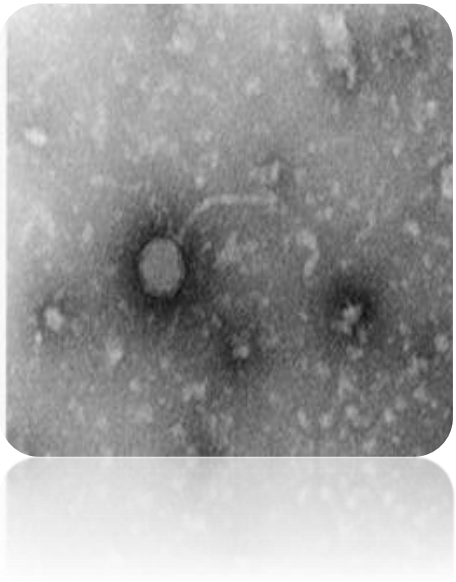
Genomic characterization and evolutionary relationships among Bacteriophages in the dairy industry and applications to detect phage Contamination



MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

By

Prianka Rajan

Supervised by

Dr. Heather L. Hendrickson

Abstract

The prevalence of bacteriophages and their pernicious effects on the Lactococcus starter culture in the dairy industry has been an ongoing problem for several decades.

The main purpose of this investigation was to understand the relationship, evolutionary history and the sources of the phages that have been isolated from the different fermentative units of Fonterra. We report the genomic comparison results of 15 phages in this study that were isolated on the host bacteria, *Lactococcus lactis cremoris*. These phages can be grouped in two clusters namely P335 and 936, commonly encountered tailed bacteriophages in the dairy industries. The majority of the phages belong to the P335 species with just one phage clustering with the 936 species. Although phages of the P335 group display a high level of synteny with one another, we report nine different types of P335 phages in this study. A prophage integrated in the host strain has been identified. The prophage and the phages show homology to the temperate P335 phage, R1T isolated in Netherlands in 1996. The genetic makeup of these phages is suggestive of their source and evolution from other prophages in strains that may have been co-cultured with the strain that was used in this study. Furthermore, we identified that horizontal gene transfer events and homologous recombination have played a role in the evolution of phages in our study.

Phage annotation was carried out for representatives in both the clusters and forty eight to fifty seven ORFs have been identified in these phages. Our analyses indicate that majority of the genes are conserved across these phages. For further detection of phages, this project also suggests rapid tools like PCR that can be used to better understand the phage species and the type of phage infecting the starter culture. With the availability of whole genome sequences, we are hoping that the genome analysis will enlighten our knowledge on the current distribution of phages and their relationship with one another in the dairy industry.

Abbreviations

DNA	Deoxyribonucleic acid
LAB	Lactic acid bacteria
RNA	Ribonucleic acid
PCR	Polymerase Chain Reaction
RPM	Revolutions per minute
bp	Base pair
CRISPR	Clustered regularly interspaced palindromic regions
TMP	Tape measure protein
MCP	Major capsid protein
MTP	Major tail protein
HGT	Horizontal gene transfer
TR	Transcriptional regulator

Acknowledgement

First and foremost I would like to thank my supervisor Dr. Heather Hendrickson for her guidance and support during my research and study at Massey University. Special thanks to her for giving me this opportunity to work on this prestigious project from Fonterra.

A sincere gratitude goes to INMS and the Head of the Institute Dianne Brunton for funding my study.

I would like to express my sincere gratitude to Dr. Evelyn Sattlegger and Dr. Weihong Ji for their immense support and guidance and for having helped me through this journey.

I would especially like to thank my Co supervisor Dr. Olin Silander for his constructive criticism and for having helped with the bioinformatics part of the research.

My sincere gratitude to Massey Genome Services especially Patrick Biggs and his team for sequencing al the phage DNA in this project and for their continuous support post sequencing.

I'm also grateful to my colleagues Jessica Fitch for isolating phage DNA and Courtney Davies, Joanna Wojitus and other lab members for their support and encouragement throughout my study.

I'm also grateful to my husband for reading my drafts multiple times without any complaints and my parents for their wise counsel throughout this journey.

Last but not the least our sincere thanks to Fonterra for giving us an opportunity to work on this project and funding a part of it.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Several concepts of the evolution of bacteriophages (commonly referred to as phages), in the dairy industry have been accumulating over the past two decades owing to their economic impact. The main purpose of this investigation was to understand the sources, evolutionary history and the relationship of a set of phages that have been isolated from different fermentative units of Fonterra on a genomic scale.

1.1 Background to the study

1.1.1 Dairy industry

New Zealand's climate, rainfall, and good pastures are ideal for cattle thus making the dairy industry the backbone of the country's economy**.** The dairy industry in New Zealand today is at its pinnacle, exporting around 95% of its milk to about 150 countries worldwide consequently becoming the biggest export earner of New Zealand (New Zealand Trade and Enterprise, Export market, n.d). This contributes to about a quarter of the total economic value of the fermented products produced in the world (Aginvest, New Zealand dairy industry, Milk products and markets section, 2013).

The dairy industry Co-operative Fonterra, traces back to 1871 when the first cheese company was opened in the Otago peninsula (Fonterra in New Zealand, Our Heritage section, 2014). Fonterra as of today produces 22 billion litres of milk annually and is a major dairy exporter contributing to around 3% of New Zealand's GDP (Fonterra, Global Dairy Trade section, 2016). With 33 manufacturing sites in New Zealand alone and

exporting about 95% of the locally produced milk, Fonterra is a leading exporter of milk and milk products in the world (Fonterra, Our Markets section, 2016).

The fermentation units of the dairy industries use huge volumes of lactic acid bacteria as starter culture for the production of dairy products like cheese, yogurt etc. The commonly used starter culture for producing dairy products is Lactococcus bacteria with other forms such as streptococcus and lactobacillus that are used occasionally (Marco et al., 2012). A continuous frozen culture collection of these bacterial strains have been maintained at the Fonterra's fermentation unit in Palmerston North, NZ. In order to maintain the growing economy of the country through dairy farming, millions of dollars are being invested by the dairy industry in conducting research aiming at improving the efficiency and quality of dairy products (Fonterra, New Zealand dairy industry section, 2014).

The major problem faced by the dairy industry today is contamination of the dairy products by bacteriophages that infect bacteria. A significant increase in the number of thermo-resistant phages in the fermentation units of the dairy industries has been observed in the past few decades and some of these phages are found to be resistant to pasteurization even at 90°C (Atamer et al., 2009; Marco et al., 2012). The phages can infect the host starter cultures that disrupt the fermentation process thus bringing huge losses to the industry.

1.1.2 Starter culture

The process of fermentation depends on the bacteria or the starter culture and the lactic acid produced by them in order to yield good quality cheese and other dairy products. Lactic acid bacteria (LAB) that are used as starter cultures are the primordial components

of any dairy fermentation industry (New Zealand institute of Chemistry, 2008., Manufacture and use of cheese products). Lactococcus bacteria are gram positive mesophiles that are being employed in the industry as starter culture for decades owing to their resistance to bacteriophages (Figure1). *Lactococcus lactis* subsp. *cremoris* are preferred as starter cultures for fermentation due to a superior product flavour compared to *Lactococcus lactis* subspecies *lactis* (Sandine, 1988).

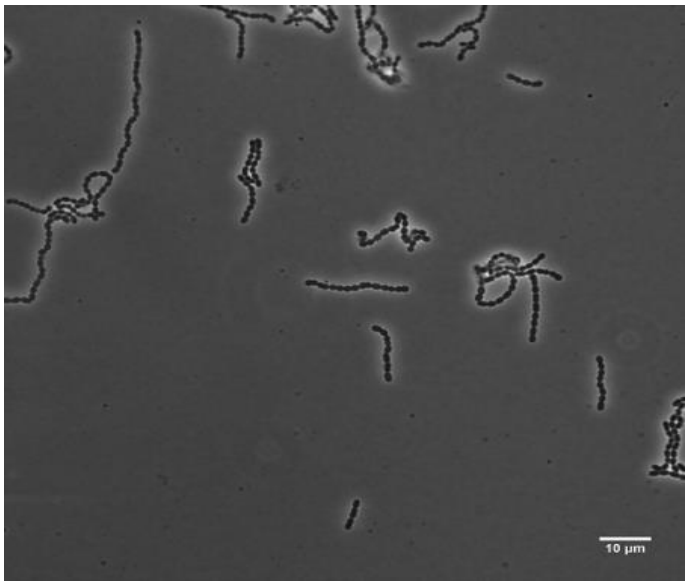


Figure 1 Image of Lactococcus lactis subspecies cremoris strain 2356 bacteria used in this study

1.1.3 Fermentation process in the dairy industry and whey disposal

The process of fermentation for manufacturing cheese starts with pasteurization of milk to kill the undesired microorganisms, followed by the addition of lactic acid bacteria for the production of lactic acid. This leads to coagulation of the protein casein in milk. Excess whey at this stage is expelled to obtain concentrated cheese.

Whey is a common by-product of fermentation. Whey produced during cheese manufacturing, are low molecular weight soluble components. Whey not only lessens the

oxygen content in water bodies but also acts as a major reservoir of phages. Hence extensive care is taken during its disposal (New Zealand institute of chemistry, n.d, Manufacture of ethanol from whey). Approximately 3 to 13 litres of whey is expelled for every kilogram of cheese produced (Atamer et al., 2013). In total, 4×10^9 litres of whey is produced annually in New Zealand. The whey produced post cheese manufacturing is finally treated to remove the phages. Though thermal treatment of whey to eliminate the phages has been common, membrane separation method has been shown to be very effective and is employed in several dairy industries. This method allows fat and casein particles to pass through the membrane of about 1.4µm while filtering microorganisms like phages (Atamer et al., 2013).

In spite of a number of strategies adopted in the dairy industries to improve the starter culture, eliminating infection causing bacteriophages that affect the starter culture has always been a challenge.

1.1.4 Strain Improvement of Starter Culture

Several strategies to improve the starter culture on a genetic level are being undertaken by the dairy industries. Among these, classical strain refinement techniques with respect to improving the texture of the products and resistance against phages are commonly adopted by the dairy industries. Precisely, some of the strain improvement techniques would include directed or adaptive evolution, dominant selection and random mutagenesis. In the directed or adaptive evolution technique, the strain used in the fermentation process is made to adapt to certain growth conditions and then this strain with the desired qualities is enriched for further use (Barrick et al., 2013). All three methods have their short comings the prominent one being the accumulation of undesired mutations (Derkx et al., 2014).

In spite of improving starter cultures, the dairy industries have been combating the problem of bacteriophage infection for several years and rely on control practices like improved sanitation, adequate ventilation, strain rotation etc. (Marco et al., 2012).

1.1.5 Bacteriophage source and infection

The ubiquitous nature of phages in the environment makes it difficult to discern their source. However, in the dairy industry, the most common source appears to be the raw milk (Madera et al., 2004). In Madera's study it was reported that 10% of the milk samples collected from different dairy units of Spain carried infectious lactococcal phages. Another common source of phage entry is through whey, a by-product of cheese manufacture. In a recent discovery, researchers were quite intrigued by the presence of genetic material from C2 and 936 groups of phages on various surfaces of the fermentation units including the floors and the pipelines (Verreault et al., 2011). On the contrary, the P335 groups of phages have been found to originate from the starter strains. This group in particular is also known to produce progeny phages in strains other than their host strains that contribute to host infection (Chopin et al., 2001).

1.1.5.1 Phage Life Cycle

The phages upon infecting the host (Lactic acid bacteria), can either takeover bacterial machinery completely or decide to co-exist with the host. The situation where the phage genome replicates and forms a number of copies, finally leading to their liberation by the destruction of the host cell is termed "lysis" (Delbruck, 1940). In contrast, lysogeny takes place if the bacteriophage integrates with the host genome i.e. the Lactic acid bacteria (Figure 2). Lysogen refers to the host/bacterium that contains phage DNA integrated with it.

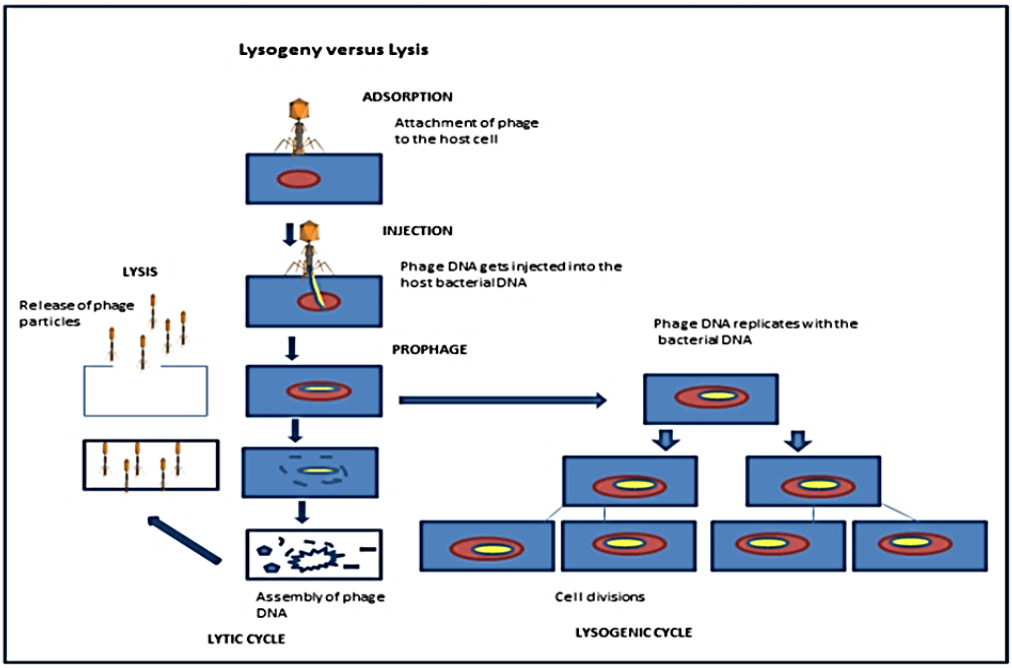


Figure 2 Diagram depicting Lysis versus Lysogeny. In the lytic cycle, the phage multiplies and lyses the host cell thereby releasing themselves. In the lysogenic cycle, phage DNA is incorporated into the host genome, and the host genome containing the pro(phage) divides subsequently (Source: Todar's online textbook of Bacteriology by Kenneth Todar., 2008).

Presence of antibiotics in raw milk, exposure to UV radiation or even a mild fluctuation in temperature have also shown to induce phages that have been an integral part of the starter strain thus contributing to lysis of the starter strain (Feirtag & McKay, 1987).

Some of the obvious symptoms of phage attack as observed during the early steps of fermentation are, low lactic acid content, a high pH and high residual lactose concentration (Marco et al., 2012). Apart from contributing to poor quality of the fermented product, phage infection can also lead to completely lost batches. In the case of fermented milk products, the termination of acid production causes a lack of coagulation, contributing to a grainy texture with excessive whey off (Marco et al., 2012).

The process of infection commences with the adsorption of the phage with the host cell. Several studies in the past have been conducted that look into the specificity of

adsorption (Lawrence et al., 1976) and phage resistance. In a research study that addressed the resistance exhibited by the lacotococcal strains to the bacteriophages, 33 out of 100 strains were found to be resistant to 34 phages (Madera et al., 2003).

1.1.6 Bacterial Resistance to phages

Several resistance strategies adopted by the bacteria to combat phage infection have been broadly studied in the past. Some of the commonly studied phage infection mechanisms include restriction/modification system, abortive infection mechanism, and CRISPR/Cas mechanism (Figure 3).

In restriction modification system, the host bacterium recognizes the phage genome and cuts them with their restriction endonucleases (Figure 3A). Its own DNA is protected by the addition of methyl groups via methyl transferases that are not recognized by the restriction endonucleases (Schouler et al., 1998).

Abortive infection immunity (Abi) is an innate immunity of bacterial strains which is triggered post phage infection (Figure 3B). This leads to the death of the infected bacterial cell itself thus stopping the spread of phage infection (Forde & Fitzgerald, 1999). More than 20 Abi systems have been reported to date (Dy et al., 2014).

CRISPR (Clustered regularly interspaced palindromic regions)/Cas system consists of the components crisper and cas (CRISPR associated proteins). The CRISPR locus consists of short DNA fragments of the phage DNA that had invaded the bacteria previously. During a phage infection, these short DNA fragments are transcribed into RNA. These RNA along with cas proteins degrade the invading foreign DNA thus providing an acquired immunity to the host (Figure 3C) (Szczepankowska, 2012).

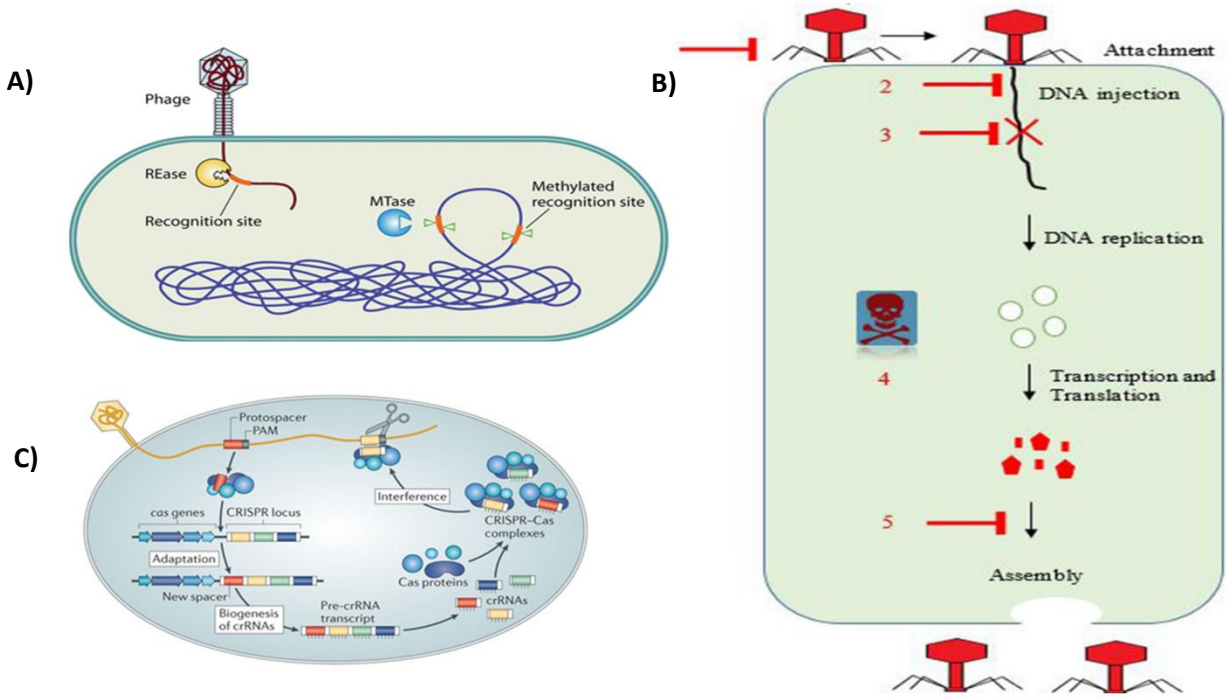


Figure 3 Illustration of Bacterial defense strategies. A) Restriction modification system "Diverse functions of restriction-modification systems in addition to cellular defense "(Source: Vasu et al., 2013). B) "Bacterial vs. Baceteriophages: Parallel Evolution of Immune Arsenals". This is a form of innate immunity that defends the host in several stages. For example, the phage is blocked from entering the host or its DNA injection is aborted (Source: Shabbir et al., 2016. C) CRISPR-cas System which is a form of acquired immunity in the host bacteria that recognize and invade the incoming phage DNA with the help of small RNAs (Source: Samson et al., 2013)

A feature that is shared among all the bacterial defense systems including those mentioned above is their genetic variability that occurs as a result of the "co evolutionary arms race" with the phages (Hoskisson and Smith, 2007). Another feature observed among the prokaryotes is their ability to undergo horizontal gene transfer which allows procurement and dispersal of novel phage defense systems (Stern & Sorek, 2011).

Phage infection may sometimes prove to be beneficial to the host strain and increase the host fitness but this comes with a cost for carrying the extra genetic material and also poses the risk of autoimmunity (Stern & Sorek, 2011). One of the effective and cost free defense systems is when prophages mediate defense against attacking phages. This type of

immunity against phages is commonly known as homoimmunity (Andam, Williams, & Gogarten, 2010). Similar to the receptors in lambda phages, lactic acid bacteriophages also contain the receptors "cro" and "cl" where the expression of the former makes them lyse the cell while the latter switches the module to lysogeny (Brussow, 2001). Some lysogenic phages express c1 repressor that stops the propagation of infection by other similar phages thereby offering immunity to the host. It has been proposed that many strains of *Lactococcus lactis* chromosomes carry prophages indicating widespread lysogeny (Davidson et al., 1990). According to a research which involved the induction of 30 probiotic strains of lactobacilli, about 25 of them contained prophages (Mercanti et al., 2011). While analysing the phage resistance mechanisms in Lactobacillus *delbrueckii,* harbouring prophages it was found that the lysogens express a complete phage resistance against several *L.delbrueckii* phages (Riipinen et al., 2007).

Another form of resistance that is established by some prophages is super infection exclusion mechanism whereby the prophage present in the host cell prevents infection by other phages by blocking phage DNA injection. This mechanism unlike the repression system is not just specific to homoimmune phages i.e phages of the same type (Donnelly et al., 1993).

In the following section I will discuss the classification of the commonly encountered lactococcal phages in the dairy industries.

1.1.7 Taxonomical classification of lactococcal phages

The structural component of the phages, their morphology, and their physiology play an important role in the classification of bacteriophages and in the construction of their

taxonomy. All the tailed phages fall under the order *Caudovirales*, which in turn is divided into three families namely, *Myoviridae, podoviridae* and *siphoviridae* (Figure 4).

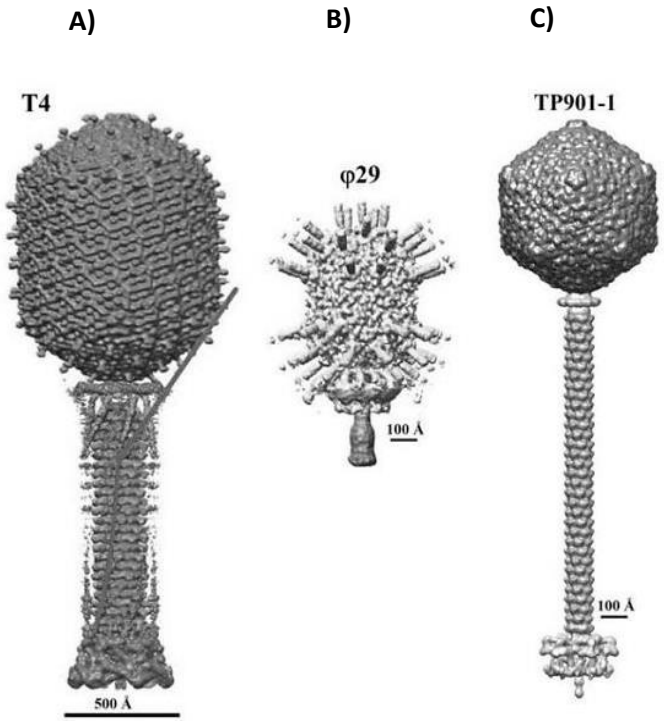


Figure 4 Structure of bacteriophages A) Myoviridae depicted by T4 phage showing long contractile tail. B) Podoviridae depicted by φ29 showing short non contractile tail. C) Siphoviridae depicted by TP901-1 showing long non contractile tail (Source: Fokine and Rossmann, 2014).

