7		1.6e-4	Lysine-rich region (45..108) Proline-rich region (41..104)	6.5e-07 5.4e-03	0.328

Sequence characteristics were analyzed by *in silico* tools in the ExPASy bioinformatics portal (Artimo *et al.*, 2012) (including MotifScan and pI calculator)[a], MEME[b] (Bailey *et al.*, 2009), and SPAAN[c] (Sachdeva *et al.*, 2005).

Table 4.11 (continued)

Cluster number	Encoded ORF	Estimated pI[a]	Motifs identified by MEME[b]	*p*- value	Sequence characteristics identified by MotifScan[a]	E value	P_{ad} score[c]
5	HDQDQEPQGVVQDPRLPHRNGLPGPHDLSEP GSDHRQADHLRHPQAPGLHRGRGPRPRPGAH AQGSCATKCRCTKSNGMLKALPSTAQMTSSS PLTIRNTPYTARSSSSTPILSGGCRSAYNGW LKQACNKQRTSQNGSSLSHDTKPY	10.0	None detected	---	None	---	0.461
6	NTNTNPRG CTNTNPRG NSNTNPRG CTNTSPRG NTNTYSRG CTNTSPRG NSNTNPRG CTNTYSRG CTNSYSRG CTNSNPRR NSNSYTRG NSNSYTRG NSN	11.1		1.7e-6	Asparagine-rich region (1..99)	5.5e-04	0.512
7	RFENILSSNYADGTKLDAVLASNDSTALGVE NALNSSYTGEWPVITGQDCDIAIMKNLVEGK QSMSVFKDTRTLASKVVEMVDALMQGKEPPV N DTETYDNG TGVIPSFLCEPVACTVENYKELLIDSGYYTM DQIGVDAAAAEEAAPAEEA	4.0		6.7e-3	None	---	0.561

Sequence characteristics were analyzed by *in silico* tools in the ExPASy bioinformatics portal (Artimo *et al.*, 2012) (including MotifScan and pI calculator)[a], MEME[b] (Bailey *et al.*, 2009), and SPAAN[c] (Sachdeva *et al.*, 2005).

4.3 Summary

In this study, we established a method to enrich for protozoa-associated methanogens, performed 16S rRNA gene sequence profiling to characterize the prokaryotic symbionts present in the sample, and created a metagenomic phage display library for adhesin screening from the isolated metagenomic DNA. The protozoa-associated symbiont community was compared with the community present in rumen contents of a sheep at the same facility on the same diet, and the rumen sample was collected in the same season (Henderson *et al.*, 2012). Ideally, the protozoa-associated symbiont community should be compared with the community derived from the same rumen contents sample, but this was not possible due to constraints in time and resources. Results from qPCR showed that DNA derived from methanogenic archaea was enriched by 54-fold in the metagenomic DNA sample, relative to its presence in DNA extracted from rumen contents; however, DNA derived from bacterial symbionts was also present in the sample. 16S rRNA gene sequence profiling by pyrosequencing revealed that the methanogenic archaeal genus *Methanobrevibacter* tended to be protozoa-associated, whereas the *Methanosphaera* and *Methanomassiliicoccales* genera were less frequently associated with protozoa. The protozoa-associated bacterial symbiont community differed from the bacterial community present in rumen contents, but showed some similarity to the community associated with the rumen epithelium. Sequences from the phylum Firmicutes dominated the protozoa-associated community. In particular, sequences from the genera *Blautia, Bulledia, Mogibacterium* and *Butyrivibrio*, as well as unclassified members of the family *Mogibacteriaceae* were over-represented. Sequences corresponding to unclassified members of the family *Coriobacteriaceae* (phylum: Actinobacteria) were also over-represented in this sample.

The metagenomic phage display library contained 1.2×10^8 clones. The library of PPs were pre-selected for PPs displaying recombinant pIII fusion proteins, and then affinity selection was performed using rumen protozoa as bait. PacBio single molecule sequencing technology was used to characterize phage display libraries for the first time. Seven polypeptide sequences were identified as potential adhesins. Four out of the seven potential adhesins contain internal repeat regions, and several contain lysine-rich and asparagine-rich regions. Affinity binding assays will be required to confirm the adhesion function of these proteins.

Chapter 5. Discussion and Conclusions

Physical associations between methanogens and protozoa are commonly found in the rumen ecosystem and contribute to ruminant enteric methane emission. The aim of this project was to identify methanogen cell surface proteins involved in attachment to protozoal hosts. The protozoa-associated methanogen community contains multiple species from several genera; however, cultured isolates are only available for a few methanogen species. The genus *Methanobrevibacter* is predominant in the protozoa-associated methanogen community and *Methanobrevibacter ruminantium* M1 was the only rumen methanogen species with published genome information at the start of this thesis; therefore, a pilot study was performed on the cultured isolate *M. ruminantium* M1 to demonstrate that phage display technology can be used to identify adhesins of archaeal origin.

5.1 Mru_1499 is a protozoa-binding adhesin with homologs in several methanogen species

Mru_1499 has previously been annotated as an adhesin-like protein based on *in silico* identification of putative domains typically associated with bacterial adhesion proteins. In this thesis, after two rounds of affinity screening, Mru_1499 was identified as a protozoa-binding adhesin from a large phage display library that represents the proteome of methanogen strain M1 (Chapter 3). The binding function of the affinity-selected clone was further verified by affinity binding assays. The functional evidence presented for protozoal cell surface binding in this study, together with cell binding and up-regulation of *mru_1499* during co-culture of M1 and H_2-producing rumen bacterium *B. proteoclasticus* B316 (Leahy *et al.*, 2010, Ng *et al.*, 2015), strongly support that Mru_1499 is a methanogen adhesin involved in cell surface attachment.

Given that as many as 62 ALPs are encoded in the genome of strain M1 and six adhesin-encoding genes were up-regulated when strain M1 interacts with *B. proteoclasticus* B316, it is possible that multiple ALPs are involved in associations between M1 and protozoal hosts. Notably, four of the six up-regulated genes (including *mru_1499*) encode ALPs that harbour Big_1 domains in the co-culture experiment (Leahy *et al.*, 2010). Additional protozoa-binding adhesins may be identified by further sequencing of the adhesin-enriched M1 library.

5.1.1 Functional domains identified in Mru_1499

Immunoglobulin-like domains, single and/or in tandem, have previously been identified in cell- and extracellular matrix-binding proteins of prokaryotic origin (Lin & Chang, 2007, Bodelon *et al.*, 2013, Gagic *et al.*, 2013). They are evolutionarily ancient protein folds found in proteins with diverse functions, ranging from cell adhesion, to immune function, to chitin-binding (Halaby & Mornon, 1998, Itoh *et al.*, 2013). These domains are found in protein sequences that have low overall levels of identity to each other, so it is unclear whether (1) their functions have diverged throughout evolutionary time, or (2) Ig-like fold is structurally stable, which is the basis of convergent evolution for proteins with different functions to adopt this fold.

The presence of tandem Big_1 domains and a TG-like domain in the protein sequence of Mru_1499 were predicted by *in silico* analysis, and, in this study, we have provided *in vitro* experimental evidence showing that a polypeptide encoding amino acids 40 to 197 of the protein (denoted Domain 1) encompassing a single Big_1 domain (amino acids 102 to 197 of Mru_1499) can be sufficient for adhesion to protozoa. The protein fragments encompassing single Domain 2 or Domain 3 did not have binding functionality. Cell binding function has been attributed to bacterial immunoglobulin-like domains in bacterial adhesins previously (Bodelon *et al.*, 2013, Gagic *et al.*, 2013, Buscetta *et al.*, 2014);

therefore, we anticipate that the Big_1 domain contained in Domain 1 is important for binding to protozoal hosts. No functional domains were predicted for the N-terminal portion of the polypeptide encoded in Domain 1 (amino acids 40 to 101 of Mru_1499); however, we cannot exclude the possibility that this region may be involved in cell binding.

Tandem bacterial immunoglobulin-like domains have been identified in cell- and extracellular matrix-binding proteins of prokaryotic origin in the past (Lin & Chang, 2007, Gagic *et al.*, 2013, Buscetta *et al.*, 2014). Avidity effects have been noted for SpcA adhesin in *L. rhamnosus*, as two Big_3 domains were required for bacterial cell binding (Gagic *et al.*, 2013). Another example of an adhesin with tandem Ig-like domains is LigB from *Leptospira* spp. which harbours 12 tandem Ig-like domains. Fibronectin-binding function was mapped to the protein region containing Ig-like domains 7 to 12 (Lin & Chang, 2007). A LigB variant containing only domain 12 exhibited low binding affinity for the gelatin-binding domain within fibronectin. However, the binding affinity of a LigB variant encompassing Ig-like domains 7 to 12 was approximately 30-fold higher than the binding affinity of a LigB variant containing Ig-like domains 7 to 11, indicating that domain 12 plays a supporting role in ligand binding (Lin *et al.*, 2010). This evidence implies that although protozoal cell binding activity was not observed for Mru_1499 domains 2 and 3, these domains may still play a supporting role in host cell binding.

Compared to bacterial proteins, little is known about adhesins in methanogenic archaea. MTH719 from *M. thermautotrophicus* is one of the well-characterized archaeal adhesins. It is a pseudomurein binding protein with three predicted PMB motifs, but the third PMB motif on its own did not exhibit binding activity (Visweswaran *et al.*, 2011). Although this latter motif was not required for binding bacterial spheroplasts, it was essential for binding archaeal cell wall. Further studies can be performed using

Mru_1499^{A} variants encoding one to three Big_1 domains and different cell types as bait to determine the effect of tandem Big_1 domains on the binding specificity of Mru_1499.

The affinity-selected sequence *mru_1499*^{A} does not encode the C-terminal TG-like domain, we therefore anticipate that this domain is not essential for protozoa binding. However, as TG-like domains in archaeal endoisopeptidases catalyse proteolytic cleavage, and Leahy et al. (2010) have proposed that TG-like proteins play a role in pseudomurein modification (Makarova *et al.*, 1999, Steenbakkers *et al.*, 2006, Leahy *et al.*, 2010), the TG-like domain in Mru_1499 may facilitate covalent linkage of the methanogen symbiont to the host cell surface by modification of the *M. ruminantium* M1 cell surface.

5.1.2 Several methanogen species harbour Mru_1499 homologs

Large repertoires of ALPs are also encoded in the genomes of other methanogen species (Table 1.4). Immunoglobulin-like domains, which have been associated with cell binding function in previous studies (Bodelon *et al.*, 2013, Gagic *et al.*, 2013), were detected in ALPs encoded in several methanogen genomes (Table 1.5). Furthermore, homologs of Mru_1499 were identified in other members of the *Methanobrevibacter* genus in this thesis. *Methanobrevibacter* species are predominant in the protozoa-associated methanogen community, which suggests that these species harbour molecular mediators that facilitate attachment to protozoal hosts. Mru_1499 homologs and additional Ig-like domain containing ALPs could facilitate the colonization of protozoal hosts by *Methanobrevibacter* species.

Although *Methanobrevibacter* species dominate the protozoa-associated methanogen community, methanogen strains outside this genus have also been detected as protozoa symbionts. In particular, a species related to *Methanosphaera stadtmanae* has been reported as a symbiont of the rumen protozoan *Eudiplodinium* (Tymensen *et al.*, 2012b).

It follows that methanogens from other genera with the ability to form physical associations with protozoa could harbour distant protein homologues involved in host recognition. A BLAST search of the NCBI non-redundant database, using the full length Mru_1499 sequence as a query, revealed a distantly-related putative protein encoded in the *Methanosphaera stadtmanae* MCB-3 genome (Figure 3.8). No other Mru_1499^{A} homologs with Big_1 domains were identified by BLASTP query of available methanogen genomes.

