Bacteriophage and Tail associated lysins of enterococci

By:

Alhassan Mansour Alrafaie

A thesis submitted for the degree of Doctor of Philosophy

March 2023

The University of Sheffield

Faculty of Medicine, Dentistry and Health

School of Clinical Dentistry

Contents

Acknowledgement

Firstly, my PhD journey in the UK would not be possible without funding from my country, Saudi Arabia, via prince Sattam bin Abdulaziz university which I am really grateful for. I would also like to thank my supervisors Prof. Graham Stafford and Dr Joey Shepherd for their constant support and guidance throughout the four years. I really enjoyed the supervisory meetings as I every time learn new things and feel motivated to continue my work. My supervisors were also so considerate and helpful during the Covid-19 pandemic which is something I really appreciate even though labs were closed for months. Special thanks also go to the microbiology lab technician Mr Jason Heath and molecular biology lab technician Ms Brenka McCabe who were so willing to help and support every time I knock on their door. The other thing that I feel fortunate about is having such amazing colleagues in the lab or in the office. Our big office has allowed us to communicate easily with each other and seek help quickly.

Being abroad is known to be difficult and studying PhD is not a simple task. During my study, I lived in Sheffield with my family who made my journey much easier. My wife, Norah, has been very supportive from the first day and her kind and encouraging words have motivated me a lot. I am also so thankful for her patience and for taking care of our children. My beloved three daughters (Lolowah, Sarah and Amal), seeing you after a long day in the lab is what makes me happy and keep me motivated. My parents back in Saudi, you have worked hard to raise me and my brothers and there is nothing we can do to pay this back. Thank you for your calls and prayers.

Abstract

Enterococcal species especially *E. faecalis* and *E. faecium* are among the most commonly isolated pathogens that can resist several classes of antibiotics. To mitigate the issue of AMR, bacterial viruses (phages) can be used to infect and kill these pathogens. The first step in this process involves adsorption to specific host receptors followed by genome ejection into the host. To achieve this step, specific phage proteins named Tail-associated lysins (TAL) facilitate phage infection to their hosts via locally degrading bacterial layers. Therefore, we first aimed to bioinformatically investigate the TALs in 506 enterococcal phage and prophage genomes. Prophage genomes were identified using the PHASTER web server. The identification of TALs was carried out using Pfam, NCBI domain database and PHYRE2. Using these tools, we identified various TALS: Endopeptidase, lytic transglycosylase, Pectinesterase, New Lipoprotein C/Protein of 60-kDa (NLPC/P60) and Glycerophosphodiester phosphodiesterase (GPDP). The most common TAL identified in both phage and prophage genomes was endopeptidase followed by lytic transglycosylase. The identified TALs have different targets with endopeptidase and lytic transglycosylase targeting the bacterial peptidoglycan structure, GPDP degrading teichoic acids and pectinesterase possibly attacking enterococcal polysaccharide antigen (EPA). Interestingly, the identified NLPC/P60 proteins were correlated to the phage genomic classifications and host specificity.

Phages are found almost everywhere and they have been isolated from several sources such as wastewater. Therefore, we collected and processed wastewater samples in our lab for phage isolation. To assess these samples, TEM was used to visualise various phage morphologies. Using these samples, 8 phages were isolated that *E. faecalis* and *E. faecium* strains in enrichment experiments, including clinical strains isolated from patients with diabetic foot ulcers. Following isolation, phage morphology was assessed via TEM with sipho-, myo- and podovirus phages observed. Since phages are very specific in terms of infecting bacteria, we have tested the ability of the isolated phages to infect 36 *E. faecalis* and *E. faecium* strains. A genomic analysis was then performed which revealed 5 novel strictly lytic phages. Since phiSHEF13 showed the broadest host range, we aimed to further characterise this phage. An experiment to assess phage host receptors was carried out which revealed that the variable region (epaV) of the *E. faecalis*

strain V583 is essential for phiSHEF13 successful infection. As three different phages were isolated using the *E. faecium* E1071 strains, killing assays were conducted which revealed quick bacterial inhibition at different MOI. Phage-resistant mutants were also observed and investigated.

The last chapter in this thesis is about assessing TALs *In vitro* by cloning and expressing candidate proteins. First, we selected one example from each identified lysin (a total of five proteins) and the candidate proteins were subcloned into vectors possessing His or GST tags and expressed using BL21(DE3) or C41(DE3). After expression, two TAL proteins namely pectinesterase and NLPC/P60 were successfully expressed and purified.

To conclude, the isolation of novel lytic phages would facilitate phage therapy targeting antibiotic-resistant enterococci. Moreover, the identification of various lytic proteins would also provide alternative use of phage to tackle antibiotic-resistant bacteria.

Abbreviations

®	**Trade mark**
α	Alpha
β	Beta
γ	Gamma
µ	Micro
Ap	Ampicillin
ATCC	American Type Culture Centre
bp	Base pairs
BHI	Brain heart infusion
BSA	Bovine serum albumin
°C	Centigrade
CFU	Colony forming unit
CO2	Carbon dioxide
DA	Domain architecture
DNA	Deoxyribonucleic acid
dsDNA	Double stranded deoxyribonucleotide
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra-acetic acid
EPA	Enterococcal polysaccharide antigen
GDPD	Glycerophosphodiester phosphodiesterase
GH	glycosyl hydrolase
GST	Glutathione S-transferase
His-tag	Histidine-tag
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kbp	kilo base pairs
kDa	kilo Daltons
L	litre
LB	Luria-Bertani
LPS	Lipopolysaccharide
LT	Lytic transglycosylase
µg	Microgram
µl	Microlitre

ml	Millilitre
min	minute
mM	Mmillimolar
MOI	Multiplicity of Infection
M	Molar
MW	Molecular weight
Ni-NTA	Nickel-nitriloacetic acid
NLPC/P60	New Lipoprotein C/Protein of 60-kDa
nm	Nanometre
OD	Optical density
PBS	Phosphate buffered saline
PEG	Polyethene glycol
PCR	Polymerase chain reaction
%	Percentage
PFU	Plaque forming units
phi	Phage
psi	Pounds per square inch (pressure)
rpm	Revolutions per minute
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
TAE	Tris-acetate-EDTA
TAEP	tail proteins associated with endopeptidase activity
TAL	Tail associated lysins
TEM	Transmission electron microscope
TEMED	Tetramethylethylenediamine
TMP	Tape measure protein
tRNA	Transfer RNA
UV	Ultraviolet light
V	Volt
VAPGH	Virion-associated peptidoglycan hydrolases

Publications and presentations

Publications:

1. Al-Zubidi M, Widziolek M, Court EK, Gains AF, Smith RE, Ansbro K, **Alrafaie A**, Evans C, Murdoch C, Mesnage S, Douglas CWI, Rawlinson A, Stafford GP. 2019. Identification of novel bacteriophages with therapeutic potential that target Enterococcus faecalis. Infect Immun 87:e00512-19. https://doi.org/10 .1128/IAI.00512-19.

2. **Alrafaie, A**., & Stafford, G. P. (2022). Enterococcal Bacteriophage: A Survey of the Tail Associated Lysin Landscape. SSRN Electronic Journal, 327(October 2022), 199073. https://doi.org/10.2139/ssrn.4251407

Presentations:

1. Poster presentation: **Alrafaie, A**., Al-Zubidi, M., Turner, R., Shepherd, J., Stafford, G. (2019) Isolation of bacteriophage targeting Enterococcus sp., Bacteriophage in Medicine, Food & Biotechnology PHAGE 2019, Oxford, United Kingdom, 12-13 September.

2. Oral presentation: **Alrafaie, A**., Shepherd, J., Stafford, G. (2020) Isolation of bacteriophage targeting Enterococcus sp. PGR Student Research Day, school of clinical dentistry, Sheffield, United Kingdom, 1 July.

3. Poster presentation: **Alrafaie, A**., Shepherd, J., Stafford, G. (2021) Enterococcal phages and prophages: isolation and TALs analysis. Phage Futures Congress, virtual conference, 24-26 February.

4. Oral presentation: **Alrafaie, A**., Shepherd, J., Stafford, G. (2021) Enterococcal phages and prophages: isolation and TALs analysis. PGR Student Research Day, school of clinical dentistry, Sheffield, United Kingdom, 18 March.

5. Oral presentation: **Alrafaie, A**., Al-Zubidi, M., Shepherd, J., Stafford, G. (2021) Enterococcal Bacteriophage: A review of the Tail Associated Lysin Landscape) Isolation of bacteriophage targeting Enterococcus sp., Bacteriophage in Medicine, Food & Biotechnology PHAGE 2021, Oxford, United Kingdom, 7-8 September.

6. Poster presentation: **Alrafaie, A**., Shepherd, J., Stafford, G. (2022) Enterococcal phages and prophages: Isolation, TAL analysis and expression. Bacteriophage in Medicine, Food & Biotechnology PHAGE 2021, Oxford, United Kingdom, 5-6 September.

7. Oral presentation: **Alrafaie, A**., Shepherd, J., Stafford, G. (2022) Enterococcal phages and prophages: Isolation, TAL analysis and expression. PGR third year presentation, school of clinical dentistry, Sheffield, United Kingdom, 29 September.

Figures

Tables

Chapter 1:

Literature review

1.1 Introduction:

Bacteriophages are a special type of virus that complete their life cycle by infecting bacterial cells. These bacteriophages or phages are obligate intracellular parasites meaning they only propagate inside living hosts (Clokie *et al*., 2011). The term "Bacteriophage" was first used by a French-Canadian scientist, Félix d'Herelle, and is derived from the Greek word phagein meaning "to devour" (Bakhshinejad & Ghiasvand, 2017). Based on several studies, phages are considered the most abundant biological entity in our ecosystem (Ackermann, 2007; Edwards, 2005; Shkoporov & Hill, 2019). On earth, it is estimated that there are about 10 phages to every bacterium and approximately 10^{31} phage particles in total (Molineux & Panja, 2013). In the ocean, there are around 10^{23} phage infections per second (Suttle, 2007) and approximately 10^{7} phages per millilitre (Ofir & Sorek, 2018). Based on metagenomic studies, bacteriophages are also part of the human microbiome and are found in several body sites such as the skin, lungs and oral cavity, thereby showing that these viruses are ubiquitous in both the human body and the environment (Navarro & Muniesa, 2017; Van Belleghem *et al*., 2019). Novel phages have also been continually discovered as a result of investigations of the environment (Ackermann, 2007).

1.2 Bacteriophage

1.2.1 History

The discovery of bacteriophages is referred to in two papers published by Frederick Twort and Félix d'Herelle. In 1915, a British scientist named Frederick Twort observed zones of clearance (plaques) on an agar plate cultivated with micrococci species (Twort, 1915). Two years later, a French-Canadian scientist named Félix d'Herelle isolated filterable agents that had antibacterial activity (against *Shigella dysenteriae*) from patients suffering from bacillary dysentery in the convalescence stage. D'Herelle found that these agents were live microorganisms and named them bacteriophages (Lourenço *et al*., 2018). After their discovery, scientists tried to take advantage of these viruses in treating bacterial diseases (Duckworth, 1976). Shortly after this,

penicillin was discovered in 1928 by Alexander Fleming and the era of antibiotics began (Fleming, 1929). As an effect of the ease and effective application of antibiotics, the bacteriophage research field has been confined to only a few countries such as Georgia (previously part of the Soviet Union) and Poland (Clokie *et al*., 2011). In 1940, the first micrograph of phages was accomplished by a German scientist Helmut Ruska, who observed *Escherichia coli* phages (Ackermann, 2011). In 1977, the first phage genome sequencing was completed by using the bacteriophage φX174, which has a small genome size of 5.3kb ssDNA (Sanger *et al*., 1977). For bacteria, the *Haemophilus influenzae* genome was the first to be sequenced as a phage host (Fleischmann *et al*., 1995). Notwithstanding the small genome size of phages, for which sequencing would be easier, the current databases contain more complete bacterial genomes than phages (Rihtman *et al*., 2016)

1.2.2 Classification

The International Committee on Taxonomy of Viruses (ICTV) is the official organization responsible for classifying viruses which was founded in 1966 (Simmonds & Aiewsakun, 2018). In the ICTV, a particular group is focused on bacteriophages called the Bacterial Viruses Subcommittee (Krupovic et al., 2021). The taxonomy of phages is primarily based on viral genome classification which tailed phages, for instance, are assigned to the class Caudoviricetes (Turner et al., 2021). In contrast to bacteria, phages lack conserved genes that are shared among all phages, which hinders building a phylogenetic tree (Simmonds & Aiewsakun, 2018). Therefore, a single-locus approach using some widely shared phage genes, such as capsid and DNA polymerase, is exploited for specific phage groups' characterisation (Edwards, 2005).

