Strategies for improving ruminant utilisation of high grain diets: Pangenome of *Streptococcus bovis*

Agri-Science Queensland Innovation Opportunity

July 2016

Rosalind Gilbert
Senior Research Scientist
Rumen Ecology Unit
Applied Biotechnology and Intensive Livestock
Animal Science

This publication has been compiled by Rosalind Gilbert of the Rumen Ecology Unit, Animal Science, Department of Agriculture and Fisheries.

Acknowledgments: The author would like to acknowledge the assistance of the Rumen Ecology staff members Diane Ouwerkerk, Cathy Minchin, Megan Vance and Carl Davis; the DSITI HPC support (Lindsay Brebber and David Diggles); and the University of Queensland Mass Spectrometry Facility (Dr Amanda Nouwens and Peter Josh) and Annette Dexter (UQ School of Agriculture and Food Sciences) for proteomics advice. The author would also like to acknowledge the assistance of Dr William Kelly and the use of genomic sequences provided by the Hungate 1000 (supported by the NZ Government, the Livestock Research Group of the Global Research Alliance on Agricultural Greenhouse Gases and the US Department of Energy Joint Genome Institute Community Science Program).

Summary

The field of microbiology is being transformed by the increased ease and reduced costs of DNA sequencing. An initiative of the United States Department of Energy (USDOE), Joint Genome Institute facilitated the establishment of a Community Sequencing Project, the Hungate 1000, an international collaborative venture which aimed to sequence the genomes of bacteria isolated from the rumen and intestinal tract of herbivores. The DAF Rumen Ecology Unit (REU) contributed 36 rumen bacteria and seven bacteriophages to this genome sequencing effort. Thirteen of these bacteria belonged to the *Streptococcus bovis/Streptococcus equinus* complex (SBSEC). These are virulent, fast-growing commensal bacteria which over-proliferate in the rumen of cattle when they transition too quickly from forage to high grain feedlot rations, causing rumen lactic acidosis. Costly management strategies based on lengthy feed transition times (up to 20 days), and the use of feed additives including rumen modifiers such as monensin and antibiotics, are currently used to control rumen acidosis. This project used the SBEC genome sequences generated by the Hungate 1000, as well as publically available SBEC genome sequences, to create a combined genome dataset or pangenome. This pangenome of 42 genome sequences was then used to identify:

1. factors which enable these organisms to rapidly proliferate on starch-rich diets, for example, carbohydrate degrading enzymes; and
2. new approaches to control these organisms, for example inhibitory peptides (bacteriocins) and enzymes, carried by naturally-occurring viruses predatory to bacteria (known as bacteriophage or phage), which can specifically target and puncture bacterial cell walls.

During the course of the project, technical capacity of the REU was also developed to facilitate the analysis and secure storage of large sequence datasets, with the expansion of high performance computing resources and installation of specialist software. Several of the phage proteins identified by the pangenome analysis were shown to be intact and functional, with the formation and release of phage particles from bacterial host cells verified by transmission electron microscopy and proteomics analysis. These phage proteins represent potential candidates for further investigation as antimicrobials to control SBEC, and will be used as background proof of concept for the development of project proposals for submission to funding agencies.

Table of contents

Table of figures

Table of images

Table of tables

Background

The *Streptococcus bovis*/*Streptococcus equinus* complex (SBSEC) is a group of animal and human-derived commensal bacteria, found in the rumen and gastrointestinal tract. *Streptococcus bovis is* a virulent, fast-growing organism which over-proliferates in the rumen of cattle when they transition too quickly from forage to high grain diets, such as those employed in the feedlot industry. *S. bovis* rapidly metabolises the simple sugars of grain to produce lactate and microbial biomass, causing the "slime" characteristic of frothy bloat. *S. bovis* and SBSEC strains are also opportunistic pathogens impacting on the health of ruminant livestock, causing mastitis in dairy cattle and laminitis in cattle and horses. There is also evidence that they may be zoonotic pathogens having been associated with infective endocarditis and colorectal cancer in humans.

The DAF Rumen Ecology Unit (REU) has amassed a collection of SBSEC strains, primarily *S. bovis* isolates from sheep, cattle and goats. Thirteen *S. bovis* isolates were sent as part of the DAF contribution towards the USDOE Joint Genome Institute Community Sequencing Project, the Hungate 1000, an international collaborative venture to sequence the genomes of 1000 rumen bacterial isolates administrated by AgResearch Ltd, New Zealand. Over AU$80 000 worth of sequencing data has been obtained through this collaboration, however, there was no provision within the Hungate1000 for further analysis and the *S. bovis* sequences currently represent an untapped genetic resource.

This project aimed to analyse the genome sequence data obtained for the DAF *S. bovis* isolates contributed to the Hungate 1000 project, alone and in combination with other related gut-derived SBEC *Streptococcus* strains. This enabled the core genetic capabilities (pangenome) of these SBEC strains to be ascertained. For example, the project sought to describe the wide array of enzymes which *S. bovis* utilizes to quickly breakdown the carbohydrate component of plant material, potentially enabling the proliferation of these organisms over other rumen bacteria. It also aimed to identify novel anti-microbial enzymes these bacteria use to inhibit the growth of other gut-associated bacteria. Sequence analysis and development of the pangenome employed modern bioinformatics programs and a high performance computing system, increasing the skills and capability within DAF to 'ensure we can access the skills and capability to support current and future organisational functions'.

Project Objectives

The project utilised genome sequences obtained through an international collaborative effort (Hungate 1000) to which DAF contributed rumen bacterial strains of Queensland origin. Through comparative analysis of the sequence data and development of a pangenome, the project sought to identify:

- novel enzymes and pathways which enable *S. bovis* to rapidly break down high grain diets
- novel antibacterial compounds (bacteriocins) which *S. bovis* may excrete in order to inhibit other rumen microbes and facilitate the over-proliferation of this organism
- new methods to inhibit and control *S. bovis* populations, for example, utilising the bacteriophage (phage) encoded lytic genes.