5.2 Mru_1499 can bind to a broad range of symbionts

Reverse panning data suggest that Mru_1499^{A} is a broad spectrum protozoa binder with a preference for binding to cells of genera in the family Ophryoscolecidae. Specifically, it exhibits strong affinity for members of the genera *Epidinium* and *Entodinium*, but it also appears to discriminate between species within the *Entodinium* genus. No binding was observed for protozoa in the Isotrichidae family (*Isotricha* spp. and *Dasytricha* spp.). Several studies have shown that endosymbiotic methanogens in the genus *Methanobrevibacter* reside in protozoa species belonging to the family Isotrichidae (Chagan *et al.*, 1999, Irbis & Ushida, 2004), which suggests that different adhesins may be responsible for interactions with this protozoal family.

Mru_1499^{A} has also been shown to bind to H_2-producing rumen bacterium *Butyrivibrio proteoclasticus* B316 (Ng *et al.*, 2015). This evidence supports Mru_1499 as a molecular mediator that enables methanogen strain M1 to bind to a broad range of symbionts by a single mechanism. This protein is, however, specific for certain bacterial species, as Mru_1499^{A} did not bind to *Ruminococcus albus* 8, *Ruminococcus flavefaciens* FD1, *Kandleria vitulina* RL 2, or *Sharpea azabuensis* RL 1 (Ng *et al.*, 2015).

It is difficult to speculate about the identity of cognate protozoal cell surface receptors for Mru_1499. The genome sequences of rumen protozoa are not yet available and very

few of their cell surface proteins have been identified. Ultrastructural studies revealed the presence of a glycocalyx layer that envelops the cell surface of *Entodinium* spp., *Epidinium caudatum*, and *Eudiplodinium maggii*, which are in the family Ophryoscolecidae (Furness & Butler, 1983, Furness & Butler, 1985a, Furness & Butler, 1985b). This feature has not been reported for the family Isotrichidae. Proteins found in the ecto-endoplasmic layer also differ between Ophryoscolecidae and Isotrichidae, with large filamentous proteins (58-96 kDa) present in the former, and smaller proteins (~22 kDa) in the latter (Lynn, 2008). Components of the glycocalyx or membrane-associated proteins produced by protozoa in the family Ophryoscolecidae may be the receptor that interacts with M1 cell surface *via* Mru_1499.

5.3 Analysis of the prokaryotic symbiont community associated with rumen protozoa

A method was developed to enrich for methanogenic archaea that are protozoa-associated. Protozoa were collected from rumen fluid, successive washes were performed to remove free-living bacteria and archaea, and then lysozyme, mutanolysin, and DNase I enzymes were applied to lyse protozoa-associated bacteria and remove the DNA released. This treatment was met with limited success, as results from qPCR analysis of 16S rRNA gene sequences indicated that bacterial DNA was present in a greater proportion than archaeal DNA in the extracted metagenomic DNA (Table 4.3). In comparison with the ratio of methanogenic archaea to bacteria found in rumen contents, archaeal DNA was enriched by 54-fold in the protozoa-associated metagenomic DNA sample in this study (Section 4.1.1). In a previous study by Belanche *et al.* (2014), qPCR analysis of 16S bacterial rRNA gene sequence and *mcr*A methanogen gene sequence indicated that there is approximately 100-fold more methanogens than bacteria in the protozoa-associated community (Belanche *et al.*, 2014). The difference in

relative amounts of methanogen to bacteria observed may be attributed to several factors. Firstly, the methanogen marker gene used for qPCR differed between studies [16S rRNA sequence was used in our study, whereas *mcr*A was used by Belanche *et al.* (2014)]. Secondly, the primers used for qPCR amplification of the bacterial 16S rRNA gene sequence also differed, which may give rise to variations in the number of bacterial 16S rRNA gene copies observed. Lastly, in our study, several incubation steps were performed to remove free-living prokaryotes from the sample, and to enzymatically remove bacteria and their DNA after protozoal fractionation. It is possible that oxygen was introduced into the sample during these steps, which would lead to death of methanogens as they are obligate anaerobes. Considering that Belanche and co-workers did not observe any significant differences in the ratio of methanogen to bacterial DNA between protozoa-associated symbionts (DNA isolated from cells that did not pass through a 5 µm filter) and free-living prokaryotic cells (DNA isolated from cells that could pass through a 5 µm filter), it is likely that the difference in proportions of protozoa-associated methanogen and protozoa-associated bacteria observed between the studies arose from differences in primers used in qPCR experiments, as bacteria are typically 100-fold more abundant than archaea in the rumen (Sirohi *et al.*, 2012, Henderson *et al.*, 2013).

Next generation sequencing of partial 16S rRNA archaeal and bacterial gene sequences was performed to characterize the microbial community associated with protozoa. It is important to note that these are preliminary results, as the protozoa-associated symbiont community should ideally be compared with rumen contents that the sample was derived from. Regarding the methanogen community associated with protozoa, we found that *Methanobrevibacter* spp. were predominant, which is in agreement with trends observed by Sanger sequencing of clone libraries in previous studies (Janssen & Kirs,

2008, Tymensen *et al.*, 2012b). In addition, within the genus *Methanobrevibacter*, members of the *Methanobrevibacter gottschalkii* clade are present in a greater proportion as protozoal symbionts, whereas *Methanobrevibacter ruminantium* clade sequences are present in a greater proportion in the rumen contents, in concordance with previous observations on the protozoa-associated methanogen community (Janssen & Kirs, 2008, Tymensen *et al.*, 2012b). Sequences corresponding to the genus *Methanomicrobium* have been reported as protozoal symbionts in other studies (Irbis & Ushida, 2004, Tymensen *et al.*, 2012b); however, none were detected in our study. The absence of this genus can be explained by the geography-dependence of *Methanomicrobium* in the rumen, as *Methanomicrobium* species are more prevalent in Asia than in New Zealand (Table 1.3) (Henderson *et al.*, 2015).

With respect to the bacterial community associated with protozoa, sequences that belong to the families *Coriobacteriaceae*, *Ruminococcaceae*, and *Mogibacteriaceae* with unassigned genus affiliations, sequences corresponding to the order Clostridiales with unassigned family and genus affiliations, and sequences assigned to the genus *Mogibacterium* were predominant (together, they account for 57.6% of total bacterial sequences). One possible explanation for the abundance of these species as protozoa-associated symbionts is that protozoa-associated bacterial symbionts work in synergy with their host to break down plant material. Cultured isolates rumen bacterium R-7 and AD3011 are relatives of several unclassified *Clostridiales* and *Mogibacteriaceae* sequences identified in this study, respectively. Both of these isolates were cultured from plant-adherent material in the rumen (Noel, 2013). *Olsenella* spp. (members of the family *Coriobacteriaceae*) have also been found to be abundant in the plant-adherent fraction in the rumen (Huws *et al.*, 2016). *Mogibacterium* species that have been cultured and characterized, however, were reported to be asaccharolytic (Schink, 2009). No activity was observed with many sugar-

containing substrates, including arabinose, cellobiose, starch, and gelatin; therefore, species related to *Mogibacterium* may only play an indirect role in the breakdown of feed. The finding that these species are predominant in the protozoa-associated bacterial community was unexpected, as results from clone library analysis of 16S rRNA gene sequences acquired from single cells of rumen protozoan *Polyplastron multivesiculatum* showed that *Ruminococcus albus* and *Streptococcus bovis* accounted for 78% (n=94) of the protozoa-associated bacterial population prior to antibiotic treatment, whereas only species that belong to the phylum Proteobacteria were observed (n=99) after protozoa were incubated in an antibiotic cocktail for 48 h (Irbis & Ushida, 2004). The rumen protozoa analyzed in this thesis belong to the Type B community, whereas *P. multivesiculatum* is only present in the Type A community, which may account for the observed differences.

Synergistic interactions in cellulose degradation in the rumen between protozoa and bacteria have been reported in the past (see Section 1.2.1.1 for details), but the role of protozoa-associated bacteria has not been specifically explored in this capacity. It is important to note, however, the possibility that bacterial strains resistant to lysozyme and mutanolysin were enriched in the metagenomic DNA sample in this study as these enzymes were used in sample preparation. The detection of a member of the *Coriobacteriaceae* family (human-derived bacterial strain *Atopobium vaginae* BAA-55) by PCR from a mock community sample was no better with the addition of lysozyme or mutanolysin for cell lysis, which suggests that this strain is not affected by lysozyme and mutanolysin; therefore, related strains might be resistant to lysozyme and mutanolysin as well (Yuan *et al.*, 2012). On the other hand, the rumen bacterium *Ruminococcus albus* (member of *Ruminococcaceae* family) was reported to be sensitive to lysis by mutanolysin (Morris & Cole, 1987), yet sequences in the *Ruminococcaceae* family without genus

affiliations were found to be almost 10-fold higher in the protozoa-associated community compared to the rumen content sample from a NZ sheep. Further studies would be required to determine whether: (1) differences in the composition of the protozoa-associated bacterial community compared to rumen contents are authentic, and not simply a bias introduced during sample preparation; (2) bacterial strains predominantly associated with protozoa are H_2 producers and/or cellulose degraders; (3) physical association with protozoa increases the efficiency of cellulose degradation.

As noted earlier, members of the family *Mogibacteriaceae* with unassigned genus affiliations as well as *Mogibacterium* genus were over-represented in the protozoa-associated bacterial community in this study. The abundance of *Mogibacterium* species have also been reported to be elevated in (1) rumen contents of high-methane emitting cattle (Wallace *et al.*, 2015), (2) rumen epithelium of cattle compared to rumen contents (Li *et al.*, 2012a), and (3) rumen epithelium of goats fed with high grain diet compared to hay (Liu *et al.*, 2015, Wetzels *et al.*, 2015). Their presence is correlated with certain rumen metabolites, including phenylacetate [a known metabolic end product of several *Mogibacterium* species (Schink, 2009)] and putrescine (Mao *et al.*, 2014). Due to the lack of a cultured representative from the rumen, little is known about species in this family.

5.3.1 Similarities between protozoa-associated symbiont community and rumen epithelium-adherent community

We noted that in both the protozoa-associated and rumen epithelium-adherent bacterial communities, the genus *Mogibacterium* tends to be more abundant compared to rumen contents, whereas *Prevotella* was less abundant. This observation suggests that species in these genera may use the same mechanism for attachment to both protozoal and ruminant hosts. In regards to methanogen species that are rumen epithelium-associated, Pei *et al.* (2010) reported a lower proportion of *Methanosphaera* spp. and an increased

proportion of *Methanobrevibacter millerae (*a member of the *Methanobrevibacter gottschalkii* clade) on the epithelium of cattle (Pei *et al.*, 2010). These trends are similar to the protozoa-associated methanogen community. However, *Methanomassilliicoccales*–affiliated sequences were more abundant in the rumen epithelium-adherent methanogen community compared to the rumen contents, which differs from the protozoa-associated community. Conceivably, *Methanobrevibacter* species, specifically members of the *Methanobrevibacter gottschalkii* clade, could be specialized in protozoa attachment, whereas *Methanomassilliicoccales*-affiliated species could be specialized in epithelium attachment.

5.4 Mining for protozoa-binding adhesins

5.4.1 Overcoming technical challenges in affinity screening of adhesins from metagenomic libraries by phage display

In recent years, very few studies demonstrated that phage display can be successfully used to screen metagenomic phage display libraries for proteins of interest (Zhang *et al.*, 2009). Several technical challenges are intrinsic to phage display technology. These challenges include: (1) background (non-specific) binding of phage during affinity selection, (2) adequate library size, and (3) selection of fast growers *versus* target.

In our study, we took precautions to prevent non-specific binding to plasticware by performing washes on PP-bound protozoal bait on a nylon mesh after incubation of PPs and bait, and then transferring the retentate (containing protozoa binders) to a low protein-binding microcentrifuge tube. With this method, PPs non-specifically bound to nylon were also excluded from further analysis, as they would be left adhered to the mesh.

A large phage display library is prerequisite for successful affinity screening against the target of interest. For example, in the study by Zhang *et al.* (2009), phage display was successfully used to identify 12 proteins that can bind to the target of interest from a metagenomic library. Their success can be attributed to a large library size (3×10^8 clones in the primary library), as well as a highly specific affinity selection system (based on the selective enzymatic modification of PPs displaying peptidyl-carrier proteins with biotinylated CoA and subsequent hybridization to streptavidin-coated surfaces) that contributes to high fold-enrichment during selection [3,600-fold enrichment of positive clones from one selection round (Yin *et al.*, 2004)]. In our study on the protozoa-associated metagenomic community, we also started with a large library size (1.2×10^8 clones).