The first taxonomical classification that identified four groups of lactococcal phages based on serological and DNA hybridization studies was conducted in 1984 (Jarvis et al., 1984). Further investigation has expanded the classification into 10 groups with the majority falling in one of the three groups namely, P335, 936 or C2 (Figure 5) (Madera et al., 2004). The majority of the temperate phages i.e. phages that exist as lysogens in the bacterial strains belong to P335 group of tailed phages.

Most of the lactococcal phages are characterized by isometric head, long non-contractile tail and belong to the family siphoviridae (Figure 5). Inspection of the structure and genome organization of these phages has revealed that only five percent of the

lactococcal phages belong to the c2 group (Spinelli et al., 2014). Lactococcal phages namely P369, KSY1 and 1138 that have been identified as belonging to the family of podoviridae differ in their tail lengths compared to other virulent phages (Deveau et al., 2006 and Spinelli et al., 2014). Phages that fall under the 936 group are the most prevalent lactococcal phages as fifty five of the eighty four sequenced phages 0000belong to this category (Murphy et al., 2016). These are known to cause most of the outbreaks in the dairy industries and have been isolated from different geographical origins. The present classification system is being inspected constantly as new lytic phages have been emerging in the dairy industries (Chopin et al., 2001). For example, suggestions have been made in the past to include phages 1483 and T187 in the P335 group of phages (Jarvis, 1995; Labrie, 2002). It has been established that phages under the P335 group constitute the most diverse group of phages in the dairy industries (Deveau et al., 2006). Phages infecting the *L. lactis* strains used in dairy starters and non-dairy starters that have been sequenced to date throughout the world are listed in Table 1.

Thus it is understood that significant variations in their structure like the head and tail morphology have so far helped to classify the phages into various groups.

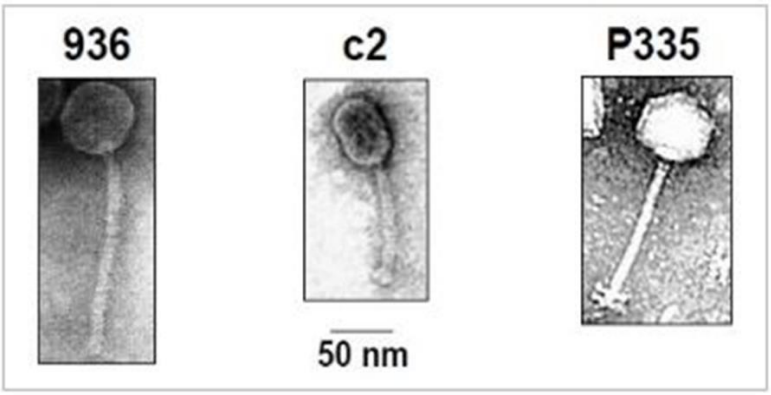


Figure 5 Types of commonly seen Lactococcal phages in the dairy industry. From left to right: Phage 936 of the 936 phage species, Phage C2 of the C2 phage species and Phage P335 of the P335 phage species (Source: Oliveira et al., 2017)

Table 1 List of phages infecting *Lactococcus lactis* strain in different geographical places that have been sequenced to date. Country and the year of isolation of the phages are also listed.

Phage	Type/Group	Host	Country of Isolation	Year
P335	P335	Dairy starter *L. lactis* subsp. *lactis* biovar diacetylactis F7/2	Germany	1979
1358	1358	*Lactococcus lactis* strain SMQ388	New Zealand	1981
936	936	Lactococcus *lactis* strain 158	New Zealand	1984
LC3	P335	*L.lactis*	Norway	1991
Ul36	P335	Dairy starter *L. lactis* subsp. *cremoris* SMQ-86	Canada	1992
Tuc2009	P335	Dairy starter *L. lactis* subsp. *cremoris* UC509	Ireland	1994
R1T	P335	Dairy starter *L. lactis cremoris* R1	Netherlands	1996
biL 310	P335	*L. lactis* IL1403 (Human vagina)	Netherlands	2001
biL 311	P335	*L. lactis* subsp. *lactis* IL1403	Netherlands	2001
biL 286	P335	*L. lactis* subsp. *lactis* IL1403	Netherlands	2001
Fd 13	*936*	*L. lactis cremoris* FD13	Denmark	2004
Phi4	*936*	*Dairy starter (Factory isolate)*	Ireland	2009
Jm2	*936*	*L.lactis cremoris*	Ireland	2010
BM 13	P335	Dairy starter *L. lactis* (cheese)	Canada	2011
63301	P335	*L.lactis*	Ireland	2016
C41431	P335	*L.lactis*	Ireland	2016
28201	P335	Dairy starter *L. lactis*	Ireland	2016
38502	P335	Dairy starter *L. lactis*	Ireland	2016
49801	P335	*L.lactis* DS68498	Germany	2016
50101	P335	Dairy starter *L. lactis*	Ireland	2016
50901	P335	Dairy starter *L. lactis*	Ireland	2016
86501	P335	Dairy starter *L. lactis*	Ireland	2016
53802	P335	*L.lactis* DS56538	Canada	2016
62503	P335	*L.lactis* DS63625	Australia	2016
98201	P335	*L.lactis* SMQ-86	Ireland	2016
98204	P335	*L.lactis* DS 64982	UK	2016
98103	P335	*L.lactis* DS64981	Japan	2016
P475	936	*L.lactis cremoris* 455	Europe	unknown

1.1.8 Structure of Lactococcal phages

Structural analyses of phages in the recent years have allowed a detailed inspection of phages and their interaction with the host. Lactococcal phages consist of a head and a long non contractile tail and have been broadly studied in the past (Spinelli et al., 2014). The head or the capsid encloses the double stranded DNA and is a solid structure that is strong enough to withstand the pressure exerted by the DNA (Bebeacua et al., 2013). Most phages isolated from the dairy industries have small isometric or prolate heads (Chandry et al., 1997). The majority of the lytic phages that belong to the 936 group are characterized by small isometric heads (Jarvis et al., 1991; Parreira 1996). The c2 group of phages have a characteristic

elongated prolate head that distinguishes them from the other two groups. Among the three groups, only phages in the c2 group have a protein receptor (phage infection protein) that is used during the host infection (Spinelli et al., 2014). Phages of the siphoviridae family have long non- contractile tail that is involved in the adsorption of the phage to the bacterial cell. The tape measure protein (TMP) controls the tail length in the tailed bacteriophages and also promotes the transfer of DNA into the host cytoplasm during infection (Mahony et al., 2016). Mutations in the TMP have shown to play a significant role in the size of the tail (Pedersen et al., 2000). The connector connects the capsid to the tail at the vertex (Spinelli et al., 2014). Electron microscopy of phage Tuc2009 has revealed a collar-like structure in the region where the tail connects with the head (Mc Grath et al., 2006). Lactococcal phages are characterized by the presence of basal plate at the end of their tail. Receptor binding proteins in the basal plates have been indicated to play a major role in host adsorption (Spinelli et al., 2014). Their primary role is the recognition of specific receptors at the cell surface for adsorption (Bebeacua et al., 2013). After analysing several receptor binding proteins of 936 and P335 group of phages, it has been proposed that the architectures of phage adsorption sites are in accordance with the evolution of the host receptors (Bebeacua et al., 2013).

In the next section I will describe what is known about dairy phages in New Zealand

1.1.9 Previous Study of Dairy phages in New Zealand

Bacteriophages of lactic acid bacteria have been the subject of immense research in New Zealand for the past several decades. The first phage specific to the lactic acid bacterium in New Zealand was isolated in 1935 (Whitehead and cox, 1935). Research in this field has been progressing ever since. Virulent phages are the major cause of fermentation failures in the dairy industry. A virulent phage Phi 949 was isolated from

cheese whey in the New Zealand Dairy Institute in 1975 and it was found to infect many lactococcal strains used in the dairy industry (Samson & Moineau, 2010). A previous study conducted by Rakonjac et al., (2005) using a new enrichment strategy on phages isolated from a New Zealand dairy plant have determined host ranges in few lactococcal species like phages c2 and biL67 successfully. Both these phages isolated from the dairy industry of New Zealand are prolate-head phages that show an overall identity of 80% with each other. Based on the results obtained from the DNA homology and morphological analysis, a classification scheme was developed for lactococcal phages that categorized twenty five lactococcal bacteriophages into four species (Jarvis, 1984).

It has been observed, that among the lactococcal phage groups, P335 is the most commonly encountered phage in dairy fermentation units (Mahony et al., 2012). Concurrent with this finding, research on KW2 strain of *Lactococcus lactis* subsp. *cremoris* obtained from fermented corn was found to contain an integrated P335 prophage. In their study which involved the comparison of 26 different P335 phages showed that the morphogenesis region comprising the head, tail and the phage structural components was highly conserved among these phages. Also, it was determined that the integration of the prophage is site specific and occurs in seven different regions of the chromosome and these phages are an integral part of the *Lactococcus lactis* chromsome (Kelly et al., 2013).

Until 1995 it was assumed that industrial starter cultures that carried temperate phages had no homology to the lytic phages. However in that year an industrial lytic phage showed genetic relatedness to temperate phages isolated from a cheese factory in New Zealand (Jarvis 1995). This in turn triggered more genome wide analysis of such phages dominating the New Zealand dairy units. At Fonterra, a repository of the bacteriophages that have infected their starter cultures have been maintained for several decades and we have

taken the initiative along with Fonterra to understand them better on a genetic level (Figure 6).

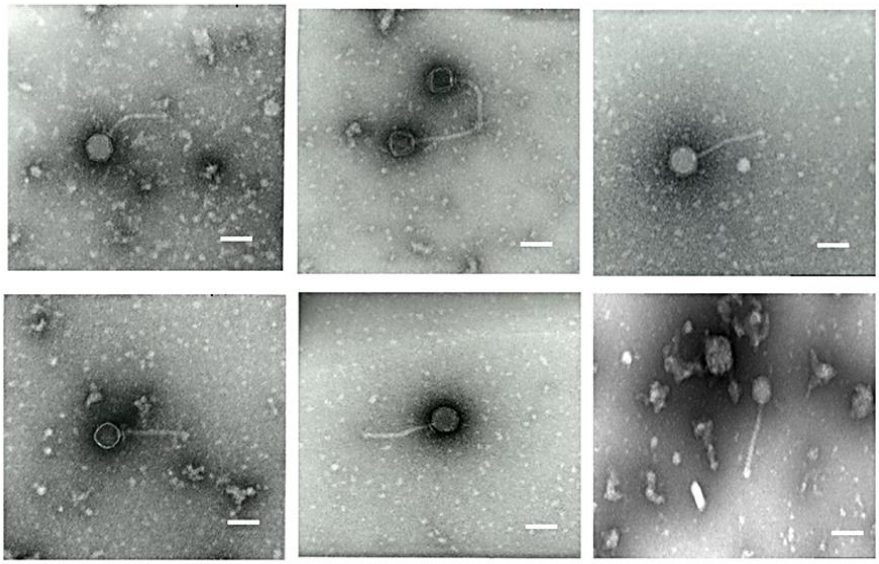


Figure 6 Electron microscopic images of phages A) Phage 5910 B) Phage 5915 C) Phage 5953 D) Phage 5914 E) Phage 5812 F) Phage 6090. (Images supplied by Fonterra. Scale bar =50 nm)

1.2 Bacteriophage genomics and its significance

A wide range of genetic research has been conducted on bacteriophages especially dairy phages because of their potential use in the dairy industries.

A recent study that involved the analysis of whole genome of 627 mycobacteriophages categorized them into 28 clusters and it was noticed many groups of genes were shared among the phages that were in different clusters (Pope et al., 2015)

Research on the lactococcal genomes has revealed very useful information that has been productive to the dairy industry. According to a 2017 report, complete genome sequences of 200 phages infecting the LAB are available (Mahony et al., 2017). A

disparity in the size of the genomes has been seen among the different groups of phages (C2, P335 and 936). Phage 949 (936 group) isolated from cheese whey in New Zealand has genome size of 11, 4768 bp (Samson & Moineau, 2010) while c2 phage genome is 22, 163 bp long (Lubbers et al., 1995). Genome sequencing data has also indicated that 936 group phages have evolved from a common ancestor and that they have the ability to multiply fast in their host strains (Samson & Moineau, 2010).

Phage genomic evolution has been largely influenced by widespread horizontal gene transfer that includes both homologous and nonhomologous DNA sequences (Rokyta et al., 2006).

1.4 Horizontal gene transfer and Homologous recombination

Horizontal gene transfer (HGT) in contrast to vertical gene transfer (parent-offspring) is the transfer of novel DNA sequences in prokaryotes and is fairly common in eukaryotes as well. HGT in prokaryotes has been found to occur via transformation, conjugation and transduction. Transformation is a form of HGT when there is uptake of genetic material from the environment. Conjugation occurs when there is transfer of genetic material between two prokaryotic cells that are in direct contact. Transduction is a form of HGT where phages carry genetic material from one bacterium to another. HGT has been considered a fast mode of evolution in prokaryotes and most of these genes can be lost as easily as they are gained (Brussow et al., 2004).

HGT has been reported in lactococcal phages as well. For example it was noted that a distinct GC content in the phage BK5-T was due to the insertion of DNA sequences from a different origin. Despite being a lactococcal phage, its closest relative was not another lactococcal phage but a *S.thermophilus* phage sfi21 (Desiere etal., 1999). Desiere et al in

2001 also confirmed the insertion of a new tail fiber gene and endonucleases in the phage biL170. The genes acquired through HGT can subsequently undergo recombination with the host chromosome and genetically be passed on or may be lost during selection (Andam et al., 2010).

Recombination among phages has been shown to occur via four major routes namely homologous recombination, site specific recombination, microhomologous recombination and non-homologous or illegitimate recombination (Casjens et al., 1992). DNA exchange that occurs as a result of homology or similarity between two genomes is termed homologous recombination. The modular theory of the evolution of phages states that each phage genome is comprised of various functional units like the capsid genes, tape measure genes and tail genes apart from the genes involved in lysogeny and replication (Botstein, 1989). These units are usually conserved across the phage genomes and are capable of undergoing recombination with phages that share homology.

Studies in the past have shown that the genetic diversity observed among many P335 group members is due to the recombination between phages or with the prophage in the host cell (Bouchard and Moineau, 2000; Deveau et al., 2006). This study describes the mode of genetic exchange between a P335 lytic phage (ul36) and its host chromosome that contains a prophage that shares homology with ul36. Homologous recombination between the phage ul36 and other P335 members have also been explained in that study (Figure 7)

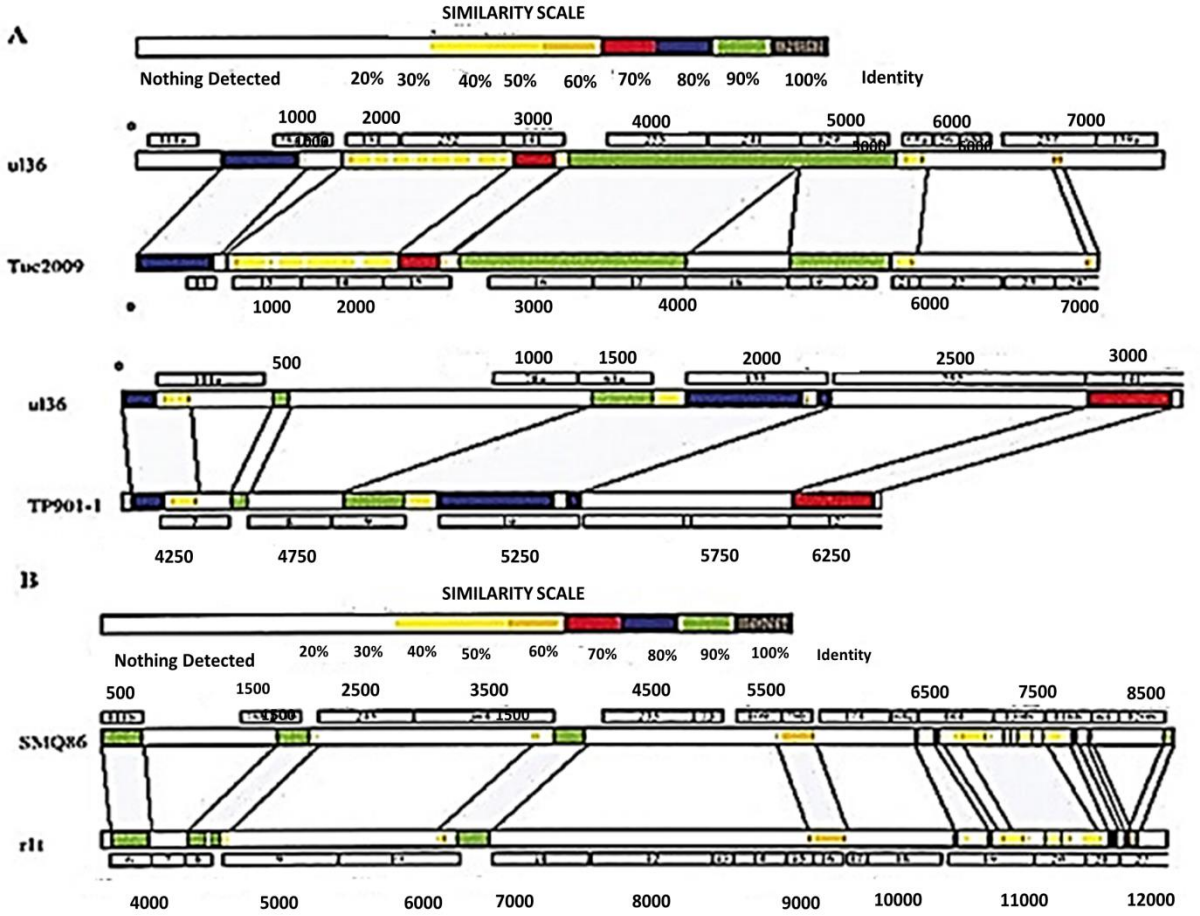


Figure 7 Homologous recombination in P335 phages. The illustration depicts the pairwise sequence alignments between the different members of the P335 phage species. Genes or the Open reading frames are indicated in the coloured boxes whose length is proportional to the length of the ORF (Source: Bouchard and Moineau, 2000)

1.5 Source of (Pro)phages in Dairy Industry

Milk and whey have shown to be a potential source of virulent phages yet most of the temperate phages (phages capable of becoming lysogenic) belong to the P335 group (Marco et al., 2012). About 10 to 20% of the all bacterial DNA is made up of prophages (Casjens et al., 2003). Whole genome sequences of many LAB strains that include both lactococci and lactobacilli are found to contain prophages (Sechaud et al., 1988). A research that studied the association of phages with the host has demonstrated that autolysis of starter strains occurs as a result of prophages integrated in them (O'Sullivan et al., 2000).

1.6 Aims and Objectives of the Study

Aim: To analyze the phages isolated from the different fermentation units of a dairy industry on a genomic scale in order to understand their relationships, evolutionary history and source which influence the quality of dairy products.

Objective 1 Isolate and purify samples obtained from the fermentation units of Fonterra in order to analyze them individually.

Positive samples from plaque tests will isolated to a single plaque in order to purify it to a single bacteriophage. To accomplish this, we will follow three rounds of purification for each phage before amplifying them.

Objective 2: Achieve higher concentration of the purified phages both qualitatively and quantitatively and extract DNA from them for further bioinformatics analysis.

Following the isolation, phages will be amplified to achieve high titer or high concentration. In order to amplify the phages, the phage lysate will be precipitated overnight at 4˚C and centrifuged the following day at about 10000 RPM for atleast an hour. DNA will be extracted from these phages using promega wizard DNA extraction kit. Extracted DNA will be sent to Massey Genome Services for sequencing. Following the arrival of the DNA sequences my aims will be to analyse the sequences on a genomic scale.

Objective 3: Address similarities and differences between the phage genomes that have been sequenced in this study from various fermentation units of Fonterra.

This will be achieved by various bioinformatic tools like MAUVE (Darling et al., 2004) and Gepard (Krumsiek and Rattei, 2007). Also, phamerator database will be used to identify the nucleotide and proteins that are shared among these phages.

Objective 4: Understand the evolutionary relationships between the phages using a whole genome approach.

In order to accomplish this, we will analyze the evolutionary relationships between all the phages under study using multiple sequence alignment software such as Clustal Omega which does multiple alignments of the sequences (Katoh and Standley, 2013). This will give us an idea about the group or the cluster to which each of the phages belong.

Objective 5: Identify the distribution of genes in these genomes focusing mainly on their location and function.

In order to achieve this, representatives of the phage genomes will be annotated using DNA master. DNA master predicts the genes present in the genome and identifies their functions. This is established by doing a Blast analysis that matches the genes under study to other genes in the NCBI database. The gene call that is made by DNAmaster is further confirmed by another gene predicting tool called genemark (Besemer and Borodovsky, 2005) that verifies the coding potential of the gene.

Aim: Use the genomic information of the sequenced phages to establish methods for rapid identification of phages attacking the dairy industry

Objective 6: Design and suggest specific PCR primers for regions that are conserved in each cluster to identify the type of phage. This would help us to rapidly distinguish between different cluster or groups of phages. Also obtain patterns of restriction digests or fragment sizes that would help us to discriminate between the two groups of phages that infect the dairy strains.

Chapter 2

Materials and Methods

2.1 Phage Isolation and amplification

2.1.1 Bacterial strain and media

The strain *Lactococcus lactis* subsp.*cremoris* (2356) was supplied by Fonterra as a host for the phages to be studied. This strain that came from the factory was a cheese isolate that was used as starter cultures in 1996. The 2356 strain was retired in 2015. The lactococcal strain used in my study was cultured in M17 media, 2% lactose and 10 mM calcium chloride incubated overnight at 30ºC between 18 to 24 hours.

2.1.2 Culture preparation

To prepare a culture of *Lactococcus lactis* subsp. *cremoris*, 40 µl of the freezer stock was added to 2 ml of M17 media followed by the addition of 2% lactose solution. This was incubated for 24 hours at 30 °C. On the following day, 200 µl of this culture was added to a vial containing 10 ml of M17, 2% lactose and 10 mM calcium chloride. This was incubated again at 30 °C for 24 hours. This culture was used in further experiments to test phage infection.

2.1.3 Plaque assay

Bacterial infection was examined by conducting spot tests for each phage at different dilutions starting from 10^{-1} to $10^{-10.}$ Briefly, 10 µl of each phage dilution was

spotted on the corresponding position of the spot plate that contained the soft agar (2.5 ml of M17 agar, 2 ml of M17, 2% lactose and 10mM calcium chloride) and the culture. The commonly used quantitative method to determine the bacteriophage infection is analysing the plaques. A clear circular zone resulting from the lysis of the bacteria is known as a plaque. Clear zones of plaques that indicated lysis were observed in all 24 phages.

2.1.4 Phage isolation

Positive samples from plaque tests were isolated to a single plaque in order to purify it to a single bacteriophage. To accomplish this, 200µl of culture was combined with 15µl of a frozen sample that gave a positive spot test. This was made to sit for 15 minutes to allow phage to infect. To this, 4.5mls of soft M17 agar was added, and poured on a M17 plate. Once dry, this was inverted and incubated overnight at 30°c. The next day, plaques of all morphologies from the plate were isolated by picking the plaque and submerging in a micro-centrifuge tube with 100µl of phage buffer. The sample has now undergone 1 round of isolation. This procedure was carried out 3 times in order to purify it to a single phage

2.1.5 Phage amplification

Once the phages were isolated, they were amplified to get a high titer or high concentration. In this method, 500 µl of the culture is infected with 10 µl of the phage and mixed with 4.5 ml of soft agar. After incubating the plates at 30ºC overnight a plaque count was conducted the following day. In order to estimate the plaque forming units (PFU) which reflects the titer or count of the phage, the number of plaques was first divided by the sample plated and then was multiplied by reciprocal of the dilution used. A

large number of overlapping plaques resulted in a webbing pattern, ideal for DNA extraction.