In addition to genome diversity, phages also have various morphologies. The four recognized morphologies of phages are tailed, polyhedral, filamentous or pleomorphic (Ackermann & Prangishvili, 2012). The tailed phages are classified under the class Caudoviricetes ("cauda" means "tail"), and it comprises three phage morphologies: myoviruses, siphoviruses, and podoviruses (Figure 1.1). All these phages are non-enveloped with a linear double-stranded DNA genome and they embrace more than 95% of all bacteriophages (Ackermann, 2007). For

myoviruses, these are characterised by a contractile tail that binds to and penetrates the bacterial cell envelope and allows phage DNA to pass into bacterial cytoplasm (Nobrega *et al*., 2018). Sheath proteins surround the tail tube causing tail contraction upon phage baseplate interaction with cell receptors (Hu *et al*., 2015). Conversely, siphophages have a long noncontractile tail, for which the binding of its distal end, tail tip complex (TTC), to cell receptors triggers genome release (Plisson *et al*., 2007; Xu & Xiang, 2017). In 2012, a study examined 6300 prokaryotic viruses under electron microscopy showed that siphoviruses constitute 57.3% of all analysed bacterial viruses (Ackermann & Prangishvili, 2012). For podoviruses, these possess a short tail that can extend to cross cell surfaces and eject their genomes inside their hosts (Hu *et al*., 2013).

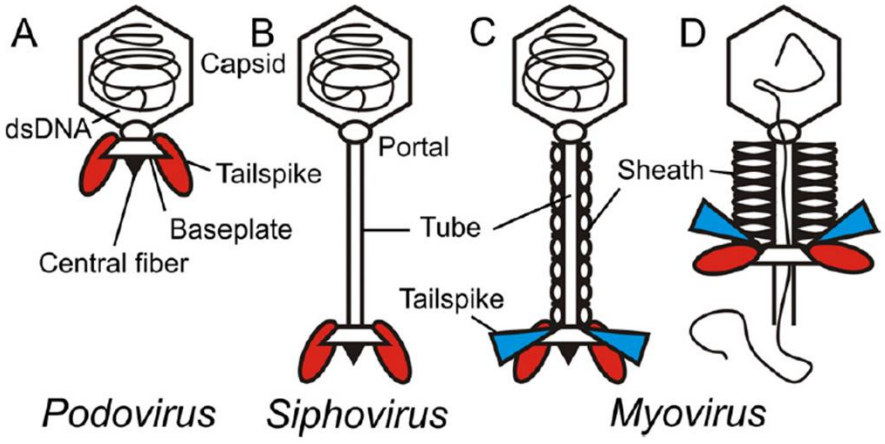


Figure 1.1 Model of tailed phages. (A) podoviruses are characterised by a short tail and a contractile tail while (B) siphoviruses phages have a long noncontractile tail. Myoviruses possess a contractile tail which can be seen as (C) uncontracted or (D) contracted form. Adapted from (Broeker et al., 2019).

1.2.3 Life cycles

Bacteria, hosts of phages, are diverse in their niches and are found almost everywhere (Dykhuizen, 2005). Bacteriophages interact with their hosts through different mechanisms such as lytic and lysogenic life cycles (Mirzaei & Maurice, 2017).

1.2.3.1 Lytic cycle

The lytic cycle is characterised by phages that lyse host cells at the end of the phage's life cycle (Sharma *et al*., 2017) (Figure 1.2). In phage infection, the first step is the adsorption of phages on the surface of their host cells which is carried out by mainly phage tail proteins, receptor-binding proteins (RBPs), that recognize bacterial receptors (Bertin *et al*., 2011). This phage binding can involve one or more receptors (Chan *et al*., 2016). A study focusing on the interaction between phage tail fibres and host receptors shows phage particles "walking" across a bacterial surface searching for a suitable receptor to adsorb (Hu *et al*., 2013). The process of phage adsorption starts with random collision followed by reversible adsorption and then irreversible attachment on a bacterial surface (Silva *et al*., 2016; Garen & Puck, 1951; Rakhuba *et al*., 2010). For example, SPP1 phage binds reversibly to cell wall teichoic acid (WTA) on *Bacillus subtilis* and irreversibly to the YueB membrane protein (Vinga *et al*., 2012). Irreversible attachment triggers phage genome ejection and delivery inside the bacterial host (Molineux & Panja, 2013). To facilitate phage adsorption and genome ejection, specific phage lysins (mostly associated with tail structure) are utilised to locally degrade bacterial layers (Fischetti, 2008). Following genome ejection, phages propagate inside the bacterial cytoplasm by exploiting the host metabolic machinery, which leads to the synthesis of new viral progeny (Drulis-Kawa *et al*., 2013). After replication, the phage capsid is first assembled then the viral genome is packaged via a packaging motor. The phage-packaged genome is digested by endonucleases either specifically by recognizing cohesive overhanging sites (*cos*) sites or non-specifically by headful mechanism (*pac*) sites (Le Marrec *et al*., 1997; Rao & Feiss, 2015). The next step in the phage life cycle involves virions' release from bacterial cells via host cell lysis. This is achieved by specific phage enzymes called holins and endolysins. Holins first degrade the bacterial cell membrane, which then allows endolysins to target the bacterial cell wall leading to cell bursts (Fischetti, 2008).

1.2.3.2 Lysogenic cycle

In the lysogenic cycle, the phage genome remains in a dormant state without propagation (Weinbauer, 2004) (Figure 1.2). It either integrates into the host genome (prophage) or remains episomally in the cytoplasm (Ofir & Sorek, 2018) and the phage genome replicates as cells divide (Chiang et al., 2019). Lysogenic or temperate phages affect their hosts in several ways: protection from invading phages, transferring new genes or modifying the surrounding community (Howard-Varona *et al*., 2017). Phage integration into a host genome may lead to modification of the bacterial phenotype, which is known as lysogenic conversion (Howard-Varona *et al*., 2017). This modification could result in antibiotic resistance or virulence factor production. For example, the cholera toxin phage (CTXΦ) carries genes encode for the cholera toxin AB which convert nontoxigenic *vibrio cholerae* to toxigenic strains after phage infection (Waldor & Mekalanos, 1996). Upon infecting bacterial cells, the temperate phages can undergo either a lytic or lysogenic cycle based on molecular determinants, known as the molecular switch (Golding, 2011; Herskowitz & Hagen, 1980). When temperate phages infect a large volume of cells, they tend to follow a lytic cycle, while multiple phages infecting a single bacterium would tend to switch to a lysogenic cycle (Ofir & Sorek, 2018). Moreover, temperate phages can transit from a lysogenic to a lytic cycle which is known as phage induction (Mirzaei & Maurice, 2017). Induction of phages can occur due to changes in nutrients, temperature, pH as well as exposure to UV radiation and antibiotics (Howard-Varona *et al*., 2017). For instance, the antibiotic mitomycin C is widely used to induce the phage lytic cycle of prophages (Cochran *et al*., 1998). Additionally, bacterial cells that carry prophages, known as lysogens, can carry more than one prophage genome (Chiang *et al*., 2019).

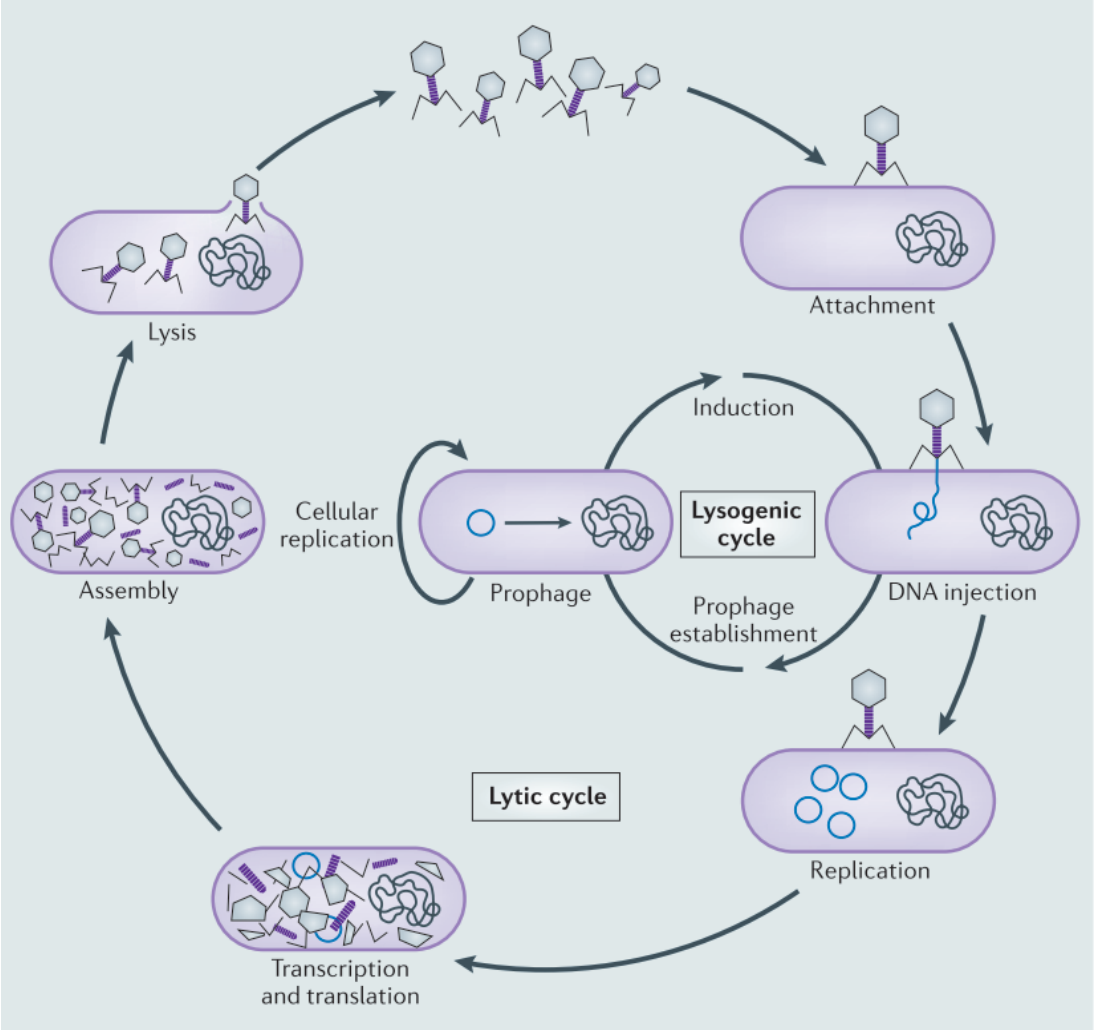


Figure 1.2 lytic and lysogenic phage life cycles. Phages are first adsorbed on the bacterial surface followed by DNA ejection into cytoplasm. For lytic cycle, phages exploit host machinery and actively multiply. When phage parts are assembled, specific phage enzymes lyse bacterial cells leading to phage particles release. For lysogenic cycle, the viral genome remains dormant inside host cell and integrates into the host chromosome. The integrated phage (prophage) can be induced to actively generate phage particles via lytic cycle. Adapted from (Salmond & Fineran, 2015) with permission.

1.2.4 Phage Diversity

Upon analysing marine viral sequences, which are dominated by phages, 63–93% of these sequences did not have a match in databases (Brum & Sullivan, 2015). This indicates that more phages still need to be identified. The unknown phages are not only limited to environmental samples since the human gut has more unidentified phages (Mirzaei & Maurice, 2017). In 2017, Aggarwala *et al* showed that from metagenomic studies of the human virome the majority of reads (75-99%) did not match identified phages in current databases (Aggarwala, 2017). This phage diversity is also seen with phage genome size as the smallest phage, Leuconostoc phage L5, contains only 2,435 bp (Dion et al., 2020) while the largest genomes, Lak phages, possess >540 kb that were identified from humans and animal gut metagenomes and they infect bacteria of the genus *Prevotella* (Devoto et al., 2019).

1.2.5 Phages and microbiota

All microorganisms that live within or on the human body are known as the microbiota (Bakhshinejad & Ghiasvand, 2017). The human gut contains at least 10^{14} microbial cells (Clemente et al., 2012; Lepage et al., 2013) and 10^{12} phage particles (Shkoporov & Hill, 2019). The microbiota is also called the "forgotten organ" (O'Hara & Shanahan, 2006) which helps in the protection from pathogens (Belkaid & Hand, 2014) and stimulation of the immune system (Clemente *et al*., 2012). An imbalance of the microbial community, known as microbiota dysbiosis, has been associated with several health issues such as Inflammatory bowel disease, obesity (Chang & Lin, 2016) and Periodontitis (Meuric *et al*., 2017). Humans are constantly exposed to phages from food consumption (Maura & Debarbieux, 2012) and these are considered part of the human and animal microbiota (Clokie *et al*., 2011). Moreover, Manrique *et al* show that more than half of 64 healthy individuals from different countries shared 23 phages in their gut. However, the presence of these shared phages was remarkably reduced in patients with gastrointestinal disease (Manrique *et al*., 2016). In addition, the commensal bacteria can utilise their prophages to attack other bacteria and preserve their niche in the gut (Duerkop *et al*., 2012).