The project also sought to build scientific capacity within the REU by developing and establishing techniques for handling and analysing large microbial genome sequence datasets. This included the installation of specialist bioinformatics programs within the DAF allocation of the DSITI High

Performance Computing facility, located at the EcoSciences Precinct, Dutton Park, to increase the capacity and broaden the potential application of this computing resource.

Methodology

Genome sequences

Genome sequences for a total of 42 *Streptococcus* strains representing primarily animal-derived isolates of *S. bovis*, *S. equinus*, *S. gallolyticus*, *S. caballi* and *S. henryi* were downloaded from the National Center for Biotechnology Information (NCBI) reference sequence database [(1) http://www.ncbi.nlm.nih.gov/refseq/) and the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) portal (http://genome.jgi.doe.gov/ (2)] which houses the results of the Hungate 1000 project (http://genome.jgi.doe.gov/TheHunmicrobiome/TheHunmicrobiome.info.html). All sequence data was saved to the DAF allocation within the DSITI high performance computer (HPC).

Bioinformatics software

To facilitate and enable genome sequence analysis the Prokka software version 1.11 (3) (http://www.vicbioinformatics.com/software.prokka.shtml) was installed on the DSITI HPC. This software incorporates several additional bioinformatics programs including BioPerl, BLAST+, HMMER, Aragon, Prodigal, tbl2asn, GNU Parallel, Infernal, Barrnap and MINCED. All additional programs can also be used outside the framework of the Prokka program for further sequence data analysis. The BLAST+ software (4) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was also updated to version 2.2.31 on the DSITI HPC and the Geneious software package update R9 (www.geneious.com) purchased and installed locally on a Mac desktop.

Genome annotation

Genes within the sequence data for the 42 genomes were identified and named (annotated). An initial genome sequence annotation was available for most sequences from the IMG genome portal, and further annotations were generated using Prokka (3).

Identification of Carbohydrate utilisation enzymes

Custom computer scripts were developed in-house to search the genome annotations for keywords, such as glycosyl hydrolase, enabling the identification of carbohydrate utilising enzymes. The HMMER program was also used to identify genes with homology to the carbohydrate utilising enzyme complexes included in the dbCAN HMMs version 5.0 (based on CAZyDB 15/07/2016 update (5) http://csbl.bmb.uga.edu/dbCAN/).

Identification of bacteriocin genes

Bacteriocin genes were detected within individual genomes using BAGEL3 software (6) and bacteriocin database (http://bagel.molgenrug.nl/). This software classifies bacteriocin gene clusters on the basis of whether the bacteriocins require post-translational modification for activity (modified peptides, class I) or can be unmodified or minimally modified (class II).The presence and sequence positioning of bacteriocin gene cassettes within genome data was verified using custom computer scripts developed in-house to search genome annotations for keywords, such as bacteriocin and lantibiotic.

Identification of prophage elements

Prophage-related regions were identified within individual genomes using the PHAST software (7) and web server (http://phast.wishartlab.com/) and custom computer scripts to search the genome annotations for keywords, such as phage and phage terminase. Prophage genomes were further annotated using a combination of open reading frame identification by Prokka and blastp of the NCBI virus reference sequence database (January 2016 update). Annotated genome maps were created and depicted using Geneious R9. Novel gene sequences for the phage lytic gene cassettes, including the holin and lysin genes, were extracted as fasta files for further analysis within the scope of the ASQ Innovation Project: *Expression Systems for Novel Enzymes* (led by Diane Ouwerkerk).

Analysis of proteins produced by *Streptococcus equinus*: Proteomic analysis

Tryptic digests of whole cell protein extracts were analysed using an Orbi-FT MS (1-3 ppm), LTQ MSMS and identified peptide fragments matched against complete genome amino acid sequences (ProteinPilot™ v5.0.1 http://sciex.com/products/software/proteinpilot-software).

Prophage characterisation: Particle purification and transmission electron microscopy

The formation of phage particles was determined for culture of *S. equinus* Sb05, 2B and Sb20 grown under anaerobic conditions in RF medium incubated at 39 °C for 18 h with cultures of 2B and Sb20 induced with ultraviolet light exposure. Viral particles were purified using previously described methods (8) and visualised by transmission electron microscopy (glutaraldehyde fixation and ammonium molybdate staining, using a Joel JEM-1400 Transmission Electron Microscope [TEM]).

Results and Discussion

Streptococcus genome sequences

A total of 42 genome sequences, representing *Streptococcus* strains isolated from the gut or faecal matter of a range of herbivores, including grazing and feedlot cattle and dairy cows, sheep, horses, camels, moose and goats, were included in the study (described in Appendix 1, Table A1.1). The majority of these genome sequences were generated by the Hungate 1000 project and all were made publically available in February 2016. Most genomes consisted of several long sequence fragments (contigs) providing near-complete sequence coverage of the entire genome. Thirteen *Streptococcus* isolates from the REU culture collection were included in the analysis.

Identification of carbohydrate utilisation enzymes

Glycoside hydrolase (GH) families were used as the basis for classifying the carbohydrate-breakdown enzymes present in each *Streptococcus* genome (Figure 1.). Eight GH families were present in all of the *Streptococcus* genomes, incorporating enzymes such as α-N-acetylgalactosaminidase (GH109), α-amylase, (GH13), β-glucosidase (GH1), peptidoglycan lyase (GH23), lysozyme (GH25 and GH73), invertase (GH32), amylomaltase or 4-α-glucanotransferase (GH77). Interestingly, genes for cellulose breakdown (for example endo-β-1,4-glucanase/cellulose GH5), an activity not usually attributed to gut-associated *Streptococcus* populations, were present in 38 genomes, indicating that the ability of the *Streptococcus* genera to break down plant structural carbohydrates may have been previously under-estimated.