In some phages, a low titer was observed. In order to increase the titer, a higher volume of the phage sample was added which subsequently resulted in more plaques. Ideally a titer value of about 10^9 was required for a good yield of DNA. In some instances, the concentration was further increased by using a combination of the PEG-NaCl (3.3M) solution.

Once a sufficiently high titer was achieved, ten plate amplification of each phage was performed. Each plate was flooded with 10ml of phage buffer and kept in room temperature for up to four hours. The lysate (phage buffer containing the phages) was filter sterilized to remove bacterial debris and stored in 30 ml vial.

2.2 DNA extraction

Phage DNA was isolated using Promega Wizard DNA clean up kit following manufacturer's instructions. Prior to the extraction, phage lysate was treated with DNase and RNase to get rid of the bacterial DNA and RNA respectively and incubated at 37ºC for 30 minutes. The lysate was finally treated with PEG (polyethylene glycol) in the ratio 2:1 to concentrate the phage particles and refrigerated overnight. Centrifugation was performed the following day at 18000rpm for 2 hours. The extracted DNA was confirmed by running a 1% agarose gel electrophoresis at 140 V. DNA concentrations was further verified by means of spectrophotometric analysis.

2.3 Restriction enzyme analysis

The isolated genomic DNA of the phages was tested for their restriction patterns with enzymes, EcoRI, HindIII and ClaI. It was identified that the enzyme, ClaI contained many restriction fragments common to the P335 group analysed in our study. This enzyme also distinguished between different groups or clusters of phages clearly.

2.4 DNA sequencing

DNA obtained from the phages was sequenced using Illumina MiSeq at Massey genome service (NZGL). Illumina MiSeq can produce 2 x 300 paired-end reads in a single run. The *de novo* assembler Velvet was used to assemble the genomes from the dataset (Zerbino and Birney, 2008). The goal was to obtain a contiguous sequence of DNA (contig) for each phage obtained by assembling overlapping DNA fragments or reads. Also, as some of the genomes generated big size assemblies, contaminating sequences (bacterial/ host) within these genomes were suspected. Hence in further analysis, the tool Kraken was chosen to extract the reads corresponding to the phages from the combined dataset (Davis et al., 2013). This tool requires a database to search the extracted reads against other genomes. The minikraken database of bacterial and viral reference genomes was used for this purpose. This finally led to the construction of a single contig for each of the 15 phage genomes and a prophage genome.

2.5 Reference alignments and software

Genome sequences for phages R1T (accession number: U38906.1), 63301 (accession number: KX456211.1), and Phi4 (accession number: KP793101.1) were obtained from NCBI and were used as reference genomes. A comparative genome analysis of the phages under study was performed using NCBI Blast (https://blast.ncbi.nlm.nih.gov/) (Altschul et al., 1990). Clustal Omega (MAFFT V.7) was used to do a multiple sequence alignment (Katoh and Standley, 2013).

2.6 Tree output software

PHYML version 3.0 was used to deduce maximum likelihood phylogeny (Guindon et al., 2010). Bootstrap support for the tree was inferred from 100 bootstrap replicates. Neighbor joining tree was obtained from Clustal Omega multiple sequence alignment program (Saitou et al., 1987).

2.7 Annotation of Phage genomes

Annotation of the predicted ORFs in the phage genomes were performed using the tool DNAmaster (http://cobamide2.bio.pitt.edu). The annotated genomes were compiled in GenBank format for phamerator database construction (Cresawn et al., 2011). Complete genome sequences of the annotated phages were deposited in GenBank (Appendix).

2.8 Phamerator database construction

All the fifteen phages and the prophage were submitted to the phamerator database that was built for these lactococcal phages. In order to construct a customized phamerator

database of phage genomes, the genome sequences are first processed in DNAmaster that consists of all the phage genomes with the same start sites. Phamerator conducts a pairwise comparison of the amino acid sequences using Blast P (Altschul et al., 1997) and Clustal W (Larkin et al., 2007). The proteins are further categorized into "phamilies" or "phams" of related sequences. Once the sequence files are fed into the database, genome maps of phages that contain nucleotide and amino acid sequence relationships, as well as genes that are conserved across the genomes are generated.

Chapter 3

Results

3.1 Host strain infection by phages across different fermentation units

Phages were collected from fermentation units in both North and South island of New Zealand and one fermentation unit in, Australia (Site S6) (Figure 8). It was confirmed that all 24 phages supplied by Fonterra's fermentation units lyse the strain 2356 that has been used as a starter culture for many years.

Following the isolation and purification of the 24 phages by our research technician Jessica Fitch, I amplified the phages by conducting 10 plate amplification of each phage in order to obtain a high titer. Clear zones of plaques that indicated lysis were observed in all the 24 phages. A webbing like pattern indicated large number of overlapping plaques which was ideal for DNA extraction. An illustration of a high titer plate is represented below (Figure 9)

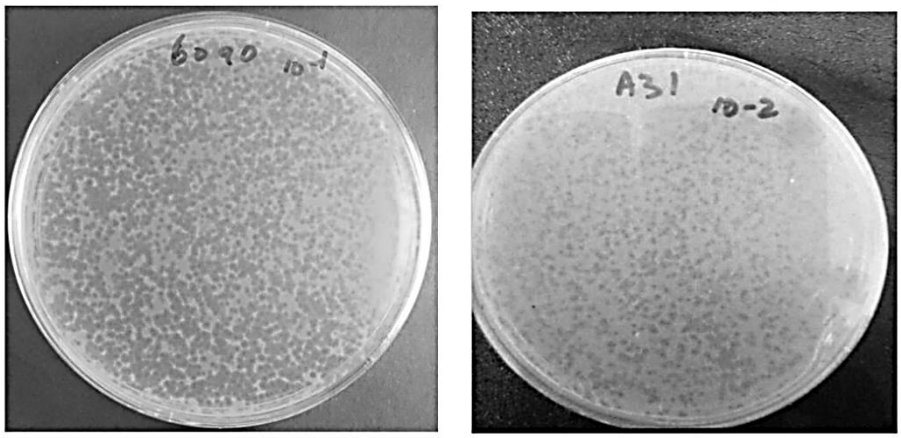


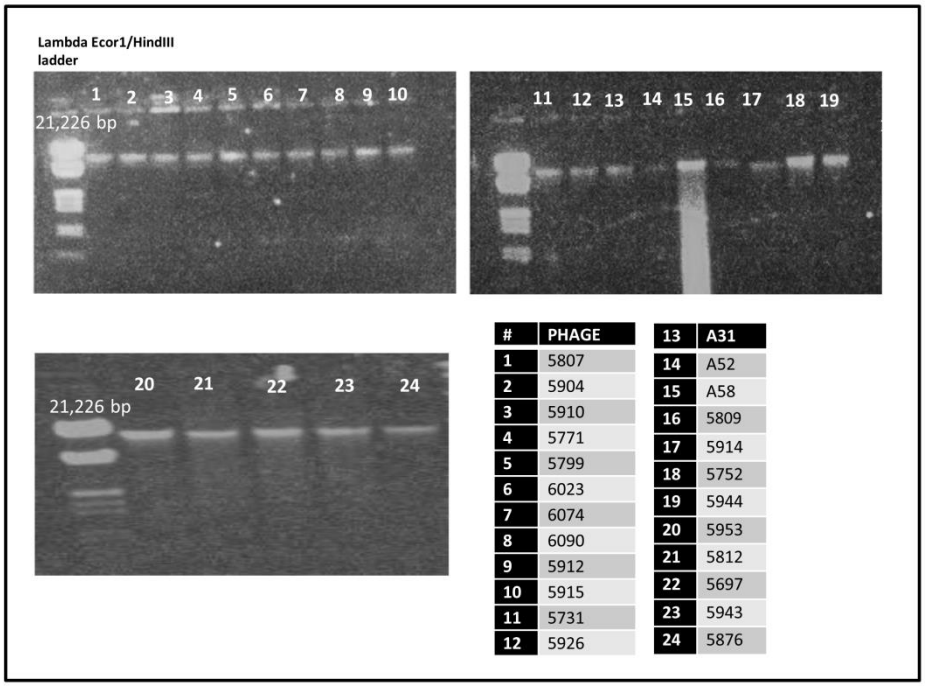
Figure 8 Plates (phages 6090 and A31) showing the formation of plaques or clear zones that appear when a phage infects the bacteria.

The phage titer concentration required for successful amplification followed by DNA extraction is listed in Table 2.

Table 2 Location of fermentation plants of Fonterra indicating each phage's source and the titer values that were obtained from the lysate. (Estimate of DNA concentration obtained using a spectrophotometer for all the phages are also included). Sites of fermentation plants are shown as (N1-N4, S1-S6).

Phage #	Location/Site	Titer	DNA conc(ng/µl)
6074	S3	1.1×10^{10}	270
5697	S2	1.05×10^{9}	180
A52	S6	2.10×10^{9}	86
5807	S3	2.17×10^{9}	152
5904	S5	1.85×10^{9}	92
5915	S1	1.77×10^{9}	158
A31	S6	2.11×10^{9}	73
5912	S3	1.20×10^{9}	386
5771	S2	1.90×10^{9}	201
5799	N3	3.10×10^{9}	189
5943	S5	5.0×10^{8}	99
6023	S3	3.12×10^{9}	108
A58	S6	2.73×10^{9}	52
5944	S3	1.53×10^{9}	69
5914	N1	1.37×10^{9}	58
5910	S4	2.55×10^{9}	98
6090	N3	2.31×10^{9}	55
5752	N4	1.97×10^{9}	53
5809	S2	2.11×10^{9}	67
5731	S4	1.91×10^{9}	53
5876	N2	2.87×10^{9}	128
5953	S4	3.21×10^{9}	25
5926	S2	1.72×10^{9}	55
5812	N3	2.97×10^{9}	28

The size and the concentration of the phage DNA was further analysed using agarose gels (Figure 10). Lambda EcoRI HindIII ladder was used as a marker to confirm the size of the DNA.



#	PHAGE
1	5807
2	5904
3	5910
4	5771
5	5799
6	6023
7	6074
8	6090
9	5912
10	5915
11	5731
12	5926

13	A31
14	A52
15	A58
16	5809
17	5914
18	5752
19	5944
20	5953
21	5812
22	5697
23	5943
24	5876

Figure 9 Visualization of all the 24 phage genomic DNA on 1% agarose gel. (Lambda EcoRI/HindIII ladder was used as the standard for estimating the size of the phage genome. Ladder fragment size is indicated above the ladder. Numbers on the gel represent the DNA of the 24 phages that were run on the gel).

The whole genome DNA of all the phages along with the host strain 2356 was submitted to Massey genome services for complete genome sequencing. In the initial round of sequencing, there were not enough reads to generate an assembly hence the phages were resequenced. For example, in the first round of sequencing, 6600 to 21300 reads were obtained for each phage genome. These numbers were three times higher during resequencing. The reads that were obtained from both the sequencing rounds were combined in order to obtain better assemblies. The phages that were considered to be complete i.e. merged to a single contig have genome sizes that range from 31,844 bp to 34,715bp (Table

3). During the sequencing of the host strain, a prophage was subsequently identified and has been included in further analyses.

3.2 Determination of similarity and differences between phages through Restriction Digests

Restriction digests were carried out while the phage DNA was being sequenced at Massey genome services. The goal of this experiment was to analyse if there was a pattern of similarity between the phages. This method helps in not only identifying the group or the type of phage infecting the starter culture based on the banding pattern, but also helps in analysing if phages isolated from the various fermentative units at different times are similar or different. Several enzymes were tested and the enzymes EcoRI, HindIII and ClaI were found to give readable banding pattern. Results of restriction digestion using the enzyme ClaI for some of the phages are provided in Figure 11. Restriction digest data of sample phages that showed clear results from our data have only been provided here.

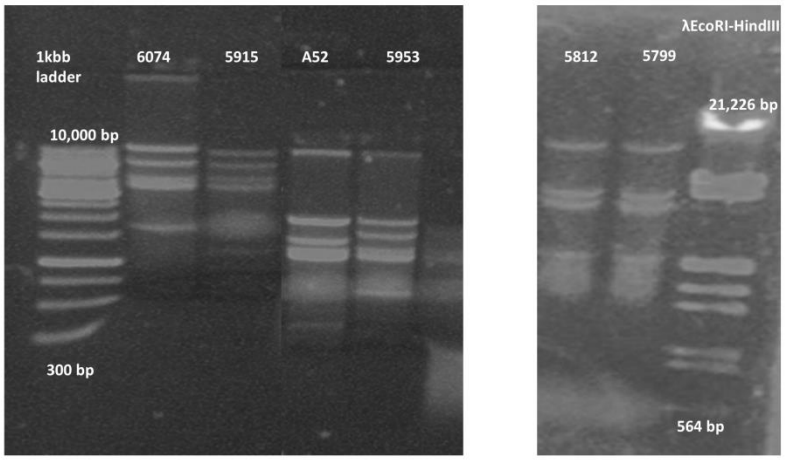


Figure 10 Restriction digestion of phages (sample) using the enzyme ClaI. (1 kbp ladder and Lambda EcoRI HindIII ladders were used for DNA size determination. High quality sequences could not be obtained for phages 6074 and A52 during DNA sequencing that were a part of this project).

Out of the 24 phage DNA samples that were sent for sequencing, a high throughput sequencing result was obtained for only 15 phage genomes. The host strain 2356 was also sequenced and a prophage was identified in that strain whose sequence was also obtained. The whole genome sequences of the fifteen phages and the prophage were subjected to further bioinformatic analyses in order to understand their genetic relatedness and their sources.

3.3 Bioinformatic analysis of the phage genomes

3.3.1 Blast Analysis

After obtaining the whole genome sequences, we performed initial identification of the closest relatives of the phages in this study using Blast (Basic local alignment sequence tool) (Altschul et al., 1990). Our goal in sequencing is to use the complete DNA of these bacteriophages in order to find out how different they are and how many clusters, or groups of closely related phages they might represent. Blast nucleotide analysis that matches the query genome to known sequences in the database indicates the percentage match between the query sequence and the subject and also indicates the percentage of similarity over the length of the coverage area. Higher coverage with a high degree of identity indicates a closer evolutionary relationship between the phages being studied and the phages in the sequenced database.

The results of this analysis indicated nucleotide sequence identity of fourteen phage genomes and the prophage to phage R1T, isolated in Netherlands which is a P335 type, temperate phage. Nucleotide identity results showed at least 97% similarity (between 78 to 89% of sequence coverage) of the phage genomes to phage R1T, phage 6090 being the notable exception (Table 3). A high blast identity of at least 96% was also noticed between

the phages and the phage 63301. Phage 63301 was isolated recently in Ireland after the phage infection of the industrial lactococcal strain *Lactococcus lactis*. Phage 63301 also has a close genetic relatedness to phage LC3 which is also a member of the P335 group (Blatny et al., 2004). Even though phage 6090 had 98% identity to phage R1T there was genome coverage of only 6%. Phage 6090 isolated from the fermentation plant located in the site N3, NZ has shown 94% similarity to phage phi4 (71% coverage) that belongs to 936 group indicating that this phage belongs to 936 group of phages.

Table 3 Bacteriophage genomes and Blast identities to known bacteriophages in the database.

Phage #	Genome size(bp)	Identity to phage R1T	Query coverage (R1T)	Identity to prophage	Query coverage	Identity to Phi4	Query Coverage
5731	34,335	98%	88%	98%	92%	83%	3%
5904	34,715	98%	83%	98%	89%	83%	3%
5910	33,401	97%	88%	98%	88%	85%	3%
5912	32,129	97%	78%	97%	77%	83%	3%
5915	31,844	99%	84%	98%	83%	85%	3%
5876	33,652	98%	78%	98%	81%	82%	3%
5943	33,019	97%	78%	97%	79%	83%	3%
5953	32,746	97%	84%	97%	79%	83%	6%
5807	33,503	98%	78%	98%	81%	82%	3%
5809	33,260	98%	89%	98%	86%	85%	6%
6023	32,419	99%	78%	97%	79%	85%	4%
5799	33,812	98%	89%	98%	87%	45%	6%
A31	33,652	98%	78%	98%	81%	82%	3%
5812	33,801	98%	83%	98%	87%	85%	3%
6090	32,834	98%	6%	91%	6%	94%	71%
Prophage	31,090	99%	100%	100%	100%	85%	4%

In order to demonstrate the presence of conserved regions or regions of high similarity between the sequenced phages that have shown close identity to the P335 and 936 group of phages and to identify their nucleotide synteny, a dot plot analysis was carried out.

Obtaining a dot plot matrix also helps to visualize the homology across the entire length of the phage genomes.

3.3.2 Dot plot analysis for detecting Synteny

Synteny refers to the order of genes located in chromosomes of organisms or species that are related through a common ancestor. Two sets of genomes or chromosomes being compared with one another are said to be syntenic if there is a conservation of gene order. In order to identify if the phages were syntenic with one another, we conducted a dot plot analysis using the tool Gepard (Krumsiek and Rattei, 2007) (Figure 12).

This method uses suffix arrays where a search is made for a word length which is often 10 bp and a dot is created whenever a match is found. Gepard analysis has several advantages over the conventional dot plot method. The main ones are the reduction in computational complexity and the ability to detect and match sequences that are even in reverse orientation.

Whole genome sequences of 12 phages that were obtained in the initial round of sequencing were analysed for synteny using this method. The figure illustrated below indicates the dot plot analysis of the 12 phage genomes that were obtained after the initial round of sequencing. The dot plot matrix seen below was obtained before a whole genome alignment of the phages was performed.

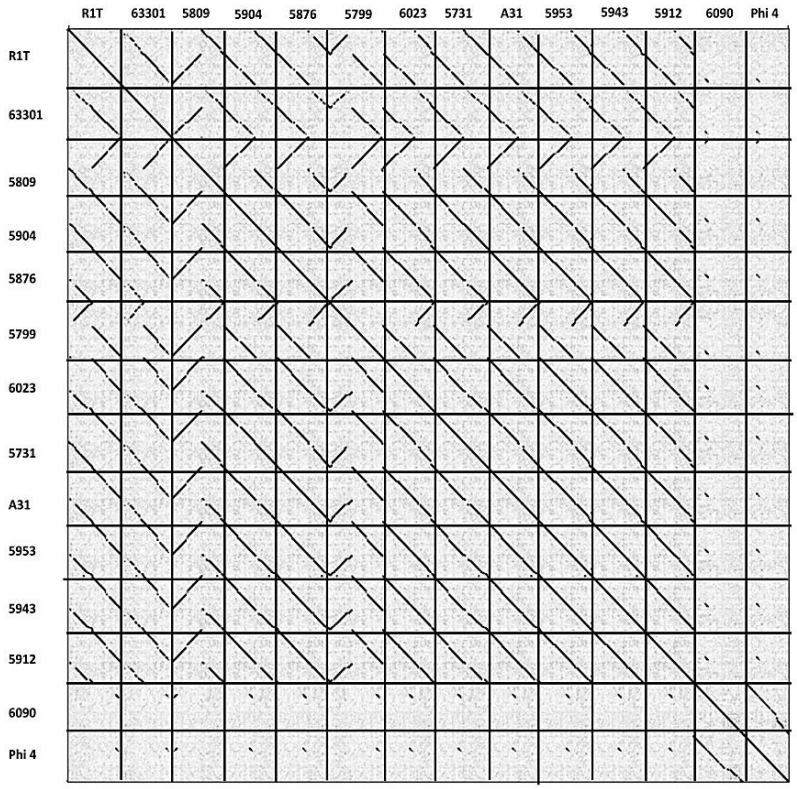


Figure 11 Dot plot analysis using the tool Gepard before whole genome alignment. Phage 6090 (936 group) and reference phage Phi4 (936 group) are not syntenic with the other P335 phages in this study can be observed.

3.3.3 Phage Whole genome alignment

After obtaining the genome sequences from the DNA sequencing center, we noticed a random order in the genome sequences. This was predicted using Mauve v.2.4.0 (Darling et al., 2004). If the genome or a part of the genome represented in blocks is in reverse orientation then it would appear below the centre line. We observed that some of the genomes after sequencing were in reverse orientation. Such blocks of sequences in reverse orientation were identified and their compliments were obtained using the reverse compliment software. In order to align the genomes form the same start point, we followed the sequence convention in the reference phages (like R1t and phage 63301) i.e all the phage genomes started with the integrase gene and ended in the gene coding for lysin. The coordinates for these sequences

coding for these genes were identified in the reference fasta file using NCBI Blast. Following this order we created a new fasta file for each phage genome starting with the sequence coding for integrase gene. This allowed us to generate the genome files for all the phage genomes. The new fasta file that consisted of all the phage genomes in the same gene order was fed into the Mauve 2.4.0 tool to generate the output.

This tool displays locally collinear blocks that are conserved across the genomes. Each collinear block has a weight attributed to it by the software. In order to identify similar regions or conserved blocks, a minimum weight has to be met. Any collinear blocks of low weight are automatically discarded from the alignment thus assuring the robustness of the conserved regions that are finally displayed. This tool was helpful to identify genomes that were in reverse orientation so that a rectification could be done. Regions not included in the blocks represent non homologous regions. Progressive mauve helps to look into the rates and patterns of genomic rearrangements.

Figure 13 demonstrates the result obtained from progressive Mauve after reorienting the phage genomes that were in reverse direction. Mauve results have revealed that a major portion of these phage genomes are conserved. No genomic rearrangements were observed. Hence we were able to align all the phage genomes from a common start point with the help of this tool which also enabled us to observe synteny among the phages. It is noteworthy to mention that we followed the reference phage genome convention of positioning the gene coding for integrase in the beginning (Van Sinderan et al., 1996).