It is possible for high proportions of out-of-frame clones to accumulate due to the use of wild-type helper phage VCSM13 for PP production. When VCSM13 helper phage is used to generate PPs, the ratio of phagemid-derived pIII to helper phage-derived pIII packaged into PPs varies from 1:9 to 1:1000 depending on the phagemid vector and the recombinant fusion protein sequence (Azzazy & Highsmith, 2002). In fact, even for a given phagemid that encodes a recombinant pIII fusion protein, the proportion of recombinant pIII fusion protein relative to all pIII present in PPs (helper phage and vector-derived pIII) was estimated at 20-30%, and it was thought that proteolytic degradation of the fusion protein in the host periplasm contributed to the low amount of pIII fusion protein incorporated into PPs (McCafferty, 1996). These data suggest that many PPs in the population may not encode recombinant pIII fusion protein, but VCSM13 helper phage-derived full length pIII protein were incorporated into their surfaces; therefore, they can still infect the *E. coli* host, thus leading to the accumulation of fast-growing non-binding PPs. Increasing proportions of these background clones

accumulated during affinity selection have been noted previously (Vodnik *et al.*, 2011). To avoid sequencing a large population of background clones, we performed affinity selection with c-myc as ligand again, on PPs eluted from Round 2 biopanning prior to sequencing, to ensure that the PP population analyzed encodes for recombinant c-myc/pIII fusion proteins.

Alternative solutions to this this issue include: (1) the use of a different helper phage for PP propagation [for example, VCSM13d3 (pIII deletion (Rakonjac *et al.*, 1997)), hyperphage (pIII truncation (Rondot *et al.*, 2001)), or AGM13 (modified pIII susceptible to trypsin cleavage (Gupta *et al.*, 2013))] or (2) abolishing the PP amplification step from the biopanning framework (Figure 1.11), i.e. directly use PPs eluted from the bait as input for the next biopanning round, and to reduce the number of biopanning rounds. Decreased library diversity caused by competition with fast-growers has been noted during amplification previously, thus the elimination of this step aids in preserving library diversity (Derda *et al.*, 2011). 't Hoen *et al.* (2012) have demonstrated that target-binding sequences can be identified from a peptide library after only one round of biopanning with next generation sequencing; therefore, amplification steps which introduce confounding factors into phage display experiments can be eliminated ('t Hoen *et al.*, 2012).

5.4.2 PacBio sequencing as a new tool for mining metagenomic phage display libraries

The combination of phage display and next generation sequencing is increasingly used to characterize affinity-selected libraries. One advantage of functional screening for adhesins by phage display over adhesin identification by *in silico* annotation of metagenomic libraries is the potential to discover moonlighting proteins and multi-functional proteins that also exhibit adhesion functions. Pyrosequencing, Illumina

sequencing, and ion torrent sequencing have been used for this purpose (Dias-Neto *et al.*, 2009, Ngubane *et al.*, 2013, Matochko & Derda, 2015); however, no reports of PacBio sequencing for phage display library screening has been published yet. The size distribution of sheared metagenomic DNA inserts in the primary libray of our study mainly ranged from 1 to 3.5 kb (Figure 4.7). As pyrosequencing, Illumina, and ion torrent sequencing technologies generate read lengths around 0.7 kb, 0.25-0.3 kb, and 0.2 kb respectively (Rhoads & Au, 2015), it would not be possible to sequence these amplicons directly. The PCR amplicons must first undergo shearing, and sequence data produced must then be assembled by computational methods. PacBio sequencing, on the other hand, can produce long sequence reads (average 2 kb), thus allowing amplicons to be sequenced without fragmentation and subsequent assembly. Therefore, we chose this method of sequencing for phage display library analysis.

The theoretical maximum number of reads that can be generated by PacBio sequencing is 150,000 (the number of zero-mode waveguides in the cell). Approximately 70,000 raw reads has been reported previously (Zhang *et al.*, 2014); however, only approximately 10,000 raw reads were generated in each of the libraries PAM1 and PAM2, and after quality filtering, ~4,600 reads per library remained. The number of reads can be enhanced by optimizing DNA loading into the SMRT cell and/or increasing the number of SMRT cells for sequencing, but this data has not been obtained due to resource constraints of the project. For the PAM_FINAL library, DNA loading was optimized, resulting in 22,527 reads after quality filtering. A new PacBio Sequel system with 1,000,000 zero-mode waveguides is now available. The number of zero-mode waveguides present in this system is 10-fold greater than the PacBio RS II system, which should result in a 10-fold increase in the number of sequencing reads. By using the new PacBio Sequel system and a greater number of SMRT cells for sequencing each sample,

it may be possible to characterize PCR amplicons from libraries of greater complexity by single molecule sequencing technology in the future.

As a low number of sequence reads is generated in a PacBio sequencing run, this technology is better suited to characterizing libraries with lower complexity, such as phage display libraries after affinity selection. In our study, the libraries PAM1 and PAM2 were sequenced as a baseline prior to affinity selection for protozoa-binding adhesins. These libraries were expected to have a diversity of 10^8 and 10^6, respectively; therefore, almost all the generated sequence reads were unique. The library PAM_FINAL has undergone two rounds of affinity selection against protozoal cells as bait, and the maximum possible diversity of this library is 10^4 (the number of PPs eluted after the second round of biopanning with protozoa). We were able to identify enriched sequences in this library using both the full dataset acquired, and the randomly generated subset of data normalized to the average number of reads acquired for the PAM1 and PAM2 libraries, thereby demonstrating that PacBio technology can be used to identify enriched sequences from a metagenomic phage display library.

In comparison to PCR amplicon sequencing by pyrosequencing, even for the average of 4,600 amplicon sequences obtained for PAM1 and PAM2 libraries in this study, this number is comparable to the the number of assembled PCR amplicons obtained by pyrosequencing of a meta-secretome phage display library generated from the fibre-adherent rumen microbial community (Ciric, 2014). Pyrosequencing of the meta-secretome library generated 691,206 raw reads, representing 379 Mb of sequence data which were assembled into 3,574 contigs, where each contig corresponds to a PCR amplicon derived from a phagemid insert. This number of amplicon sequences is close to the average of 4,600 obtained in our study. For the PAM_FINAL library, where DNA

loading into the SMRT cell has been optimized, six-fold more amplicon sequences were generated compared to the pyrosequenced fibre-adherent metagenomic sample.

Although only a small portion of the protozoa-associated metagenome phage display primary library (PAM1) amplicons was sequenced by PacBio sequencing, we were able to identify trends that are concordant with those previously reported for rumen metagenomes. As discussed in Section 4.2.1.1, ABC transporter domain (Pfam00005) was previously identified as the most abundant in the rumen metagenome (0.8%) (Li *et al.*, 2014). We found that this domain was the second most abundant domain in the protozoa-associated symbiont community, occupying 1.2% of identified domains. In the study published by Wallace *et al.* (2015), high coverage was attained for the rumen samples investigated, but the functions encoded in the majority of reads remained unknown due to the limited knowledge about genes found in the rumen metagenome (Wallace *et al.*, 2015). Similarly, Ciric *et al.* (2014) found that Pfam domains were only assigned to 35% of meta-secretome reads (Ciric *et al.*, 2014). In our study, Pfam domains were assigned to approximately 30% of the amplicons in the libraries.

5.4.3 Common characteristics of putative adhesins identified in the PAM_FINAL library

PacBio sequencing of the PAM_FINAL library revealed that seven nucleotide sequences together accounted for over 60% of all sequence reads, thus this approach was successfully used to identify enriched sequences present in the metagenomic library after two rounds of biopanning against protozoal bait. One of the proteins is likely derived from protozoa (cluster 2), three from bacteria (clusters 3, 6, 7), and the origins of the remaining three are unclear due to low levels of similarity between the enriched sequences and known sequences in the databases. These sequences could potentially belong to methanogens that have not been sequenced. However, due to the low

representation of sequenced archaeal genomes in public databases, it is not possible to ascertain the origins of these sequences. Other factors that may contribute to the lack of methanogen adhesins identified include: (1) low proportion of archaeal sequences present in the primary phage display metagenomic library, and (2) differences between codon usage in methanogenic archaea and the *E. coli* bacterial host, which can contribute to slower propagation of PPs encoding archaeal sequences (Nakamura *et al.*, 2000).

Similarities to annotated surface-associated proteins were identified for five out of the seven encoded proteins by BLAST (Clusters 2, 3, 4, 6, 7; E-value < 0.001). Other common features were noted in the enriched sequences by analyses based on amino acid sequence characteristics. For example, six out of the seven encoded proteins have an estimated pI > 8, which indicates that these proteins have an overall positive charge under physiological conditions. Positively charged adhesins from protozoa and bacteria have been implicated in cell binding in previous studies (Engbring & Alderete, 1998, Formosa-Dague *et al.*, 2016). Repeated motifs were detected for four out of seven sequences (Clusters 2, 3, 4, 6), which is highly suggestive of adhesion function as plant-colonizing microbes (eukaryotic and prokaryotic) and human bacterial pathogens have been reported to harbour repeat-containing effectors and tetratricopeptide repeat-containing proteins, respectively, that play a role in host colonization (Cerveny *et al.*, 2013, Mesarich *et al.*, 2015). Four out of seven sequences (Clusters 2, 3, 6, 7) were assigned a P_{ad} score > 0.5 by SPAAN analysis based on amino acid sequence characteristics, which also suggests adhesion function. Lastly, lysine-rich regions (Clusters 1, 4), glutamine-rich regions (Cluster 3), asparagine-rich regions (Clusters 2, 6), and proline-rich regions (Cluster 4) were detected. Adhesins with over-represented lysine, glutamine, asparagine, and proline residues have been reported previously (Mai & Samuelson, 1998, Pethe *et al.*, 2000, Koba *et al.*, 2009, Levine, 2011, Mesarich *et al.*, 2015). The sequence

encoded by Cluster 5 had the least amount of information associated with it. A breakdown of its amino acid composition revealed that it contained ~9% of glycine residues, ~11% of proline residues, and ~12% of serine residues, with two homopolymer stretches of serine residues. However, *in silico* analysis tools did not reveal any features strongly indicative of adhesion function.

5.5 Conclusions

As the presence of protozoa can account for up to 37% of methane emitted by ruminants (Finlay *et al.*, 1994) and meta-analysis has shown positive correlations between protozoa abundance and methane emission, defaunation was proposed as a potential method for methane mitigation (Martin *et al.*, 2010, Guyader *et al.*, 2014, Newbold *et al.*, 2015). However, defaunation did not always successfully mitigate methane in *in vivo* experiments (Machmuller *et al.*, 2003, Morgavi *et al.*, 2012). Apart from diet-dependent variation, methane variability after defaunation may be attributed to the alternate H_2 producers in the rumen, such as H_2-producing bacteria, thus allowing methanogens to continue producing methane even though they are not attached to protozoa. In a study exploring relationships between rumen microbes within three ruminant host species, co-occurrence analysis failed to show strong correlations between specific ciliate protozoal and methanogen genera (Kittelmann *et al.*, 2013), suggesting that these associations are facultative, further supporting the idea that methanogens can sequester H_2 from sources other than protozoa. The results in the first part of this thesis have demonstrated that Mru_1499, a methanogen cell surface protein, can bind a variety of protozoa. Related experiments performed in our lab have shown that Mru_1499 can also bind to the H_2-forming rumen bacterium *B. proteoclasticus* B316. The binding of Mru_1499^A to widely different microorganisms provides evidence of a single molecular mechanism that

would allow a methanogen to attach to a wide range of very diverse H_2-producing microbes. The promiscuous binding observed for Mru_1499 may be an example of ecological redundancy, as this protein can bind to at least one other H_2-producing species in the absence of protozoa. Therefore, it is important to consider the resilience of the microbial community in future efforts to reduce ruminant methane production by manipulating protozoa in the rumen.