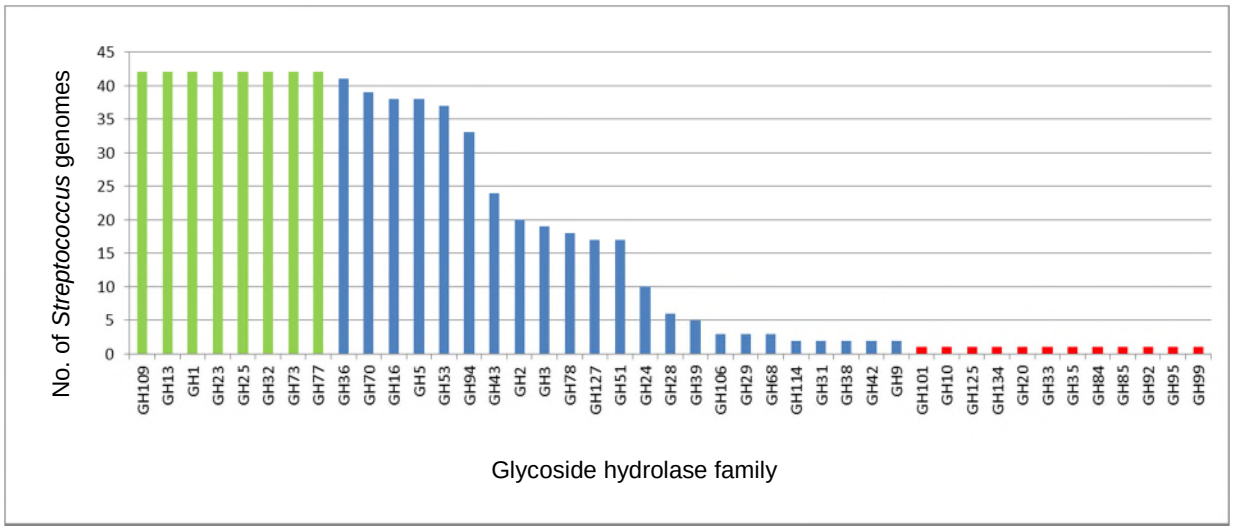

Figure 1: Abundance of glycoside hydrolase (GH) families identified in 42* Streptococcus *genomes. The GH families present in all* Streptococcus *strains are shaded in green and GH families only present in a single* Streptococcus *genome are shaded in red.

Identification of bacteriocin genes

Bacteriocin activity usually requires clusters of genes which contribute functional roles such as regulation, cleavage, modification, immunity and transport and secretion. Bacteriocin gene clusters were detected in all but four of the *Streptococcus* isolates (*S. bovis* JB1, *S. equinus* ATCC 9812, *S. equinus* Sb18 and *S. equinus* Sb17). A complete listing of the bacteriocin gene clusters identified and their classification into known bacteriocin classes, are presented in Appendix 1, Table A1.2.

Of the 33 identified gene clusters with genetic similarity to previously identified modified peptides belonging to bacteriocin class I, approximately half (54.55%) were further classified as lanthipeptide class II peptides (Figure 1). Lanthipeptide class I and sactipeptide classes also detected and glyocin (glycocins) and linear azol(in)e-containing peptides (LAPs) were infrequently observed. LAPs were only detected in strains of *S. gallolyticus,* and glyocin-related peptides only detected in two *S. equinus* strains originating from goats (strains Sb09 and Ye01).

Of the gene clusters identified as class I, Nukacin A and Nisin U-related peptides were the most abundant. These peptides are also classified as lantibiotics (lanthionine-containing antibiotics) which are small peptides (19-38 amino acids in length) that undergo enzyme-mediated post-translational modification, and can be distinguished by the presence of the rare thioether amino acids, lanthionin or β-methyllathionine (9). This type of bacteriocin interacts with the lipid II molecules of the bacterial cell membrane to form pores, resulting in ionic imbalance and disruption of normal cell wall formation (10, 11) and causing cell death.

The most abundant bacteriocin genes identified in 22 of the *Streptococcus* strains examined were class II bacteriocins, related to the bovicin 255 peptide (protein family (Pfam) ID 10439.4). These class II bacteriocins are small (<10 kDa), heat-stable, unmodified peptides, and have been previously shown to be secreted by rumen Streptococcal species in order to destroy closely-related competitor Gram-positive bacteria (12). This study however, revealed the presence of several class II bacteriocins not previously reported in gut-associated *Streptococcus* strains, and the presence of

large bacteriocins (<10 kDa, class III bacteriocins) related to Zoocin A and dysgalacticin which both target the cell wall in a lytic and non-lytic manner, respectively (13).