1.3 Enterococci

1.3.1 Introduction

Enterococci are Gram-positive facultative anaerobic bacteria that normally inhabit human gastrointestinal tracts (De Been *et al*., 2013). Enterococci have been associated with hospital-acquired infections (HAIs) since the 1980s (Emori & Gaynes, 1993). Of all enterococcal species, *E. faecalis* and *E. faecium* are the most clinically important (Galloway-Peña *et al*., 2009). Both *E. faecalis* and *E. faecium* are the causative agents in various clinical issues including urinary tract infection (UTI), bacteraemia, wound infections (especially DFUs- Diabetic Foot Ulcers) and endocarditis infections (Huycke *et al*., 2014). Following staphylococci, enterococci are the second leading cause of Gram-positive nosocomial infections (Miller *et al*., 2014). Weiner *et al* analysed over 350,000 HAIs and found that *E. faecalis* (7.4%) and *E. faecium* (3.7%) are among the most common isolated pathogens causing nosocomial infections (Weiner *et al*., 2016). Additionally, this study analysed the antimicrobial resistance of the isolated pathogens from the urinary tract, bloodstream and surgical sites which showed that *E. faecium* strains are more antibiotic-resistant than *E. faecalis* although the number of *E. faecalis* isolates was higher (Weiner *et al*., 2016). The genus *Enterococcus* have become resistant to various antibiotics such as vancomycin, with vancomycin-resistant enterococci (VRE) rapidly disseminating worldwide (Fiore, 2019).

1.3.2 Clades

Enterococci as well as other bacteria are usually identified and assigned to different genetic lineages using multilocus sequence typing (MLST) which in the case of the enterococci is based on seven housekeeping genes (Lee *et al*., 2019). For *E. faecium* strains, these are divided into two main clades: Hospital associated clade A, which includes clinical isolates from humans (A1) and animals (A2) and community-associated clade B for human commensal isolates (Cattoir, 2022; Palmer et al., 2012; Qin et al., 2012). *E. faecium* ST17 is likely the original clone of clade A1, which is also named clonal complex 17 (CC17) (Gouliouris et al., 2019; T. Lee et al., 2019).

The hospital-acquired complex CC17 is characterised by intrinsic ampicillin and quinolone resistance (Top *et al*., 2008), the presence of Enterococcal Surface Protein (ESP) (T. Lee et al., 2019), genome adaptability (Miller *et al*., 2014) and vancomycin resistance (Gouliouris et al., 2019). In addition to *E. faecium* ST17, other particular enterococcal lineages are linked to HAIs such as *E. faecium* ST18, ST78, and *E. faecalis* ST203 and ST40 (Palmer *et al*., 2012).

1.3.3 Enterococcal cell wall

An enterococcal cell envelope consists of a cell membrane as well as a cell wall (Hancock et al., 2014b)(Figure 1.3). The enterococcal cell wall mainly comprises three parts: the peptidoglycan layer, anionic polymers and a few proteins (Bhavsar & Brown, 2006). Both the peptidoglycan layer and anionic polymers (teichoic acids and cell wall polysaccharides) comprise 90% of cell wall constituents, whereas the remaining percentage is proteins anchored in or associated with the cell wall (Hancock *et al*., 2014b). Peptidoglycan is composed of N-acetylmuramic acid MurNAc (NAM) and N-acetylglucosamine GlcNAc (NAG), via which these repeating disaccharide sugars are linked by β-1,4-glycosidic bonds (Navarre & Schneewind, 1999) and cross-linked by a stem peptide chain anchored to the NAM sugar (Hancock *et al*., 2014b). Amide linkages connect the stem peptide to NAM residue (the amino group of the stem peptide L-alanine binds to the carboxyl group of the D- lactyl moiety of NAM) and consist of five amino acids: L-Ala-D-Glu-L-Lys-D-Ala-D-Ala. The third amino acid, L-Lys, binds to the fourth amino acid, D-Ala, in an adjacent peptidoglycan strand to create an interpeptide bridge and this results in the loss of the fifth amino acid (Figure 1.4) (Hancock *et al*., 2014a). The interpeptide bridge, also called a cross-bridge, consists of a short peptide sequence in most Gram-positive bacteria, whereas there is direct linking (no peptide sequence) in most Gram-negative bacteria (Vollmer *et al*., 2008). For Gram-positive bacteria, the composition of the cross bridges varies from five Gly in *S.aureus*, a single D-Asn or D-Asp residue in *E. faecium* or two L-Ala residues in *E. faecalis* (Bouhss *et al*., 2002).

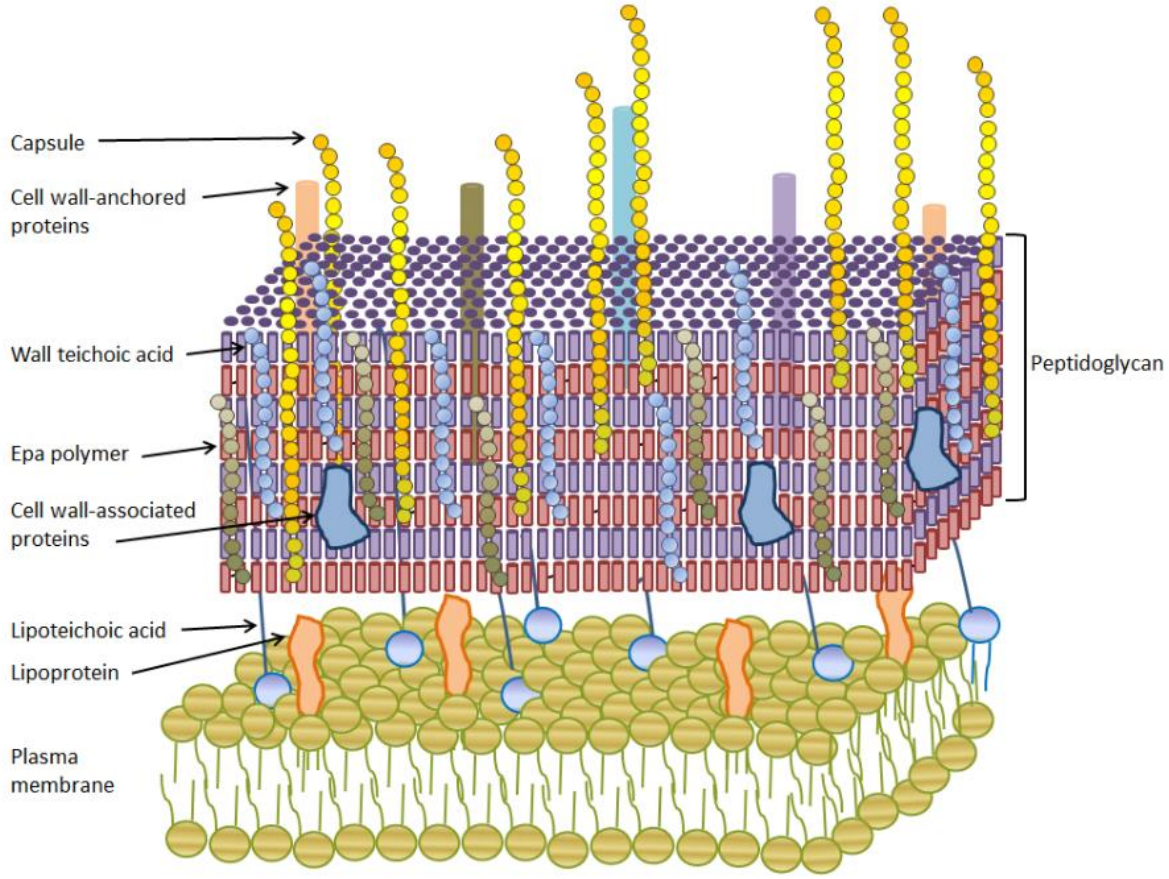


Figure 1.3 Schematic representation of the enterococcal cell wall. The plasma membrane with bound lipoteichoic acids and lipoproteins are depicted below the cell wall. The cell wall consists of the peptidoglycan layer, wall-anchored and associated proteins, wall teichoic acid and EPA polymers. Adapted from (Hancock et al., 2014a) - Creative Commons.

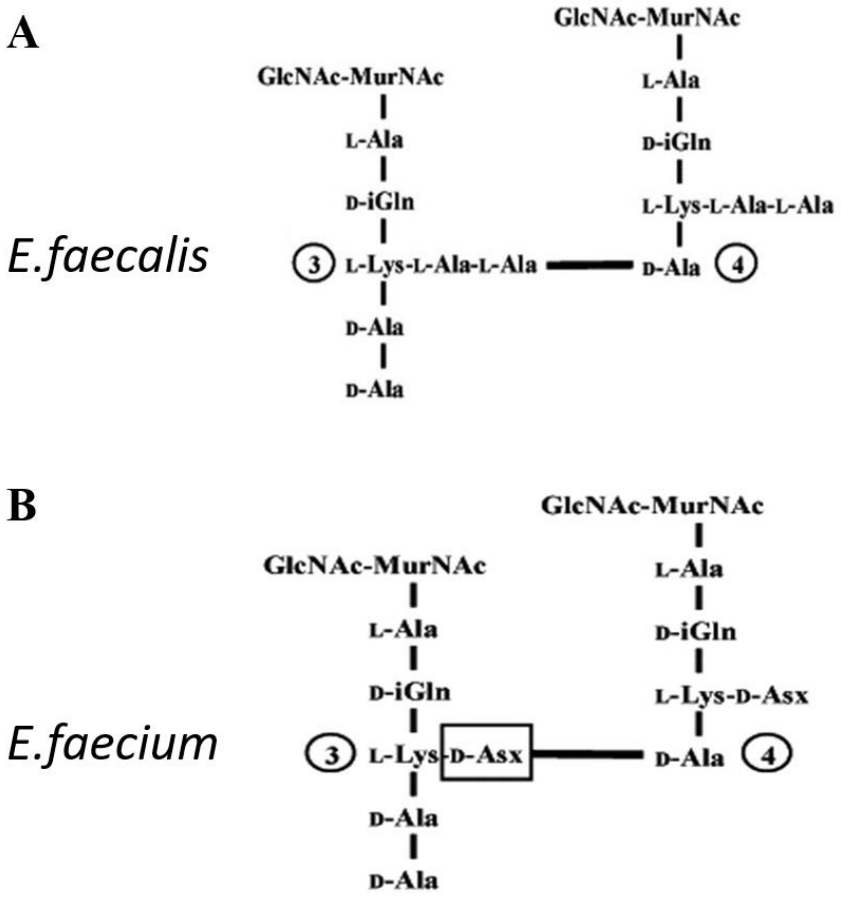


Figure 1.4 Schematic representation of the peptidoglycan cross-linking in *E. faecalis* (A) and *E. faecium* (B). The interpeptide bridge is formed by binding the third amino acid (L-lys) to the fourth amino acid (D-Ala). *E. faecalis* cross bridge is composed of two L-Ala residues while D-ASX (D-Asn or D-Asp) is for *E. faecium*. Adapted from (Arbeloa et al., 2004) with permission.

1.3.4 Lipoteichoic acid (LTA)

Streptococci can be identified by using a serological test that detects polysaccharides in the cell wall which is called Lancefield grouping (Lancefield, 1933). Enterococci belong to the Lancefield group D, along with other streptococci such as *Streptococcus bovis*, and their differentiation is based on lipoteichoic acid (LTA) (Burger, 1966; Coia & Cubie, 1995; Hancock & Gilmore, 2002; A. J. Wicken et al., 1963). LTA is composed of a chain of glycerolphosphate residues (linked by phosphodiester bonds and substituted with various groups such as glucose or

D-Alanine) that are covalently bound to a glycolipid moiety and are anchored to the cell membrane (A. Wicken & Knox, 1975) (Figure 1.3 and 5).

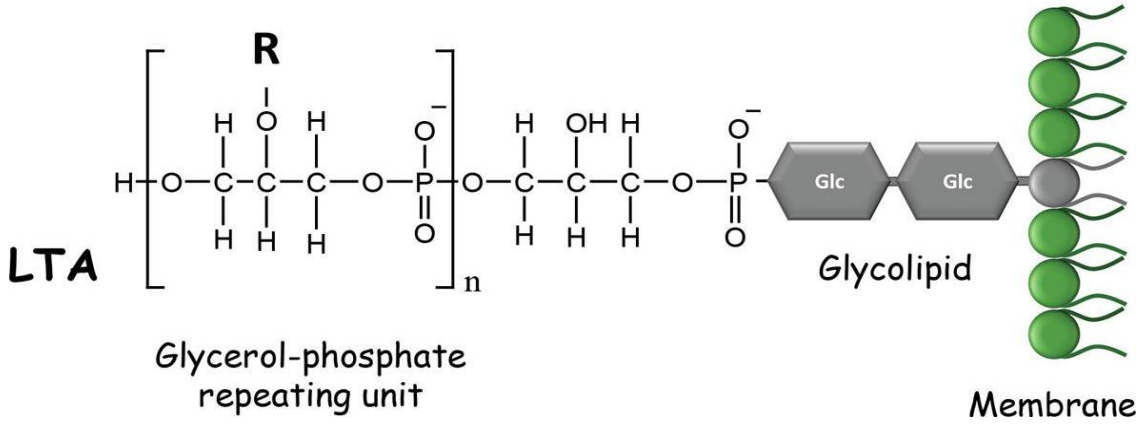


Figure 1.5 Schematic representation of the structure of LTA. LTA is composed of a repeating unit of glycerol-phosphate which is bound to the cell membrane via a glycolipid which is diglucosyldiacylglycerol. "R" indicates substituents (e.g., D-Ala, Glc, Gal, or GlcNAc). Adapted from (Chapot-Chartier & Kulakauskas, 2014) – Creative Commons

1.3.5 Cell-Wall teichoic acid (WTA)

Wall teichoic acids (WTAs) bind to peptidoglycan (PG) via NAM and consist of disaccharide sugars and multiple linked glycerolphosphate molecules (Chapot-Chartier & Kulakauskas, 2014) (Figure 1.3). The composition of WTs is complex which the *E. faecium* U0317 WTAs are composed of two residues of N-Acetylgalactosamine, glycerol (Gro), and phosphate (Bychowska *et al*., 2011), whereas the *E. faecalis* 12030 WTAs consist of D-glucose, D-galactose, N-Acetylgalactosamine, N-Acetylglucosamine, D-ribitol, and phosphate (Theilacker *et al*., 2012). In the cell wall, the most abundant polymer bound to PG is WTA (Brown *et al*., 2013; Hancock *et al*., 2014b) and both WTA and LTA confer a negative charge to the cell wall (Hancock *et al*., 2014b).