The second part of this thesis focused on the protozoa-associated symbiont community, and it was demonstrated that the protozoa-associated bacterial community is diverse and appears to harbour H_2-producing bacterial species, which, to our knowledge, has not been reported before. The functional role of protozoa-associated bacteria has yet to be characterized. We hypothesize that some of these species may participate in the synergistic degradation of plant material with protozoa, as two of the identified bacterial taxa are related to isolates prominent in the fibre-adherent microbial community. Regarding the protozoa-associated methanogen community, large scale 16S rRNA gene sequence analysis of methanogen species associated with protozoa are in agreement with previous small scale sequence analysis studies which showed that members of *Methanobrevibacter gottschalkii* clade are more prevalent on the protozoal surface than in the rumen contents.

Affinity screening of the protozoa-associated metagenomic phage display library was performed to identify adhesins involved in protozoal attachment, and large scale sequencing data was obtained by PacBio sequencing of amplicons in the phage display library before and after biopanning against protozoal bait. Seven potential adhesin-encoding ORFs were identified, but experimental verification of adhesion function is required. From the results of this study, we conclude that the current PacBio RS II

platform can be used for characterizing metagenomic phage display libraries after enriching for proteins that can bind to the ligand/bait of interest.

5.6 Future directions

A large shotgun phage display library was created from *Methanobrevibacter ruminantium* M1 genomic DNA. This library can be used as a resource to explore interactions between the M1 strain and other biotic components of the rumen microbial ecosystem. The M1 strain is known to form physical associations with the H_2-producing bacterium *B. proteoclasticus* B316 (Leahy *et al.*, 2010); therefore, the library can be screened for adhesins involved in binding this species, or related bacterial species (such as *Pseudobutyrivibrio* spp.). Methanogens have also been shown to attach to the rumen epithelium (Shin *et al.*, 2004, Pei *et al.*, 2010), and this association may be a strategy for methanogens to stay in the anaerobic rumen habitat long enough for cell proliferation to occur. Affinity screening of the M1 phage display library against different cell types as bait (rumen epithelial cells, bacteria, and protozoa) would reveal whether the same cell surface adhesins are used for attachment to various symbiotic partners. This information would be beneficial in developing effective biotechnology strategies to disrupt physical associations between methanogens and their symbiotic partners for enteric methane mitigation.

We have successfully identified a protozoa-binding adhesin from M1 strain; however, the cognate receptors on protozoal cell surfaces have yet to be discovered. Complete genome information is not yet available for any rumen protozoal species due to their recalcitrance to long term maintenance as axenic cultures and their complex genome organization (Newbold *et al.*, 2015). An experimental strategy that could be used to discover protozoal receptors is affinity screening of a phage display library generated

from protozoal cDNA with M1 strain whole cells or a specific adhesin as bait. It may be beneficial to perform cell surface display of recombinant protozoal proteins using eukaryotic hosts, such as *Tetrahymena thermophila* or *Saccharomyces cerevisiae*, to avoid differential codon usage issues. Like rumen protozoa, *T. thermophila* is also a ciliated protozoan, and it is the closest relative to rumen protozoa that can be genetically manipulated. Their codon usage indices are more likely to be similar because they are more closely related; however, transformation efficiency for *T. thermophila* is low (30 to 200 transformants/μg of DNA by biolistic bombardment) (Gaertig *et al.*, 1999). Yeast surface display is commonly used to generate recombinant protein libraries for eukaryotes (Pepper *et al.*, 2008). At 10^6 transformants/μg of DNA (reported for transformation by electroporation) (Kawai *et al.*, 2010), the transformation efficiency of *S. cerevisae* is higher than for *T. thermophila* which would be more suitable for generating larger libraries.

The role of bacteria in the rumen protozoa-associated ecosystem has not been well-studied. In this study, preliminary data from a protozoa-associated bacterial community showed interesting trends; however, these results may be affected by the use of lysozyme/mutanolysin treatment. Protozoa-associated bacteria could be isolated without the use of these selective agents, and multiple biological replicates should be sequenced to confirm the initial findings. If the same results are observed, further experiments can be executed to test the hypothesis that H_2-producing bacteria are present in the community as these bacteria participate in synergistic degradation of plant material with protozoa. Cultured isolates can be tested for fibrolytic activity on various substrates as well as H_2 production. The rate of cellulose degradation between bacterial mono-cultures and bacterial co-culture with protozoa can be compared to determine whether these bacterial isolates work synergistically with protozoa in fibre degradation.

To verify that the seven enriched amplicon sequences from the protozoa-associated metagenomic phage display library encode for proteins with protozoa-binding function, these sequences will be cloned into the phage display vector, and PPs displaying the recombinant proteins will be tested for protozoa-binding activity. As many of these proteins did not share significant similarity with proteins with known functions, adhesin identification by sequence homology was of limited utility; therefore, the computational tool SPAAN was used to assess adhesion potential of these proteins based on amino acid sequence characteristics. This tool has, however, only been used for putative proteomes of sequenced organisms in the past. The experimental data generated from testing protozoa-binding activity of the enriched recombinant proteins can also be used to assess the efficacy of SPAAN for adhesin identification for proteins encoded in metagenome data.

Chapter 6. References

[1] 't Hoen PA, Jirka SM, Ten Broeke BR, Schultes EA, Aguilera B, Pang KH, Heemskerk H, Aartsma-Rus A, van Ommen GJ & den Dunnen JT (2012) Phage display screening without repetitious selection rounds. *Analytical Biochemistry* 421: 622-631.

[2] Akhmanova A, Voncken F, Van Alen T, Van Hoek A, Boxma B, Vogels G, Veenhuis M & Hackstein JHP (1998) A hydrogenosome with a genome. *Nature* 396: 527-528.

[3] Altermann E & Klaenhammer TR (2003) GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *Omics* 7: 161-169.

[4] Artimo P, Jonnalagedda M, Arnold K, *et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* 40: W597-603.

[5] Atarashi K, Tanoue T, Shima T, *et al.* (2011) Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* 331: 337-341.

[6] Azzazy HM & Highsmith WE, Jr. (2002) Phage display technology: clinical applications and recent innovations. *Clinical Biochemistry* 35: 425-445.

[7] Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW & Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* 37: W202-208.

[8] Barlag B & Hensel M (2015) The giant adhesin SiiE of *Salmonella enterica*. *Molecules* 20: 1134-1150.

[9] Bassler J, Hernandez Alvarez B, Hartmann MD & Lupas AN (2015) A domain dictionary of trimeric autotransporter adhesins. *International Journal of Medical Microbiology* 305: 265-275.

[10] Belanche A, de la Fuente G & Newbold CJ (2014) Study of methanogen communities associated with different rumen protozoal populations. *FEMS Microbiology Ecolology* 90: 663-677.

[11] Berisio R & Vitagliano L (2012) Polyproline and triple helix motifs in host-pathogen recognition. *Current Protein & Peptide Science* 13: 855-865.

[12] Bober M, Morgelin M, Olin AI, von Pawel-Rammingen U & Collin M (2011) The membrane bound LRR lipoprotein Slr, and the cell wall-anchored M1 protein from *Streptococcus pyogenes* both interact with type I collagen. *PLoS ONE* 6: e20345.

[13] Bodelon G, Palomino C & Fernandez LA (2013) Immunoglobulin domains in *Escherichia coli* and other enterobacteria: from pathogenesis to applications in antibody technologies. *FEMS Microbiology Reviews* 37: 204-250.

[14] Boeke JD, Model P & Zinder ND (1982) Effects of bacteriophage f1 gene III protein on the host cell membrane. *Molecular & General Genetics* 186: 185-192.

[15] Bokhari H, Bilal I & Zafar S (2012) BapC autotransporter protein of *Bordetella pertussis* is an adhesion factor. *Journal of Basic Microbiology* 52: 390-396.

[16] Bonhomme A (1990) Rumen ciliates: their metabolism and relationships with bacteria and their hosts. *Animal Feed Science and Technology* 30: 203-266.

[17] Booyse D & Dehority BA (2011) Rumen protozoa in South African sheep with a summary of the worldwide distribution of sheep protozoa. *Onderstepoort Journal of Veterinary Research* 78: 307.

[18] Brady LJ, Maddocks SE, Larson MR, Forsgren N, Persson K, Deivanayagam CC & Jenkinson HF (2010) The changing faces of *Streptococcus* antigen I/II polypeptide family adhesins. *Molecular Microbiology* 77: 276-286.

[19] Buddle BM, Denis M, Attwood GT, Altermann E, Janssen PH, Ronimus RS, Pinares-Patino CS, Muetzel S & Neil Wedlock D (2011) Strategies to reduce methane emissions from farmed ruminants grazing on pasture. *Veterinary Journal* 188: 11-17.

[20] Buscetta M, Papasergi S, Firon A, *et al.* (2014) FbsC, a novel fibrinogen-binding protein, promotes *Streptococcus agalactiae*-host cell interactions. *Journal of Biological Chemistry* 289: 21003-21015.

[21] Caporaso JG, Kuczynski J, Stombaugh J, *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335-336.

[22] Cerveny L, Straskova A, Dankova V, Hartlova A, Ceckova M, Staud F & Stulik J (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infection & Immunity* 81: 629-635.

[23] Chagan I, Tokura M, Jouany JP & Ushida K (1999) Detection of methanogenic archaea associated with rumen ciliate protozoa. *Journal of General and Applied Microbiology* 45: 305-308.

[24] Chahales P & Thanassi DG (2015) Structure, Function, and Assembly of Adhesive Organelles by Uropathogenic Bacteria. *Microbiology Spectrum* doi: 10.1128/microbiolspec.UTI-0018-2013.

[25] Chan PP, Holmes AD, Smith AM, Tran D & Lowe TM (2012) The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Research* 40: D646-D652.

[26] Cheeseman SL, Hiom SJ, Weightman AJ & Wade WG (1996) Phylogeny of oral asaccharolytic Eubacterium species determined by 16S ribosomal DNA sequence comparison and proposal of *Eubacterium infirmum* sp. nov. and *Eubacterium tardum* sp. nov. *International Journal of Systematic Bacteriology* 46: 957-959.

[27] Ciric M (2014) Metasecretome phage display: A new approach for mining surface and secreted proteins from microbial communities. Thesis, Massey University, Palmerston North, NZ.

[28] Ciric M, Moon CD, Leahy SC, Creevey CJ, Altermann E, Attwood GT, Rakonjac J & Gagic D (2014) Metasecretome-selective phage display approach for mining the functional potential of a rumen microbial community. *BMC Genomics* 15: 356.

[29] Cole JN, Ramirez RD, Currie BJ, Cordwell SJ, Djordjevic SP & Walker MJ (2005) Surface analyses and immune reactivities of major cell wall-associated proteins of group a streptococcus. *Infection & Immunity* 73: 3137-3146.

[30] Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R & Wittwer P (2005) Metagenomic gene discovery: past, present and future. *Trends in Biotechnology* 23: 321-329.

[31] Creevey CJ, Kelly WJ, Henderson G & Leahy SC (2014) Determining the culturability of the rumen bacterial microbiome. *Microbial Biotechnology* 7: 467-479.

[32] Dehority BA (2006) Studies on the variation in caudal spination of *Epidinium*. *Zootaxa* 1305: 33-39.

[33] Denton BL, Diese LE, Firkins JL & Hackmann TJ (2015) Accumulation of reserve carbohydrate by rumen protozoa and bacteria in competition for glucose. *Applied and Environmental Microbiology* 81: 1832-1838.

[34] Derda R, Tang SK, Li SC, Ng S, Matochko W & Jafari MR (2011) Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16: 1776-1803.

[35] Devillard E, Goodheart DB, Karnati SK, Bayer EA, Lamed R, Miron J, Nelson KE & Morrison M (2004) *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *Journal of Bacteriology* 186: 136-145.

[36] Dias-Neto E, Nunes DN, Giordano RJ, Sun J, Botz GH, Yang K, Setubal JC, Pasqualini R & Arap W (2009) Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. *PLoS ONE* 4: e8338.

[37] Dunn AK (2012) *Vibrio fischeri* metabolism: symbiosis and beyond. *Advances in Microbial Physiology* 61: 37-68.