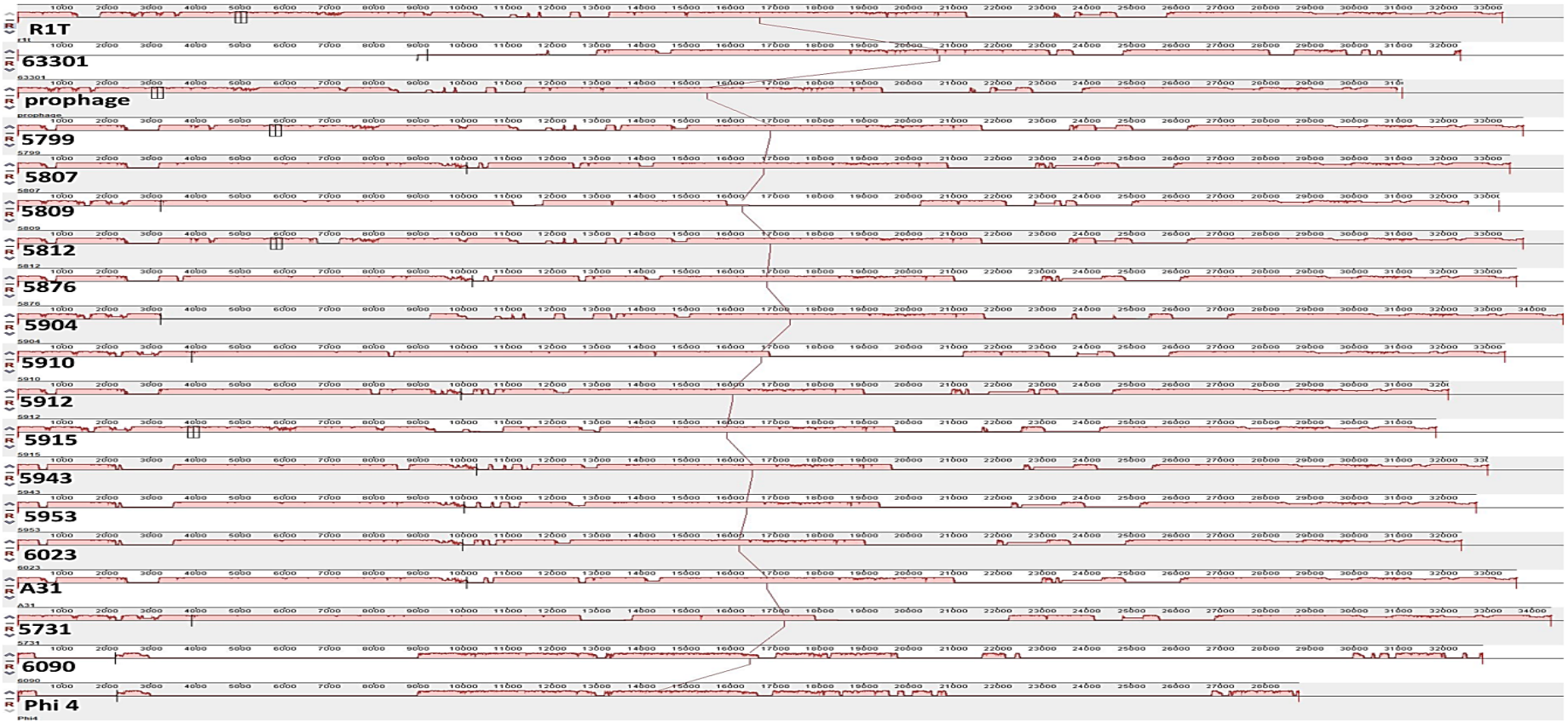


Figure 12 Illustration of multiple alignment of whole genome sequences of the 15 phages and the prophage along with a reference genome R1T , 63301 (P335 group) and Phi4 (936 group) using Mauve V2.4.0

The following dot plot analysis (Figure 14) illustrates synteny observed after whole genome alignment was completed using progressive Mauve for all the 15 phages and the prophage in this study.

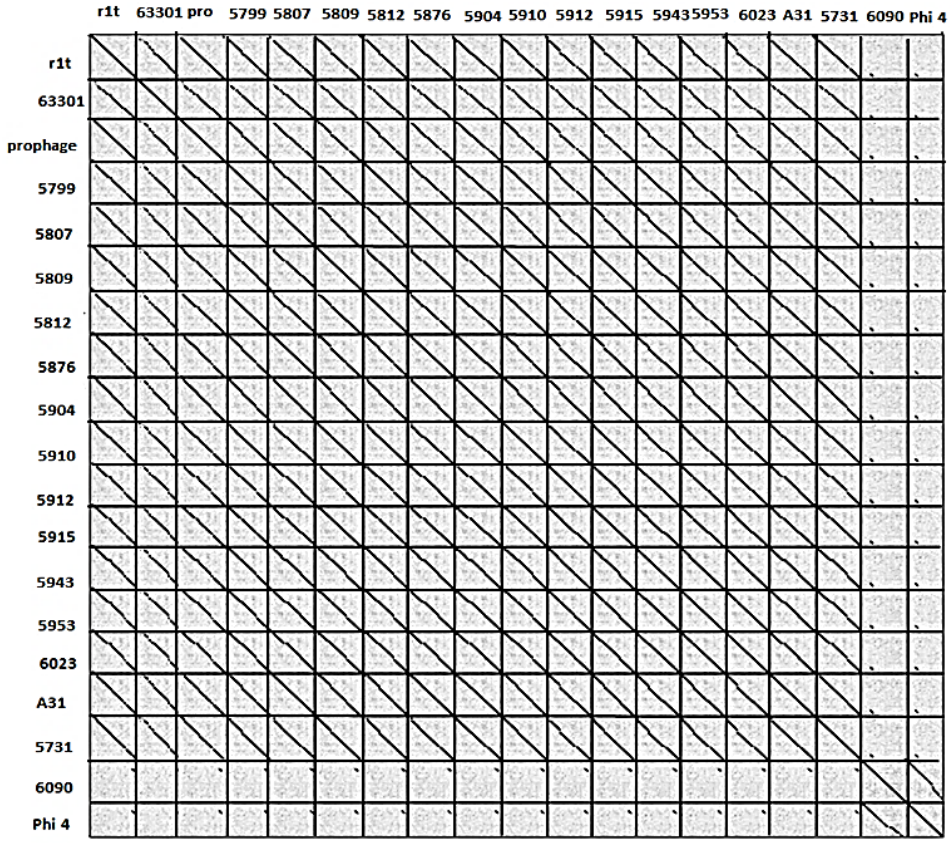


Figure 13 Dotplot analysis (using Gepard) of fifteen phages and one prophage along with reference phage genomes R1T and 63301. (Diagonal lines indicate genomic regions in which base pairs match between two genomes. A continuous diagonal line indicates that the phages align with themselves while noncontiguous lines show the dissimilarity between different phage genomes).

From out dot plot analysis we interpret that phages and the prophage that show identity to R1t (P335 group) are syntenic with one another as represented in continuous diagonal lines and significantly divergent from phage 6090 and its reference phage phi4 (936 group) as no strong homology lines were detected. This confirms the conservation of gene

order in these phages that share homology with each other and with their respective reference phages.

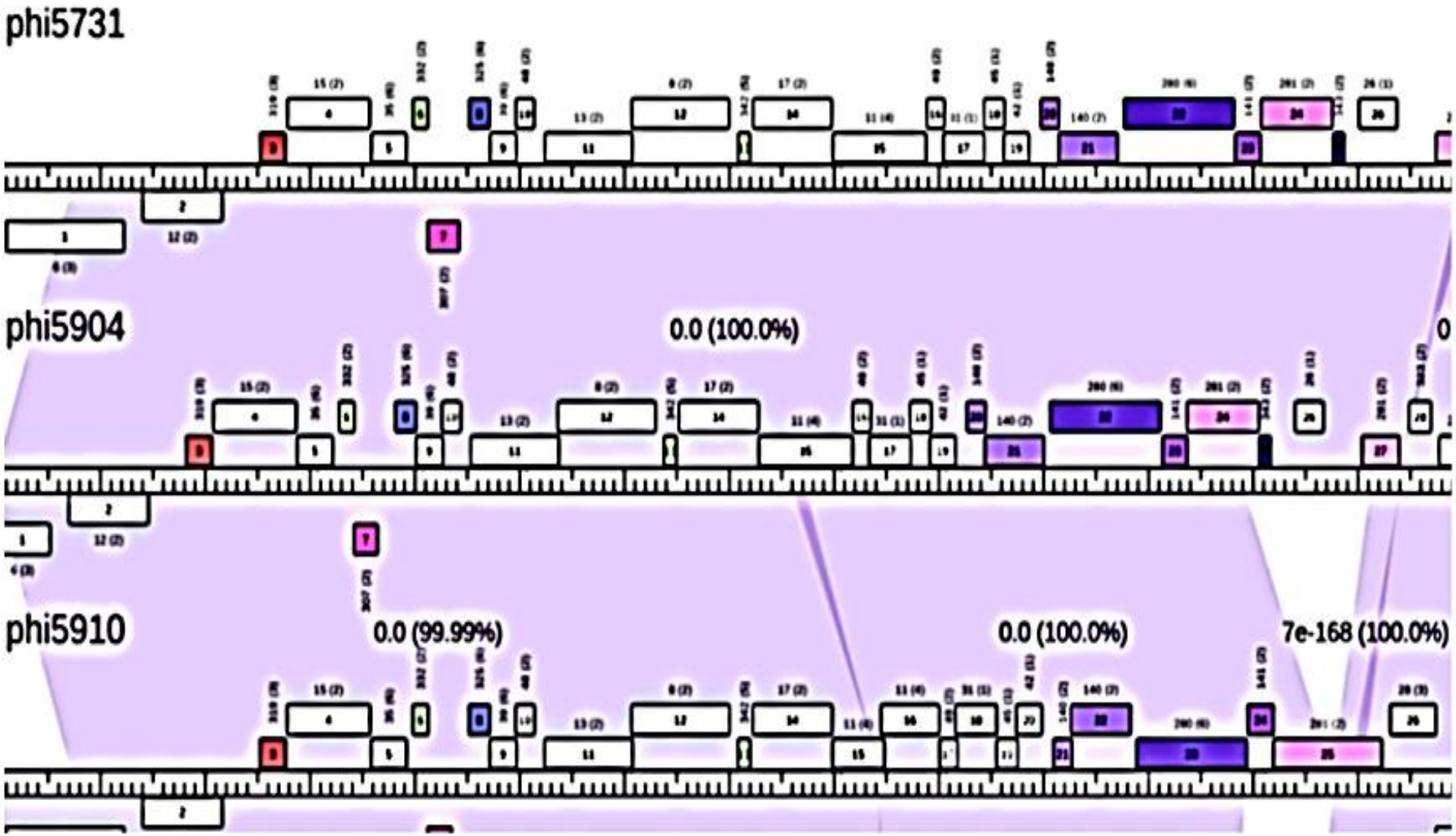
After observing synteny in these phages we wanted to identify regions of high similarities and differences both at the nucleotide and the protein level. Hence we extended our analysis further to understand the relationship of genes and the proteins coded by these genes by using the phamerator database.

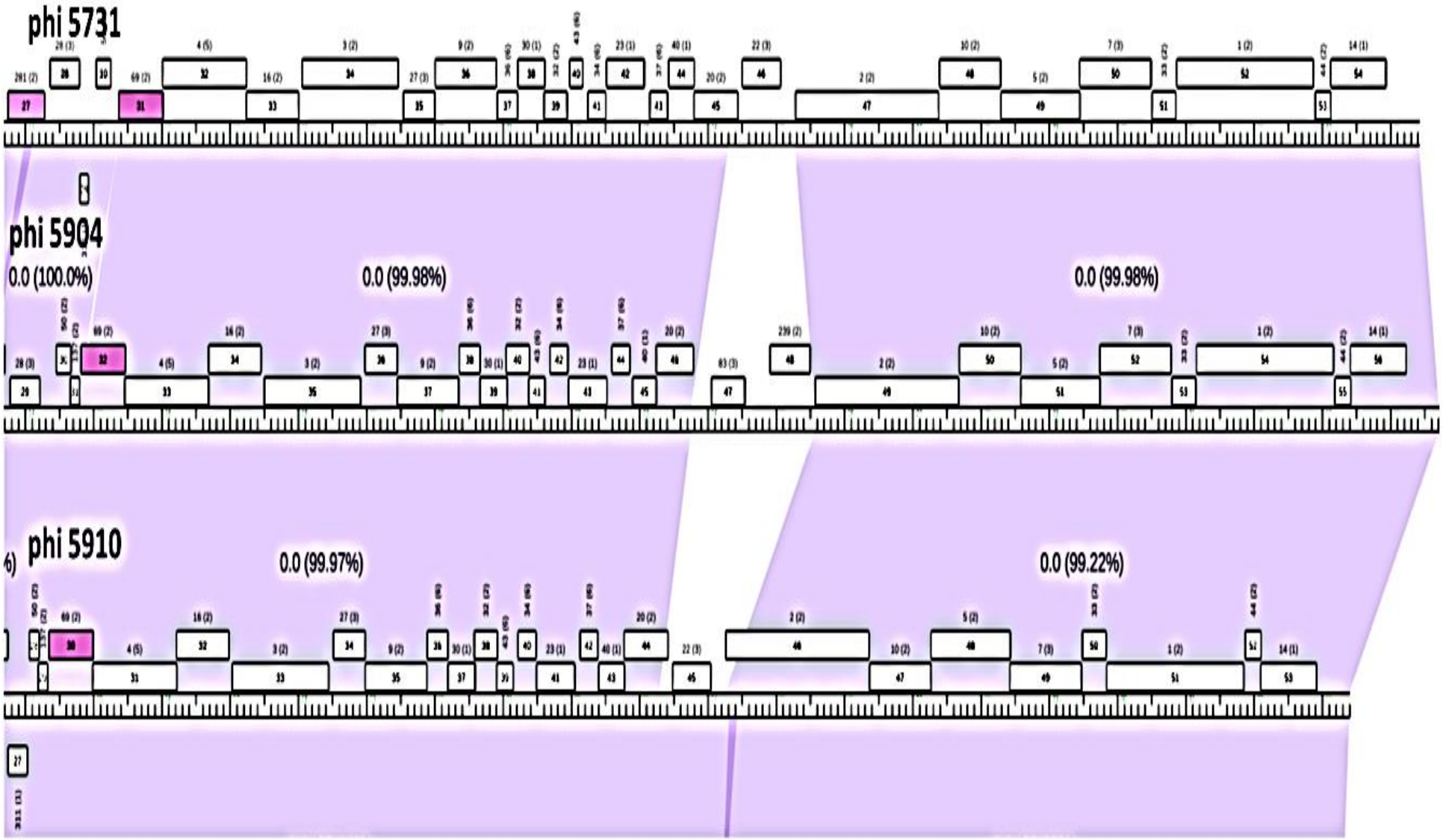
3.4 Visualization of protein 'phamilies' in prophage and phage genomes using phamerator

In this study, we have used the tool Phamerator to assign the protein coding genes of related sequences into specific groups or "phamilies" / "phams" (Cresawn et al., 2011). This tool elucidates the genetic relationship at both the nucleotide and the amino acid level by generating a genome map based on the relationships between the phages in the database. The nucleotide homology between the phage genomes is considered high if shown in purple and no homology if shown in white. The numbers inside each box represents the gene number and genes with a score of at least 25% amino acid identity and/or an E value of 0.0001 were placed in the same pham. The numbers on top of the genes indicate the phamily with which the gene is associated (Figure 15). The phams allow the relationships between the different phages to be analysed using genome maps that enable us to observe the mosaic nature of the phages.

From our phamerator data we observed that a number of phams are shared among the P335 phages indicating the conservation of those genes in these phages. Phamerator data also has provided evidences of horizontal gene transfer among the

P335 phages. We also interpret that the P335 phages display a high nucleotide identity with one another and very less identity with the phage 6090 that belongs to the 936 group.


phi5731
phi5904
0.0 (100.0%)
phi5910
0.0 (99.99%)
0.0 (100.0%)
7e-168 (100.0%)


phi 5731
phi 5904
0.0 (100.0%)
0.0 (99.98%)
0.0 (99.98%)
phi 5910
0.0 (99.97%)
0.0 (99.22%)

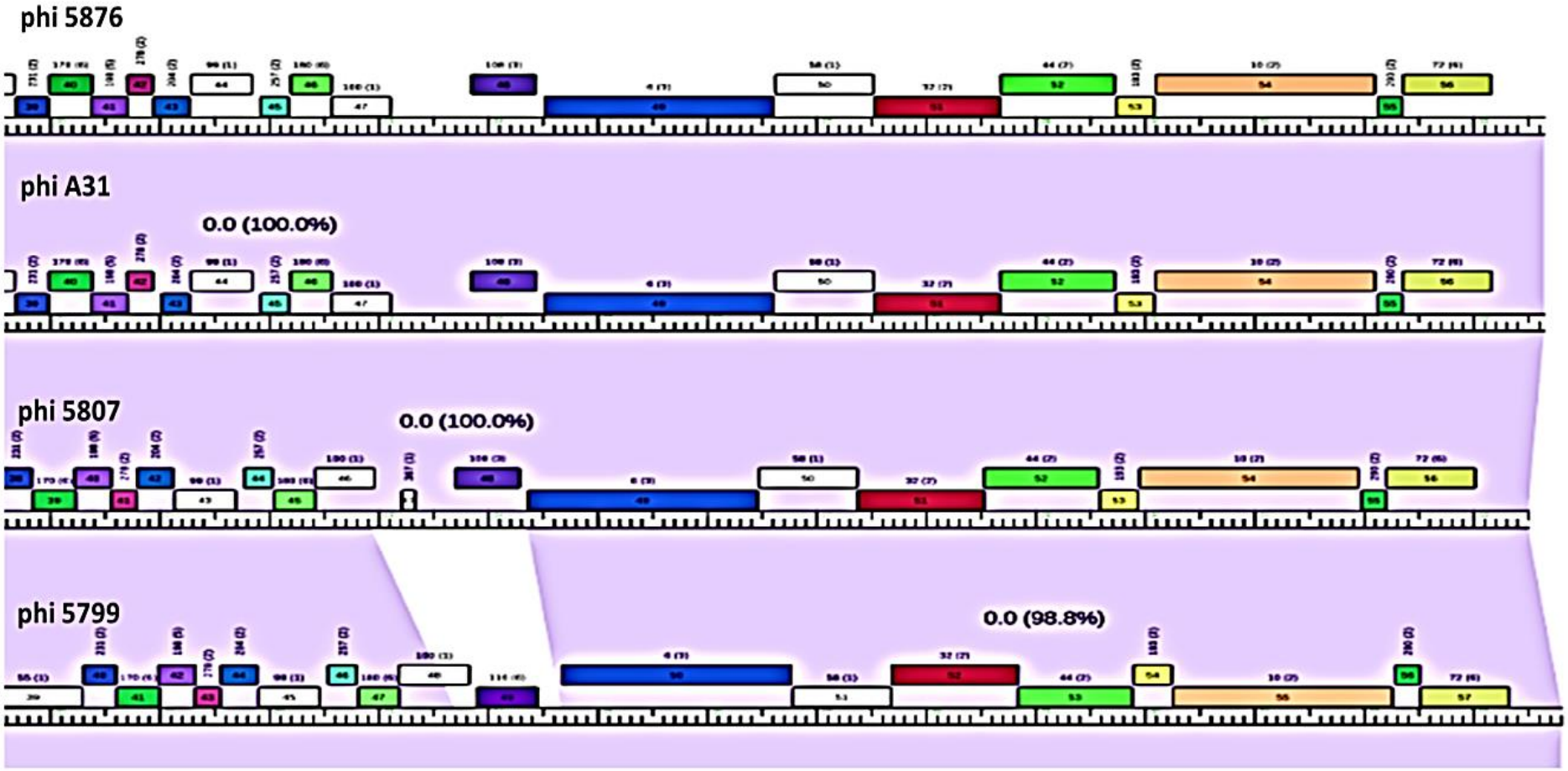


Figure 14 Schematic representation of phamerator maps presented in three tiers for an expanded view and clarity. A) Genome maps of representative phages (5731, 5904,5910) showing the former half of the genome. B) Genome map of the phages (5731, 5904, 5910) showing the latter half. C) Genome map of representative phages (A31, 5876, 5807, 5799, 5812). The predicted ORFs/genes are placed on top and bottom of the genomes in boxes. The number inside the boxes indicates the gene number. The numbers above the boxes represent the pham number or the protein family to which the gene belongs. Purple colour signifies region of high similarity and while colour indicates lack of similarity.

Despite obtaining ample evidence on the relationship of the phages at the level of nucleotide and proteins from phamerator data, in order to understand their evolutionary history we decided to use phylogenetic approach. Initial BLAST searches from the sequenced phage genomes indicated that the structural genes and replication machinery of all phages (and prophage) expect 6090 are more similar to those of the phage R1T of the P335 group than to any other phage group. Phage 6090 indicated a high similarity to the phage Phi4 of the 936 group. In order to place these phages and the prophage in a phylogenetic context with these two groups, we used their whole genome sequences to construct a phylogenetic tree

3.5 Phylogenetic analysis

3.5.1 Phylogenetic tree and clustering of phages

The 15 phage genomes approximately 31-34 kbp were submitted to the Clustal Omega software in order to obtain a phylogenetic tree using neighbour joining tree algorithm (Saitou et al., 1987) (Figure 16). Based on our Blast sequence similarity results, temperate lactococcal phages R1T and 63301 both belonging to the P335 group and virulent phage Phi4 belonging to the 936 group have been included in this study for comparative analysis. The alignment obtained from Clustal Omega also helps us to cluster them based on their distribution in the tree and also analyse their evolutionary history. Though an initial neighbour-joining tree was obtained to observe the clusters in these phages, for further speculative analysis involving recombination, comparative maximum likelihood trees were obtained using the PHYML software (Figure 17).

The percentage identity matrix between the reference phages and the phages under study generated a pairwise comparison output between the phage genomes (Table 4). These results along with the phylogenetic tree that was generated helped us to obtain two clusters namely Cluster P335 and Cluster 936. These clusters can also be related to their locations of isolation.

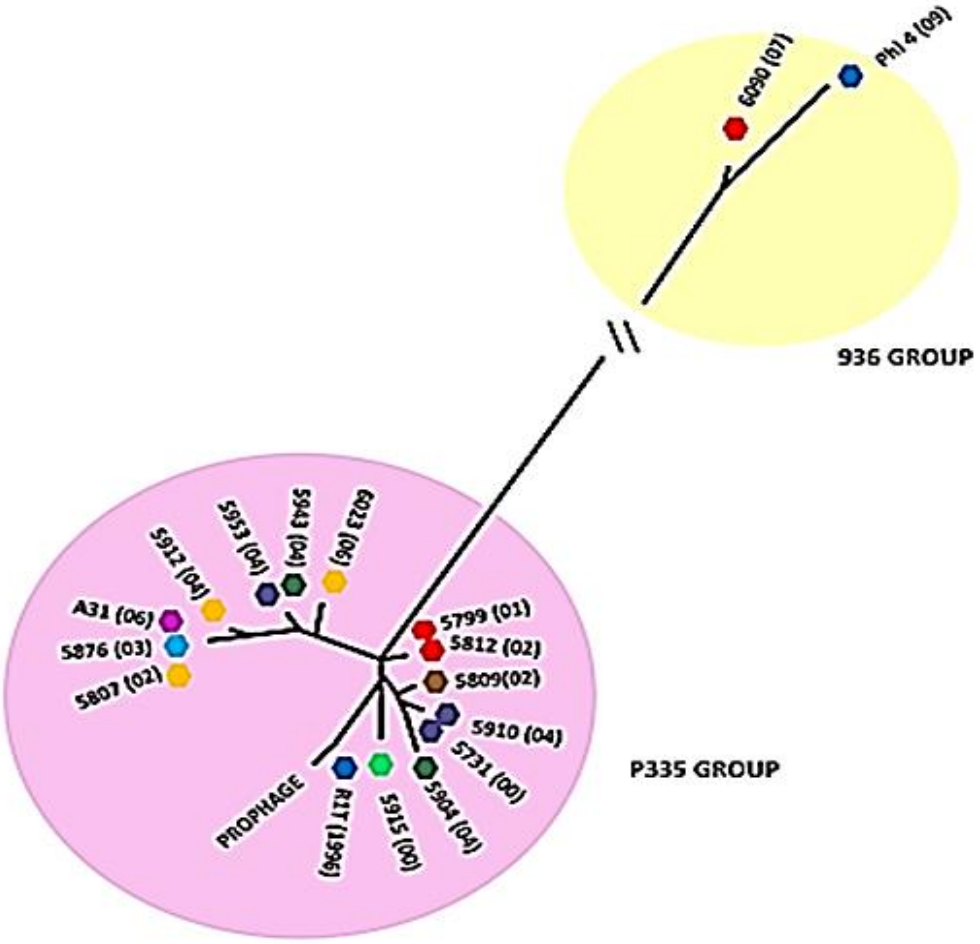


Figure 15 Phylogenetic tree (unrooted) obtained from whole genome sequences of phages in this study. Clusters showing bacteriophages along with Reference bacteriophages (R1t and Phi4). (Whole genome alignment was performed by (Clustal Omega) and a tree using the neighbour-joining method was subsequently generated. Tree visualization was performed using the software Figtree. Clusters (P335 and 936) are indicated in pink and yellow shades respectively. An illustrative geographical map is also provided to indicate the regions of isolation of phages from different units of Fonterra). Small colour coded hexagons used to represent the fermentation sites are as follows: Site N1 (blue), Site N2 (red), Site S1 (light green), Site S2 (brown), Site S3 (yellow), Site S4 (purple), Site S5 (Dark green), Site S6 (pink). Reference phages (R1T and phi 4) are shown in dark blue.