1.3.6 Enterococcal polysaccharide antigen (EPA)

EPA is part of the enterococcal cell wall and is also known as rhamnopolysaccharide (Palmer *et al*., 2012). The EPA has been shown to be important in biofilm formation (Almohamad *et al*., 2014) and resistance to phagocyte killing (Teng *et al*., 2002). Additionally, EPA has an essential role in protecting cell envelope integrity (Ramos et al., 2021). The best characterised EPA is arguably that of *E. faecalis* OG1RF which is composed of glucose, galactose, rhamnose, N-acetyl glucosamine and N-acetyl galactosamine (Teng *et al*., 2009). For the *E. faecalis* V583 strain, the EPA analysis showed to contain glucose, rhamnose, N-acetyl glucosamine and N-acetyl galactosamine (Guerardel et al., 2020). The synthesis of the EPA is encoded by the EPA genes which ,in *E. faecalis*, consists of 18 highly conserved genes (Teng et al., 2009) followed by (10-20) EPA variable genes (Guerardel et al., 2020; Smith et al., 2019). For the *E. faecium* strains, the conserved region contains 15 genes and a set of variable EPA genes (Palmer et al., 2012). Three EPA genes in the conserved region (*epaI*, *epaJ*, and *epaK*) are only found in *E. faecalis* strains while the *epaP* and *epaQ* are positioned differently between the two species (Ramos et al., 2021) (Figure 1.6). The EPA is considered as a promising target for antimicrobial agents focusing on conserved genes in both *E. faecalis* and *E. faecium* (Palmer *et al*., 2012).

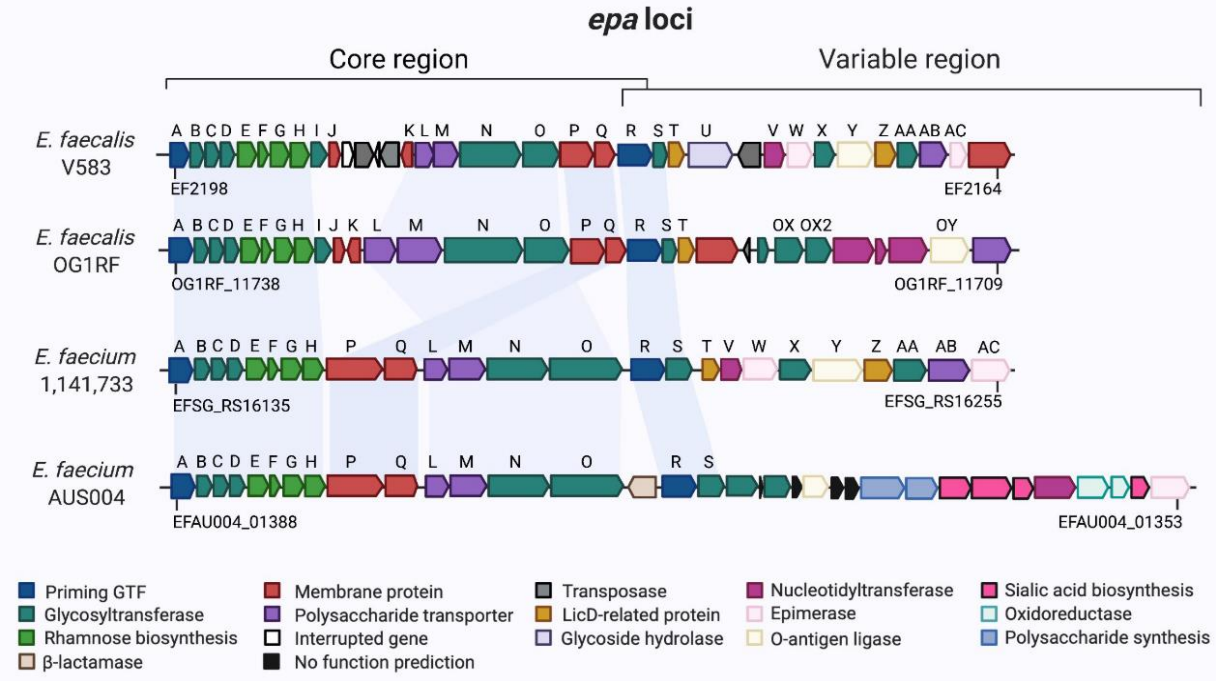


Figure 1.6 Schematic representation of the epa loci within *E. faecalis* and *E. faecium* strains. For *E. faecalis* V583 and OGIRF, EPA loci are divided into two main regions: a conserved region (18 genes) and a variable region (10-20 genes). For the *E. faecium* strains, the core regions have a different organisation than *E. faecalis* which lacks homology of *epaI*, *epaJ*, and *epaK* genes. The *epaP* and *epaQ* are also located differently between *E. faecalis* and *E. faecium* strains. homologs of different strains are indicated by blue shade. All genes are drawn to scale. Coloured boxes (bottom) are referred to as arrow boxes. Adapted from (Ramos et al., 2021).

Furthermore, the analysis of the *epa*-like locus in 34 *E. faecium* strains revealed four different variants (De Been et al., 2013). The conserved EPA regions are shared among all variants while the difference is in the variable downstream region. The four variants are represented by the following strains: Com15 (variant one), Com12 (variant two), 1,231,408 (variant three) and Aus0004 (variant four) (Figure 1.7).

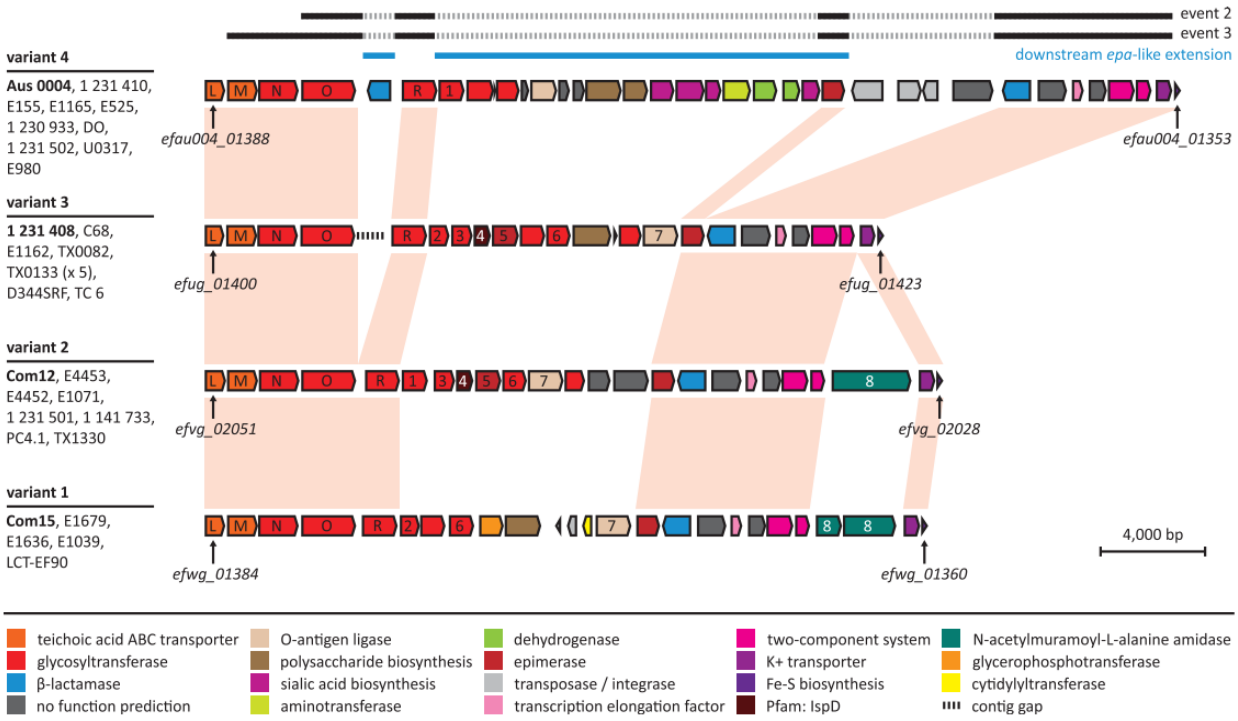


Figure 1.7 The four epa-like locus variants of *E. faecium*. Genes of a representative strain from each variant are shown as follows: Aus0004 (variant four), 1,231,408 (variant three), Com12 (variant two), and Com15 (variant one). The pink shades connect core genomic genes. Part of the conserved epa-like (epaL–epaO and epaR) are indicated by L, M, N, O and R and they are part of the conserved upstream epa-like locus. Orthologous and paralogous genes are indicated by numbers. Genes with the specific functional group are colour-coded. Drawings are to scale. Adapted from (De Been et al., 2013) with permission.

1.3.7 Capsule

Bacterial cells can be targeted and eliminated by immune cells via recognizing specific surface receptors. However, bacterial capsules can mask these receptors and circumvent the immune response (Thurlow *et al*., 2009). Both *E. faecium* and *E. faecalis* have a capsular polysaccharide (CPS) operon that regulates the synthesis of the capsular polysaccharide (Hancock *et al*., 2014a). *E. faecalis* have four different capsule serotypes (A, B, C, or D) with serotypes C and D showing more resistance to neutrophil phagocytosis (Thurlow *et al*., 2009). In 2012, a new capsule-like region was identified in eight *E. faecium* which encodes proteins that resemble CpsBCD proteins of *Streptococcus pneumoniae* involved in capsule production (Palmer *et al*., 2012).

1.4 Virulence factors of *E. faecalis* and *E. faecium*

Although Enterococci are found normally in the GIT of humans and animals, they also could have the ability to cause diseases and overcome host immunity (Johnson, 1994). Both *E. faecalis* and *E. faecium* utilise virulence factors to promote bacterial pathogenesis (Sava *et al*., 2010). These virulence factors are summarised in (Table 1.1)

Table 1.1 Enterococcal virulence factors

Virulence factor	Role in virulence	References
Adhesin to collagen of *E. faecalis* (ACE) or *E. faecium* (ACM)	Bacterial colonization to extracellular matrix proteins (collagen types I and IV and laminin)	(Nallapareddy et al., 2000) (Nallapareddy et al., 2003)
***E. faecalis* antigen A (EfaA)**	Associate with infective endocarditis	(Lowe et al., 1995) (Singh et al., 1998)
Aggregation substance (AS)	Bind to extracellular matrix (ECM) proteins and facilitate mating between donor and recipient cells for conjugation	(Rozdzinski et al., 2001) (Clewell, 1993)
Enterococcal surface protein (ESP)	bacterial adhesion and biofilm formation	(Klare et al., 2005)
Cytolysin (Cyl)	Bactericidal effects against both Gram-positive and negative bacteria as well as being toxic toward eukaryotic cells (erythrocytes, leukocytes and macrophages)	(Chajęcka-Wierzchowska et al., 2017) (Coburn & Gilmore, 2003)
Gelatinase (GeIE)	Degradation of gelatin, elastin, collagen, and haemoglobin. It also has a role in biofilm formation.	(Archimbaud et al., 2002) (Eaton & Gasson, 2001) (Thomas et al., 2008)
Hyaluronidase (hyl)	facilitates bacterial dissemination by degrading hyaluronic acids (Hyaluronate) found in various body sites such as ECM, muscle, cartilage and skin.	(Hynes & Walton, 2000) (Gulhan et al., 2015)

1.5 Bacterial defence mechanisms

Bacterial cells have been exposed and adapted to various stressors such as changes in temperature, pH, osmolarity and the concentration of nutrients (Marles-Wright & Lewis, 2007). Additionally, bacterial resistance and survival of antibacterial treatment have been addressed in several studies (Davies, 1996; Prestinaci *et al*., 2015). Of these antibacterials, antibiotics and phage resistance are described in the following sections.