[38] Dziallas C, Allgaier M, Monaghan MT & Grossart HP (2012) Act together-implications of symbioses in aquatic ciliates. *Frontiers in Microbiology* 3: 288.

[39] Eadie JM (1967) Studies on ecology of certain rumen ciliate protozoa. *Journal of General Microbiology* 49: 175-194.

[40] Easton S (2009) Functional and metagenomic analysis of the human tongue dorsum using phage display. Thesis, University College London, London, UK.

[41] Egea L, Aguilera L, Gimenez R, Sorolla MA, Aguilar J, Badia J & Baldoma L (2007) Role of secreted glyceraldehyde-3-phosphate dehydrogenase in the infection mechanism of enterohemorrhagic and enteropathogenic *Escherichia coli*: interaction of

the extracellular enzyme with human plasminogen and fibrinogen. *International Journal of Biochemistry & Cell Biology* 39: 1190-1203.

[42] Eid J, Fehr A, Gray J, *et al.* (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323: 133-138.

[43] Engbring JA & Alderete JF (1998) Three genes encode distinct AP33 proteins involved in *Trichomonas vaginalis* cytoadherence. *Molecular Microbiology* 28: 305-313.

[44] Eschenlauer SC, McEwan NR, Calza RE, Wallace RJ, Onodera R & Newbold CJ (1998) Phylogenetic position and codon usage of two centrin genes from the rumen ciliate protozoan, *Entodinium caudatum*. *FEMS Microbiology Letters* 166: 147-154.

[45] Eugene M, Archimede H, Michalet-Doreau B & Fonty G (2004) Effects of defaunation on microbial activities in the rumen of rams consuming a mixed diet (fresh *Digitaria decumbens* grass and concentrate). *Animal Research* 53: 187-200.

[46] Evangelista KV, Hahn B, Wunder EA, Jr., Ko AI, Haake DA & Coburn J (2014) Identification of cell-binding adhesins of *Leptospira interrogans*. *PLoS Neglected Tropical Diseases* 8: e3215.

[47] Ezer A, Matalon E, Jindou S, Borovok I, Atamna N, Yu Z, Morrison M, Bayer EA & Lamed R (2008) Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. *Journal of Bacteriology* 190: 8220-8222.

[48] Findley SD, Mormile MR, Sommer-Hurley A, Zhang XC, Tipton P, Arnett K, Porter JH, Kerley M & Stacey G (2011) Activity-based metagenomic screening and biochemical characterization of bovine ruminal protozoan glycoside hydrolases. *Applied and Environmental Microbiology* 77: 8106-8113.

[49] Finlay BJ, Esteban G, Clarke KJ, Williams AG, Embley TM & Hirt RP (1994) Some rumen ciliates have endosymbiotic methanogens. *FEMS Microbiology Letters* 117: 157-162.

[50] Formosa-Dague C, Feuillie C, Beaussart A, Derclaye S, Kucharikova S, Lasa I, Van Dijck P & Dufrene YF (2016) Sticky matrix: Adhesion mechanism of the staphylococcal polysaccharide intercellular adhesin. *ACS Nano* 10: 3443-3452.

[51] Foroozandeh AD, Rezaeian M, Balaly GR & Alikhani M (2009) Relative contributions of ruminal bacteria, protozoa and fungi to degradation of forage fiber fractions. *Journal of Animal and Veterinary Advances* 8: 603-607.

[52] Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC & Huttenhower C (2015) Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nature Reviews Microbiology* 13: 360-372.

[53] Frederiksen RF, Paspaliari DK, Larsen T, Storgaard BG, Larsen MH, Ingmer H, Palcic MM & Leisner JJ (2013) Bacterial chitinases and chitin-binding proteins as virulence factors. *Microbiology* 159: 833-847.

[54] Frese SA, Mackenzie DA, Peterson DA, *et al.* (2013) Molecular characterization of host-specific biofilm formation in a vertebrate gut symbiont. *PLoS Genetics* 9: e1004057.

[55] Frese SA, Benson AK, Tannock GW, *et al.* (2011) The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genetics* 7: e1001314.

[56] Fricke WF, Seedorf H, Henne A, Kruer M, Liesegang H, Hedderich R, Gottschalk G & Thauer RK (2006) The genome sequence of *Methanosphaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H_2 for methane formation and ATP synthesis. *Journal of Bacteriology* 188: 642-658.

[57] Friedrich MW (2005) Methyl-coenzyme M reductase genes: unique functional markers for methanogenic and anaerobic methane-oxidizing Archaea. *Methods in Enzymology* 397: 428-442.

[58] Furness DN & Butler RD (1983) The cytology of sheep rumen ciliates. 1. Ultrastructure of *Epidinium caudatum* Crawley. *Journal of Protozoology* 30: 676-687.

[59] Furness DN & Butler RD (1985a) The cytology of sheep rumen ciliates. 2. Ultrastructure of *Eudiplodinium maggii*. *Journal of Protozoology* 32: 205-214.

[60] Furness DN & Butler RD (1985b) The cytology of sheep rumen ciliates. 3. Ultrastructure of the genus *Entodinium* (Stein). *Journal of Protozoology* 32: 699-707.

[61] Gaertig J, Gao Y, Tishgarten T, Clark TG & Dickerson HW (1999) Surface display of a parasite antigen in the ciliate *Tetrahymena thermophila*. *Nature Biotechnology* 17: 462-465.

[62] Gagic D, Wen W, Collett MA & Rakonjac J (2013) Unique secreted-surface protein complex of *Lactobacillus rhamnosus*, identified by phage display. 2: 1-17.

[63] Gagic D, Maclean PH, Li D, Attwood GT & Moon CD (2015) Improving the genetic representation of rare taxa within complex microbial communities using DNA normalization methods. *Molecular ecology resources* 15: 464-476.

[64] Gagic D, Ciric M, Wen WX, Ng F & Rakonjac J (2016) Exploring the secretomes of microbes and microbial communities using filamentous phage display. *Frontiers in Microbiology* 7: 429.

[65] Gallotta M, Gancitano G, Pietrocola G, *et al.* (2014) SpyAD, a moonlighting protein of group A Streptococcus contributing to bacterial division and host cell adhesion. *Infection & Immunity* 82: 2890-2901.

[66] Gast RJ, Sanders RW & Caron DA (2009) Ecological strategies of protists and their symbiotic relationships with prokaryotic microbes. *Trends in Microbiology* 17: 563-569.

[67] Girard V & Mourez M (2006) Adhesion mediated by autotransporters of Gram-negative bacteria: Structural and functional features. *Research in Microbiology* 157: 407-416.

[68] Gupta A, Shrivastava N, Grover P, Singh A, Mathur K, Verma V, Kaur C & Chaudhary VK (2013) A novel helper phage enabling construction of genome-scale ORF-enriched phage display libraries. *PLoS ONE* 8: e75212.

[69] Guyader J, Eugene M, Noziere P, Morgavi DP, Doreau M & Martin C (2014) Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. *Animal* 8: 1816-1825.

[70] Halaby DM & Mornon JP (1998) The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. *Journal of Molecular Evolution* 46: 389-400.

[71] Hansen EE, Lozupone CA, Rey FE, *et al.* (2011) Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1: 4599-4606.

[72] Hartmann MD, Grin I, Dunin-Horkawicz S, Deiss S, Linke D, Lupas AN & Hernandez Alvarez B (2012) Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109: 20907-20912.

[73] Hegarty RS (1999) Reducing rumen methane emissions through elimination of rumen protozoa. *Australian Journal of Agricultural Research* 50: 1321-1327.

[74] Henderson B (2014) An overview of protein moonlighting in bacterial infection. *Biochemical Society transactions* 42: 1720-1727.

[75] Henderson G, Cox F, Ganesh S, Jonker A, Young W & Janssen PH (2015) Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 5: 14567.

[76] Henderson G, Cox F, Kittelmann S, Miri VH, Zethof M, Noel SJ, Waghorn GC & Janssen PH (2013) Effect of DNA extraction methods and sampling techniques on the apparent structure of cow and sheep rumen microbial communities. *PLoS ONE* 8: e74787.

[77] Hill GB, Ayers OM & Kohan AP (1987) Characteristics and sites of infection of *Eubacterium nodatum*, *Eubacterium timidum*, *Eubacterium brachy*, and other asaccharolytic eubacteria. *Journal of Clinical Microbiology* 25: 1540-1545.

[78] Hobson PN & Stewart CS (1997) *The Rumen Microbial Ecosystem*.

[79] Huberts DH & van der Klei IJ (2010) Moonlighting proteins: an intriguing mode of multitasking. *Biochimica et Biophysica Acta* 1803: 520-525.

[80] Hunter S, Jones P, Mitchell A, *et al.* (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research* 40: D306-D312.

[81] Huws SA, Edwards JE, Creevey CJ, Rees Stevens P, Lin W, Girdwood SE, Pachebat JA & Kingston-Smith AH (2016) Temporal dynamics of the metabolically active rumen bacteria colonizing fresh perennial ryegrass. *FEMS Microbiology Ecology* doi: 10.1093/femsec/fiv137

[82] Irbis C & Ushida K (2004) Detection of methanogens and proteobacteria from a single cell of rumen ciliate protozoa. *Journal of General and Applied Microbiology* 50: 203-212.

[83] Ishaq SL & Wright AD (2015) Wild ruminants. *Rumen microbiology: From evolution to revolution,* (Puniya BL, Singh R & Kamra DN, eds.), p. 37-45. Springer, India.

[84] Ishida K, Sekizuka T, Hayashida K, *et al.* (2014) Amoebal endosymbiont *Neochlamydia* genome sequence illuminates the bacterial role in the defense of the host amoebae against *Legionella pneumophila*. *PLoS ONE* 9: e95166.

[85] Ishii S, Kosaka T, Hori K, Hotta Y & Watanabe K (2005) Coaggregation facilitates interspecies hydrogen transfer between *Pelotomaculum thermopropionicum* and *Methanothermobacter thermautotrophicus*. 71: 7838-7845.

[86] Itoh T, Hibi T, Fujii Y, Sugimoto I, Fujiwara A, Suzuki F, Iwasaki Y, Kim JK, Taketo A & Kimoto H (2013) Cooperative degradation of chitin by extracellular and cell surface-expressed chitinases from *Paenibacillus* sp. strain FPU-7. *Applied and Environmental Microbiology* 79: 7482-7490.

[87] Ivan M, Veira DM & Kelleher CA (1986) The alleviation of chronic copper toxicity in sheep by ciliate protozoa. *British Journal of Nutrition* 55: 361-367.

[88] Jacobsson K, Rosander A, Bjerketorp J & Frykberg L (2003) Shotgun Phage Display - Selection for Bacterial Receptins or other Exported Proteins. *Biological Procedures Online* 5: 123-135.

[89] Jami E, White BA & Mizrahi I (2014) Potential role of the bovine rumen microbiome in modulating milk composition and feed efficiency. *PLoS ONE* 9: e85423.

[90] Jankovic D, Collett MA, Lubbers MW & Rakonjac J (2007) Direct selection and phage display of a Gram-positive secretome. *Genome Biology* 8. R266.

[91] Janssen PH & Kirs M (2008) Structure of the archaeal community of the rumen. *Applied and Environmental Microbiology* 74: 3619-3625.

[92] Jarocki VM, Tacchi JL & Djordjevic SP (2015) Non-proteolytic functions of microbial proteases increase pathological complexity. *Proteomics* 15: 1075-1088.

[93] Jarrell KF, Stark M, Nair DB & Chong JPJ (2011) Flagella and pili are both necessary for efficient attachment of *Methanococcus maripaludis* to surfaces. *FEMS Microbiology Letters* 319: 44-50.

[94] Jeffery CJ (1999) Moonlighting proteins. *Trends in Biochemical Sciences* 24: 8-11.

[95] Jeffery CJ (2015) Why study moonlighting proteins? *Frontiers in Genetics* 6: 211.

[96] Jia B, Cheong GW & Zhang S (2013) Multifunctional enzymes in archaea: promiscuity and moonlight. *Extremophiles* 17: 193-203.