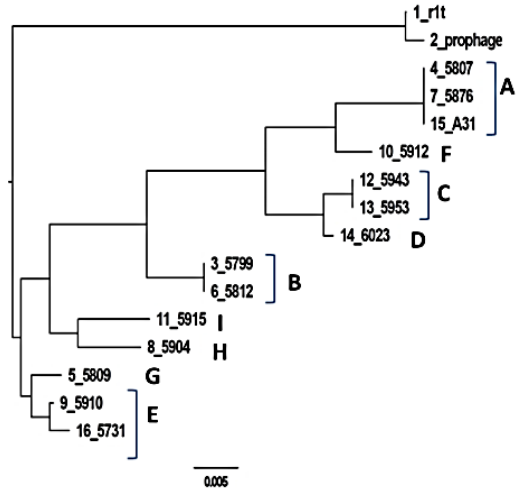
Table 4 Percentage identity matrix showing pairwise distance between the phages generated by Clustal Omega. (Colour code indications: Red: 99.6-100% identity, Green: 95-99.5% identity, Yellow: 90-94.9% identity and Grey: 53-89% identity).

	R1t	prophage	A31	5807	5876	5912	5943	5953	6023	5799	5812	5915	5904	5809	5910	5731	6090	Phi4
R1t	100	99.76	91.43	91.42	91.43	91.77	90.78	90.78	91.95	94.11	94.11	93.3	93.75	94.18	94.7	94.49	57	53.81
prophage	99.76	100	91.17	91.16	91.17	91.54	90.68	90.68	91.9	93.95	93.95	93.03	93.52	93.88	94.43	94.21	57.15	53.92
A31	91.43	91.17	100	100	100	98.15	97.25	97.24	97.84	96.93	96.92	95.46	96.38	95.01	95.98	95.2	56.6	53.09
5807	91.42	91.16	100	100	100	98.15	97.26	97.25	97.84	96.93	96.92	95.46	96.38	95.01	95.98	95.19	56.5	53.09
5876	91.43	91.17	100	100	100	98.15	97.25	97.24	97.84	96.93	96.92	95.46	96.38	95.01	95.98	95.2	56.5	53.09
5912	91.77	91.54	98.15	98.15	98.15	100	97.67	97.68	98.48	97.75	97.74	96.53	96.7	95.83	95.45	95.68	57.35	53.82
5943	90.78	90.68	97.25	97.24	97.25	97.67	100	100	98.74	96.78	96.77	95.49	95.82	95.06	95.92	95.1	56.98	53.54
5953	90.78	90.68	97.24	97.25	97.24	97.68	100	100	98.74	96.79	96.78	95.5	95.82	95.07	95.92	95.1	56.98	53.54
6023	91.95	91.9	97.84	97.84	97.84	98.48	98.74	98.74	100	97.18	97.18	95.87	95.44	95.41	95.82	95.59	57.25	54
5799	94.11	93.95	96.93	96.93	96.93	97.75	96.78	96.79	97.18	100	99.99	95	95.86	95.85	96.87	96.66	56.98	53.76
5812	94.11	93.95	96.92	96.92	96.92	97.74	96.77	96.78	97.18	99.99	100	95	95.86	95.84	96.87	96.65	56.98	53.76
5915	93.3	93.03	95.46	95.46	95.46	96.53	95.49	95.5	95.87	95	95	100	97.69	99.04	97.82	97.62	56.98	53.71
5904	93.75	93.52	96.38	96.38	96.38	96.7	95.82	95.82	95.44	95.86	95.86	97.69	100	97.52	98.84	98.64	57.18	53.9
5809	94.18	93.88	95.01	95.01	95.01	95.83	95.06	95.07	95.41	97.52	95.84	99.04	97.52	100	98.84	98.64	57.18	53.9
5910	94.7	94.43	95.98	95.98	95.98	95.45	95.92	95.92	95.82	96.87	96.87	97.82	98.84	98.84	100	99.78	56.97	53.67
5731	94.49	94.21	95.2	95.19	95.2	95.68	95.1	95.1	95.59	96.66	96.65	97.62	98.64	98.64	99.78	100	56.86	53.46
6090	57	57.15	56.5	56.5	56.5	57.35	56.98	56.98	57.25	56.98	56.98	56.98	56.82	57.18	56.97	56.86	100	86.03
Phi4	53.81	53.92	53.09	53.09	53.09	53.82	53.54	53.54	54	53.76	53.76	53.71	53.44	53.9	53.67	53.46	86.03	100

3.5.2 Classification of P335 type phages based on pairwise comparison

Based on our analysis using whole genome pairwise comparison and percentage identity matrix output, we report 9 (A-I) different P335 phages isolated from the different fermentation units of Fonterra. The table below indicates that the type of a phage is location independent. For example, though type A phages are from different locations, exhibit similarity, while phages 5943 and 5904 both from site S5 are not identical thus falling under different types (Table 5). Whole genome pairwise comparison of the phages helped us to evaluate the percentage of relatedness between them. Also, the neighbour joining tree that was generated illustrates the evolutionary distance between the phages thus categorizing them as a particular type. The high level of identity observed among some phages thus identifying them as the same type leads to a plausible conclusion that these phages have coexisted at some point. It is noteworthy to mention here that most of the Lactococcal strains (lactis and cremoris) display strong homology in their genomes (Fernandez et al., 2011).

Table 5 **Phage type characterization based on location/site, their distribution on the phylogenetic tree (shown on the left) and percentage identity matrix results**



Phages	Location/site	Type
5807,5876,A31	S6, N2, S3	A
5799, 5812	N3	B
5943, 5953	S5, S4	C
6023	S3	D
5731, 5910	S4	E
5912	S3	F
5809	S2	G
5904	S5	H
5915	S1	I

3.6 Evolutionary history and Strategies adopted for evolution

3.6.1 Homologous recombination between the P335 phages

In order to address the similarity identified among the phages (P335 phages in this study) and the prophage, that may be due to homologous recombination events, we have attempted to provide supporting evidence using the variations or incongruence seen in the phylogenetic trees obtained from the whole genomes of the phages in this study.

We would expect to observe incongruence in the phylogenetic trees if genetic exchange has occurred via homologous recombination. To address this, we have used a model that was explained by Boni et al (2010) while describing the phylogeny of influenza viruses. If a phage is represented in different clusters when different regions of the genomes are being analysed, then we consider this to be a recombination signal that has occurred. As a means to establish this, we aligned the phage genomes (P335 group) to make sure that the gene order displayed synteny. This was followed by a multiple genome alignment using the MAFFT software in Clustal Omega. The results were fed into the program PHYML to generate maximum likelihood trees that were used for further comparisons (Figure 17). Phage 5904 clusters with different clades in both the trees (Bootstrap values of 100% and 89%). This phage clusters with 5910 and 5731, 5809 and 5915 in the first half of the genome (Figure 17B) and with A31, 5876 and 5807 in the second second half of the genome (Figure 17C). Phage 5809 which shares a more common ancestor with 5915 in the first tree appears to jump clades in the second tree and this is supported by a bootstrap value of 100%. Such

discordant topologies are indicative of homologous recombination events.

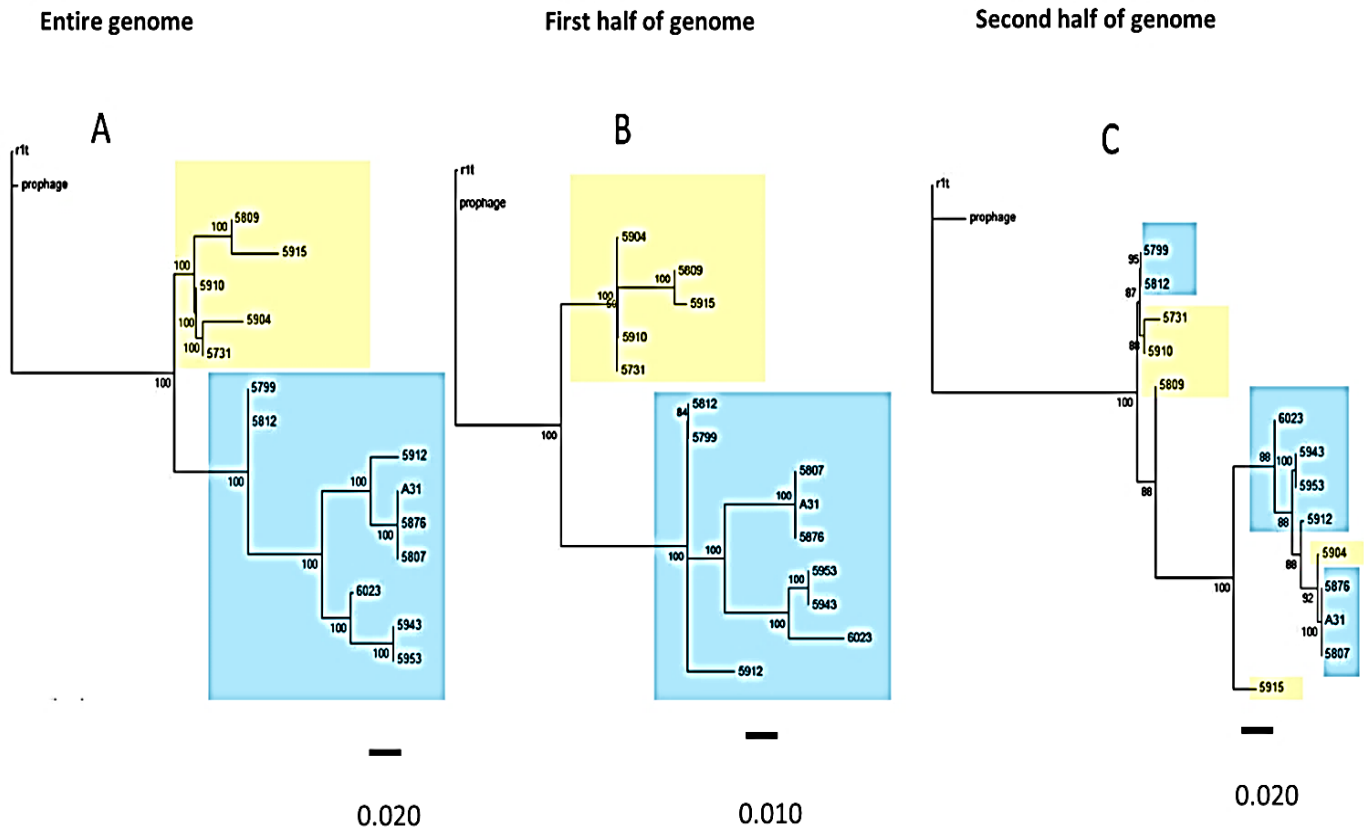


Figure 16 Bootstrapped Maximum Likelihood trees for the whole genome and parts of the genomes. Tree A represents the entire genome of the P335 group phages in this study. Tree B was inferred for the first half of all the phage genomes and Tree C was inferred for the second half of the genomes. Trees inferred for the different regions with PHYML are different thus supporting the hypothesis of homologous recombination (Yellow and blue colours indicate cluster 1 and cluster 2. Phages specific to the clusters in trees A and B are shuffled in tree C indicating homologous recombination).

3.6.2 Horizontal gene transfer and phage evolution

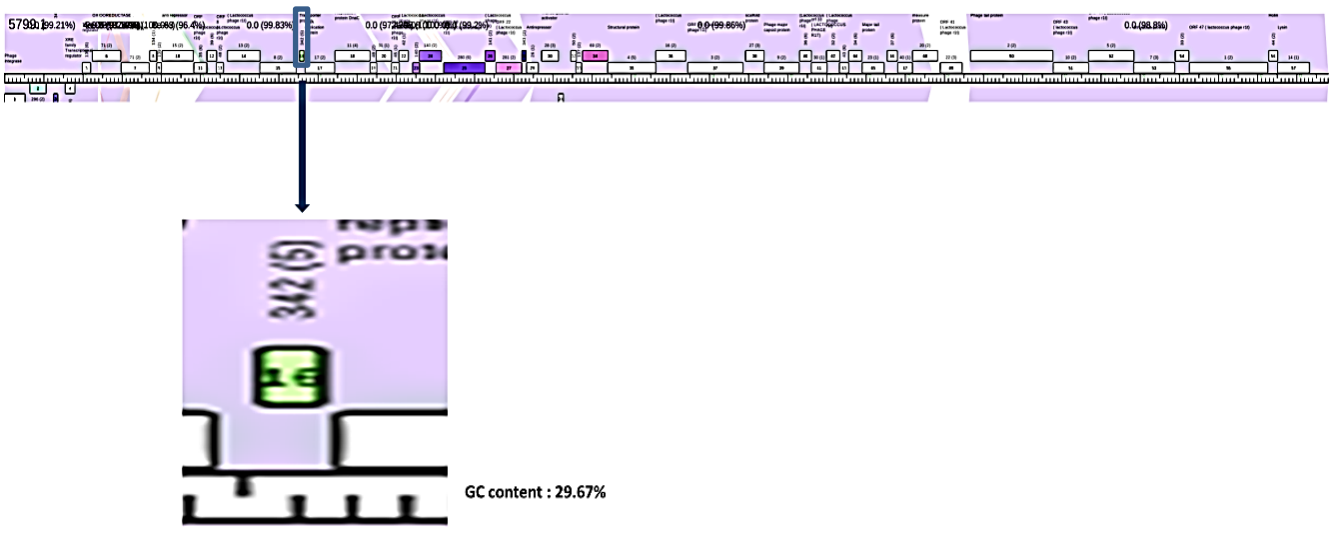
In this section, we have attempted to address mosaicism observed in our P335 phages. Mosaicism genomes consist of regions that share homology with other similar genomes interspersed with unrelated regions. According to research, phage populations exhibit mosaicism in their genetic makeup as they are constantly involved in exchanging genes (Hendrix et al., 1999 and Brussow et al., 2001). In this study we observed that genes coding for ABC binding ATP transporter protein that is only 39 amino acids long is identified in

only certain phages. The GC content of this gene is 29.67% which is different from the GC content of genes in all the phage genomes in this study (Figure 18A).

Similarly the gene coding for the endonuclease str-PAP-1 identified in phages 5904 and 5912 are only 72 amino acids long with a GC content of 40% which is higher than the GC content of the rest of the genes in this study

Based on these results we infer that these genes may have been acquired via HGT from another infecting phage. Apart from a variable GC content, evidence to substantiate non homologous horizontal gene transfer events is provided by the lack of sequence similarity both upstream and downstream of this gene with the other phages (Figure 18 B). Therefore it is evident that HGT has played a significant role in the evolution of these phages.

A)



B)

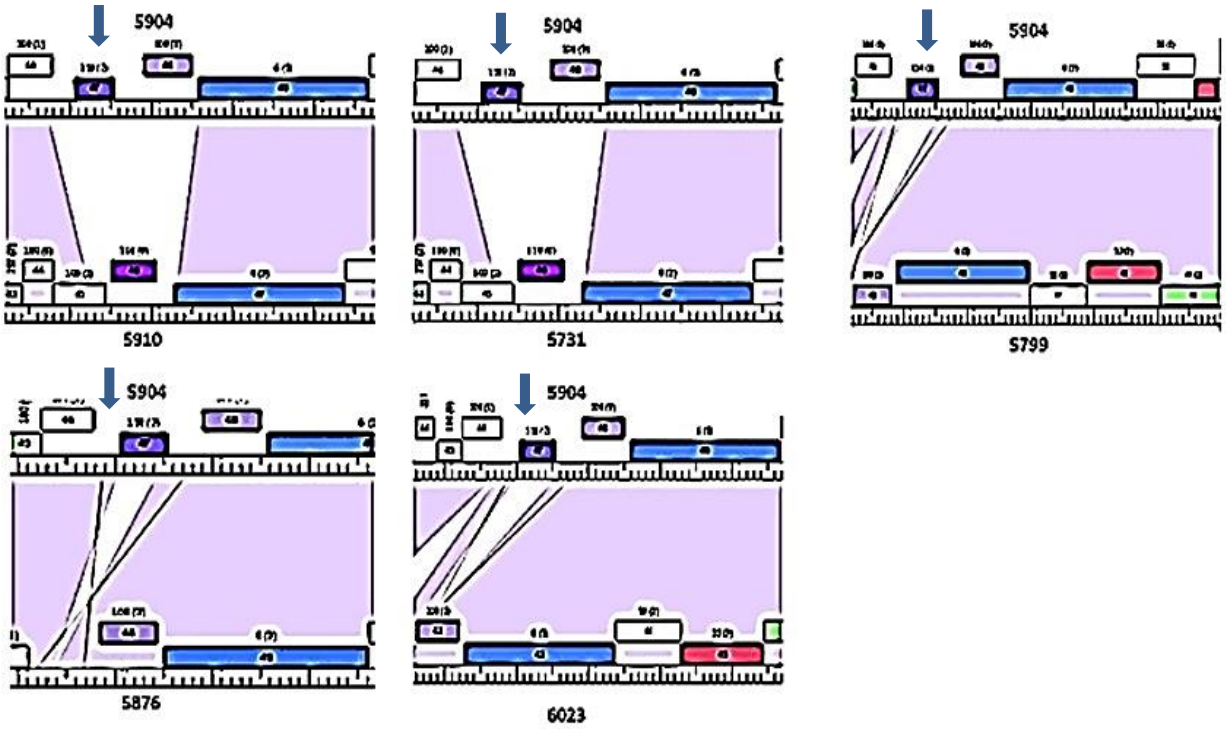


Figure 17 A) Illustration of Horizontal gene transfer in specific phages (Phage 5799) depicted by the gene coding for ATP transporter. B) Horizontal gene transfer of the gene coding for endonuclease str-PAP-1 observed in phage 5904 and absent in other phages. The arrow indicates the presence of the gene coding for str-PAP-1 seen in phage 5904 while missing in other phage genomes. Specific regions showing HGT have been obtained and highlighted from phamerator genome maps **(**Lack of nucleotide similarity represented in white colour confirms nonhomologous gene transfer of this gene).

3.7 Prophage in the strain 2356 and its significance

The *L. lactis* subsp.*cremoris* strain 2356 consists of an integrated prophage of size 31090 bp. The prophage contains 50 ORFs. The prophage showed at least 90% nucleotide identity to the other phages in this study. The prophage also revealed 99% nucleotide identity with 100% query coverage to the phage R1T isolated from Netherlands. It is noteworthy to mention here that the phage R1T that is used as a reference P335 phage in this study owing to its similarity with the phages, was first isolated in the year 1996 in Netherlands and the strain *L. lactis* subsp.*cremoris* (2356) was also introduced in the same

year in Fonterra. The prophage isolated from the strain 2356 has been annotated and will be submitted to Genbank (Appendix). The presence of a prophage in the host strain instigated us to probe further into its relationship with the sequenced phages.

3.7.1 Origin of phages (from prophages)

A number of studies on dairy strains have revealed the presence of one or more temperate bacteriophages in them and these can be induced by various factors (Huggins and Sandine, 1977). When I started my investigation of the whole genome sequences regarding the origin and the evolutionary relationship of the phages under study, I considered various hypotheses.

A) We first hypothesized that these phages may share a more common ancestry with the prophage in the strain 2356 used in this study, and have undergone homologous recombination with other infecting lytic phages.

Phamerator data suggests that though some of these phages exhibit a high sequence similarity with the prophage, a huge variation is observed in the lysogeny module of the prophage and the phages (Figure 19).

Also the differences in percentage identity matrix between the prophage and the phages obtained after pairwise sequence comparison is not suggestive of a recent common ancestor between these phages and the prophage in the strain 2356.

B) We also considered the hypothesis that the phages in this study especially the phages in the P335 may be independent or free living phages. The strong nucleotide similarity observed between these phages gives us an indication that these are not independent phages and have evolved from a common ancestor. The presence of genes coding for integrase in all the

isolated fourteen P335 phages and their striking similarity with the temperate phage R1T imply that these are not free living phages but are capable of integrating with a host strain

C) Based on the phamerator data that revealed the presence of lysogeny modules that were common to certain P335 phages in this study, we decided to test the hypothesis that the similarities observed among some P335 phages in this study may be a result of a prophage in a strain (part of the multiple starters) that underwent a lytic cycle infected new strains carrying prophage(s).

Majority of the lactococcal strains carry prophages that exist as lysogens (Casjens et al., 2003). Lysogeny module of the P335 phages comprises of genes that allows a phage to integrate with its host chromosome (Bruttin et al., 1997). In our study we noticed that a 4.5kbp DNA segment covers the putative lysogeny module of the P335 phages. The lysogeny module was chosen in our analysis because it contained genes that were similar across the P335 phages that were investigated in this study. Phage integrases that are a part of the lysogeny module mediate integration between the phage and the host chromosome attachment sites (attP and att B respectively) (Groth et al., 2000).

3.7.2 Genetic variations surrounding the lysogeny module

Characterization of the lytic phages from their temperate (lysogenic) homologues has shown that the difference is mainly due to the inactivation of the lysogeny module (Mikkonen et al., 1996). Transcription of the c1 gene coding for the repressor promotes an autoregulatory function that maintains the phage in a lysogenic state. A comparative analysis of this gene in the prophage and a few representative phages in this study is illustrated in Figure 19. A similarity in the gene coding for repressor was observed in the phages (5731, 5809, 5915,

5904 and 5910). This gene is not only homologous across the above mentioned phages but also differs from the prophage in the strain 2356 that was infected by these phages. In addition, a transcriptional activator that was common to just the above mentioned phages was also observed. Transcriptional activators and regulators are known to play an important role in DNA packaging and lysis in many Lactococcal bacteria (Quiles et al., 2013).

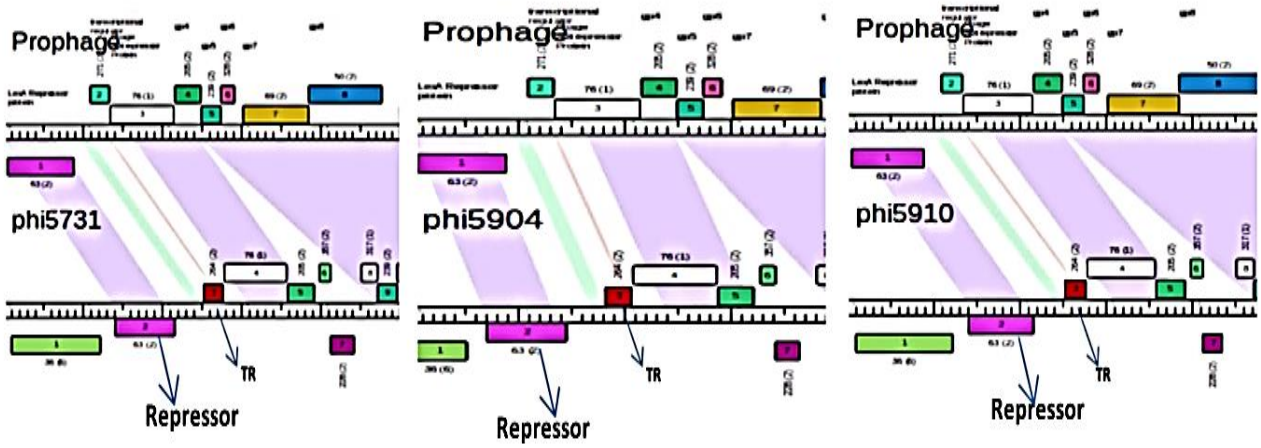


Figure 18 Illustration of a portion of the phamerator map showing the difference in the lysogeny module of representative phages and the prophage (White colour signifies difference in the nucleotide sequences and purple colour signifies the similarity. TR indicates transcriptional regulator found in these phages that is absent in the prophage.