1.6 Antibiotic resistance

The resistance of bacteria is now almost observed in all antibiotic classes (Mobarki *et al*., 2019)(Figure 1.8). In 2017, the WHO published a list of antibiotic-resistant pathogens to prioritize the research and discovery of new antibiotics (WHO, 2017). The list was divided into three categories (critical, high and medium) and the pathogens were selected based on several factors, including the ease of transmission, availability of treatment options and frequency of acquiring resistance. *E. faecium* (vancomycin-resistant) was placed in the high category indicating its global significance and the need for new treatment options (WHO, 2017). In addition to acquiring resistant genes, enterococci can transmit genes to other bacteria. For example, the vanA gene is carried on a transposon Tn*1546* which can be transferred from *Enterococcus* to *S.aureus* strains (Périchon & Courvalin, 2009). Enterococci have developed resistance to even new antibiotics: linezolid, quinupristin/dalfopristin [Q/D], Daptomycin [DAP] and Tigecycline (Miller *et al*., 2014).

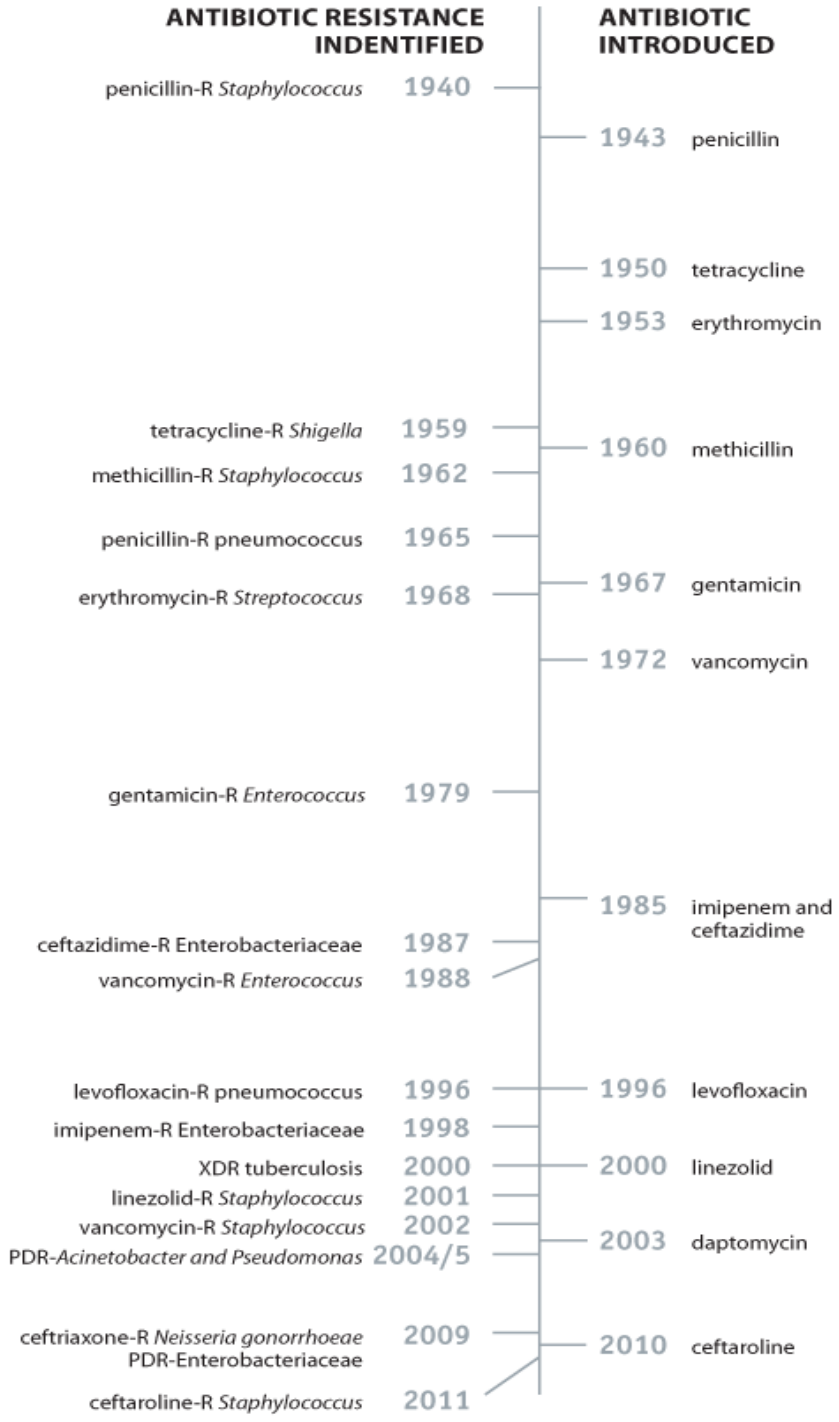


Figure 1.8 Timeline of antibiotic discovery and antibiotic resistance. Adapted from (CDC, 2018)

1.6.1 Antibiotic Resistance mechanisms in enterococci

In 1928, the first antibiotic, penicillin, was discovered by Alexander Fleming which showed bactericidal activity against *S.aureus* (Fleming, 1929). Penicillin and ampicillin are β–lactam antibiotics that inhibit cell wall synthesis (Miller *et al*., 2014). The β–lactam molecule closely resembles the pentapeptide structure in peptidoglycan causing covalent binding between the β–lactam ring and penicillin-binding-proteins (PBPs) that leads to aborting cell wall synthesis and eventually cell death (Tomasz, 1979). For enterococci, intrinsic or acquired resistance mechanisms are used against several β–lactam antibiotics (Moon *et al*., 2018)(Figure 1.9). An intrinsic resistance mechanism involves PBP4 and PBP5 proteins in *E. faecalis* and *E. faecium*, respectively, which have less affinity toward β–lactam antibiotics and confer low-level resistance (Moon *et al*., 2018). For acquired resistance, a mutation in PBP5 has been shown to increase the level of resistance of *E. faecium* strains (Rice *et al*., 2001). Another class of antibiotics that affect the cell wall is the glycopeptides, which include vancomycin and teicoplanin that target the terminal residues D-alanine-D-alanine in peptidoglycan precursors and block transpeptidation. However, changing to D-alanine-D-lactate or D-alanine-D-serine confers resistance (Dutka-Malen & Courvalin, 1990) (Figure 1.9). Vancomycin resistance genes are classified into nine clusters with the most frequent resistance mechanism used by enterococci is based on the *vanA* gene cluster (Miller *et al*., 2014).

The bacterial cell membrane is targeted by daptomycin, an antimicrobial peptide, which is a cyclic lipopeptide that oligomerizes within the cell membrane leading to the leakage of ions and eventually cell death (Steenbergen *et al*., 2005). The development of daptomycin resistance in both *E. faecalis* and *E. faecium* is attributed to mutations in two sets of genes: a three-component regulatory system, liaFSR, that regulates the cell envelope response to stress and genes encoding proteins that have a role in the phospholipid metabolism (cls and gdpD)(Munita et al., 2013) (Figure 1.9).

Aminoglycosides are a class of bactericidal antibacterial agents that interfere with the protein synthesis process (Krause *et al*., 2016). Low-level resistance to aminoglycosides is caused by the enterococcal intrinsic characteristic of limited drug uptake. However, using cell-wall targeting antibiotics has shown a synergetic effect which has caused efficient enhancement in aminoglycoside uptake (Moellering & Weinberg, 1971). Linezolid belongs to Oxazolidinones antimicrobials which affect the initiation of bacterial protein synthesis by binding to the 23S rRNA and interfering with aminoacyl tRNA docking at the A site of the ribosome (Leach *et al*., 2007). As Linezolid resistance in enterococci is caused by a G2576U mutation in the 23S rRNA, Marshall *et al* showed that the presence of this mutation in multiple 23s rRNA genes is associated with higher levels of resistance (Marshall *et al*., 2002). Other antibiotics that inhibit protein synthesis by interfering with the large ribosomal subunit (50S) are streptogramin A (dalfopristin) and B (quinupristin), which are the first FDA-approved antibiotics for VRE infections (Arias & Murray, 2012). Both dalfopristin and quinupristin have bacteriostatic activity but their synergetic action of changing the ribosome conformation results in a bactericidal effect (Gurk-Turner, 2000). In contrast to *E. faecium*, *E. faecalis* is intrinsically resistant to clindamycin and streptogramin A because of the lsa (lincosamide and streptogramin A resistance) protein, which is predicted to be involved in the efflux pump mechanism (Isnard et al., 2013; Singh et al., 2002). For E. faecium resistance, streptogramin A is modified with acetyltransferases (VatD and VatE) (Werner et al., 2002) and streptogramin B is cleaved by virginiamycin B lyase (Vgb) (Korczynska et al., 2007) (Figure 1.9).

Quinolones are a class of antibiotics that inhibit DNA replication and transcription by targeting two enzymes: DNA gyrase (composed of GyrA and GyrB) and topoisomerase IV (ParC and ParE). The gyrase enzyme helps in creating negative supercoil, while topoisomerase IV relaxes negative supercoiled DNA and separates linked daughter chromosomes at cell division (Redgrave *et al*., 2014). Enterococci resistance to quinolones takes place via mutating target sites (such as gyrA and parC), expelling antibiotics by an efflux pump system (Yasufuku *et al*., 2011) or possessing the qnr protein, which prevents quinolines from binding to its target (Arsène & Leclercq, 2007). For RNA synthesis, Rifampicin targets the beta subunit of RNA polymerase and

inhibits bacterial transcription. However, a mutation in the gene (*rpoB*) that encodes the beta subunit of RNA polymerase confers drug resistance (Kristich & Little, 2012).

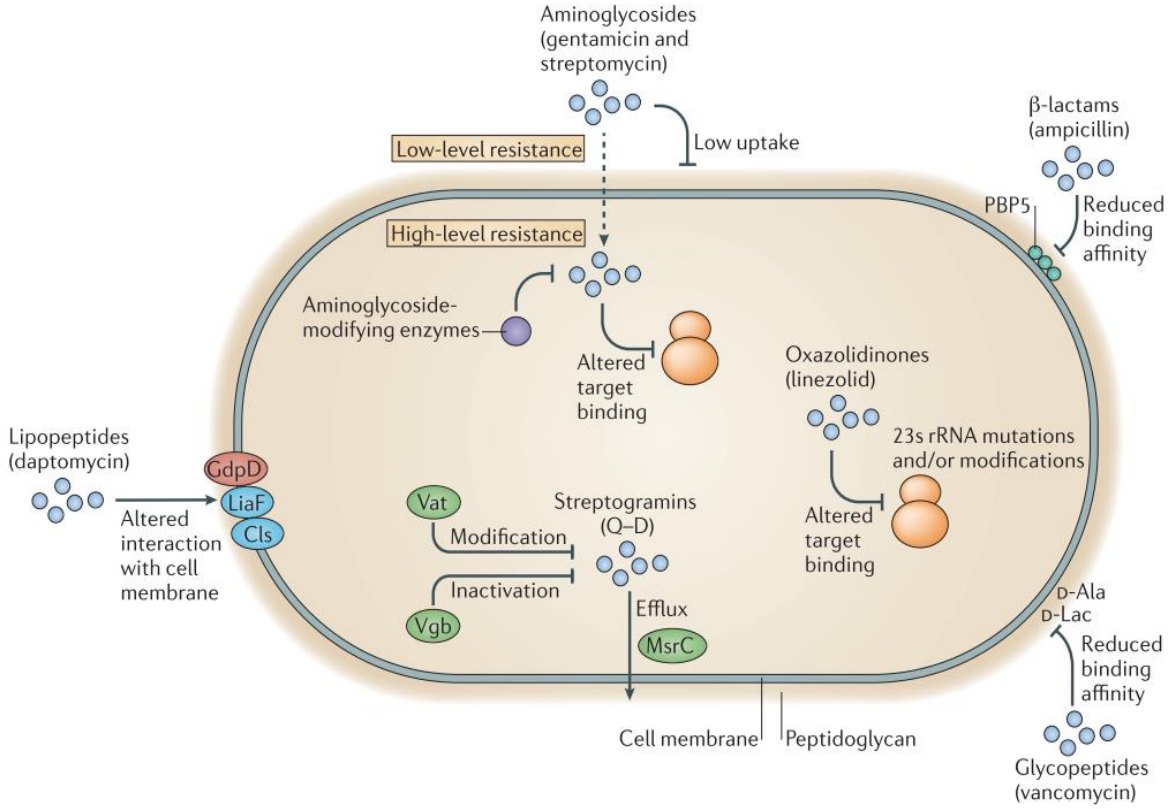


Figure 1.9 Enterococcal antibiotic resistance mechanisms. Enterococci confer low-level resistance to aminoglycosides via low drug uptake and high-level resistance through aminoglycoside-modifying enzymes and alteration of the target site. low-affinity PBPs also lead to B-lactams resistance. Resistance to vancomycin is achieved through the alteration of cell wall residues. Three mechanisms that enterococci utilise to resist streptogramins quinupristin–dalfopristin (Q–D): modification of drug via virginiamycin acetyltransferase (Vat), inactivation of drug by virginiamycin B lyase (Vgb) and drug efflux via macrolide–streptogramin resistance protein (MsrC). Enterococcal daptomycin resistance is accomplished by the alteration of cell membrane interactions which involve the membrane protein Liaf, the glycerophosphoryl diester phosphodiesterase family (GdpD) and cardiolipin synthase (Cls). For oxazolidinones, the most common mechanism for linezolid resistance is a mutation in the target site. Adapted from (Arias & Murray, 2012) with permission.