[97] Jouany JP (1996) Effect of rumen protozoa on nitrogen utilization by ruminants. *Journal of Nutrition* 126: S1335-S1346.

[98] Kainulainen V & Korhonen TK (2014) Dancing to another tune-adhesive moonlighting proteins in bacteria. *Biology* 3: 178-204.

[99] Kawai S, Hashimoto W & Murata K (2010) Transformation of *Saccharomyces cerevisiae* and other fungi: methods and possible underlying mechanism. *Bioengineered Bugs* 1: 395-403.

[100] Kelly WJ, Pacheco DM, Li D, Attwood GT, Altermann E & Leahy SC (2016a) The complete genome sequence of the rumen methanogen *Methanobrevibacter millerae* SM9. *Standards in Genomic Sciences* 11: 49.

[101] Kelly WJ, Li D, Lambie SC, Cox F, Attwood GT, Altermann E & Leahy SC (2016b) Draft genome sequence of the rumen methanogen *Methanobrevibacter olleyae* YLM1. *Genome Announcements* 4: e00232-00216.

[102] Kenters N, Henderson G, Jeyanathan J, Kittelmann S & Janssen PH (2011) Isolation of previously uncultured rumen bacteria by dilution to extinction using a new liquid culture medium. *Journal of Microbiological Methods* 84: 52-60.

[103] Khan I, Chen Y, Dong T, Hong X, Takeuchi R, Mori H & Kihara D (2014) Genome-scale identification and characterization of moonlighting proteins. *Biology Direct* 9: 30.

[104] Kiessling KH, Pettersson H, Sandholm K & Olsen M (1984) Metabolism of aflatoxin, ochratoxin, zearalenone, and 3 trichotecenes by intact rumen fluid, rumen protozoa, and rumen bacteria. *Applied and Environmental Microbiology* 47: 1070-1073.

[105] Kišidayová S, Houserova P, Váradyová Z, Mihaliková, K., Pristaš P & Javorský P (2010) Bacterial-protozoal interactions in a microbial community of rumen ciliate *Entodinium caudatum* culture under mercury stress. *Canadian Journal of Microbiology* 56: 202-208.

[106] Kisiela D, Laskowska A, Sapeta A, Kuczkowski M, Wieliczko A & Ugorski M (2006) Functional characterization of the FimH adhesin from *Salmonella enterica* serovar Enteritidis. *Microbiology* 152: 1337-1346.

[107] Kittelmann S & Janssen PH (2011) Characterization of rumen ciliate community composition in domestic sheep, deer, and cattle, feeding on varying diets, by means of PCR-DGGE and clone libraries. *FEMS Microbiology Ecology* 75: 468-481.

[108] Kittelmann S, Devente SR, Kirk MR, Seedorf H, Dehority BA & Janssen PH (2015) Phylogeny of intestinal ciliates, including *Charonina ventriculi*, and comparison of microscopy and 18S rRNA gene pyrosequencing for rumen ciliate community structure analysis. *Applied and Environmental Microbiology* 81: 2433-2444.

[109] Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI & Janssen PH (2013) Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS ONE* 8: e47879.

[110] Kittelmann S, Pinares-Patino CS, Seedorf H, Kirk MR, Ganesh S, McEwan JC & Janssen PH (2014) Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLoS ONE* 9: e103171.

[111] Kline KA, Falker S, Dahlberg S, Normark S & Henriques-Normark B (2009) Bacterial adhesins in host-microbe interactions. *Cell Host & Microbe* 5: 580-592.

[112] Koba H, Okuda K, Watanabe H, Tagami J & Senpuku H (2009) Role of lysine in interaction between surface protein peptides of *Streptococcus gordonii* and agglutinin peptide. *Oral Microbiology and Immunology* 24: 162-169.

[113] Koch EJ, Miyashiro T, McFall-Ngai MJ & Ruby EG (2014) Features governing symbiont persistence in the squid-vibrio association. *Molecular Ecology* 23: 1624-1634.

[114] Kouzuma A, Kato S & Watanabe K (2015) Microbial interspecies interactions: recent findings in syntrophic consortia. *Frontiers in Microbiology* 6: 477.

[115] Kraatz M, Wallace RJ & Svensson L (2011) *Olsenella umbonata* sp. nov., a microaerotolerant anaerobic lactic acid bacterium from the sheep rumen and pig jejunum, and emended descriptions of *Olsenella*, *Olsenella uli* and *Olsenella profusa*. *International Journal of Systematic and Evolutionary Microbiology* 61: 795-803.

[116] Krishnan V (2015) Pilins in gram-positive bacteria: A structural perspective. *IUBMB Life* 67: 533-543.

[117] Kumar S, Puniya BL, Parween S, Nahar P & Ramachandran S (2013) Identification of novel adhesins of *M. tuberculosis* H37Rv using integrated approach of multiple computational algorithms and experimental analysis. *PLoS ONE* 8: e69790.

[118] Kyriakis JM (2014) In the beginning, there was protein phosphorylation. *Journal of Biological Chemistry* 289: 9460-9462.

[119] Leahy SC, Kelly WJ, Li D, Li Y, Altermann E, Lambie SC, Cox F & Attwood GT (2013) The complete genome sequence of *Methanobrevibacter* sp. AbM4. *Standards in Genomic Sciences* 8: 215-227.

[120] Leahy SC, Kelly WJ, Altermann E, *et al.* (2010) The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLoS ONE* 5: e8926.

[121] Lebeer S, Vanderleyden J & De Keersmaecker SC (2010) Host interactions of probiotic bacterial surface molecules: comparison with commensals and pathogens. *Nature Reviews Microbiology* 8: 171-184.

[122] Leo JC & Skurnik M (2011) Adhesins of human pathogens from the genus *Yersinia*. *Advances in Experimental Medicine and Biology* 715: 1-15.

[123] Letunic I, Doerks T & Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* 40: D302-305.

[124] Leung TLF & Poulin R (2008) Parasitism, commensalism, and mutualism: Exploring the many shades of symbioses. *Vie et Milieu* 58: 107-115.

[125] Levine M (2011) Susceptibility to dental caries and the salivary proline-rich proteins. *International Journal of Dentistry* 2011: Article 953412.

[126] Li M, Zhou M, Adamowicz E, Basarab JA & Guan LL (2012a) Characterization of bovine ruminal epithelial bacterial communities using 16S rRNA sequencing, PCR-DGGE, and qRT-PCR analysis. *Veterinary Microbiology* 155: 72-80.

[127] Li RW, Connor EE, Li C, Baldwin Vi RL & Sparks ME (2012b) Characterization of the rumen microbiota of pre-ruminant calves using metagenomic tools. *Environmental Microbiology* 14: 129-139.

[128] Li RW, Giarrizzo JG, Wu S, Li W, Duringer JM & Craig AM (2014) Metagenomic insights into the RDX-degrading potential of the ovine rumen microbiome. *PLoS ONE* 9: e110505.

[129] Li W & Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658-1659.

[130] Li YF, Poole S, Rasulova F, McVeigh AL, Savarino SJ & Xia D (2007) A receptor-binding site as revealed by the crystal structure of CfaE, the colonization factor antigen I fimbrial adhesin of enterotoxigenic *Escherichia coli*. *Journal of Biological Chemistry* 282: 23970-23980.

[131] Li YF, Poole S, Rasulova F, McVeigh AL, Savarino SJ & Xia D (2009) Crystallization and preliminary X-ray diffraction analyses of several forms of the CfaB major subunit of enterotoxigenic *Escherichia coli* CFA/I fimbriae. *Acta Crystallographica Section F, Structural Biology and Crystallization Communications* 65: 242-247.

[132] Lima SS, Ching ATC, Favaro RD, Da Silva JB, Oliveira MLS, Carvalho E, Abreu PAE, Vasconcellos SA & Ho PL (2013) Adhesin activity of *Leptospira interrogans* lipoprotein identified by *in vivo* and *in vitro* shotgun phage display. *Biochemical and Biophysical Research Communications* 431: 342-347.

[133] Lin YP & Chang YF (2007) A domain of the *Leptospira* LigB contributes to high affinity binding of fibronectin. *Biochemical and Biophysical Research Communications* 362: 443-448.

[134] Lin YP, McDonough SP, Sharma Y & Chang YF (2010) The terminal immunoglobulin-like repeats of LigA and LigB of *Leptospira* enhance their binding to gelatin binding domain of fibronectin and host cells. *PLoS ONE* 5: e11301.

[135] Liu JH, Bian GR, Zhu WY & Mao SY (2015) High-grain feeding causes strong shifts in ruminal epithelial bacterial community and expression of Toll-like receptor genes in goats. *Frontiers in Microbiology* 6: 167.

[136] Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L & Law M (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology* 2012: 251364.

[137] Lizcano A, Sanchez CJ & Orihuela CJ (2012) A role for glycosylated serine-rich repeat proteins in gram-positive bacterial pathogenesis. *Molecular Oral Microbiology* 27: 257-269.

[138] Lloyd D, Williams AG, Amann R, Hayes AJ, Durrant L & Ralphs JR (1996) Intracellular prokaryotes in rumen ciliate protozoa: Detection by confocal laser scanning

microscopy after *in situ* hybridization with fluorescent 16S rRNA probes. *European Journal of Protistology* 32: 523-531.

[139] Loimaranta V, Hytonen J, Pulliainen AT, Sharma A, Tenovuo J, Stromberg N & Finne J (2009) Leucine-rich repeats of bacterial surface proteins serve as common pattern recognition motifs of human scavenger receptor gp340. *Journal of Biological Chemistry* 284: 18614-18623.

[140] Lunder M, Bratkovič T, Urleb U, Kreft S & Štrukelj B (2008) Ultrasound in phage display: a new approach to nonspecific elution. *BioTechniques* 44: 893-900.

[141] Lynch MD & Neufeld JD (2015) Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology* 13: 217-229.

[142] Lynn DH (2008) Subphylum 2. Intramacronucleata: Class 3. Litostomatea – Simple ciliates but highly derived. *The Ciliated Protozoa,*(Lynn DH, ed.) p. 197-201. Springer Netherlands.

[143] Lyskowski A, Leo JC & Goldman A (2011) Structure and biology of trimeric autotransporter adhesins. *Advances in Experimental Medicine and Biology* 715: 143-158.

[144] Machmuller A, Soliva CR & Kreuzer M (2003) Effect of coconut oil and defaunation treatment on methanogenesis in sheep. *Reproduction Nutrition Development* 43: 41-55.

[145] Mai Z & Samuelson J (1998) A new gene family (ariel) encodes asparagine-rich *Entamoeba histolytica* antigens, which resemble the amebic vaccine candidate serine-rich *E. histolytica* protein. *Infection & Immunity* 66: 353-355.

[146] Makarova KS, Aravind L & Koonin EV (1999) A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein Science* 8: 1714-1719.

[147] Mao SY, Huo WJ & Zhu WY (2016) Microbiome-metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain in a goat model. *Environmental Microbiology* 18:525-541.

[148] Mardis ER (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics* 9: 387-402.

[149] Markowitz VM, Chen IM, Chu K, Pati A, Ivanova NN & Kyrpides NC (2015) Ten years of maintaining and expanding a microbial genome and metagenome analysis system. *Trends in Microbiology* 23: 730-741.

[150] Markowitz VM, Chen IM, Palaniappan K, *et al.* (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Research* 40: D115-122.

[151] Martin C, Morgavi DP & Doreau M (2010) Methane mitigation in ruminants: from microbe to the farm scale. *Animal* 4: 351-365.

[152] Massey RC, Kantzanou MN, Fowler T, Day NP, Schofield K, Wann ER, Berendt AR, Hook M & Peacock SJ (2001) Fibronectin-binding protein A of *Staphylococcus aureus* has multiple, substituting, binding regions that mediate adherence to fibronectin and invasion of endothelial cells. *Cellular Microbiology* 3: 839-851.

[153] Matochko WL & Derda R (2015) Next-generation sequencing of phage-displayed peptide libraries. *Methods in Molecular Biology* 1248: 249-266.