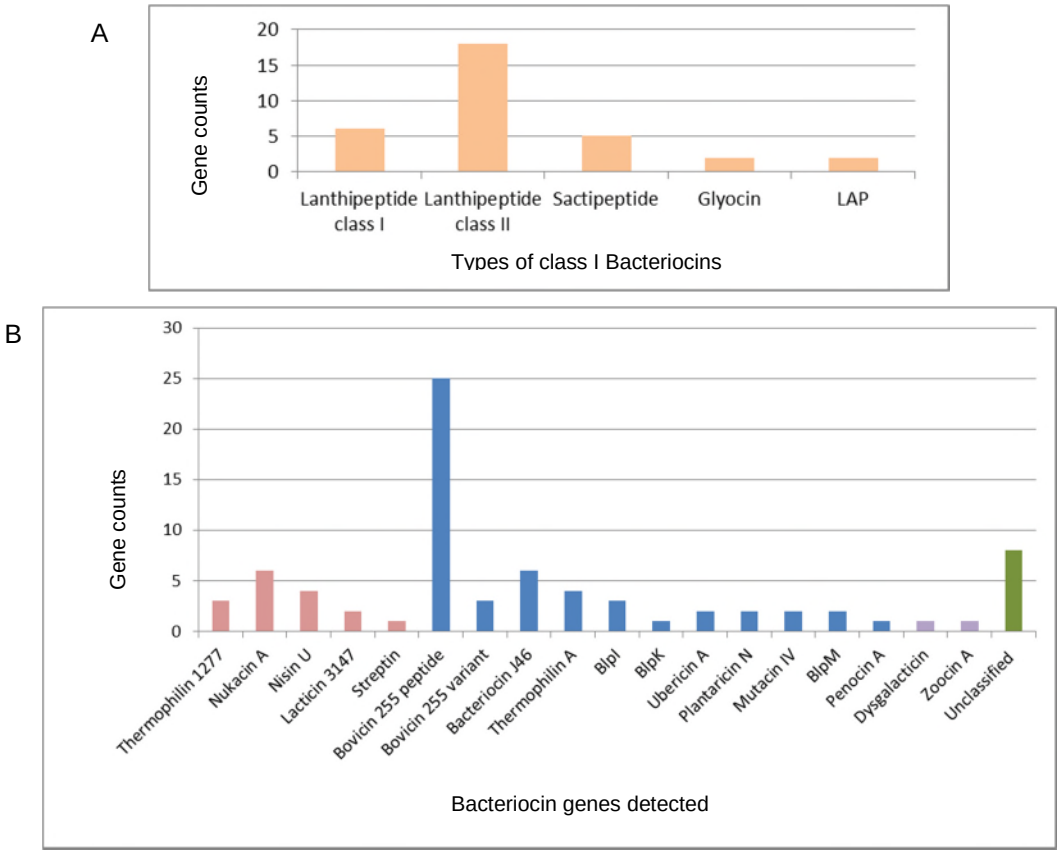

Figure 2: Classification of bacteriocin gene clusters and specific bacteriocin genes within 42* Streptococcus *genome sequences indicated by the gene or open reading frame (orf) counts. (A) Abundance of class I bacteriocin gene cluster subclasses; (B) Abundance of specific bacteriocin gene elements including the BAGEL3 groupings Bacteriocin I ; Bacteriocin II ; Bacteriocin III and unclassified bacteriocin-like elements .

Identification of prophage elements

Phage are bacterial viruses which can infect their host cells and integrate their DNA genome into that of the host, thus forming a stable, heritable association with the host (17). When a phage has successfully formed this genetic insertion, their genome is described as a prophage element. A total of 25 prophage-associated sequences were identified in 20 genome sequences, encoding a sufficient complement of phage genes to be designated as 'intact' prophages. When compared on a genetic basis, all of these novel prophage elements appeared to encode for dsDNA tailed phages of the Siphoviridae family. The prophage genomes were up to approximately 65 kb in size and were modular in nature, with phage genes sequentially organised to encode structural and non-structural proteins (Appendix 1, Figure A1.3). Alignment of prophage elements and comparison of selected phage genes (integrase, terminase large subunit and tail-host specificity genes) indicated the presence of several distinct sub-groups within the *Streptococcus* pangenome.

Phage genes involved in host cell lysis have previously been shown to have antimicrobial properties, representing a source of novel antimicrobials (14). The identity of prophage genes predicted to be responsible to causing host cell lysis (holin and lysin genes) were identified within the *Streptococcus* prophage sequences. Sequence data was made available to the ASQ Innovation Project: *Expression Systems for Novel Enzymes* (led by Diane Ouwerkerk) for testing of the expression system protocols, and determine whether these genes represented functional enzymes and could inhibit bacterial growth.

Three of the *Streptococcus* strains stored in the REU collection whose genomes contained intact prophage elements (*S. equinus* Sb05, Sb20 and *S. bovis* 2B) were cultured and two induced with UV light. Intact phage particles were purified and examined using electron microscopy with each of the three strains producing tailed phage particles (Figure 3). The expression of phage-encoded genes was also confirmed by proteomics analysis (example shown for *S. equinus* Sb05, Table 1), demonstrating that phage proteins derived from an integrated prophage element may be shed by cells within a growing *Streptococcus* culture, even in the absence of an inducing agent.

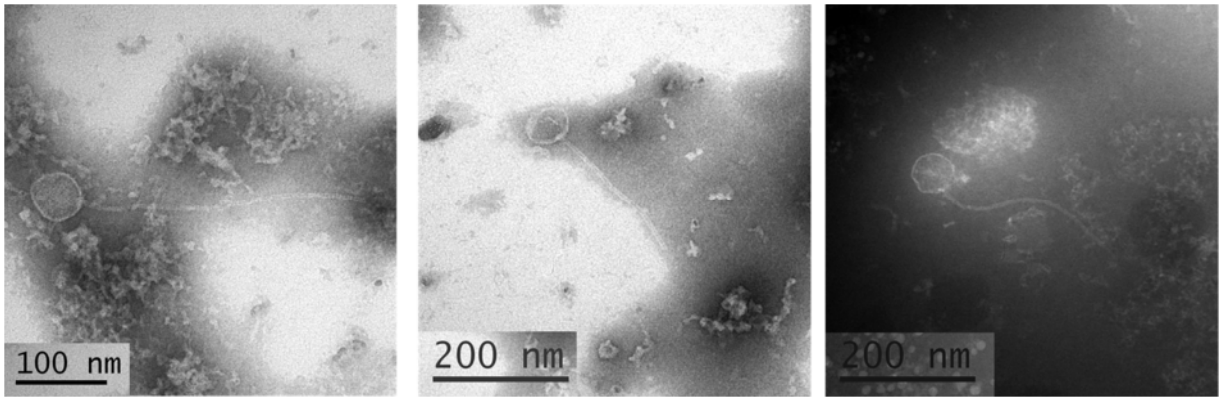

Image 1: TEM micrographs of phage particles of the Siphoviridae family morphotype produced cells containing integrated prophage elements (A) un-induced culture of* S. equinus *Sb05; (B) UV-induced culture of* S. equinus *Sb20; (C) UV-induced culture of* S. bovis *2B. Images obtained with the assistance of Dr Kathy Crew, Horticulture and Forestry Science, DAF.

Table 1: Prophage genes identified within the genome of* S. equinus *Sb05 also detected in culture supernatants following proteomic analysis (Orbi-FT MS (1-3 ppm), LTQ MSMS).

Gene name	Gene type (structural/non-structural protein)	Protein sequence coverage (%)
Scaffold protein	non-structural	90.40
Phage tail fibre protein	structural	70.40
Phage major head protein	structural	66.05
Sensor protein (YopX family)	non-structural	21.57
Phage tail protein	structural	21.12
N-acetylmuramoyl-L-alanine amidase (sle)	non-structural	18.68
Phage portal protein	non-structural	17.45
DNA adenine methyltransferase (YhdJ)	non-structural	5.94
Tail-host specificity protein	structural	3.85

Conclusions/Significance/Recommendations

Genome sequences from 42 *Streptococcus* strains classified within the BSEC and isolated from the gut or faecal matter of a range of herbivores, including grazing and feedlot cattle and dairy cows, sheep, horses, camels, moose and goats, were utilised in the study. The majority of these genome sequences, including those for the 13 *Streptococcus* strains contributed by the REU have not been previously published or made publically available and, therefore, represent a novel, comprehensive dataset.