1.6.2 Phage resistance

Similar to antibiotic resistance, bacterial cells have several mechanisms that lead to unsuccessful phage infection. In contrast, phages also have shown various ways to circumvent bacterial

defence to ensure lifecycle completion. The battle between phages and their hosts has "co-evolved" for billions of years (Ofir & Sorek, 2018) and Bacterial defence mechanisms interrupt phages in every step of their lifecycle (Hoskisson & Smith, 2007)(Figure 1.10). The bacterial/phage "arms race" mechanisms are discussed in the following section based on the phage lifecycle.

1.6.2.1 Phage adsorption inhibition

The Phage adsorption to bacterial cells can be inhibited by several mechanisms, such as mutating or masking receptors, using receptor competitive inhibitors or producing an extracellular matrix (Dy *et al*., 2014). Furthermore, bacterial cells can avoid phage adsorption through transiently repressing genes encoding phage receptors, which is known as "phase variation" (Hoskisson & Smith, 2007). However, phages can alter the tail fibres to adapt to either a mutated receptor or target a new receptor. For example, the lambda phage J protein targets the LamB receptor on the *E.coli* outer membrane. However, when the LamB receptor is mutated, the J protein mutates to bind to another receptor which appears to be the OmpF (Meyer *et al*., 2012). In addition, phages can degrade the exopolysaccharide matrix by lyase enzyme action, a class of bacteriophage-encoded depolymerases, which facilitates phage infection (Sutherland, 1995).

1.6.2.2 DNA ejection blocking

After a phage successfully attaches to its receptor, the viral genome is delivered through the cell envelope and ejected into the bacterial cytoplasm (Xu & Xiang, 2017). Phages that have cell-wall degrading enzymes can locally lyse the peptidoglycan layer to facilitate DNA ejection (Moak & Molineux, 2004). Bacterial cells can block phage DNA ejection via the superinfection exclusion (SIE) systems, which prevent secondary infection by exact or closely related phages. To illustrate, the T4 phage confers SIE to its host against new infections by blocking the DNA ejection by the Imm phage protein and inhibiting the lysozyme enzyme by the Sp protein (Lu & Henning, 1994).

1.6.2.3 Nucleic acid degradation

Bacterial cells can target viral nucleic acid in the cytoplasm with two main systems leading to phage genome degradation: Restriction-modification (RM) and CRISPR-Cas systems (Dy *et al*., 2014). The RM system is known as the bacterial innate immune system that identifies foreign DNA. As the term suggests, the RM system carries two main functions. First, it modifies specific nonmethylated DNA sequences by a methyltransferase enzyme. Secondly, if specific DNA sequences are nonmethylated, restriction endonuclease (REase) enzymes identify this DNA as foreign and subsequently cleave the sequence, leading to DNA degradation (Tock & Dryden, 2005). Methylation of DNA is the way that bacteria distinguish self (methylated) from non-self (nonmethylated). There are four types of RM (Type I, II, IV, V) which differ in some features such as restriction site recognition, cleavage location and cofactor usage (Tock & Dryden, 2005). Phages can escape from the RM defence systems in various ways. One is by altering the sequence of or temporally occluding (using phage proteins that bind to DNA) restriction sites, which minimises bacterial RM systems' identification and eventually cleavage (Tock & Dryden, 2005). In addition, phages can carry enzymes involved in degrading restriction endonucleases (Krüger & Bickle, 1983) or inactivating essential cofactors (Studier & Movva, 1976). Furthermore, some phages may carry methyltransferase genes, which can be used to methylate phage genomes and inhibit the RM system (Krüger & Bickle, 1983).

The clustered, regularly interspaced short palindromic repeats (CRISPR) systems are considered the adaptive immune system in bacteria and archaea (Bhaya *et al*., 2011). CRISPR-Cas proteins are involved in the processing of foreign DNA recognition and cleavage. When foreign DNA is ejected into a bacterial cell, specific DNA sequences (called protospacers) are recognized which leads to DNA cleavage. In addition to protospacers, the short sequences adjacent to target DNA sequences, known as the Protospacer-associated motifs (PAM), are important for spacer acquisition and later interference (Datsenko *et al*., 2012; Yosef *et al*., 2013). Also, specific sequences of the cleaved DNA (called spacers) are added to the bacterial genome to allow future

recognition of the invaded DNA. RNA transcripts are generated and processed into small sequences called CRISPR RNA (crRNA). Lastly, a Cas protein along with crRNA forms a complex that identifies and cleaves new invading phage DNA (Richter *et al*., 2012). As expected, phages have developed a mechanism to evade this system by having anti-CRISPR (acr) genes which were discovered in 2013 by using *Pseudomonas aeruginosa* phages (Bondy-Denomy *et al*., 2013). Moreover, phages also can escape the CRISPR-Cas system through a single nucleotide mutation in viral sequences that are recognized by the host defence system (Deveau *et al*., 2008)

1.6.2.4 Phage exclusion systems

Another mechanism that bacteria use to impair phage infection is called abortive infection (Abi) which a bacterium sacrifices itself to inhibit phage progeny dissemination (Dy et al., 2014) (Figure 1.10). Bacterial suicide is also considered one of the outcomes of a biological module known as the toxin-antitoxin (TA) systems which an antitoxin neutralizes a toxin in normal conditions (Harms *et al*., 2018). However, if the antitoxin is lost, the toxin can cause a determinantal effect leading to cell death (Schuster & Bertram, 2013). Antitoxins can be proteins or RNAs which either directly inhibit toxins or interfere with toxins' expression whereas toxins are proteins (Harms *et al*., 2018). Bacteriophages have developed mechanisms to counteract Abi and TA systems which render phage infection achieved. For example, phiTE phage evaded Abi in *Pectobacterium atrosepticum* by mimicking an antitoxin which renders Abi ineffective (Blower *et al*., 2012). Also, a single nucleotide mutation in a bacteriophage ⏀M1 gene, *M1-23*, allowed the phage to escape from both systems (Blower *et al*., 2017).

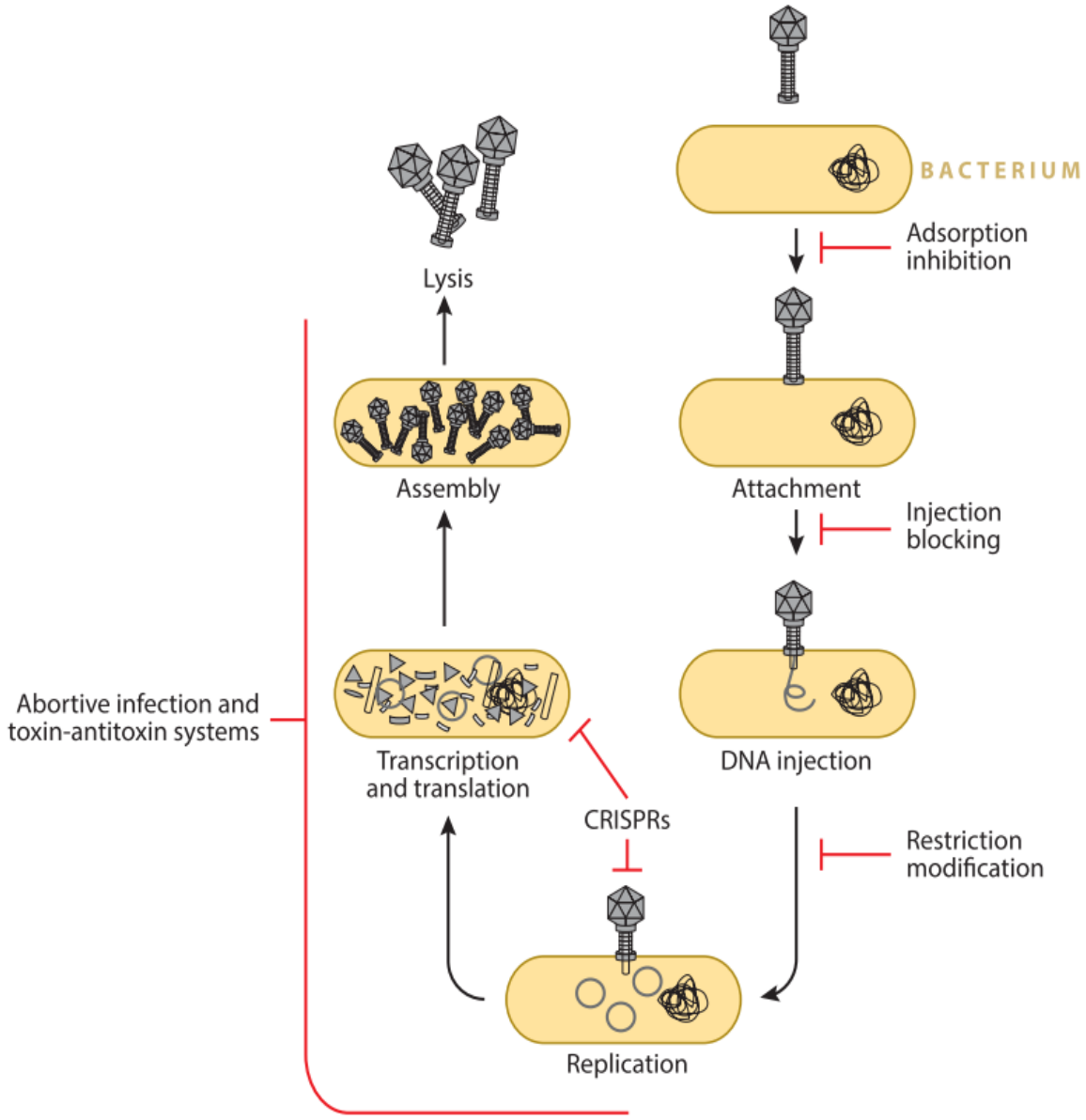


Figure 1.10 Bacterial defense mechanisms during phage lytic cycle. Bacteria can prevent the binding of phages to bacteria's surface and block phage DNA ejection into the cytoplasm. Also, restriction-modification (RM) and CRISPR systems can identify and degrade phage genomes. Lastly, bacterial suicide can be accomplished via abortive infection and Toxin-antitoxin systems to prevent phage dissemination. Adapted from (Dy et al., 2014) with permission.

1.7 Enterococci and diabetic foot infection

The skin is the largest organ in the human body. It prevents the colonisation of pathogens to the subcutaneous tissues (Bowler *et al*., 2001) and it is the most exposed to injury (Negut *et al*., 2018). Damage to skin integrity (loss of epithelia and connective tissue) reduces its protective role, which requires a healing process to restore its main function (Negut *et al*., 2018). Wound infections are often polymicrobial and lead to wound healing failure, which increases the cost of treatment and patient trauma (Bowler *et al*., 2001). Regarding diabetes, approximately 33% of diabetic patients develop foot ulcers (Armstrong *et al*., 2017) and more than half of these ulcers become infected (Armstrong *et al*., 2017). Diabetic foot ulcers (DFU) can lead to amputation, ranging from a toe to a full extremity (Walsh *et al*., 2016). In the U.S., the evaluation of 784 million ambulatory care visits by diabetic patients from 2007 to 2013 indicated that DFU and diabetic foot infection (DFI) increased the risk of in-patient admission and emergency department visits (Skrepnek *et al*., 2017). Moreover, patients with DFU are at higher risk of death compared with diabetic patients with no DFU and it is most likely that the presence of DFU is a sign of other major factors that collectively increase mortality rate (Walsh *et al*., 2016). Upon examining 745 patients with DFI, Son *et al* found that the most isolated causative organism was Methicillin-resistant *Staphylococcus aureus* (MRSA) (13.7%) followed by *E. faecalis* (12.6%) (Son *et al*., 2017). In addition, The most isolated pathogen group in DFU is Gram-positive aerobic bacteria (Kwon & Armstrong, 2018; Nelson et al., 2018; Son et al., 2017) and enterococci species are one of the frequently isolated bacteria from foot ulcers (Hancock *et al*., 2014b). In 2018, a UK study focused on identifying the etiologic agents in 395 patients with DFI found that 43.8% and 14.9% of the patients were infected with *S.aureus* and *Enterococcus* pathogens, respectively (Nelson et al., 2018).