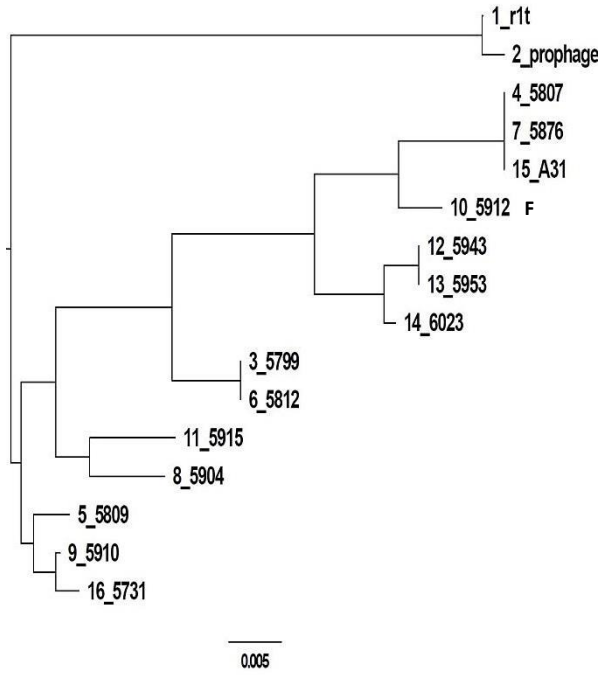
The XRE family transcriptional regulators and the oxidoreductase gene clusters in the lysogeny modules of the phages (5943, 5953, 6023 A31, 5876, 5807, 5812, 5799 and 5912) is also suggestive of a prophage containing these genes had undergone lysis and infected other strains. Interestingly, the XRE family of transcriptional regulators are known to be involved in a broad range of regulatory processes like controlling transcription, stress response, plasmid copying including its role in restriction modification system (Santos et al., 2009). This gene has recently been identified in one of the two prophages PLg-IPLA3145a in the *L. garvieae* strain. The other prophage PLg-IPLA31405b in this strain shows a high identity to the strain R1T which is closely related to the phages in this project. *L. garvieae* which

colonizes diverse environmental niches is known to share more than 900 genes with *Lactococcus lactis* strains used in the fermentation industries (Eraclio et al., 2017). Therefore the consistent differences observed in the lysogeny module of these phages and the prophage in the strain 2356 demonstrates that prophages in other strains that have become lytic and infected new strains share a higher homology with these phages.

The oxidoreductase gene represents the phage antirepressor genes. In the phages 5943, 5953, 6023, there is only partial representation of these genes (Figure 20). Another interesting feature observed in this set of three phages is the duplication of genes coding for HNH homing endonuclease intervened by a second tape measure protein that is only 39 amino acids long (Figure 20).

Majority of the lactococcal strains used in the dairy industries carry one or more prophages in them. Most of these temperate phages that exist as lysogens belong to the P335 group of phages (Madera et al., 2004). In this study, a high degree of similarity is observed among certain P335 phages primarily sharing genes in their lysogeny module.

With the evidence obtained from our phamerator database, we interpret that one or more prophages that have lysed from a strain (that has been a part of the starter cultures) have infected new strains that exist as lysogens. Homology among these phages has allowed recombination to occur between them that has significantly contributed to the genetic architecture of these phages. Hence isolation and detection of such phages in the different fermentation units have clearly helped us to understand their relationship with each other.



Phage #	Rep *	TR_1 *	TR_ 2*	XRE *	TR_3 *	Oxidoreductase * Ant A/Ant b	ATP Transporte r *	HNH Pham 116(6)	HNH Pham 108(3)	TMP (SHORT TAIL)
Prophage	■	■								
5807(02)▲				■	■	■	■		■	
5876(01)▲				■	■	■	■		■	
A31 (06)▲				■	■	■	■		■	
5912 (04)				■	■	■			■	
5943 (04)▲				■	■	PARTIAL			DUPLICATED	Additional TMP
5953 (04)▲				■	■	PARTIAL			DUPLICATED	Additional TMP
6023 (06)				■	■	PARTIAL			DUPLICATED	Additional TMP
5799 (03)▲				■	■	■	■			
5812 (02)▲				■	■	■	■			
5915 (00)	■		■						■	
5904 (04)	■		■				■	■		
5809 (02)	■		■					■		
5910 (04)	■		■				■	■		
5731 (00)	■		■				■	■		

Figure 19 Similarities and differences between the phages of the P335 group based on the distribution of genes in the lysogeny module and genes involved in horizontal gene transfer.

Black solid boxes (■) indicate the presence of a gene and white boxes (▭) indicate the absence of that gene. Grey box indicates either that gene is represented partially or is duplicated or is an additional gene. Phylogenetic tree generated by Clustal omega (MAFFT V.7) is provided for a clear interpretation of the gene distribution that signify diversity and clustering of phages largely due to the differences in their lysogeny module. "TR" represents transcription regulator. Pham (short form for phamily) number is provided based on the phamerator genome map data. Grey triangles represent identical phages (in each type) that are suspected of having derived by lytic cycle of a prophage within a dairy starter lysogen. ***** represents genes in the lysogeny module.

3.7.3 Identical phages through prophage induction from same strains

Based on the strong identity and synteny observed among the phages A31 (A) 5807(A) 5876(A), 5799 (B) 5812 (B) 5943 (C) 5953 (C) and 5731(E) 5910(E), we hypothesize that the source of these identical phages could be location dependent or independent.

(i) To start with we hypothesized that the similarity observed in the phages A31, 5876, 5807 (A), 5799, 5812 (B), 5943, 5953, 6023 (C) and 5731, 5910 (E) could be due to the region (or proximity between the regions) of their isolation. Hence investigation on their geographical regions of isolation (location) was carried out.

Based on our analysis, we report that the phages that belong to the same type were not from the same location. Also, it was observed that phages that were isolated from the same locations did not display a high level of synteny (Figure 21). These results allowed us to dismiss this hypothesis and suggested that the phages of the same type had a shared source that was location independent.

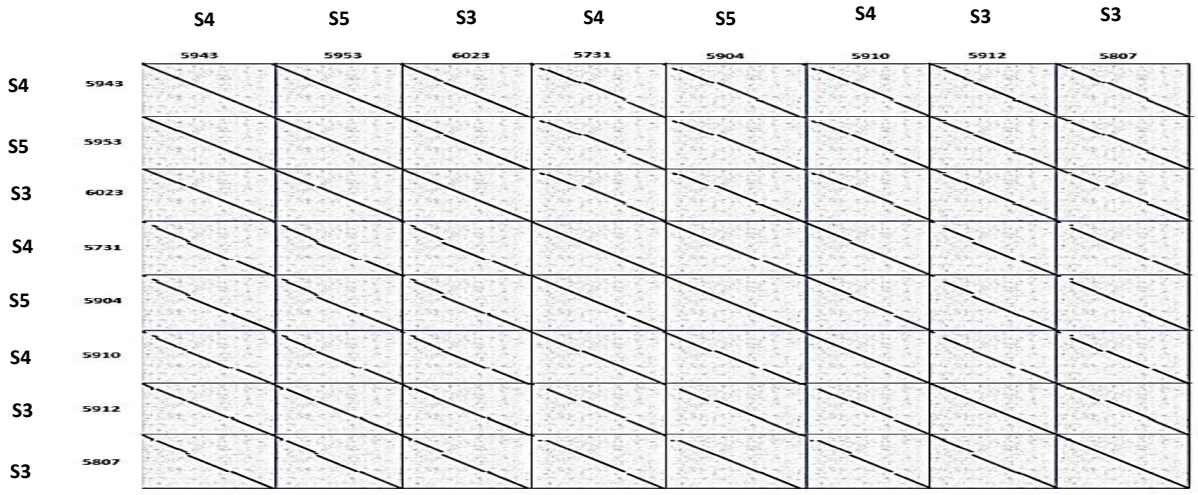


Figure 20 Gepard dot plot analysis displaying synteny among the P335 phages from different locations. Solid diagonal lines without breaks throughout the length of their genome observed among some phages that have been isolated from different locations(Sites S3, S4, S5) is indicative of strong identity among them.

(ii) In order to test our alternative hypothesis that the phages are location independent and to demonstrate that induction of prophages from the same strains that have been rotated in the different fermentation units has resulted in the detection of identical phages [5876 (A), A31 (A), 5807(A) and 5799, 5812(B) and 5943, 5953(C) and 5731, 5910(E)], we have taken the evidence from phamerator data.

Examination of these phage genomes with the help of the phamerator database revealed identity in their nucleotide and amino acid architecture. No insertion or deletion of a specific gene was observed. We also did not notice any recombination or horizontal gene transfer in these phages (Figure 22). Clustering of the specific phages (Type A, B, C, E) earlier in the phylogenetic tree further revealed their close evolutionary distance with each other. Based on these evidences. we have recognized that phages that belong to the same type are identical phages.

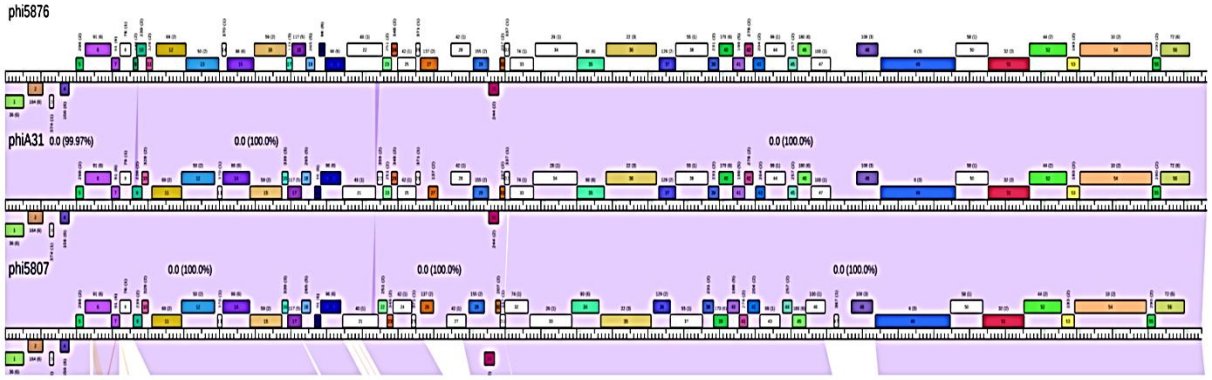


Figure 21 Illustration of phamerator data for representatives of identical phages (5876, A31, 5807). A strong nucleotide and amino acid identity can be observed between these identical phages. The nucleotide homology between the phage genomes is considered high if shown in purple and no homology if shown in white.

As previously reported, most of the starter cultures have one or more prophages in them. The conditions that would allow the induction of a prophage from the starter culture has always been

of importance and questioned in the dairy industries (Lunde et al., 2005). Induction of prophages from LAB has been found to occur under various environmental conditions like heat, starvation, temperature and antibiotics (Marco et al., 2012).

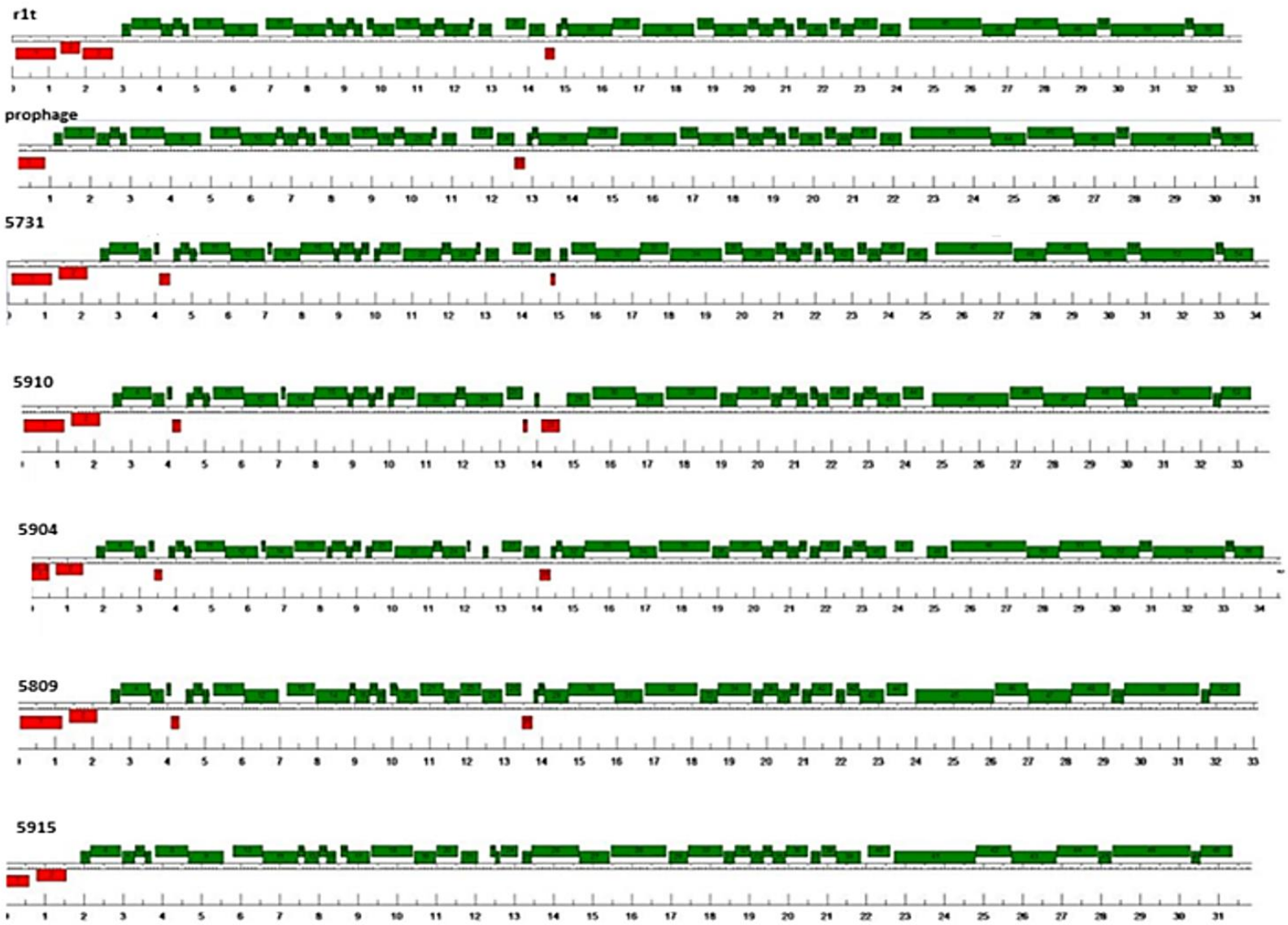
Hence we propose that the identical phages in each type are prophages that have been induced from the same starter strain that has been rotated in the different fermentation units.

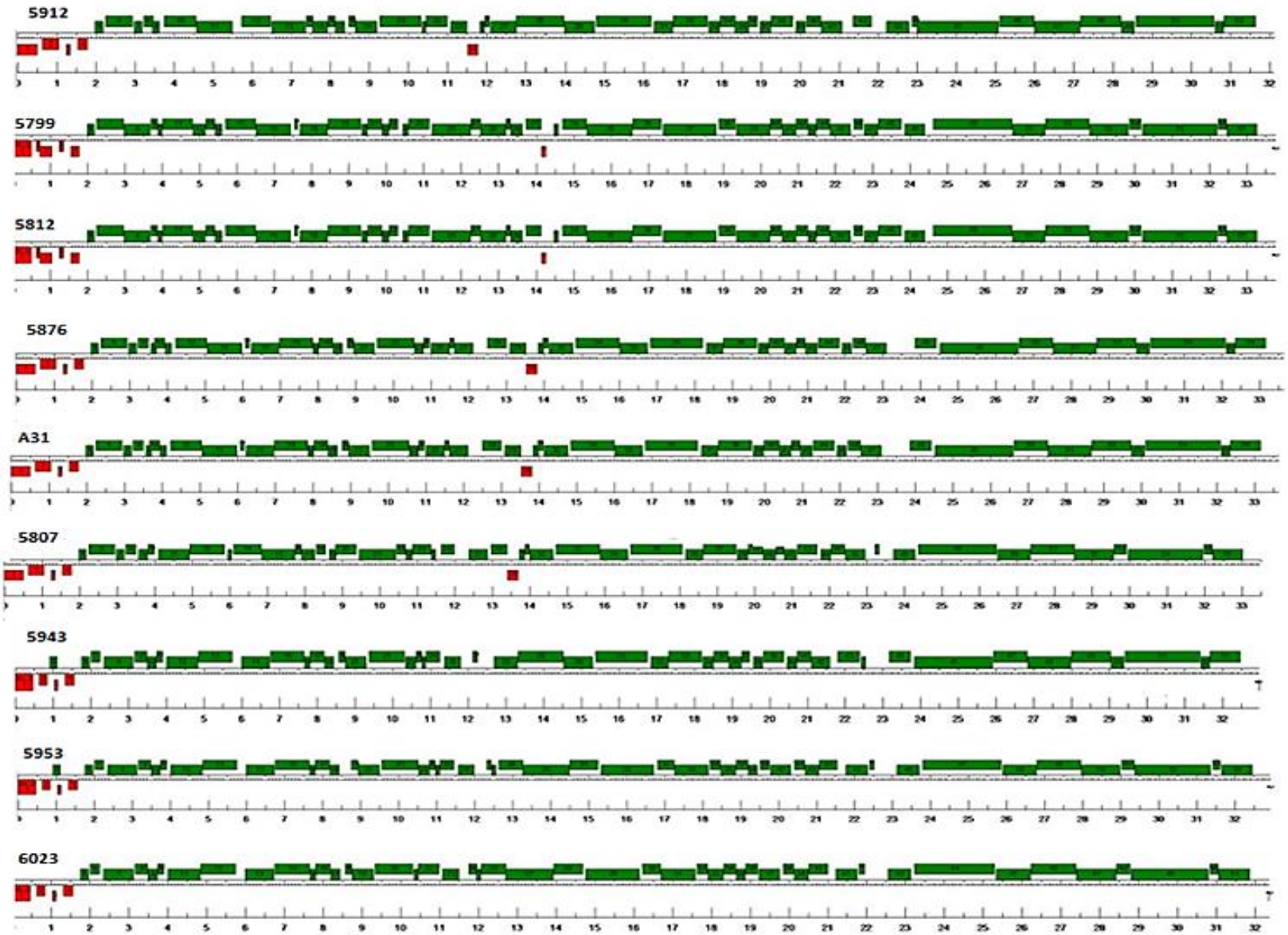
To better understand the ORFs (open reading frame) that are shared among these similar phages and to identify the genes that have been acquired from other phages, we decided to annotate selective phage gnomes from our study.

3.8 Functional Annotation of Phage Genomes

Open reading frames were identified in both the strands. Forty eight to fifty seven open reading frames were identified in all the phages. Putative genes and the appropriate start codons were identified using Genemark and Glimmer (Delcher et al., 1999) that are associated with DNAmaster. DNA master is a tool for locating specific genes that are required for further analysis. Information about a gene can also be obtained by this tool by clicking on the gene to obtain its size or its amino acid composition and looking into its function that is obtained by Blast search.

In order to identify the similarity or relatedness of the phages under study with other sequenced phages in the database, DNAmaster searches for proteins that are identical or similar to the ORFs using the tool Blast P. Blast results revealed high degree of similarity of several genes in our phage genomes to the phages R1T (Van Sinderen et al., 1996) and the phage Phi4 (Murphy et al., 2016) in the database. Representative phages 5943, 5912, 5799, A31 and 6090 and the prophage were selected and have been annotated. Finally our phamerator database for our phage genomes was customized with the help of DNAmaster. Figure 23 demonstrates the genome maps generated for each phage genome along with the genome maps of the reference phages used in this study.


r1t
prophage
5731
5910
5904
5809
5915


5912
5799
5812
5876
A31
5807
5943
5953
6023

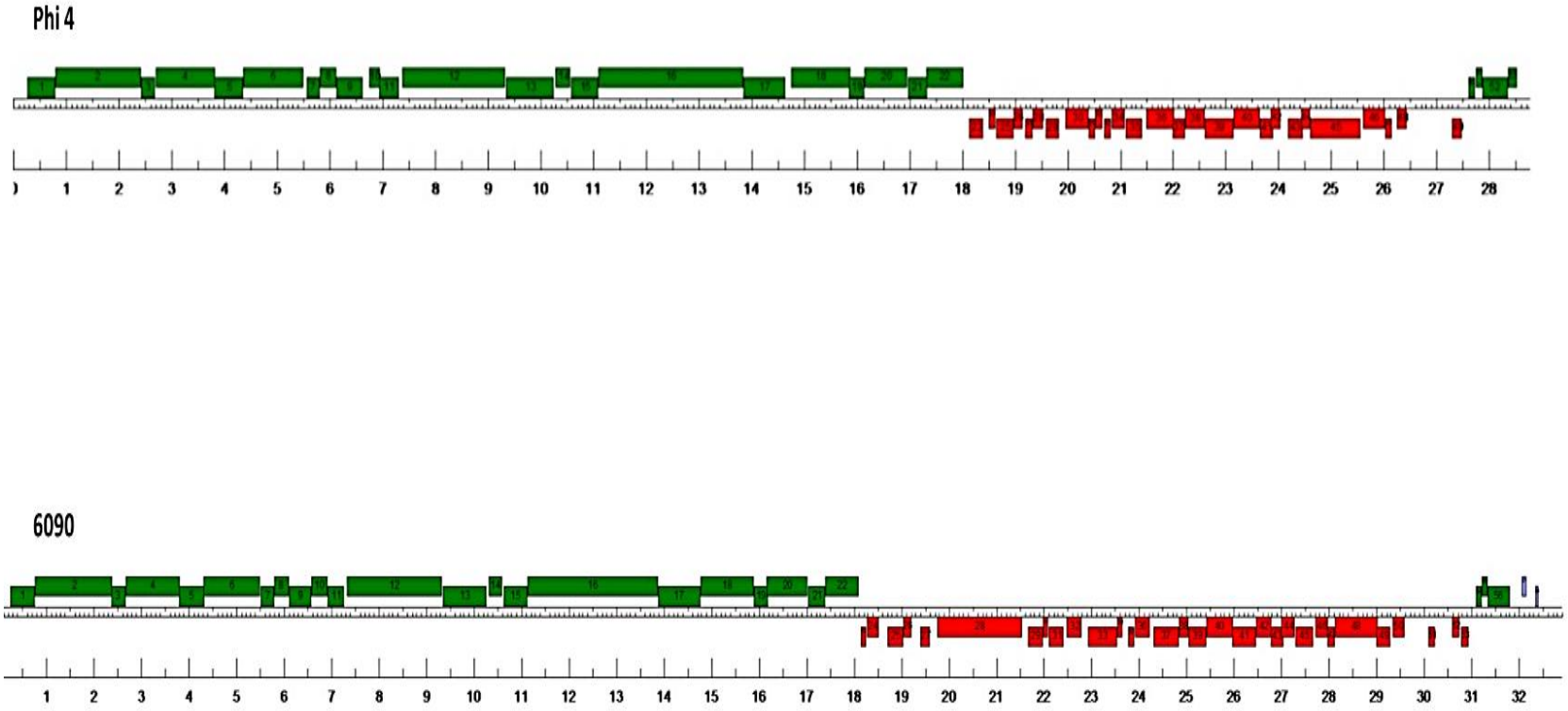


Figure 22 Schematic representations of annotated phage genomes in this project using DNAmaster along with their most closely related reference strains (Red boxes indicate genes that are transcribed in reverse orientation and green boxes indicate forward orientation).