This sequence dataset was used in the creation of a pangenome to characterise the core enzymes involved in the breakdown of plant carbohydrates by these organisms. Microbial carbohydrate breakdown usually involves modules of genes encoding enzymes which can attach to and digest plant material (15), and are classified on the basis of their function, for example, the glycoside hydrolase (GH) family includes genes which hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate compound. A core of eight GH families were present in all of the *Streptococcus* genomes examined, indicating that there is a common enzymatic approach used by gut-associated BSEC strains to the breakdown plant carbohydrates. The study however, revealed the presence of a diverse array of GH families with many strain-specific enzymes identified. Enzymes of GH families not usually attributed to gut-associated *Streptococcus* populations were also identified, providing new genetic insights into of the mechanisms employed by the *Streptococcus* genera to break down plant carbohydrates.

All but four of the *Streptococcus* genomes analysed possessed an array of bacteriocin-like genes, some of which were related to known lantibiotic peptides, such as Nukacin and Nisin. Bacteriocins such as these have been shown to exhibit high specific activity against Gram-positive bacteria, and have been previously investigated with a view to their application in food and medicine (16). There were also several bacteriocin-like elements identified which were unlike any of the bacteriocins currently present in bacteriocins databases. All of the bacteriocin-like elements detected during the course of this pangenome investigation represent an untapped resource of novel antimicrobials. Further in-vitro experiments will be required to validate their function and full spectrum of bacteriocin activity prior to use in ruminant production systems. The information generated in this project can therefore, form the basis of funding applications to further investigate the potential of these bacteriocins.

A total of 25 prophage elements were identified in the 42 *Streptococcus* genomes examined, with a sufficient complement of phage-like genes to be designated as ‘intact’. Further testing (TEM and proteomics) of the three host *Streptococcus* strains with designated intact prophage elements showed that the prophages were biologically functional and could produce tailed phage particles and prophage proteins. Testing also showed that these phage particles could be shed by actively growing *Streptococcus* cultures. These findings contribute to our understanding of the biological role of phages in the rumen, indicating the extent to which phages can integrate into the genomes of their gut-associated *Streptococcus* hosts. The study also verified that prophage elements identified using a sequence-based approach can produce viable, functional proteins.

There is an increasing interest in using naturally-occurring, host-specific phage-encoded enzymes as alternatives to conventional broad-spectrum antibiotics for the control of microbial pathogens (14, 18-20) and other microbial populations, such as rumen methanogens (21). To further develop this approach, sequences for the phage-encoded lysis proteins, including the holin and lysin genes of *S. equinus* strain Sb05 were supplied to a second ASQ Innovation project (Di Ouwerkerk) for gene

expression and testing. This work represents a proof-of-concept approach where genome sequence analysis can deliver gene sequences of interest, which can then be assessed for their viability and functional capacity. One distinct advantage of using phage-encoded genes for protein discovery is that the production of viable phage particles from intact prophage elements enables researchers to quickly assess whether phage-encoded gene sequences can produce functional proteins.

The technology developed during the course of this project and the sequence information and outcomes produced will be used to form the basis of further research proposals, targeting the use of novel proteins, including carbohydrate-degrading enzymes, and antimicrobials (bacteriocins and phage lytic enzymes). Funding will be sought from industry bodies and companies with an interest in developing feed enzymes and additives to increase the productivity and health of ration-fed livestock.

Key Messages

The Australian Feedlot industry production represents a significant ‘value adding and finishing’ mechanism to the Australian beef industry. The industry’s production is valued at ‘approximately $2.7 billion annually and is estimated to employ 2,000 people directly and 7,000 indirectly’ (Australian Lot Feeders’ Association 2011). Nationally there are 700 accredited feedlots, with the majority of these located in Queensland and New South Wales (https://futurebeef.com.au). As the drought in Northern Australia continues, increasing numbers of cattle are finished on high grain diets employed in the intensive feedlot industry. Prolonging the time taken to transition from forage-based to high grain diets is the main management practice employed to avoid the development of frothy bloat. Any measures which can reduce this dietary transition time and reduce the need for antibiotic interventions will have positive consequences for animal production efficiency, and directly address the DAF strategic plan priority to ‘improve sustainability of agriculture’, with the proposed project driving ‘productivity through research’ and ‘improve management practices to reduce the impacts of agriculture on the environment’.

This study used a genetic approach to characterise isolates of the *Streptococcus bovis*/*Streptococcus equinus* complex (SBSEC), a group of animal and human-derived commensal bacteria, found in the rumen and gastrointestinal tract. These bacteria are virulent, fast-growing organisms which over-proliferate in the rumen of cattle when they transition too quickly from forage to high grain diets, such as those employed in the feedlot industry. This study identified individual enzymes and profiled the combinations of enzymes used by these organisms to attach to and rapidly break down the simple sugars and more complex plant carbohydrates contained in high grain diets. Analysis of genome sequences also identified the compounds (bacteriocins) these organisms use to inhibit the growth of other bacteria. Enzymes capable of selectively targeting *Streptococcus* cell walls (phage-encoded holin and lysin genes) were also identified during the course of the investigation. Both the bacteriocin and phage-encoded lytic enzyme genes represent the basis for the development of new enzyme-based approaches of the control of rumen *Streptococcus* populations.

The development of the *Streptococcus* pangenome and subsequent dataset analysis ‘exploited new technologies’ by utilising modern computing and bioinformatics methodologies. This has also increased the skills and capability within DAF to process and securely archive large DNA and protein sequence datasets ‘ensuring we can access the skills and capability to support current and future organisational functions’.