1.8 Phage therapy

The application of phage in the treatment of bacterial infections is known as phage therapy. In hospitals, antibiotic sensitivity tests are routinely performed to check the most suitable antibiotics for patients. However, the increasing number of antibiotic-resistant bacteria necessitates looking for alternative treatments (Górski *et al*., 2018). Phage therapy is a promising treatment to tackle resistant bacteria and this approach has been used in several phage therapy centres. One of these well-known centres is the Eliava Institute of Bacteriophages, Microbiology and Virology (IBMV) in Georgia which was founded by George Eliava in 1923 (Międzybrodzki *et al*., 2018). In 1952, another phage centre, The Institute of Immunology and Experimental Therapy, was established in Poland (Gordillo Altamirano & Barr, 2019) and thousands of patients have been treated with phages in these two centres (Debarbieux *et al*., 2016). Although phages were discovered before antibiotics, the practice of phage therapy was confined to limited countries due to the wide use and efficacy of antibiotics as well as the uncertainty of phage preparations (Debarbieux *et al*., 2016). In 2017, the Department of Health & Human Services (HHS) in the U.S. released the "National Action Plan for Combating Antibiotic-Resistant Bacteria" in which bacteriophage and phage lysins were mentioned as a potential new therapy (HHS, 2017). In 2019, The U.S. Food and Drug Administration approved the first clinical trial of an intravenous route phage therapy carried out by the Centre for Innovative Phage Applications and Therapeutics (IPATH), which is the first centre in North America specializing in phage therapy (LaFee & Buschman, 2019). Even though phage therapy is not yet approved in many countries, Unlicensed medicine can be agreed in special circumstances (Górski *et al*., 2018; Health Canada, 2013).

One of the main concerns regarding phage therapy is about safety and possible adverse effects. Solid evidence is required to allow the notion of introducing a living virus to the human body to be widely accepted. Several studies have concluded that no safety concerns regarding phages when applied to clear bacterial infections. For instance, 39 patients with chronic venous leg ulcers (VLUs) were treated with a phage cocktail (8 phages) for 12 weeks and no adverse effects were observed (Rhoads *et al*., 2014). In the food industry, a virulent phage (P100) targeting *Listeria monocytogenes* was approved by the FDA as a food biopreservative and granted GRAS

(Generally Recognised As Safe) indicating the P100 phage is not harmful to be consumed (Iacumin *et al*., 2016). To further ascertain phage safety, phages used in therapy must lack any virulence or toxin genes as well as genes involved in the lysogenic cycle such as integrase (Henein, 2013). Additionally, phage preparations are monitored for bacterial contaminants, especially the endotoxin which can cause toxic shock (Bonilla *et al*., 2016). Certain rules and guidelines, known as Good Manufacturing Practice (GMP), must be followed to ensure phage sterility and purification (Kakasis & Panitsa, 2019).

Although considered generally safe, phages administrated into a human body could lead to a host immune response. To clarify, phages can be cleared from circulation using the reticuloendothelial system (spleen and liver) by recognizing phage proteins (Abedon & Thomas-Abedon, 2010). Also, antiphage antibodies have been detected in some studies which their activity is based on the route of phage administration and phage type; however, these antibodies did not affect the treatment outcome (Łusiak-Szelachowska *et al*., 2014; Zaczek *et al*., 2016). Some studies showed that the oral administration of the T4 phage led to weak antiphage antibody responses (Majewska *et al*., 2015) while others reported no immune reaction (Bruttin & Bru, 2005).

1.9 The emergence of phage resistance

Although phage therapy has been practised since 1919 (Summers, 2012), higher interest in this treatment has recently regained due to the crisis of AMR. Like antibiotics, the development of phage resistance is inevitable because of the dynamic nature of bacteria (Kortright *et al*., 2019). Therefore, multiple phages (phage cocktails) with different targets can be used to reduce the emergence of resistant mutants (Duplessis *et al*., 2018). For example, Fabijan *et al* treated patients with severe *S. aureus* infections with an intravenously administered phage cocktail (3 phages) which showed no emergence of phage-resistant cells (Petrovic Fabijan *et al*., 2020). In addition, the phage-resistant mutant population can be anticipated based on which bacterial

receptors are being targeted by phages. To illustrate, targeting the capsule of *Acinetobacter baumannii* may lead to the emergence of an uncapsulated population. Based on this notion, Regeimbal *et al* successfully treated mice infected with *Acinetobacter baumannii* by using a phage cocktail (one phage that targets the parent capsulated cells and four phages that infect uncapsulated emerged cells) (Regeimbal *et al*., 2016).

Nonetheless, the development of bacterial resistance to phages can be exploited in the favour of increasing antibiotic sensitivity and decreasing bacterial invasiveness i.e. exerting a cost to fitness in terms of virulence and survival in the environment. To illustrate, a study shows that the MDR *P. aeruginosa* was targeted with a phage, OMKO1, that targets one of the efflux system proteins, the outer membrane porin M (OprM), as a bacterial receptor. As a result of phage selection, the MDR *P. aeruginosa* lost the OprM to survive phage infection and rendered less resistant to different antibiotics classes (Chan *et al*., 2016). Similarly, a highly invasive *L. monocytogenes* pathogen showed a significant reduction in invasiveness (internalization of pathogens into host cells) when it developed resistance to a phage that targets cell wall teichoic acids (Sumrall *et al*., 2019)

1.10 Phage therapy targeting enterococcal infections

Of thirty-five species of enterococci, *E. faecalis* and *E. faecium* are considered the most clinically isolated organisms (Bolocan *et al*., 2019) and in the last two decades, these two species have been targeted to assess phage therapy using various models. For example, Khalifa *et al* tested the effect of a myovirus phage (EFDG1) on different antibiotic-resistant *E. faecalis* and *E. faecium* isolates which showed a significant reduction in the *in vitro* planktonic and biofilm cultures (Khalifa *et al*., 2015). In the same study, the EFDG1 phage was able to prevent an *ex vivo E. faecalis* root canal infection (Khalifa *et al*., 2015). In addition, the number of VRE isolates has increased in the past decade which necessitates investigating other treatment options (Vehreschild *et al*., 2019). In 2021, Kevin et al showed the effect of a phage cocktail in a One-

Year-Old child case in which an abdominal infection caused by a vancomycin-resistant *E. faecium* strain was treated with two phages administered intravenously (Paul et al., 2021). No safety concerns were reported as well as neutralizing anti-phage antibodies were not detected. Moreover, vancomycin-susceptible *Enterococcus faecium* were also isolated following phage therapy indicating a possible cost of fitness mechanism.

1.11 Phage lytic enzymes

During the phage lifecycle, different lytic enzymes are used to make their host infection successful. These lysins include endolysins, holin and tail-associated lysins. For endolysins, phage utilises these enzymes to lyse bacterial cells at the end of the lytic cycle leading to virions release (Fischetti, 2010). As endolysins target the peptidoglycan layer in the cell wall, another phage enzyme, holin, is required to hydrolyse the cell membrane to facilitate endolysin translocation to its target (Wang *et al*., 2003). In 2004, Yoong *et al* purified the lysin (PlyV12) from a phage infecting the *E. faecalis* strain V12 which showed activity against both *E. faecalis* and *E. faecium* strains as well as other Gram-positive bacteria (Yoong *et al*., 2004). Exogenous application of endolysins is effective only against Gram-positive bacteria whereas Gram-negative bacteria are protected due to the outer membrane (Fischetti, 2010). Endolysins attack the peptidoglycan structure using various mechanisms (Elbreki et al., 2014) (Figure 1.11).

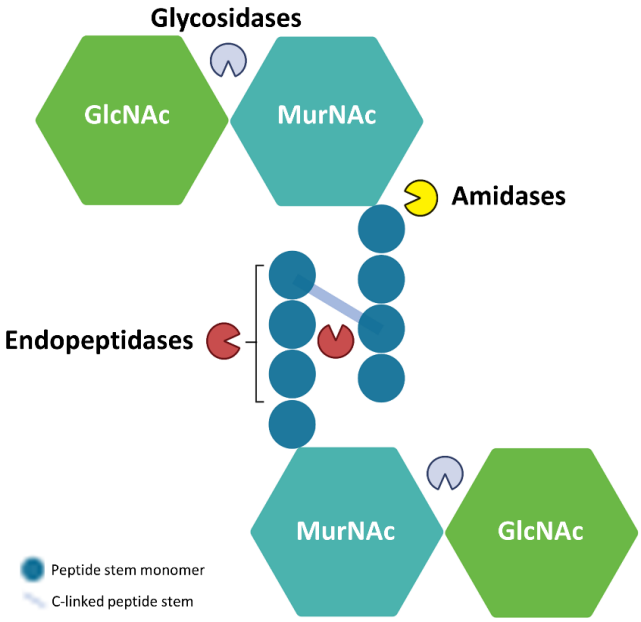


Figure 1.11 Schematic representation of the bacterial peptidoglycan structure and its lysins. These lysins are Glycosidases (grey), Amidases (yellow) and Endopeptidases (red). MurNAc: N-acetylmuramic acids GlcNAc: N-acetylglucosamine.

For domain structure, endolysins are characterised by generally having two domains: The N-terminal domain which contains catalytic activity and the C-terminal which possesses a cell binding domain (Gutiérrez *et al*., 2018). The C-terminal domain differs among endolysins which confer binding specificity to bacterial cell walls (Fischetti, 2010). Other domain organizations are also found such as two EDA and no CBD (Figure 1.12). To assess the activity of endolysins, Loeffler *et al* reported lytic activity of the purified bacteriophage enzyme (PAL) against 15 commonly isolated *Streptococcus pneumoniae* strains, of which some are penicillin-resistant strains (Loeffler *et al*., 2001). Djurkovic *et al* also showed that the endolysin (Cpl-1) can be used synergistically with penicillin to kill penicillin-resistant Streptococcus pneumoniae strains (Djurkovic et al., 2005). Additionally, the lysins showed broader specificity compared with whole phage particles and the lysins have activity at the genus level (Young & Gill, 2015).

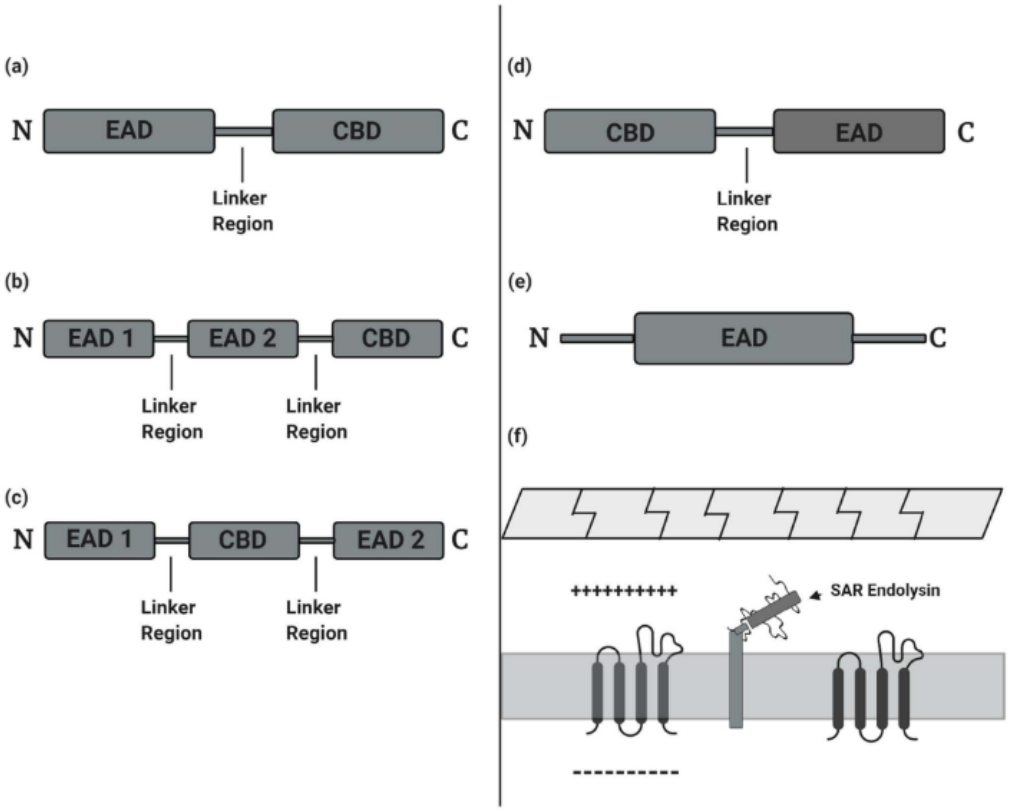


Figure 1.12 Domains structure of phage endolysins. A) structure with a single N-terminal enzymatically active domain (EAD) and a C-terminal cell wall-binding domain (CBD). B) two N-terminal enzymatically active domains (EAD) with a CBD. C) two enzymatically active domains located at N- and C terminals while CBD is suited in the middle. D) N-terminal CBD while EDA is at the C-terminal side. E) A single EDA domain. F) A tethered signal-arrest-release (SAR) endolysin that is anchored to the periplasmic membrane before being released and turning on the lysis function. adapted from (Abdelrahman et al., 2021)

Another type of phage lysins that is also exploited during the first steps of phage infection is generally known as tail-associated lysins (TAL). These TALs can also be named VAPGH or deploymerases based on their targets (Figure 1.13). TALs are found in phages infecting both Gram-positive and negative bacteria (Latka et al., 2017). One of the main targets of these lysins is the cell wall peptidoglycan which can be degraded via Glycosidases (grey), Amidases (yellow) or Endopeptidases (Rodríguez-Rubio et al., 2013) (Figure 1.11). Other bacterial components can also be targeted such as wall teichoic acids (Shi et al., 2008), lipopolysaccharide (Plattner et al., 2019), capsule (Born et al., 2014) or exopolysaccharide (EPS) in biofilms (Gutiérrez et al.,

2015). For domain organizations, these lysins can possess one or multiple EAD alone or with CBD, similar to endolysins (Latka et al., 2017; Roach & Donovan, 2015) (Figure 1.12).