3.9 Rapid Identification of phages

3.9.1 Identification of phage group using PCR

PCR is a simple and a quick method that has been commonly employed in different fermentation industries to experimentally identify phage groups. In order to detect the P335 and the 936 phage members in our study, a PCR method was devised by targeting certain conserved regions.

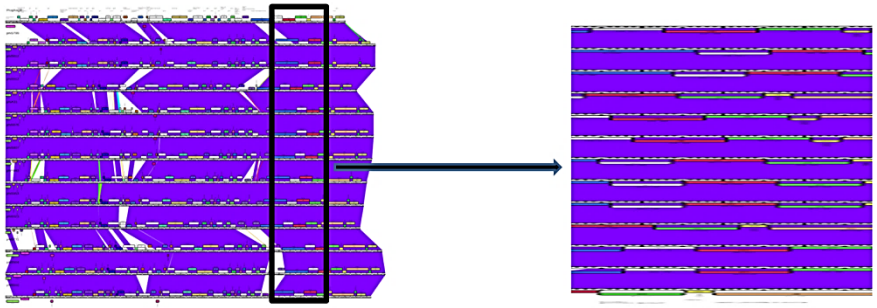
Analysis of the 15 phage genomes showed the presence of many conserved regions in both the groups. The region corresponding to Major tail protein, Tape measure protein and Major capsid protein was found to be conserved in all our phages that have shown strong identity to the P335 phage R1T. Specific primers for the major tail protein and major capsid protein unique to each cluster have been designed to identify phages that belong to these groups. The primer specifications are mentioned in Table 6. The specificity of the primers was verified using primer blast (Ye et al., 2012). Significant band differences could be observed between the two clusters hence we suggest that this approach would aid in the rapid identification of the phage species infecting the starter culture in the dairy industry. Colony PCR is a quick and easy method for determining the presence or absence of specific DNA sequences or gene of interest. In this method, the host bacterium is infected with the phage that has been isolated and incubated overnight. Phage infection leads to the formation of plaques that can be picked. These plaques can be either lysed in water following a brief heating step or added directly to the PCR mixture and lysed during the initial denaturation step. Heating during denaturation causes the release of DNA from the phage, which serves as a template for the amplification reaction. As PCR tests are sensitive and rapid, they can be used to investigate the type of phage that is isolated. An illustration of the primer specific to

the major tail protein of all the phages in this study pertaining to the P335 group has been provided (Figure 24)

Table 6 PCR primer specifications for Tape measure protein and Major tail protein of P335 and 936 group of phages under study

Primer name	Sequence	Primer length	Region amplified	Product size
MCP-936.F	TGAAAGCCGAAACGCTTTGAA	21	Major capsid protein	1167
MCP-936.R	CAACCAAAATCATGTTGC	18		
MTP-936.F	GCAGTAATCGGTTCAACA	18	Major Tail Protein	842
MTP-936.R	CCTTAATCACTGATGTTAC	19		
MCP-P335.F	GACGGAAAAGATGGAGCT	18	Major capsid protein	970
MCP-P335. R	CACGTCTTCCTTATTCAGC	19		
MTP-P335.F	TCAGCGCCAAAAGGTACAGC	20	Major Tail Protein	552
MTP-P335.R	TAGAGTTTTGCGTCCCCTACG	21		

A)



B)

```
prophage ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5799     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5812     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5912     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
A31      ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5876     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5807     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5943     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5953     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
6023     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5731     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5915     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5809     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5904     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA
5910     ATGGCACAAGTAGAAAATGTAACTACTGCAAAGCCCAAAATTGATGGTGCTATTTACTCAGCGCCAAAAGGTACAGCTTTACCAACTGA

prophage ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5799     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5812     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5912     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
A31      ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5876     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5807     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5943     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5953     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
6023     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5731     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5915     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5809     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5904     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
5910     ACTAATGACCCCGTAGTCGGTCAAGCCGCCGTAGGGGACGCAAAACTCTAAATATTAAATAAAGAAAGCGAGAAATATGTTAAAAGGA
```

Phi4

GCAGTAATCGGTTCAACAACGTACTGCATAAAAAAATAAAAAGGAAAATAAAAAATGAAATTAGATTATAACTCACGTGAG
ATTTTCTTTGGTAATGAAGCTCTAATCGTAGCTGATATGGCTAAGGGAAGCAACGGAAAACCAGAGTTCACTAACCATAAAA
TCGTAACTGGTTTAGTATCAGTTGGCGAAATGGAAGACCAAGCGGAGACAAACAGCTATCCTGCTGATGACGTGCCAGACC
ATGGAGTAAAAAAAGGTGCTACCTTGCTTCAAGGCGAAATGGTATTTATTCAAACAGACCAAGCACTTAAAGAAGATATTTT
AGGTCAACAAAGAACAGCGAATGGTTTGGGTTGGTCTCCTACTGGTAATTGGAAAACGAAATGTGTTCAGTACCTTATTAAA
GGTCGCAAGCGTGATAAAGTTACAGGAGAATTTATTGACGGTTACCGTGTAGTCGTTTATCCAAGTTTGAGACCAACAGCAG
AAGCTACAAAAGAATCAGAAACAGATTCAGTAGACGGTGTAGACCCTATTCAATGGACTCTGGCAGTACAAGCGACTGAAT
CAGATATTTATTTGAATGGCGATAAAAAAGTTCCTGCTATTGAGTATGAAATTTGGGGAGAACAAGCTAAAAGCTTTGCTAA
GAAAATGGAAGCAGGCTTATTCATCATGCAACCTGACACAGTTCTAGCTGACACATTTACACTTGTACCTCCTGTTATTCCTAA
TACGATTACTGCTAAACATGGAGGAAATGACGGAGCAATCATAGTACCTACCACTTTGAAAGACTCTAATGGTGAAACTGTA
AAAGTAACATCAGTGATTAAGGACGCACATGGAAAAGAAGCAACAAATGGGAAACTTGCGCCCGGTATCTATCTCGTAACG

6090

GCAGTAATCGGTTCAACAACGTACTGCATAAAAATATAAAAAGGAAAATAAAAAATGAAATTAGATTATAATTCACGTGAG
ATTTTCTTTGGTAATGAAGCTCTAATCGTAGCTGATATGGCCAAGGGAAGTAACGGAAAACCAGAGTTCACTAACCATAAAA
TCGTAACTGGTTTGGTATCAGTTAGCGAAATGGAAGACCAAGCGGAAACTAACAGTTATCCAGCTGATGACGTGCCAGACC
ATGGAGTTAAAAAAGGCGCTACCTTACTTCAAGGCGAAATGGTATTTATTCAAACAGACCAAGCGCTCAAAGAAGACATTTT
AGGTCAACAAAGAACAGCAAATGGTTTGGGTTGGTCTCCAACTGGTAATTGGAAAACGAAATGTGTTCAGTACCTTATTAAA
GGGCGCAAACGTGATAAAGTTACAGGAGAATTTATTGACGGTTACCGTGTAGTCGTTTATCCAAATTTGAGACCAACAGCAG
AAGCTACAAAAGAATCAGAAACAGATTCAGTAGACGGTGTAGACCCTATCCAATGGACTTTGGCAGTACAAGCGACTGATT
CAGATATTTATTTGAATGGAAATAAAAAAGTCCCTGCTATTGAGTACGAAATTTGGGGAGAACAAGCAAAAGATTTCGCCA
AAAAAATGGAAAGCGGACTGTTCATCATGCAACCTGATACAGTTCTAGCTGGTGCAATTACACTTGTAGCTCCTGTTATTCCT
AATGTAACTACTGCTACAAAGGGTAATAATGACGGAACAATCGTAGTGCCTGACACTTTGAAAGATTCTAATGGTGGAACTG
TAAAAGTAACATCAGTGATTAAGGACGCACATGGAAAAGCAGAAACAAATGGACACCTTGCGCCCGGTGTTCATATCGTAA

Figure 23 Illustration of PCR primer specificity showing the placement of primers for the gene coding for Major tail protein in P335 and 936 group members. A) Genome map generated by phamerator for the gene in P335 members along with a zoomed region showing the Major tail protein. B) DNA sequence alignment of the gene showing the placement of primers. (The primers were designed using primer blast. Green colour represents forward primer and red represents reverse primer).

3.9.2 In silico Restriction Digests

We also attempted to perform in silico restriction digests using the whole genome of the phages with the enzyme ClaI that gave identifiable bands in some phages during manual restriction digests in our study. This was mainly performed to identify bands that differentiated the P335 group from 936 group and to observe any patterns of similarity among the P335 group of phages. This method also helped us to observe bands that could not be visualized in manual restriction digests. There were at least three fragments sizes that were similar across the P335 members in this study (Figure 25). The single representative of the 936 group, phage 6090 that clustered with the phage Phi4 has just one restriction site identified by ClaI.

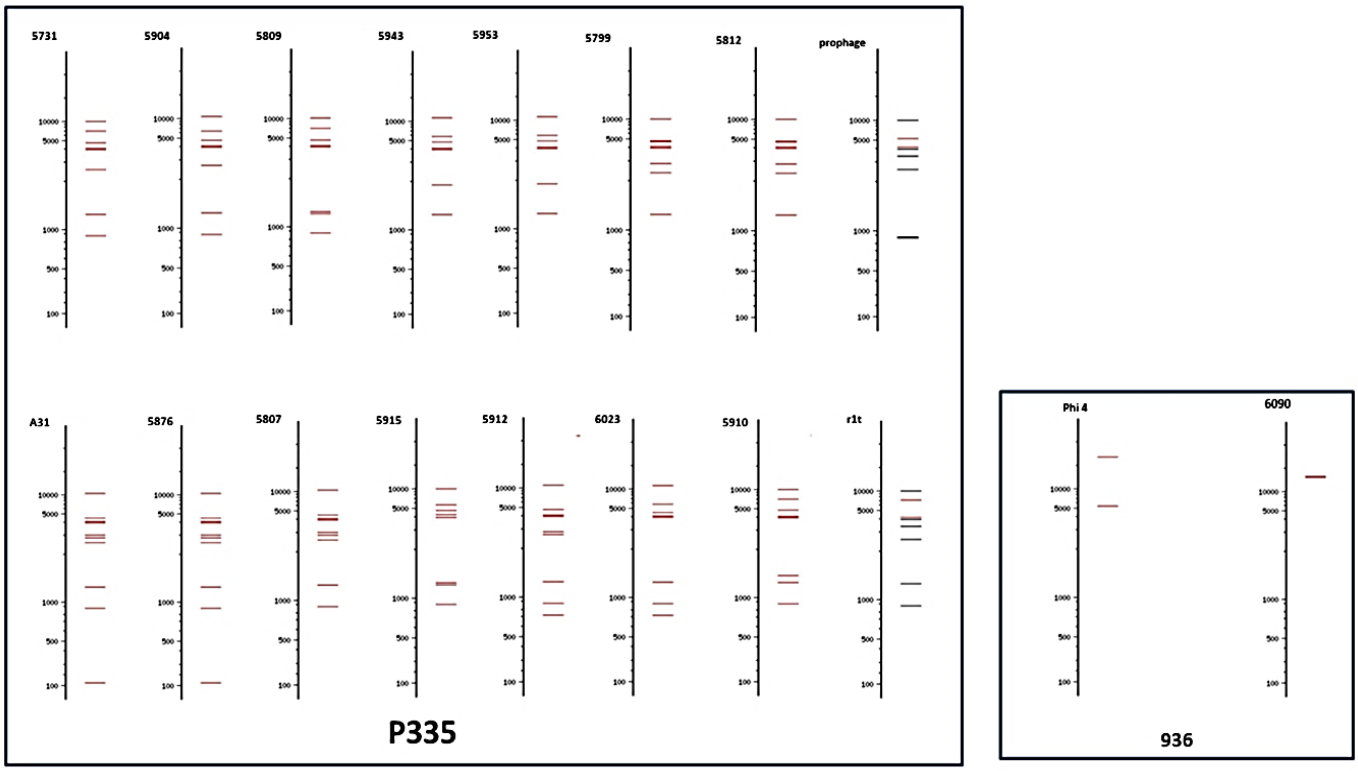


Figure 24 Depiction of Insilico Restriction endonuclease digestion of phage DNA using enzyme Cla 1. (Fragments obtained from NEB V2.0. DNA fragment patterns unique to each cluster are observed between the two phage clusters P335 and 936. R1T and Phi4 (reference phages) have been included as they emphasize the cluster specificity).

Chapter 4

Discussion

4.1 Major findings

Whole genome sequencing based approach was informative in not only revealing the relatedness between the phages but also helped in clustering the phages based on their evolutionary history. Two clusters or groups (P335 and 936) have been generated based on the fifteen genome sequences. A high degree of similarity is observed among the phages obtained from the different units of Fonterra that are categorized into the P335 cluster. Blast results show high degree of relatedness (at least 97%) of 14 phages to phage R1T (isolated from Netherlands). The P335 cluster (containing reference phage R1T found in Netherlands) is widespread in New Zealand and has been isolated in Fonterra plants from both North and south islands. The cluster 936 represents phage 6090 and this single representative was isolated in the north island (Site N3). Despite the prevalence of P335 members in our study, it has been widely accepted that the 936 group of phages are the major causal organisms of fermentation failures (Marco et al., 2012). Although a high degree of diversity may not be common in the 936 group of phages, a genetic divergence has been observed in this group due to horizontal transfer events (Murphy et al., 2016).

From our analysis, it is quite evident that P335 phages in this study have been found to emerge from the starter culture themselves. This is substantiated by doing a comparative analysis of these phages. Distribution of identical genes in the lysogeny module in certain phages that showed similarity and the similarity observed in their amino acid alignment demonstrates that prophages from other strains have been lysed and infected new lysogens.

However, the heterogeneity or the mosaicism observed among the P335 members is attributed to various factors.

4.2 Investigations into mosaicism in P335 phages

Previous research (Bouchard et al., 2000) and the results obtained from our study indicate that homologous recombination between the prophages in the starter strain and the incoming lytic phages plays a significant role contributing to a high degree of genetic diversity or mosaicism among the P335 group. Previous work on these phages involves division of this group into four sub groups owing to their mosaic nature (Deveau et al., 2006). As prophages have been shown to be an integral part of the host cell, the probability of a phage to undergo recombination with a prophage is quite high (Kelly et al., 2013). This kind of rearrangement can not only lead to the formation of new phages but can also lead to the emergence of new host strains (Labrie et al., 2007). The impact of prophages on their hosts has given a clear understanding that prophages have coevolved with their host genome. As a result both the prophage and the *Lactococcus lactis* strains are subjected to the same selection (Kelly et al., 2013).

Mosaicism in the P335 phages can be attributed to recombination that occurs not only between genomes that share homology but also between genomes that require little or no homology like horizontal gene transfer of novel genes. Their presence in just one or few phages may represent the loss of that gene in other phages or could indicate acquirement of the gene by horizontal gene transfer. For example, the presence of certain phams in only some phages in our study represents genes that participate highly in the evolutionary flux and have been transferred horizontally. A software tool like phamerator is extremely useful to study the mosaicism in phage genomes by analysing at the level of amino acids thereby showing any evidence of horizontal gene transfer.

Though novel genes are introduced from time to time through evolutionary pathways, they may also be homologous to the genes that already exist in their genome. These genes can subsequently undergo recombination and genetically be passed on or may be lost during selection (Andam et al., 2010). Genetic mosacism of the P335 phages has been discussed in several studies and it has been postulated that the flexibility in their genome is a provision to adapt to a new host (Botstein, 1980; Labrie et al., 2008).

The presence of multiple endonucleases in lactococcal phages like HNH endonuclease has been reported to play a significant role in phage DNA recombination (Crutz et al., 2002). In this study phages that belong to the P335 group consist of genes coding for HNH endonucleases commonly seen in other phages of this group. HNH/Homing endonucleases have also shown to drive horizontal gene transfer and promote genetic exchange of not just their genes but also genes that are localized around them (Belfort and Bonocora, 2014). Also, HGT augments a strong similarity between members that share a more common ancestry (Andam et al., 2010).

Other than horizontal gene transfer, factors like point mutations and gene disruption have also shown to play a big part that have led to the diversification of the P335 group of phages thus justifying the formation of subgroups or sub clusters (Labrie, 2008). Also, natural phage defense systems like the abortive infection mechanism have been influential in the evolution of lytic phages in the P335 group (Labrie et al., 2007).

Thus the genetic diversity common to lactococcal phages studied previously and phages in our study suggests that homologous recombination and horizontal gene transfer events have contributed to a significant portion of their mosaic architecture. Despite the mosaicism observed in the P335 group, in order to identify the presence of these phages, it has

been suggested in the past to target the dUTPase gene using PCR detection techniques as it is conserved in this entire species (Labrie and Moineau, 2002).

4.3 Phage detection techniques and primer recommendations

4.3.1 Detection and identification of phage groups through PCR

Basic detection and enumeration of phages in the dairy industries, like plaque assays, spot tests have not been very helpful in identifying the exact phage species. In the recent years however microbiological assays and PCR techniques have been able to provide a great source of information about the phages infecting the starter cultures (Mahony et al., 2012). As discussed above, the dUTPase gene was suggested to be used as a target for detection of these phages, however a lot of the P335 species phages including several phages in this study lack this gene. The C2 species could be easily identified by this technique by targeting the major capsid protein (Labrie and Monieu, 2000).

In this study, based on the gene conservation, we have selected the major capsid protein and major tail protein to distinguish between the 2 groups P335 and 936. PCR analysis has the benefit of simultaneously detecting the presence as well as the species to which the lactococcal phages belong directly from samples. The method however, is incapable of indicating the viability of the phages and if they are able to attack the starter culture. Thus, this method may be used to confirm the presence of phages in a positive sample and most significantly to rapidly identify the group of the phages that are currently infecting the starter strains. This method can also be useful in identifying any new species that infects the starter culture.

4.3.2 Insilico restriction digests and experimental suggestions

Our manual restriction digestion data has so far not yielded thorough results that would help us to identify the phages in each cluster yet we have provided in silico digests indicative of patterns that distinguish the two clusters or groups we observed in this study. This method is not only rapid but also helps to identify recurring phages. The reappearance of similar phages can be suggestive of their source and this can be useful to prevent further contamination. Therefore we suggest that restriction digestion is a quick method to differentiate between the clusters.

However, with the availability of whole genome sequences of these phages, it is now not only possible to know the diversity of the phages but also understand the phage host interaction, necessary for alleviating the ordeal faced by the dairy industry.

4.4 Whole genome sequencing and Contributions from this study

In dairy industries where commercial strains are used, the presence of prophages has always been considered a threat, as lysis of the cells leading to phage excision leading to infection can happen at any time.

Despite several measures being undertaken to eradicate phages from the fermentation units of the dairy industries, their repeated occurrence has been of great concern. Analysing lysogeny in bacteria gives useful information on the phage recombinants that play a major role in the process of infection inhibition. DNA homology studies in streptococci strains conducted using hybridization techniques in the past have provided a good source of information about genetic relatedness between the prophages in the host chromosome and the temperate phages that were induced from the host strain (Jarvis, 1984). However, whole genome sequencing which is more specific and rapid than the conventional techniques opens up a plethora of possibilities that can cope with the phage diversity as well as cater to the demands of the dairy

industry today. It allows precise characterisation of the phages and their virulence nature at the genetic level that can help in preventing fermentation outbreaks.

Conclusion

Comparative sequence analysis of phages obtained from Fonterra has provided information about the evolutionary history of the phages and has also revealed the predominance of P335 phages infecting a single strain in the fermentation units. This project has also helped us to identify 9 types of P335 phages in this study and an R1T type prophage.

After a thorough analysis we suggest that the striking similarity found among some phages in this group especially with respect to their lysogeny module reflects on the prophages that became lytic and further infecting new strains that exist as lysogens. This could suggest that many of the lactococcal phages exist as prophages in the dairy strains. Some of them are inducible and have been identified in this study. Homologous recombination between the lysogen and the infecting phage has contributed to the similarities observed among these phages. In order to identify if a phage isolated from the dairy industry shares a close relatedness to other similar phages that have been identified earlier, PCR can be a very helpful tool and the primers suggested in this project would be suggestive of the type and the probable source of phages that exist in the dairy industry.

Future Directions

In our research we have annotated a set of Fonterra phage genomes. Future directions for the project include:

1. More such lactococccal phage annotations should be carried out.
2. In the future, other strains that were used in the different fermentation units at the time of the phage isolation will be analysed in order to identify prophage(s) since these phages may have evolved from one or more prophages present in starter strains. This will also help in addressing the horizontal transfer events that are reported in this study.
3. The primers that we have suggested in this study can be tested with more phages and if required additional primers targeting other conserved regions can be designed for rapid identification of phages in the New Zealand Dairy Industry.
4. The whole genome sequence data gives us perspicacity about the relatedness among the phages in the P335 group in this study. Albeit a strong relatedness between the members of the P335 group in this study with the temperate phage R1T, typically for lysis to take place, the genes coding for the lysogeny module have to be repressed. In order to identify the lytic lifestyle of the P335 members in the genes coding for the replication module and the genes that are involved in the lytic/lysogenic switch can be investigated further
5. Despite several efforts taken in the dairy industry to eliminate these phages, only modest progress has been observed. More genetic screening is required to understand phage biology and phage multiplication in order to increase productivity and find better ways to benefit New Zealand's dairy industry.

References

AG Invest. (2013). *New Zealand Dairy industry*. Retrieved from www.aginvest.com/research-for-investing/nz-dairy-industry/

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research, 25*(17), 3389-3402.

Andam, C. P., Williams, D., & Gogarten, J. P. (2010). Biased gene transfer mimics patterns created through shared ancestry. *Proceedings of the National Academy of Sciences, 107*(23), 10679-10684. doi:10.1073/pnas.1001418107

Atamer, Z., Dietrich, J., Müller-Merbach, M., Neve, H., Heller, K. J., & Hinrichs, J. (2009). Screening for and characterization of Lactococcus lactis bacteriophages with high thermal resistance. *International Dairy Journal, 19*(4), 228-235. doi:http://dx.doi.org/10.1016/j.idairyj.2008.10.012

Atamer, Z., Samtlebe, M., Neve, H., Heller, K., & Hinrichs, J. (2013). Review: elimination of bacteriophages in whey and whey products. *Frontiers in Microbiology, 4*(191). doi:10.3389/fmicb.2013.00191

Barrick, J. E., & Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nature Reviews. Genetics, 14*(12), 827.