Where to next

The lytic (holin and lysin) genes identified within an intact prophage element of *S. equinus* Sb05, will be produced and tested in a proof-of-concept experiment using the gene expression technology developed as part of Di Ouwerkerk's Agri-Science Queensland Innovation Opportunity Project *'New strategies for improving ruminant feed digestibility: developing expression systems for novel enzymes*'. Results arising from both of these projects will be further compiled into manuscripts for publication in peer-reviewed international scientific journals.

This project has made significant progress in establishing the methodology, technical capacity and background data required to enable the development of project proposals for larger, longer projects to formulate novel enzymes with enhanced feed breakdown capability and antimicrobials suitable for on-farm application.

Budget Summary

The budget allocation of $2500 was spent prior to the end of the 2015/2016 financial year on two major items:

- special purpose software, Geneious R9 upgrade
- annual fee for data storage and use of the DSITI High Performance Computing Facility.

Remaining funds were spent on laboratory consumables for the proteomics analysis.

Appendix 1 Additional Data

Table A1.1: Table of genome sequences utilised in the study with strains from the REU culture collection highlighted.

Organism full name	JGI genome portal ID or GenBank assembly accession No.	No. of Contigs	Source
Streptococcus henryi A-4	2593339212.fna	63	cow rumen
Streptococcus bovis AG46 (1)	2562617049.fna	1	sheep rumen
Streptococcus bovis AG46 (2)	2582580713.fna	1	sheep rumen
Streptococcus bovis C277	2623621021.fna	8	sheep rumen
Streptococcus bovis ES1	2623620585.fna	10	sheep rumen
Streptococcus bovis JB1	2593339156.fna	28	cow rumen
Streptococcus equinus pGA-7	2608642168.fna	6	cow rumen
Streptococcus equinus pR-5	2608642181.fna	11	cow rumen
Streptococcus bovis B315	2524614859.fna	9	cow rumen
Streptococcus equinus GA-1	2593339267.fna	11	cow rumen
Streptococcus equinus HC5	GCF_000731085.1_ASM73108v1_genomic.fna	8	cattle rumen
Streptococcus equinus ATCC 33317	GCF_000747195.1_ASM74719v1_genomic.fna	17	cow faeces
Streptococcus sp. NLAE-zl-C503	2654588205.fna	19	cow faeces
Streptococcus caballi DSM 19004	2515154033.fna	62	horse faeces
Streptococcus equinus ATCC 9812	GCF_000187265.1_ASM18726v1_genomic.fna	20	horse faeces
Streptococcus gallolyticus	GCF_000723985.1_SGAL_genomic.fna	260	calf faeces
Streptococcus henryi DSM 19005	2518645609.fna	37	horse caecum
Streptococcus gallolyticus LMG 15572	2606217751.fna	24	goat rumen
Streptococcus gallolyticus ATCC 700065	2599185152.fna	17	goat rumen
Streptococcus equinus MPR1	2651870309.fna	23	camel forestomach
Streptococcus equinus MPR2	2651870307.fna	22	camel forestomach
Streptococcus equinus MPR4	2654588204.fna	14	camel forestomach
Streptococcus sp. 45	2654588198.fna	14	camel forestomach
Streptococcus bovis SN033	2524614842.fna	13	deer faeces
Streptococcus gallolyticus VTM1R27	2595698218.fna	19	moose rumen
Streptococcus gallolyticus VTM1R29	2606217753.fna	25	moose rumen
Streptococcus gallolyticus VTM2R47	2606217755.fna	33	moose rumen

Organism full name	JGI genome portal ID or GenBank assembly accession No.	No. of Contigs	Source
Streptococcus gallolyticus VTM3R24	2608642166.fna	13	moose rumen
Streptococcus gallolyticus VTM3R42	2608642220.fna	14	moose rumen
Streptococcus bovis 2B	2561511223.fna	9	sheep rumen
Streptococcus equinus AR3	2654588192.fna	16	sheep rumen
Streptococcus equinus H24	2654588151.fna	11	sheep rumen
Streptococcus equinus Sb04	2651870306.fna	21	cattle rumen
Streptococcus equinus Sb05	2654588139.fna	6	cattle rumen
Streptococcus equinus Sb10	2654588134.fna	6	cattle rumen
Streptococcus equinus Sb13	2654588137.fna	14	cattle rumen
Streptococcus equinus Sb17	2654588136.fna	13	cattle rumen
Streptococcus equinus Sb18	2654588130.fna	27	cattle rumen
Streptococcus equinus Sb20	2654588135.fna	9	cattle rumen
Streptococcus equinus SI	2654588207.fna	11	sheep rumen
Streptococcus equinus Ye01	2654588209.fna	19	goat rumen
Streptococcus equinus Sb09	2654588197.fna	18	goat rumen

Table A1.2: Table of bacteriocin-like genes or Open Reading Frames (ORF) identified with the BAGEL3 software. Within each listed strain of* Streptococcus*, genome sequence regions were identified with homology to known bacteriocin-associated genes. Within these regions, specific open ORFs (genes) were classified according to their relatedness to known bacteriocin classes based on the degree of modification and size, and the relatedness to previously characterised bacteriocin determined. Where no specific, previously characterized bacteriocin-like element could be identified, the ORF is described as Unclassified. Strains from the REU culture collection are highlighted

Strain Name	Small ORF class	ORF ID	Bacteriocin I	Bacteriocin II	Bacteriocin III	Bacteriocin-associated region
S. bovis 2B		orf012		Bovicin 255 peptide		
		orf018		Bovicin 255 peptide		
S. henryi A4	Lanthipeptide class II	orf006	Thermophilin 1277			
	Lanthipeptide class II	small ORF 1	Thermophilin 1277			
S. bovis AG46		orf012		Bovicin 255 peptide		
		orf022		Bovicin 255 peptide		
S. equinus AR3	Lanthipeptide class II	orf002				
	Lanthipeptide class II	orf006	Nukacin A	bacteriocin J46		
	Sactipeptides	small ORF 11				
	Sactipeptides	orf009				
S. equinus ATCC 33317		small ORF 10				Unclassified
		small ORF 1				Unclassified
		orf008		Bovicin 255 peptide		
S. gallolyticus ATCC 700065	LAPs	small ORF 6				
		small ORF 10				Unclassified
		orf017		BlpI		
		orf019		Thermophilin A		
S. bovis B315	Lanthipeptide class I	orf005	Nisin U			
	Lanthipeptide class II	orf002				
	Lanthipeptide class II	orf006	Nukacin A	bacteriocin J46		
S. bovis C277	Lanthipeptide class II	orf002				
	Lanthipeptide class II	orf006	Nukacin A	bacteriocin J46		
		orf015		Bovicin 255 peptide		
Streptococcus. sp. NLAE zl C503		orf001				Unclassified
S. caballi DSM 19004		orf005		Thermophilin A		