[154] McCafferty J (1996) Phage display: Factors affecting panning efficiency. *Phage Display of Peptides and Proteins - A Laboratory manual,* (Kay BK, Winter J & McCafferty J, eds.), p. 261-276. Academic Press Limited, London, U.K.

[155] McCann JC, Wickersham TA & Loor JJ (2014) High-throughput methods redefine the rumen microbiome and its relationship with nutrition and metabolism. *Bioinformatics and Biology Insights* 8: 109-125.

[156] McFall-Ngai MJ (2014) The importance of microbes in animal development: lessons from the squid-*Vibrio* symbiosis. *Annual Review of Microbiology* 68: 177-194.

[157] Mei S, Zhang J, Zhang X & Tu X (2015) Solution structure of leptospiral LigA4 Big domain. *Biochemical and Biophysical Research Communications* 467: 288-292.

[158] Merckel MC, Tanskanen J, Edelman S, Westerlund-Wikstrom B, Korhonen TK & Goldman A (2003) The structural basis of receptor-binding by *Escherichia coli* associated with diarrhea and septicemia. *Journal of Molecular Biology* 331: 897-905.

[159] Mesarich CH, Bowen JK, Hamiaux C & Templeton MD (2015) Repeat-containing protein effectors of plant-associated organisms. *Frontiers in Plant Science* 6: 872.

[160] Monteils V, Rey M, Silberberg M, Cauquil L & Combes S (2012) Modification of activities of the ruminal ecosystem and its bacterial and protozoan composition during repeated dietary changes in cows. *Journal of Animal Science* 90: 4431-4440.

[161] Morgavi DP, Martin C, Jouany JP & Ranilla MJ (2012) Rumen protozoa and methanogenesis: not a simple cause-effect relationship. *British Journal of Nutrition* 107: 388-397.

[162] Morris BE, Henneberger R, Huber H & Moissl-Eichinger C (2013) Microbial syntrophy: interaction for the common good. *FEMS Microbiology Reviews* 37: 384-406.

[163] Morris EJ & Cole OJ (1987) Relationship between cellulolytic activity and adhesion to cellulose in *Ruminococcus albus*. *Journal of General Microbiology* 133: 1023-1032.

[164] Mosher JJ, Bowman B, Bernberg EL, Shevchenko O, Kan J, Korlach J & Kaplan LA (2014) Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *Journal of Microbiological Methods* 104: 59-60.

[165] Mullen LM, Nair SP, Ward JM, Rycroft AN & Henderson B (2006) Phage display in the study of infectious diseases. *Trends in Microbiology* 14: 141-147.

[166] Muller M (1993) The hydrogenosome. *Journal of General Microbiology* 139: 2879-2889.

[167] Myllykangas S, Buenrostro J & Ji HP (2012) Overview of Sequencing Technology Platforms. *Bioinformatics for High Throughput Sequencing,* (Rodríguez-Ezpeleta N, Hackenberg M & Aransay AM, eds.), p. 11-25. Springer-Verlag New York.

[168] Nakamura Y, Gojobori T & Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* 28: 292.

[169] New Zealand Ministry for the Environment (2015) Snapshot April 2015 - New Zealand's Greenhouse Gas Inventory 1990-2013. Ministry for the Environment, Wellington.

[170] Newbold CJ, de la Fuente G, Belanche A, Ramos-Morales E & McEwan NR (2015) The role of ciliate protozoa in the rumen. *Frontiers in Microbiology* 6: 1313.

[171] Ng F, Kittelmann S, Patchett ML, Attwood GT, Janssen PH, Rakonjac J & Gagic D (2015) An adhesin from hydrogen-utilizing rumen methanogen *Methanobrevibacter ruminantium* M1 binds a broad range of hydrogen-producing microorganisms. *Environmental Microbiology* doi: 10.1111/1462-2920.13155.

[172] Ngubane NA, Gresh L, Ioerger TR, Sacchettini JC, Zhang YJ, Rubin EJ, Pym A & Khati M (2013) High-throughput sequencing enhanced phage display identifies peptides that bind mycobacteria. *PLoS ONE* 8: e77844.

[173] Noel SJ (2013) Cultivation and community composition analysis of plant-adherent rumen bacteria. Thesis, Massey University, Palmerston North, New Zealand.

[174] Ogimoto K & Imai S (1981) *Atlas of rumen microbiology*. Japan Scientific Societies Press, Tokyo, Japan.

[175] Orpin CG & Hall FJ (1983) Surface structures of the rumen holotrich protozoon *Isotricha intestinalis* with particular reference to the attachment zone. *Current Microbiology* 8: 321-325.

[176] Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, Arvanitidis C & Iliopoulos I (2015) *Bioinformatics and Biology Insights.* 9: 75-88.

[177] Ozutsumi Y, Tajima K, Takenaka A & Itabashi H (2005) The effect of protozoa on the composition of rumen bacteria in cattle using 16S rRNA gene clone libraries. *Bioscience, Biotechnology, and Biochemistry* 69: 499-506.

[178] Paul K, Nonoh JO, Mikulski L & Brune A (2012) "*Methanoplasmatales*," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Applied and Environmental Microbiology* 78: 8245-8253.

[179] Pei CX, Mao SY, Cheng YF & Zhu WY (2010) Diversity, abundance and novel 16S rRNA gene sequences of methanogens in rumen liquid, solid and epithelium fractions of Jinnan cattle. *Animal* 4: 20-29.

[180] Pepper LR, Cho YK, Boder ET & Shusta EV (2008) A decade of yeast surface display technology: where are we now? *Combinatorial Chemistry & High Throughput Screening* 11: 127-134.

[181] Perras AK, Wanner G, Klingl A, *et al.* (2014) Grappling archaea: ultrastructural analyses of an uncultivated, cold-loving archaeon, and its biofilm. *Frontiers in Microbiology* 5: 397.

[182] Perras AK, Daum B, Ziegler C, *et al.* (2015) S-layers at second glance? Altiarchaeal grappling hooks (hami) resemble archaeal S-layer proteins in structure and sequence. *Frontiers in Microbiology* 6: 543.

[183] Pethe K, Aumercier M, Fort E, Gatot C, Locht C & Menozzi FD (2000) Characterization of the heparin-binding site of the mycobacterial heparin-binding hemagglutinin adhesin. *Journal of Biological Chemistry* 275: 14273-14280.

[184] Prakash T & Taylor TD (2012) Functional assignment of metagenomic data: challenges and applications. *Briefings in Bioinformatics* 13: 711-727.

[185] Prosdocimi EM, Mapelli F, Gonella E, Borin S & Crotti E (2015) Microbial ecology-based methods to characterize the bacterial communities of non-model insects. *Journal of Microbiological Methods* 119: 110-125.

[186] Pruesse, E., Peplies, J. and Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28: 1823-1829

[187] Qi H, Lu H, Qiu HJ, Petrenko V & Liu A (2012) Phagemid vectors for phage display: properties, characteristics and construction. *Journal of Molecular Biology* 417: 129-143.

[188] Qiao Y, Zhou L, Wang Y, Liu F, Xie P, Chen Y, Zhang D & Zhao X (2012) Screening for antibodies against intact cancer cells with a naive large phage antibody library. *International Journal of Molecular Medicine* 29: 37-46.

[189] Rakonjac J, Jovanovic G & Model P (1997) Filamentous phage infection-mediated gene expression: construction and propagation of the gIII deletion mutant helper phage R408d3. *Gene* 198: 99-103.

[190] Rakonjac J, Feng JN & Model P (1999) Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *Journal of Molecular Biology* 289: 1253-1265.

[191] Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D & Russel M (2011) Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Current Issues in Molecular Biology* 13: 51-76.

[192] Regensbogenova M, McEwan NR, Javorsky P, Kisidayova S, Michalowski T, Newbold CJ, Hackstein JHP & Pristas P (2004) A re-appraisal of the diversity of the methanogens associated with the rumen ciliates. *FEMS Microbiology Letters* 238: 307-313.

[193] Rego AT, Johnson JG, Gelbel S, Enguita FJ, Clegg S & Waksman G (2012) Crystal structure of the MrkD1P receptor binding domain of *Klebsiella pneumoniae* and identification of the human collagen V binding interface. *Molecular Microbiology* 86: 882-893.

[194] Rhoads A & Au KF (2015) PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* 13: 278-289.

[195] Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42: W320-324.

[196] Rogers EA, Das A & Ton-That H (2011) Adhesion by pathogenic corynebacteria. *Advances in Experimental Medicine and Biology* 715: 91-103.

[197] Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Research* 11: 3-11.

[198] Rondot S, Koch J, Breitling F & Dubel S (2001) A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnology* 19: 75-78.

[199] Rosander A, Guss B, Frykberg L, Bjorkman C, Naslund K & Pringle M (2011) Identification of immunogenic proteins in *Treponema phagedenis*-like strain V1 from digital dermatitis lesions by phage display. *Veterinary Microbiology* 153: 315-322.

[200] Rothberg JM, Hinz W, Rearick TM, *et al.* (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348-352.

[201] Russell JB & Rychlik JL (2001) Factors that alter rumen microbial ecology. *Science* 292: 1119-1122.

[202] Sachdeva G, Kumar K, Jain P & Ramachandran S (2005) SPAAN: a software program for prediction of adhesins and adhesin-like proteins using neural networks. *Bioinformatics* 21: 483-491.

[203] Samuel BS, Hansen EE, Manchester JK, Coutinho PM, Henrissat B, Fulton R, Latreille P, Kim K, Wilson RK & Gordon JI (2007) Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proceedings of the National Academy of Sciences of the United States of America* 104: 10643-10648.

[204] Sauret C, Severin T, Vetion G, Guigue C, Goutx M, Pujo-Pay M, Conan P, Fagervold SK & Ghiglione JF (2014) 'Rare biosphere' bacteria as key phenanthrene degraders in coastal seawaters. *Environmental Pollution* 194: 246-253.

[205] Schink B (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology and Molecular Biology Reviews* 61: 262-280.

[206] Schink B (2009) *Bergey's Manual of Systematic Bacteriology*. Springer-Verlag, New York, NY.

[207] Schink B & Thauer RK (1988) *Energetics of syntrophic methane formation and the influence of aggregation*. Pudoc, Wageningen.

[208] Schopf S, Wanner G, Rachel R & Wirth R (2008) An archaeal bi-species biofilm formed by *Pyrococcus furiosus* and *Methanopyrus kandleri*. *Archives of Microbiology* 190: 371-377.

[209] Scott JK & Barbas CFr (2001) Phage-display vectors. *Phage Display: A Laboratory Manual,*(Barbas CFr, Burton DR, Scott JK & Silverman GJ, eds.), p. 2-8. Cold Spring Harbor Press, Cold Spring Harbor, NY.

[210] Seedorf H, Kittelmann S & Janssen PH (2015) Few highly abundant operational taxonomic units dominate within rumen methanogenic archaeal species in New Zealand sheep and cattle. *Applied and Environmental Microbiology* 81: 986-995.

[211] Seedorf H, Kittelmann S, Henderson G & Janssen PH (2014) RIM-DB: a taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ* 2: e494.

[212] Shi PJ, Meng K, Zhou ZG, Wang YR, Diao QY & Yao B (2008) The host species affects the microbial community in the goat rumen. *Letters in Applied Microbiology* 46: 132-135.

[213] Shi W, Moon CD, Leahy SC, *et al.* (2014) Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Research* 24: 1517-1525.

[214] Shimoyama T, Kato S, Ishii S & Watanabe K (2009) Flagellum mediates symbiosis. *Science* 323: 1574-1574.

[215] Shin EC, Choi BR, Lim WJ, *et al.* (2004) Phylogenetic analysis of archaea in three fractions of cow rumen based on the 16S rDNA sequence. *Anaerobe* 10: 313-319.

[216] Sievers F, Wilm A, Dineen D, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7: 539.

[217] Siqueira Jé F, Fouad AF & Rôças IN (2012) Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of Oral Microbiology* 4. doi: 10.3402/jom.v4i0.10743

[218] Sirohi SK, Singh N, Dagar SS & Puniya AK (2012) Molecular tools for deciphering the microbial community structure and diversity in rumen ecosystem. *Applied Microbiology and Biotechnology* 95: 1135-1154.