Bebeacua, C., Tremblay, D., Farenc, C., Chapot-Chartier, M. P., Sadovskaya, I., van Heel, M., . . . Cambillau, C. (2013). Structure, adsorption to host, and infection mechanism of virulent lactococcal phage p2. *Journal of Virology, 87*(22), 12302-12312. doi:10.1128/jvi.02033-13

Belfort, M., & Bonocora, R. P. (2014). Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Homing Endonucleases: Methods and Protocols*, 1-26.

Besemer, J., & Borodovsky, M. (2005). GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research, 33*, W451-W454. doi:10.1093/nar/gki487

Bissonnette, F., Labrie, S., Deveau, H., Lamoureux, M., & Moineau, S. (2000). Characterization of mesophilic mixed starter cultures used for the manufacture of aged cheddar cheese. *Journal of Dairy Science, 83*(4), 620-627

Blatny, J. M., Godager, L., Lunde, M., & Nes, I. F. (2004). Complete genome sequence of the Lactococcus lactis temperate phage φLC3: comparative analysis of φLC3 and its relatives in lactococci and streptococci. *Virology, 318*(1), 231-244.

Boni, M. F., de Jong, M. D., van Doorn, H. R., & Holmes, E. C. (2010). Guidelines for identifying homologous recombination events in influenza A virus. *PLoS ONE, 5*(5), e10434. doi:10.1371/journal.pone.0010434

Botstein, D. (1980). A theory of modular evolution for bacteriophages. *Annals of the New York Academy of Sciences, 354*. doi:10.1111/j.1749-6632.1980.tb27987.x

Bouchard, J. D., & Moineau, S. (2000). Homologous Recombination between a Lactococcal Bacteriophage and the Chromosome of Its Host Strain. *Virology, 270*(1), 65-75. doi:http://dx.doi.org/10.1006/viro.2000.0226

Bradley, D. E. (1967). Ultrastructure of bacteriophage and bacteriocins. *Bacteriological reviews, 31*(4), 230.

Braun Jr, V., Hertwig, S., Neve, H., Geis, A., & Teuber, M. (1989). Taxonomic differentiation of bacteriophages of Lactococcus lactis by electron microscopy, DNA-DNA hybridization, and protein profiles. *Microbiology, 135*(9), 2551-2560.

Brussow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews, 68*(3), 560-602. doi:10.1128/MMBR.68.3.560-602.2004

Brussow, H., & Desiere, F. (2001). Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. *Molecular Microbiology, 39*(2), 213-223. doi:10.1046/j.1365-2958.2001.02228.x

Bruttin, A., Desiere, F., Lucchini, S., Foley, S., & Brussow, H. (1997). Characterization of the lysogeny DNA module from the temperate Streptococcus thermophilus bacteriophage phi Sfi21. *Virology, 233*(1), 136-148. doi:10.1006/viro.1997.8603

Bylund, G. (1995). Dairy Processing Handbook: Tetra Pak Processing Systems AB S 221 86 Lund: Sweden.

Casjens, S., Hatfull, G., & Hendrix, R. (1992). *Evolution of dsDNA tailed-bacteriophage genomes.* Paper presented at the Seminars in Virology.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology, 49*(2), 277-300. doi:10.1046/j.1365-2958.2003.03580.x

Chandry, P. S., Moore, S. C., Boyce, J. D., Davidson, B. E., & Hillier, A. J. (1997). Analysis of the DNA sequence, gene expression, origin of replication and modular structure of the Lactococcus lactis lytic bacteriophage sk1. *Molecular Microbiology, 26*(1), 49-64.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D., & Chopin, M. (2001). Analysis of six prophages in Lactococcus lactis IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Research, 29*(3), 644-651.

Cresawn, S. G., Bogel, M., Day, N., Jacobs-Sera, D., Hendrix, R. W., & Hatfull, G. F. (2011). Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics, 12*, 395. doi:10.1186/1471-2105-12-395

Crutz-Le Coq, A.-M., Cesselin, B., Commissaire, J., & Anba, J. (2002). Sequence analysis of the lactococcal bacteriophage bIL170: insights into structural proteins and HNH endonucleases in dairy phages. *Microbiology, 148*(4), 985-1001.

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research, 14*(7), 1394-1403. doi:10.1101/gr.2289704

Davidson, B. E., Powell, I. B., & Hillier, A. J. (1990). Temperate bacteriophages and lysogeny in lactic acid bacteria. *FEMS Microbiology Reviews, 7*(1-2), 79-90. doi:10.1111/j.1574-6968.1990.tb04880.x

Davis, M. P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A. J. (2013). Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods, 63*(1), 41-49. doi:10.1016/j.ymeth.2013.06.027

Delbruck, M. (1940). THE GROWTH OF BACTERIOPHAGE AND LYSIS OF THE HOST. *The Journal of General Physiology, 23*(5), 643-660.

Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research, 27*(23), 4636-4641.

Derkx, P. M., Janzen, T., Sørensen, K. I., Christensen, J. E., Stuer-Lauridsen, B., & Johansen, E. (2014). The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. *Microbial Cell Factories, 13*(1), S5.

Desiere, F., Lucchini, F., and Brussow, H. (1999). Comparative sequence analysis of the DNA packaging, head and tail morphogenesis modules in the temperate cos-site Streptococcus thermophilus bacteriophage Sfi21. Virology 260, 244–253.

Desiere, F., Mahanivong, C., Hillier, A. J., Chandry, P. S., Davidson, B. E., & Brüssow, H. (2001). Comparative genomics of lactococcal phages: Insight from the complete genome sequence of lactococcus lactis phage BK5-T. *Virology, 283*(2), 240-252. doi:10.1006/viro.2001.0857

Deveau, H., Labrie, S. J., Chopin, M.-C., & Moineau, S. (2006). Biodiversity and Classification of Lactococcal Phages. *Applied and Environmental Microbiology, 72*(6), 4338-4346. doi:10.1128/AEM.02517-05

Donnelly-Wu, M. K., Jacobs, W. R., & Hatfull, G. F. (1993). Superinfection immunity of mycobacteriophage L5: applications for genetic transformation of mycobacteria. *Molecular Microbiology, 7*(3), 407-417.

Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C., & Fineran, P. C. (2014). A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. *Nucleic Acids Research*. doi:10.1093/nar/gkt1419

Eraclio, G., Fortina, M. G., Labrie, S. J., Tremblay, D. M., & Moineau, S. (2017). Characterization of prophages of Lactococcus garvieae. *Scientific Reports, 7*(1), 1856. doi:10.1038/s41598-017-02038-y

Fernández, E., Alegría, Á., Delgado, S., Martín, M. C., & Mayo, B. (2011). Comparative Phenotypic and Molecular Genetic Profiling of Wild Lactococcus lactis subsp. lactis Strains of the L. lactis subsp. lactis and L. lactis subsp. cremoris Genotypes, Isolated from Starter-Free Cheeses Made of Raw Milk. *Applied and Environmental Microbiology, 77*(15), 5324-5335. doi:10.1128/AEM.02991-10

Feirtag, J. M., & McKay, L. L. (1987). Thermoinducible Lysis of Temperature-Sensitive Streptococcus cremoris Strains1. *Journal of Dairy Science, 70*(9), 1779-1784.

Fonterra (2014). *Our heritage*. Retrieved from http://galaxy.fonterra.com/nz/en/About/Our+History

Fonterra (2014). *The New Zealand dairy industry*. Retrieved from http://galaxy.fonterra.com/nz/en/About/Research+and+Innovation

Fonterra (2016). *Global dairy trade*. Retrieved from https://www.globaldairytrade.info/en/gdt-events/gdt-events-sellers-and-products/fonterra/

Fonterra (2016). Our markets. Retrieved from https://www.fonterra.com/nz/en/our-markets.html

Fokine, A., & Rossmann, M. G. (2014). Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage, 4*(2), e28281.

Forde, A., & Fitzgerald, G. F. (1999). Bacteriophage defence systems in lactic acid bacteria. *Antonie Van Leeuwenhoek, 76*(1-4), 89-113.

Freifelder, D., & Kirschner, I. (1971). The formation of homoimmune double lysogens of phage λ and the segregation of single lysogens from them. *Virology, 44*(3), 633-637. doi:http://dx.doi.org/10.1016/0042-6822(71)90378-3

Garneau, J. E., & Moineau, S. (2011). Bacteriophages of lactic acid bacteria and their impact on milk fermentations. *Microbial Cell Factories, 10*(Suppl 1), S20-S20. doi:10.1186/1475-2859-10-S1-S20

Gasson, M. J., & Davies, F. L. (1980). Prophage-cured derivatives of Streptococcus lactis and Streptococcus cremoris. *Applied and Environmental Microbiology, 40*(5), 964-966.

Groth, A. C., Olivares, E. C., Thyagarajan, B., & Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences, 97*(11), 5995-6000. doi:10.1073/pnas.090527097

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology, 59*(3), 307-321. doi:10.1093/sysbio/syq010

Hardies, S. C., Thomas, J. A., & Serwer, P. (2007). Comparative genomics of Bacillus thuringiensis phage 0305φ8-36: defining patterns of descent in a novel ancient phage lineage. *Virology journal, 4*(1), 97.

Hatfull, G. F., Cresawn, S. G., & Hendrix, R. W. (2008). Comparative genomics of the mycobacteriophages: insights into bacteriophage evolution. *Research in Microbiology, 159*. doi:10.1016/j.resmic.2008.04.008

Hatfull, G. F., Jacobs-Sera, D., Lawrence, J. G., Pope, W. H., Russell, D. A., & Ko, C. C. (2010). Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. *Journal of Molecular Biology, 397*. doi:10.1016/j.jmb.2010.01.011

Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America, 96*(5), 2192-2197.

Hoskisson, P. A., & Smith, M. C. M. (2007). Hypervariation and phase variation in the bacteriophage 'resistome'. *Current Opinion in Microbiology, 10*(4), 396-400. doi:http://dx.doi.org/10.1016/j.mib.2007.04.003

Huggins, A., & Sandine, W. (1977). Incidence and properties of temperate bacteriophages induced from lactic streptococci. *Applied and Environmental Microbiology, 33*(1), 184-191.

Jarvis, A. W. (1984). Differentiation of lactic streptococcal phages into phage species by DNA-DNA homology. *Applied and Environmental Microbiology, 47*(2), 343-349.

Jarvis, A. W. (1995). Relationships by DNA homology between lactococcal phages 7–9, P335 and New Zealand industrial lactococcal phages. *International Dairy Journal, 5*(4), 355-366. doi:http://dx.doi.org/10.1016/0958-6946(94)00013-F

Jarvis, A. W., Fitzgerald, G. F., Mata, M., Mercenier, A., Neve, H., Powell, I. B., . . . Teuber, M. (1991). Species and type phages of lactococcal bacteriophages. *Intervirology, 32*(1), 2-9.

Kelly, W., Altermann, E., Lambie, S., & Leahy, S. (2013). Interaction between the genomes of Lactococcus lactis and phages of the P335 species. *Frontiers in Microbiology, 4*(257). doi:10.3389/fmicb.2013.00257

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution, 30*(4), 772-780. doi:10.1093/molbev/mst010

Kenneth, T. (2012). *Todar's online textbook of bacteriology*. Madison, Wisconsin.

Krumsiek, J., Arnold, R., & Rattei, T. (2007). Gepard. *Bioinformatics, 23*(8), 1026-1028.

Labrie, S. J., Josephsen, J., Neve, H., Vogensen, F. K., & Moineau, S. (2008). Morphology, genome sequence, and structural proteome of type phage P335 from Lactococcus lactis. *Applied Environmental Microbiology, 74*(15), 4636-4644. doi:10.1128/AEM.00118-08

Labrie, S., & Moineau, S. (2000). Multiplex PCR for detection and identification of lactococcal bacteriophages. *Appl Environ Microbiol, 66*(3), 987-994.

Labrie, S., & Moineau, S. (2002). Complete genomic sequence of bacteriophage ul36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. *Virology, 296*(2), 308-320. doi:10.1006/viro.2002.1401

Labrie, S. J., & Moineau, S. (2007). Abortive infection mechanisms and prophage sequences significantly influence the genetic makeup of emerging lytic lactococcal phages. *Journal of Bacteriology, 189*(4), 1482-1487.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948. doi:10.1093/bioinformatics/btm404

Lawrence, R., Thomas, T., & Terzaghi, B. (1976). Cheese starters. *Journal of Dairy Research, 43*(1), 141-193.

Lubbers, M. W., Waterfield, N. R., Beresford, T., Le Page, R., & Jarvis, A. W. (1995). Sequencing and analysis of the prolate-headed lactococcal bacteriophage c2 genome and identification of the structural genes. *Applied and Environmental Microbiology, 61*(12), 4348-4356.

Lunde, M., Aastveit, A. H., Blatny, J. M., & Nes, I. F. (2005). Effects of diverse environmental conditions on {phi}LC3 prophage stability in Lactococcus lactis. *Appl Environ Microbiol, 71*(2), 721-727. doi:10.1128/aem.71.2.721-727.2005

Madera, C., García, P., Janzen, T., Rodriguez, A., & Suárez, J. E. (2003). Characterisation of technologically proficient wild Lactococcus lactis strains resistant to phage infection. *International Journal of Food Microbiology, 86*(3), 213-222. doi:10.1016/S0168-1605(03)00042-4

Madera, C., Monjardín, C., & Suárez, J. E. (2004). Milk contamination and resistance to processing conditions determine the fate of Lactococcus lactis bacteriophages in dairies. *Applied and Environmental Microbiology, 70*(12), 7365-7371.

Marco, M. B., Moineau, S., & Quiberoni, A. (2012). Bacteriophages and dairy fermentations. *Bacteriophage, 2*(3), 149-158. doi:10.4161/bact.21868

Mahony, J., Murphy, J., & van Sinderen, D. (2012). Lactococcal 936-type phages and dairy fermentation problems: from detection to evolution and prevention. *Front Microbiol, 3*, 335. doi:10.3389/fmicb.2012.00335

Mahony, J., Cambillau, C., & van Sinderen, D. (2017). Host recognition by lactic acid bacterial phages. *FEMS Microbiology Reviews*. doi:10.1093/femsre/fux019

Mc Grath, S., Neve, H., Seegers, J. F., Eijlander, R., Vegge, C. S., Brondsted, L., . . . van Sinderen, D. (2006). Anatomy of a lactococcal phage tail. *Journal of Bacteriology, 188*(11), 3972-3982.

Mercanti, D. J., Carminati, D., Reinheimer, J. A., & Quiberoni, A. (2011). Widely distributed lysogeny in probiotic lactobacilli represents a potentially high risk for the fermentative dairy industry. *International Journal of Food Microbiology, 144*(3), 503-510.

Mikkonen, M., Dupont, L., Alatossava, T., & Ritzenthaler, P. (1996). Defective site-specific integration elements are present in the genome of virulent bacteriophage LL-H of Lactobacillus delbrueckii. *Applied and Environmental Microbiology, 62*(5), 1847-1851.

Moineau, S. (1999). Applications of phage resistance in lactic acid bacteria. In W. N. Konings, O. P. Kuipers, & J. H. J. H. In 't Veld (Eds.), *Lactic Acid Bacteria: Genetics, Metabolism and Applications: Proceedings of the Sixth Symposium on lactic acid bacteria: genetics, metabolism and applications, 19–23 September 1999, Veldhoven, The Netherlands* (pp. 377-382). Dordrecht: Springer Netherlands.

Murphy, J., Bottacini, F., Mahony, J., Kelleher, P., Neve, H., Zomer, A., . . . van Sinderen, D. (2016). Comparative genomics and functional analysis of the 936 group of lactococcal Siphoviridae phages. *Scientifc Reports, 6*, 21345. doi:10.1038/srep21345

New Zealand Trade and Enterprise. (n.d). *Export market*. Retrieved from https://www.nzte.govt.nz/en/export/export-markets/

New Zealand institute of chemistry (2008). *Manufacture and use of cheese*. Retrieved from http://nzic.org.nz/ChemProcesses/dairy/

New Zealand institute of chemistry (n.d). *Manufacture of Ethanol from Whey.* Retrieved http://nzic.org.nz/ChemProcesses/dairy/

Ochman, H., Lawrence, J. G., & Grolsman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature, 405*(6784), 299-304. doi:10.1038/35012500

Oliveira, J., Mahony, J., Lugli, G. A., Hanemaaijer, L., Kouwen, T., Ventura, M., & van Sinderen, D. (2016). Genome Sequences of Eight Prophages Isolated from Lactococcus lactis Dairy Strains. *Genome Announc, 4*(6). doi:10.1128/genomeA.00906-16

Oliveira, J., Mahony, J., Hanemaaijer, L., Kouwen, T. R. H. M., Neve, H., MacSharry, J., & van Sinderen, D. (2017). Detecting Lactococcus lactis Prophages by Mitomycin C-Mediated Induction Coupled to Flow Cytometry Analysis. *Frontiers in Microbiology, 8*(1343). doi:10.3389/fmicb.2017.01343

O'Sullivan, D., Ross, R. P., Fitzgerald, G. F., & Coffey, A. (2000). Investigation of the Relationship between Lysogeny and Lysis of Lactococcus lactis in Cheese Using Prophage-Targeted PCR. *Applied and Environmental Microbiology, 66*(5), 2192-2198.

Parreira, R., Valyasevi, R., Lerayer, A., Ehrlich, S. D., & Chopin, M.-C. (1996). Gene organization and transcription of a late-expressed region of a Lactococcus lactis phage. *Journal of Bacteriology, 178*(21), 6158-6165.

Pedersen, M., Ostergaard, S., Bresciani, J., & Vogensen, F. K. (2000). Mutational Analysis of Two Structural Genes of the Temperate Lactococcal Bacteriophage TP901-1 Involved in Tail Length Determination and Baseplate Assembly. *Virology, 276*(2), 315-328. doi:http://dx.doi.org/10.1006/viro.2000.0497

Pope, W. H., Bowman, C. A., Russell, D. A., Jacobs-Sera, D., Asai, D. J., Cresawn, S. G., . . . Hatfull, G. F. (2015). Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife, 4*, e06416.

Quiles-Puchalt, N., Tormo-Más, M. Á., Campoy, S., Toledo-Arana, A., Monedero, V., Lasa, Í., . . . Penades, J. R. (2013). A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. *Nucleic Acids Research, 41*(15), 7260-7275.

Rakonjac, J., O'Toole, P. W., & Lubbers, M. (2005). Isolation of lactococcal prolate phage-phage recombinants by an enrichment strategy reveals two novel host range determinants. *J Bacteriol, 187*(9), 3110-3121. doi:10.1128/jb.187.9.3110-3121.2005

Riipinen, K. A., Raisanen, L., & Alatossava, T. (2007). Integration of the group c phage JCL1032 of Lactobacillus delbrueckii subsp. lactis and complex phage resistance of the host. *Journal of Applied Microbiology, 103*(6), 2465-2475.

Rokyta, D. R., Burch, C. L., Caudle, S. B., & Wichman, H. A. (2006). Horizontal gene transfer and the evolution of microvirid coliphage genomes. *J Bacteriol, 188*(3), 1134-1142. doi:10.1128/jb.188.3.1134-1142.2006

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), 406-425.

Samson, J. E., Magadan, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Micro, 11*(10), 675-687. doi:10.1038/nrmicro3096

Samson, J. E., & Moineau, S. (2010). Characterization of Lactococcus lactis Phage 949 and Comparison with Other Lactococcal Phages. *Applied and Environmental Microbiology, 76*(20), 6843-6852. doi:10.1128/AEM.00796-10

Sandine, W. E. (1988). New nomenclature of the non-rod-shaped lactic acid bacteria. *Biochimie, 70*(4), 519-521. doi:http://dx.doi.org/10.1016/0300-9084(88)90087-9

Santos, C. L., Tavares, F., Thioulouse, J., & Normand, P. (2009). A phylogenomic analysis of bacterial helix-turn-helix transcription factors. *FEMS Microbiology Reviews, 33*(2), 411-429. doi:10.1111/j.1574-6976.2008.00154.x

Schouler, C., Gautier, M., Ehrlich, S. D., & Chopin, M.-C. (1998). Combinational variation of restriction modification specificities in Lactococcus lactis. *Molecular Microbiology, 28*(1), 169-178. doi:10.1046/j.1365-2958.1998.00787.x

Sechaud, L., Cluzel, P.-J., Rousseau, M., Baumgartner, A., & Accolas, J.-P. (1988). Bacteriophages of lactobacilli. *Biochimie, 70*(3), 401-410. doi:http://dx.doi.org/10.1016/0300-9084(88)90214-3

Seed, K. D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLoS Pathog, 11*(6), e1004847. doi:10.1371/journal.ppat.1004847

Shabbir, M. A., Hao, H., Shabbir, M. Z., Wu, Q., Sattar, A., & Yuan, Z. (2016). Bacteria vs. Bacteriophages: Parallel Evolution of Immune Arsenals. *Front Microbiol, 7*, 1292. doi:10.3389/fmicb.2016.01292

Spinelli, S., Veesler, D., Bebeacua, C., & Cambillau, C. (2014). Structures and host-adhesion mechanisms of lactococcal siphophages. *Frontiers in Microbiology, 5*(3). doi:10.3389/fmicb.2014.0000

Stern, A., & Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. *Bioessays, 33*(1), 43-51. doi:10.1002/bies.201000071

Stewart, C. R., Casjens, S. R., Cresawn, S. G., Houtz, J. M., Smith, A. L., Ford, M. E., . . . Pedulla, M. L. (2009). The genome of Bacillus subtilis bacteriophage SPO1. *Journal of Molecular Biology, 388*(1), 48-70. doi:10.1016/j.jmb.2009.03.009

Szczepankowska, A. (2012). Role of CRISPR/cas system in the development of bacteriophage resistance. *Advances in Virus Research, 82*, 289-338. doi:10.1016/b978-0-12-394621-8.00011-x

Todar, K. (2009). Todar's Online Textbook of Bacteriology. 2008. *Kenneth Todar University of Wisconsin-Madison Dept. of Bacteriology*.

Van Sinderen, D., Karsens, H., Kok, J., Terpstra, P., Ruiters, M. H. J., Venema, G., & Nauta, A. (1996). Sequence analysis and molecular characterization of the temperate lactococcal bacteriophage R1T. *Molecular Microbiology, 19*(6), 1343-1355. doi:10.1111/j.1365-2958.1996.tb02478.x

Vasu, K., & Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and Molecular Biology Reviews, 77*(1), 53-72. doi:10.1128/mmbr.00044-12

Verreault, D., Gendron, L., Rousseau, G. M., Veillette, M., Masse, D., Lindsley, W. G., . . . Duchaine, C. (2011). Detection of airborne lactococcal bacteriophages in cheese manufacturing plants. *Appl Environ Microbiol, 77*(2), 491-497. doi:10.1128/aem.01391-10

Vincze, T., Posfai, J., & Roberts, R. J. (2003). NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Research, 31*(13), 3688-3691.

Wang, W.-B., Jiang, T., & Gardner, S. (2013). Detection of Homologous Recombination Events in Bacterial Genomes. *PLoS ONE, 8*(10), e75230. doi:10.1371/journal.pone.0075230

Whitehead, H. R., Cox, G. A., Dairy Research, I., New, Z., Department of, S., & Industrial, R. (1935). *The occurrence of bacteriophage in cultures of lactic streptococci : a preliminary note*. Wellington, N.Z.: Govt. Printer.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13*, 134. doi:10.1186/1471-2105-13-134

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research, 18*(5), 821-829. doi:10.1101/gr.074492.107

Appendix

Supplementary Data File

Description:

The information present in the CD consists of GenBank files for six phage genomes that have been annotated for this project. The phages include 5912, 5799, 5943, A31, 6090 and the prophage in the strain 2356.

Filename:

Genbank files of Fonterra phages.mp4