Strain Name	Small ORF class	ORF ID	Bacteriocin I	Bacteriocin II	Bacteriocin III	Bacteriocin-associated region
		orf013		BlpK		
		orf014		Bovicin 255 variant		
S. henryi DSM 19005	Lanthipeptide class II	small ORF 4	Lacticin 3147 A1			
	Lanthipeptide class II	orf004	Lacticin 3147 A1			
		orf010		ubericin-A		
S. bovis ES1	Lanthipeptide class II	orf002				
	Lanthipeptide class II	orf006	Nukacin A	bacteriocin J46		
		orf015		Bovicin 255 peptide		
S. equinus GA1	Lanthipeptide class I	orf007				
		orf012		Bovicin 255 peptide		
		orf002				Unclassified
S. equinus H24		orf001				Unclassified
		orf008		Bovicin 255 peptide		
S. equinus strain HC5		orf008		Bovicin 255 variant		
		orf018		Bovicin 255 peptide		
	Lanthipeptide class I	small ORF 11	Streptin			
S. gallolyticus LMG 15572		orf010		Thermophilin A		
		orf011		BlpI		
		small ORF 4				Unclassified
	LAPs	small ORF 6				
S. gallolyticus		orf007		Bovicin 255 peptide		
		orf014		Thermophilin A		
		orf015		BlpI		
		orf005			Dysgalacticin	
		orf001				Unclassified
S. equinus MPR1		orf013		Bovicin 255 peptide		
S. equinus MPR2		orf013		Bovicin 255 peptide		
S. equinus MPR4		orf022		Bovicin 255 peptide		
S. equinus pGA7		orf008		Bovicin 255 peptide		
		orf014		Bovicin 255 variant		
	Lanthipeptide class II	orf002				
	Lanthipeptide class II	orf007	Nukacin A	bacteriocin J46		
S. equinus pR5	Lanthipeptide class I	orf006	Nisin U			

Strain Name	Small ORF class	ORF ID	Bacteriocin I	Bacteriocin II	Bacteriocin III	Bacteriocin-associated region
		orf006		Bovicin 255 peptide		
S. equinus Sb04		orf009		Bovicin 255 peptide		
S. equinus Sb05		orf008			Zoocin A	
S. equinus Sb09		orf001				Unclassified
	Glyocin	small ORF 6				
		orf017		Plantaricin N		
S. equinus Sb10		orf009		Bovicin 255 peptide		
		orf011		Mutacin IV		
		orf012		BlpM		
S. equinus Sb13		orf001				Unclassified
S. equinus Sb20		orf005		BlpM		
		orf006		Mutacin IV		
		orf008		Bovicin 255 peptide		
S. equinus SI		orf012		Bovicin 255 peptide		
	Sactipeptides	small ORF 11				
	Sactipeptides	orf010				
	Lanthipeptide class I	orf005	Nisin U			
		orf001				Unclassified
S. bovis SN033	Lanthipeptide class I	orf006	Nisin U			
	Lanthipeptide class II	orf002				
	Lanthipeptide class II	orf006	Nukacin A	bacteriocin J46		
Streptococcus. sp 45	Sactipeptides	orf003				
S. gallolyticus VTM1R29	Lanthipeptide class II	small ORF 1				
		orf011		Bovicin 255 peptide		
S. gallolyticus VTM2R47		orf008		Penocin A		
		orf010		ubericin-A		
	Lanthipeptide class II	small ORF 1				
S. gallolyticus VTM3R24		orf008		Bovicin 255 peptide		
S. gallolyticus VTM3R42		orf008		Bovicin 255 peptide		
S. equinus Ye01	Glyocin	small ORF 8				
		orf018		Plantaricin N		