[219] Skillman LC, Evans PN, Strompl C & Joblin KN (2006) 16S rDNA directed PCR primers and detection of methanogens in the bovine rumen. *Letters in Applied Microbiology* 42: 222-228.

[220] Smith DR, Doucette-Stamm LA, Deloughery C, *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparative genomics. *Journal of Bacteriology* 179: 7135-7155.

[221] Smith GP (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228: 1315-1317.

[222] Smith TF, Gaitatzes C, Saxena K & Neer EJ (1999) The WD repeat: a common architecture for diverse functions. *Trends in Biochemical Sciences* 24: 181-185.

[223] Snelling TJ, Genç B, McKain N, Watson M, Waters SM, Creevey CJ, Wallace RJ (2014) Diversity and community composition of methanogenic archaea in the rumen of Scottish upland sheep assessed by different methods. *PLoS ONE* 9:e106491

[224] Steenbakkers PJ, Geerts WJ, Ayman-Oz NA & Keltjens JT (2006) Identification of pseudomurein cell wall binding domains. *Molecular Microbiology* 62: 1618-1630.

[225] Stumm CK, Gijzen HJ & Vogels GD (1982) Association of methanogenic bacteria with ovine rumen ciliates. *British Journal of Nutrition* 47: 95-99.

[226] Sylvester JT, Karnati SK, Yu Z, Newbold CJ & Firkins JL (2005) Evaluation of a real-time PCR assay quantifying the ruminal pool size and duodenal flow of protozoal nitrogen. *Journal of Dairy Science* 88: 2083-2095.

[227] Thoma C, Frank M, Rachel R, Schmid S, Nather D, Wanner G & Wirth R (2008) The Mth60 fimbriae of *Methanothermobacter thermoautotrophicus* are functional adhesins. *Environmental Microbiology* 10: 2785-2795.

[228] Thorasin T, Hoyles L & McCartney AL (2015) Dynamics and diversity of the '*Atopobium* cluster' in the human faecal microbiota, and phenotypic characterization of '*Atopobium* cluster' isolates. *Microbiology* 161: 565-579.

[229] Tjhung KF, Deiss F, Tran J, Chou Y & Derda R (2015) Intra-domain phage display (ID-PhD) of peptides and protein mini-domains censored from canonical pIII phage display. *Frontiers in Microbiology* 6: 340.

[230] Tokura M, Ushida K, Miyazaki K & Kojima Y (1997) Methanogens associated with rumen ciliates. *FEMS Microbiology Ecology* 22: 137-143.

[231] Tokura M, Chagan I, Ushida K & Kojima Y (1999) Phylogenetic study of methanogens associated with rumen ciliates. *Current Microbiology* 39: 123-128.

[232] Travers KJ, Chin CS, Rank DR, Eid JS & Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research* 38: e159.

[233] Tymensen L, Barkley C & McAllister TA (2012a) Relative diversity and community structure analysis of rumen protozoa according to T-RFLP and microscopic methods. *Journal of Microbiological Methods* 88: 1-6.

[234] Tymensen LD, Beauchemin KA & McAllister TA (2012b) Structures of free-living and protozoa-associated methanogen communities in the bovine rumen differ according to comparative analysis of 16S rRNA and *mcrA* genes. *Microbiology* 158: 1808-1817.

[235] Upadhyay SK, Mahajan L, Ramjee S, Singh Y, Basir SF & Madan T (2009) Identification and characterization of a laminin-binding protein of *Aspergillus fumigatus*: extracellular thaumatin domain protein (AfCalAp). *Journal of Medical Microbiology* 58: 714-722.

[236] Valle ER, Henderson G, Janssen PH, Cox F, Alexander TW & McAllister TA (2015) Considerations in the use of fluorescence in situ hybridization (FISH) and confocal laser scanning microscopy to characterize rumen methanogens and define their spatial distributions. *Canadian Journal of Microbiology* 61: 417-428.

[237] Velasquez-Manoff M (2015) Gut microbiome: the peacekeepers. 518: S3-S11.

[238] Vigues B & Groliere CA (1985) Evidence for a Ca^{2+}-binding protein associated to non-actin microfilamentous systems in two ciliated protozoans. *Experimental Cell Research* 159: 366-376.

[239] Visweswaran GRR, Dijkstra BW & Kok J (2011) A minimum of three motifs is essential for optimal binding of pseudomurein cell wall-binding domain of *Methanothermobacter thermautotrophicus*. 6: e21582.

[240] Vodnik M, Zager U, Strukelj B & Lunder M (2011) Phage display: selecting straws instead of a needle from a haystack. *Molecules* 16: 790-817.

[241] Vogel J (2008) Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* 11: 301-307.

[242] Vogels GD, Hoppe WF & Stumm CK (1980) Association of methanogenic bacteria with rumen ciliates. *Applied and Environmental Microbiology* 40: 608-612.

[243] Wagner C & Hensel M (2011) Adhesive mechanisms of *Salmonella enterica*. *Advances in Experimental Medicine and Biology* 715: 17-34.

[244] Wallace RJ, Rooke JA, McKain N, Duthie CA, Hyslop JJ, Ross DW, Waterhouse A, Watson M & Roehe R (2015) The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 16: 839.

[245] Wang T, Zhang J, Zhang X, Xu C & Tu X (2013) Solution structure of the Big domain from *Streptococcus pneumoniae* reveals a novel Ca^{2+}-binding module. *Scientific Reports* 3: 1079.

[246] Watanabe T, Asakawa S, Nakamura A, Nagaoka K & Kimura M (2004) DGGE method for analyzing 16S rDNA of methanogenic archaeal community in paddy field soil. *FEMS Microbiology Letters* 232: 153-163.

[247] Westerlund-Wikstrom B & Korhonen TK (2005) Molecular structure of adhesin domains in *Escherichia coli* fimbriae. *International Journal of Medical Microbiology* 295: 479-486.

[248] Wetzels SU, Mann E, Metzler-Zebeli BU, Wagner M, Klevenhusen F, Zebeli Q & Schmitz-Esser S (2015) Pyrosequencing reveals shifts in the bacterial epimural community relative to dietary concentrate amount in goats. *Journal of Dairy Science* 98: 5572-5587.

[249] Williams AG & Coleman GS (1992) *The Rumen Protozoa.* Springer-Verlag New York Inc., New York.

[250] Williams AG, Withers S & Sutherland AD (2013) The potential of bacteria isolated from ruminal contents of seaweed-eating North Ronaldsay sheep to hydrolyse seaweed components and produce methane by anaerobic digestion *in vitro*. *Microbial Biotechnology* 6: 45-52.

[251] Wright AD (2015) Rumen Protozoa. *Rumen microbiology: From evolution to revolution,*(Puniya AK, Singh R & Kamra DN, eds.), p. 113-120. Springer, India.

[252] Wuichet K & Sogaard-Andersen L (2015) Evolution and diversity of the Ras superfamily of small GTPases in prokaryotes. *Genome Biology and Evolution* 7: 57-70.

[253] Xia Y, Kong YH, Seviour R, Forster RJ, Kisidayova S & McAllister TA (2014) Fluorescence *in situ* hybridization probing of protozoal *Entodinium* spp. and their methanogenic colonizers in the rumen of cattle fed alfalfa hay or triticale straw. *Journal of Applied Microbiology* 116: 14-22.

[254] Yin J, Liu F, Schinke M, Daly C & Walsh CT (2004) Phagemid encoded small molecules for high throughput screening of chemical libraries. *Journal of the American Chemical Society* 126: 13570-13571.

[255] Yuan S, Cohen DB, Ravel J, Abdo Z & Forney LJ (2012) Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE* 7: e33865.

[256] Zavialov A, Zav'yalova G, Korpela T & Zav'yalov V (2007) FGL chaperone-assembled fimbrial polyadhesins: anti-immune armament of Gram-negative bacterial pathogens. *FEMS Microbiology Reviews* 31: 478-514.

[257] Zened A, Combes S, Cauquil L, Mariette J, Klopp C, Bouchez O, Troegeler-Meynadier A & Enjalbert F (2013) Microbial ecology of the rumen evaluated by 454 GS FLX pyrosequencing is affected by starch and oil supplementation of diets. *FEMS Microbiology Ecology* 83: 504-514.

[258] Zhang K, He J, Yang M, Yen M & Yin J (2009) Identifying natural product biosynthetic genes from a soil metagenome by using T7 phage selection. *Chembiochem* 10: 2599-2606.

[259] Zhang W, Ciclitira P & Messing J (2014) PacBio sequencing of gene families - a case study with wheat gluten genes. *Gene* 533: 541-546.

[260] Zhang Y, Baranov PV, Atkins JF & Gladyshev VN (2005) Pyrrolysine and selenocysteine use dissimilar decoding strategies. *Journal of Biological Chemistry*. 280: 20740-20751.

[261] Zhang YQ, Gao W & Meng QX (2007) Fermentation of plant cell walls by ruminal bacteria, protozoa and fungi and their interaction with fibre particle size. *Archives of Animal Nutrition* 61: 114-125.

[262] Zheng Y, Kahnt J, Kwon IH, Mackie RI & Thauer RK (2014) Hydrogen formation and its regulation in *Ruminococcus albus*: involvement of an electron-bifurcating [FeFe]-hydrogenase, of a non-electron-bifurcating [FeFe]-hydrogenase, and of a putative hydrogen-sensing [FeFe]-hydrogenase. *Journal of Bacteriology* 196: 3840-3852.

[263] Zhou M, Hernandez-Sanabria E & Guan LL (2009) Assessment of the microbial ecology of ruminal methanogens in cattle with different feed efficiencies. *Applied and Environmental Microbiology* 75: 6524-6533.

[264] Zhou M, Chung YH, Beauchemin KA, Holtshausen L, Oba M, McAllister TA & Guan LL (2011) Relationship between rumen methanogens and methane production in dairy cows fed diets supplemented with a feed enzyme additive. *Journal of Applied Microbiology* 111: 1148-1158.

Chapter 7. Appendices

Table 7.1. Distribution of bacterial phyla present in the DNA sample derived from protozoa-associated symbionts *versus* rumen contents.

Phylum	Sequence reads (%)	
	Total rumen contents (Henderson *et al.*, 2013)	Protozoa-associated
Actinobacteria	0.027	8.3
Bacteroidetes	71.6	12
Chloroflexi	0	1.3
Cyanobacteria	0.025	0.06
Elusimicrobia	0	0.07
Fibrobacteres	3.2	0.02
Firmicutes	22.1	75
Lentisphaerae	0	0.13
Planctomycetes	0	0.24
Proteobacteria	0.012	1.1
Spirochaetes	0.8	0.35
Tenericutes	2.0	0.86
TM7	0.15	0.19
Verrucomicrobia	0	0.31
Others (unassigned or <0.01% of sequence reads)	0.086	0.07

7.1 Statement of contribution to doctoral thesis containing publications (DRC16)

DRC 16

MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION TO DOCTORAL THESIS CONTAINING PUBLICATIONS

(To appear at the end of each thesis chapter/section/appendix submitted as an article/paper or collected as an appendix at the end of the thesis)

We, the candidate and the candidate's Principal Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of Candidate: Filomena Ng

Name/Title of Principal Supervisor: Dragana Gagic / Senior lecturer

Name of Published Research Output and full reference:

Ng F, Kittelmann S, Patchett ML, Attwood GT, Janssen PH, Rakonjac J, Gagic D. (2015) An adhesin from hydrogen-utilizing rumen methanogen Methanobrevibacter ruminantium M1 binds a broad range of hydrogen-producing microorganisms. Environmental Microbiology. doi: 10.1111/1462-2920.13155

In which Chapter is the Published Work:

Please indicate either:

- The percentage of the Published Work that was contributed by the candidate:

  and / or

- Describe the contribution that the candidate has made to the Published Work:

  FN carried out 95% of hands-on experimental work and bioinformatics analyses. Manuscript was written by FN and DG.

Candidate's Signature — 21 August 2016 (Date)

Gagic Dragana — 25 August 2016 (Date)

Principal Supervisor's signature

GRS Version 3– 16 September 2011