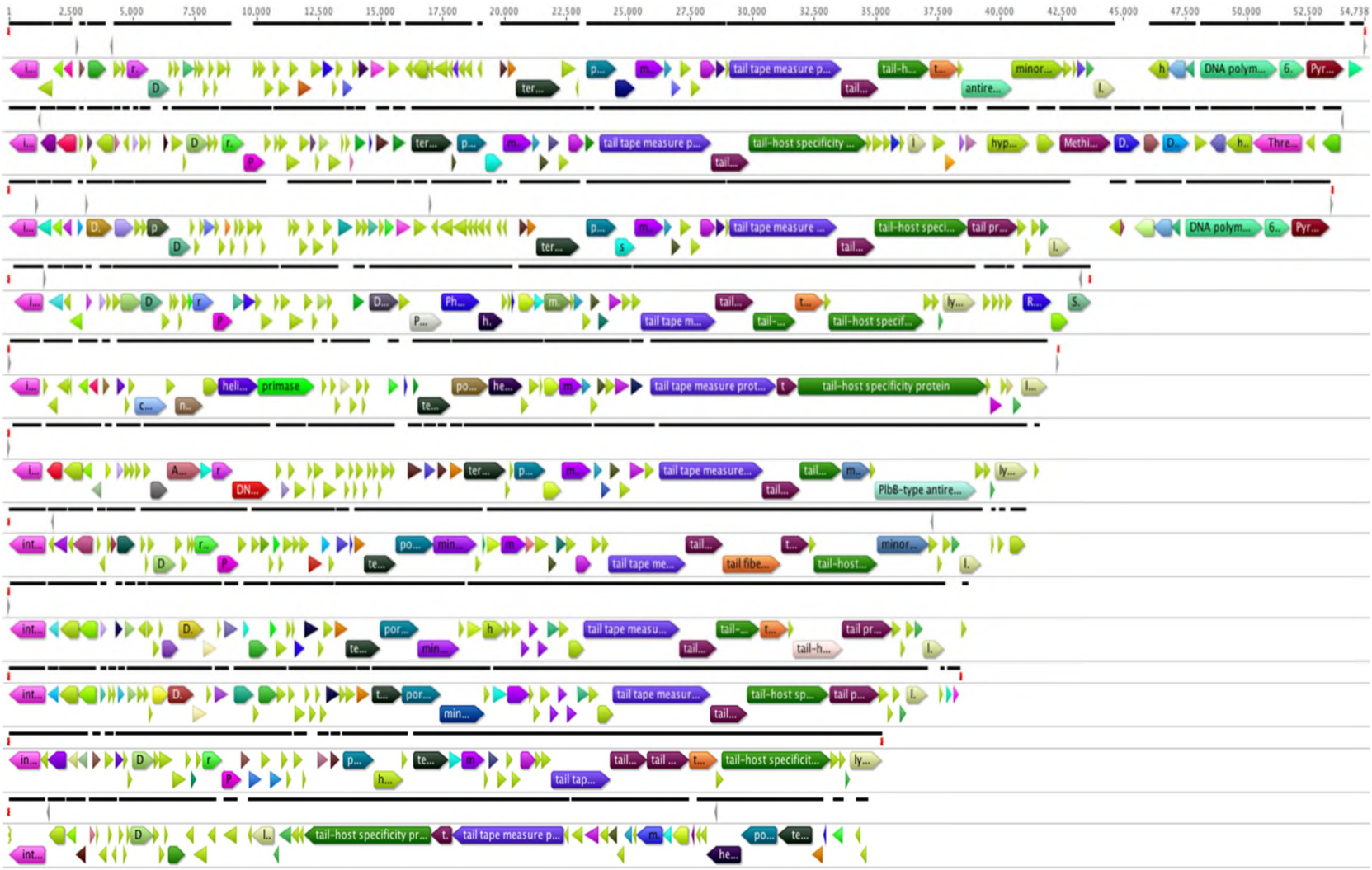
tail tape measure p...
tail-h...
minor...
DNA polym...
Pyr...
tail-host specificity ...
hyp...
Methi...
tail tape measure ...
tail-host speci...
tail pr...
tail tape m...
tail-host specif...
heli...
primase
tail tape measure prot...
tail-host specificity protein
tail tape measure...
PlbB-type antire...
tail fibe...
tail-host...
tail tape measu...
tail tape measur...
tail-host sp...
tail p...
tail-host specificit...
tail tap...
tail-host specificity pr...
tail tape measure p...

Figure A1.3: Gene maps of intact prophage elements identified within Streptococcus genomes with maps aligned with integrase genes positioned on the left of the figure showing modular, conserved gene arrangements.

References

1. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. 2014;42(Database issue):D553-9.
2. Markowitz V M, Chen I M, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42(Database issue):D560-7.
3. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
4. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):1-9.
5. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40(Web Server issue):W445-51.
6. van Heel A J, de Jong A, Montalbán-López M, Kok J, Kuipers OP. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic Acids Research. 2013;41(Web Server issue):W448-W53.
7. Zhou Y, Liang Y, Lynch K H, Dennis J J, Wishart D S. PHAST: A Fast Phage Search Tool. Nucleic acids research. 2011;39:W347-W52.
8. Klieve A V. 2.2. Bacteriophages. In: Makkar HPS, McSweeney CS, editors. Methods in gut microbial ecology for ruminants: International Atomic Energy Agency. Springer Academic Press; 2005. p. 39-46.
9. Cotter P D, Hill C, Ross R P. Bacteriocins: developing innate immunity for food. Nat Rev Micro. 2005;3(10):777-88.
10. Cotter P D, Ross R P, Hill C. Bacteriocins - a viable alternative to antibiotics? Nat Rev Micro. 2013;11(2):95-105.
11. Roy U, Islam M R, Nagao J, Iida H, Mahin A A, Li M, et al. Bactericidal activity of nukacin ISK-1: an alternative mode of action. Bioscience, biotechnology, and biochemistry. 2014;78(7):1270-3.
12. Whitford M F, McPherson M A, Forster R J, Teather R M. Identification of Bacteriocin-Like Inhibitors from Rumen Streptococcus spp. and Isolation and Characterization of Bovicin 255. Applied and Environmental Microbiology. 2001;67(2):569-74.
13. Alvarez-Sieiro P, Montalban-Lopez M, Mu D, Kuipers O P. Bacteriocins of lactic acid bacteria: extending the family. Appl Microbiol Biotechnol. 2016;100(7):2939-51.
14. Roach D R, Donovan D M. Antimicrobial bacteriophage-derived proteins and therapeutic applications. Bacteriophage. 2015;5(3):e1062590.
15. Fontes C M, Gilbert H J. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem. 2010;79:655-81.
16. Field D, Cotter P D, Ross R P, Hill C. Bioengineering of the model lantibiotic nisin. Bioengineered. 2015;6(4):187-92.
17. Salmond G P C, Fineran P C. A century of the phage: past, present and future. Nat Rev Micro. 2015;13(12):777-86.
18. Callaway T R, Anderson R C, Edrington T S, Genovese K J, Harvey R B, Poole T L, et al. Novel methods for pathogen control in livestock pre-harvest: an update. Sofos J, editor 2013. 275-304 p.
19. Shen Y, Mitchell M S, Donovan D M, Nelson D C. Phage-based enzybiotics. In: Hyman P, Abedon S T, editors. Bacteriophages in health and disease2012. p. 217-39.
20. Young R. Phage lysis: Three steps, three choices, one outcome. Journal of Microbiology. 2014;52(3):243-58.
21. Gilbert R A, Ouwerkerk D, Klieve A V. Phage therapy in livestock methane amelioration. In: Malik P K, Bhatta R, Takahashi J, Kohn R A, Prasad C S, editors. Livestock production and climate change. Wallingford, Oxfordshire, UK: CABI; 2015. p. 318-35.