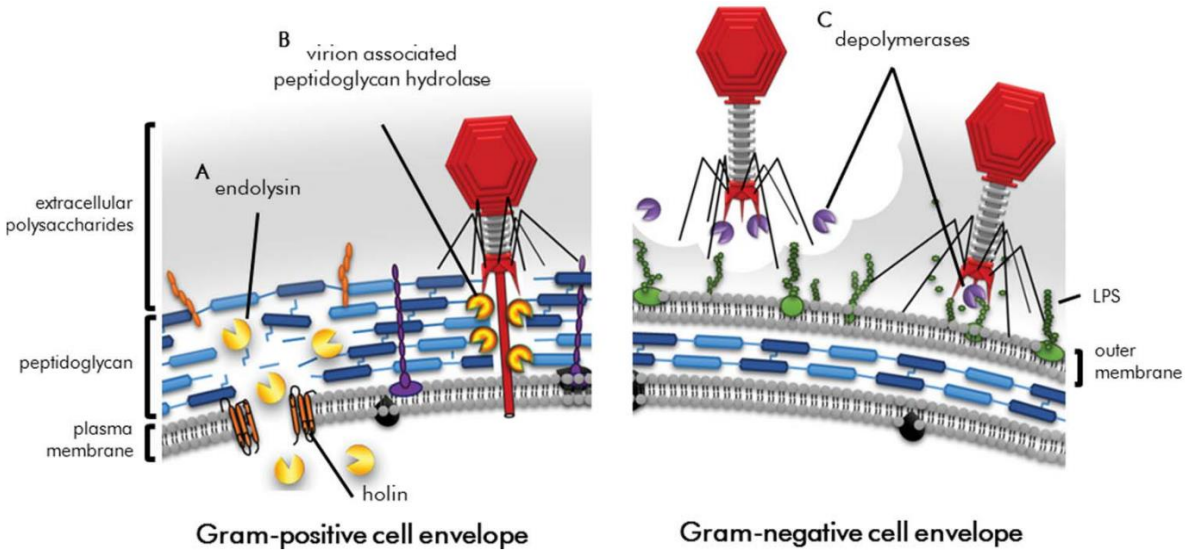


Figure 1.13 Schematic representation of endolysins, holin, VAPGH and depolymerase attacking bacterial cell envelope. Adapted from (Roach & Donovan, 2015)

1.12 Hypothesis and aims

1.12.1 Hypothesis

Novel bacteriophages can be isolated and characterised from wastewater targeting *E. faecalis* and *E. faecium* strains that might have therapeutic potential or contain novel lytic enzymes or cell wall binding proteins.

1.12.2 Aims

Aims are divided into three parts:

Aim 1: Bioinformatic analysis of Tail- associated lysins:

- Obtaining enterococcal phage and prophage genomes
- Reannotating phage genomes using RASTtk
- Scanning tail modules for proteins with lytic domains using Pfam, CDD and Phyre2

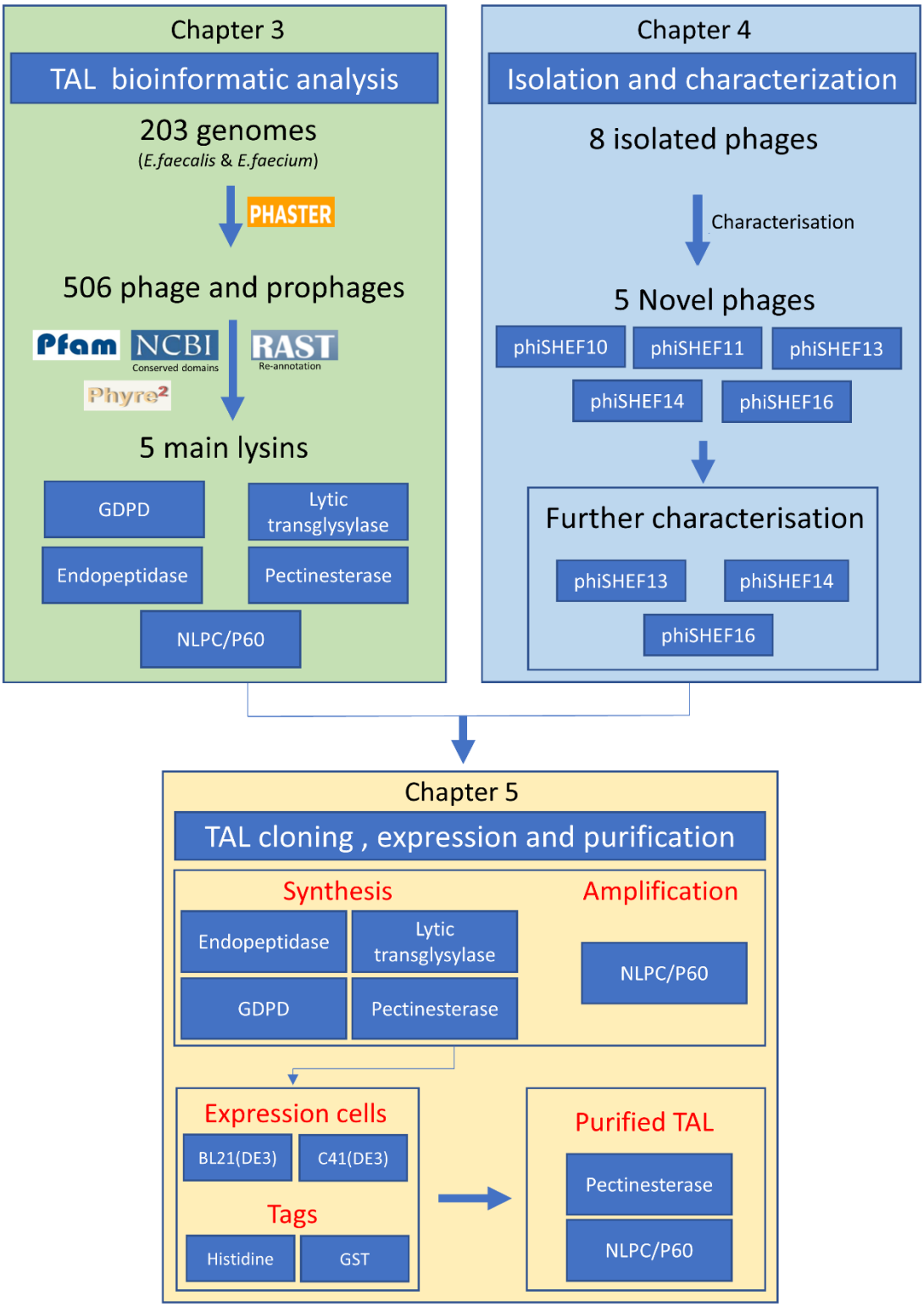
Aim 2: Isolation and characterisation of phages targeting enterococci through:

- Using environmental sources such as wastewater for phage isolation
- Determining the morphological and genomic features of the isolated phages
- Performing host range analysis against various *E. faecalis* and *E. faecium* strains
- Isolating and testing phage resistance mutants
- Investigating potential phage receptor regions

Aim3: *In vitro* investigation of putative TAL proteins

- Selecting candidate lysins
- Cloning of selected genes either by synthesis or amplification
- Expressing and purifying selected proteins using *E.coli* strains

Thesis story



Chapter 2:

Material and Methods

All reagents used in this work were from Sigma-Aldrich Company Ltd, Dorset, UK unless otherwise stated.

2.1 Wastewater collection and processing

Wastewater is one of the major environmental sources for phage isolation. The collection and processing of wastewater were performed according to the protocol of Al-Zubidi *et al*., (2019). Preparation of wastewater samples began with on-site filtration using 3M filter paper to remove large debris. In the lab, the wastewater filtrate was centrifuged (8,000 xg, 10 min) to pellet bacteria and other large particles. The supernatant was filtered with a 0.45 μm syringe filter (Sartorius, Germany) or filter units 0.45 μm and then centrifuged (35,000 xg, 90 min) to pellet phages. The pellet was then suspended in 1 ml SM buffer and stored at 4 °C.

Amicon® Ultra-15 Centrifugal Filter Units (100,000 MWCO) were also used as an alternative method to the ultracentrifugation step (Figure 2.1). These concentrator units were used to concentrate phages in wastewater into a small volume (Bonilla *et al*., 2016). Briefly, the filter units were loaded with the filtered wastewater and centrifuged at (4000 xg, 5 min). The retentate was aspirated and stored at 4 °C.

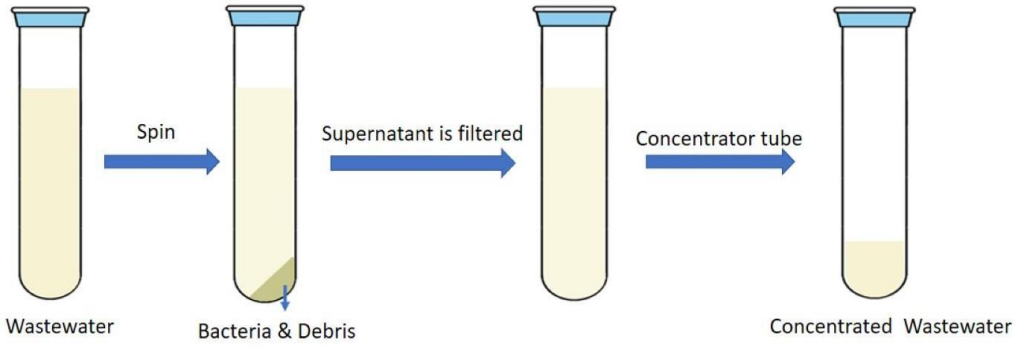


Figure 2.1. Schematic representation of wastewater processing and concentration using concentrator units.

2.2 Bacterial strains used in this study

Enterococcus strains used in this study are described in Table 2.1, 2.2 and 2.3. Eight *E. faecium* strains with various EPA types and an *E. faecalis* OG1RF strain were obtained from Dr Stéphane Mesnage, School of BioSciences, University of Sheffield, UK. Moreover, nine clinical isolates from patients with diabetic foot ulcers were obtained from Dr David Partridge, Northern General Hospital, Sheffield, UK. In addition, several *E. faecalis* strains used in this study named OS16, EF1 ,EF2, EF3 , EF54, JH2-2 were donated by ACTA university, Amsterdam, Holland. Another *E. faecalis* strain named OMGS3919 strain was donated by the department of Oral Microbiology and Immunology, Institute of Odontology, Sahlgrenska Academy, University of Gothenburg, Sweden. All *Enterococcus* strains were cultured aerobically at 37°C using BHI agar culture media (OXOID, UK). Other bacterial strains (*E. coli*) were also used (Table 2.4) as well as plasmid vectors (Table 2.5) and primers (Table 2.6).

Table 2.1 *Enterococcus faecium* strains with distinct EPA variants used in this study

EPA[a] Variants	Bacterial strain	Reported Resistance and virulence genes[b]	Source	MLST
1	E1636	AMP	Clinical isolate (blood)	106
	E1679	AMP, VAN; esp+	Clinical isolate (vascular catheter)	114
2	E1071	VAN	Commensal isolate (faeces)	32
	E4452	AMP	Commensal isolate (faeces)	266
3	E1162	AMP; esp+	Clinical isolate (blood)	17
4	E980	-	Commensal isolate (faeces)	94
	Aus0004	VAN; esp+	Clinical isolate (blood)	17
	U0317	AMP; esp+ hyl+	Clinical isolate (urine)	78

[a] Enterococcal polysaccharide antigen (EPA). [b] Ampicillin (AMP), Vancomycin (VAN), Enterococcal surface protein (esp), hyaluronidase (hyl). Table adapted from (De Been, 2013) with permission.

Table 2.2 Enterococcal clinical isolates from patients with diabetic foot ulcers at the Northern General Hospital, Sheffield, used in this study

Strain	MALDI ID (MLST)	Antibiotic Susceptibilities[a]		
		AMX	TEC	VAN
dp1	*E. faecalis* (197)	S	S	S
dp2	*E. faecalis*	S	S	S
dp3	*E. faecalis* (179)	S	S	S
dp4	*E. faecalis*	S	S	S
dp5	*E. faecalis* (16)	S	S	S
dp6	*E. faecium*	R	S	S
dp7	*E. faecium* (17)	R	S	S
dp8	*E. faecium*	R	S	S
dp9	*E. faecium* (787)	R	R	R

[a]Amoxicillin (AMX), Teicoplanin (TEC), Vancomycin (VAN), Resistant (R), Sensitive (S). The table was obtained from Dr David Partridge, Northern General Hospital, Sheffield, UK.

Table 2.3 Other *Enterococcus Faecalis* strains used in this study

Bacterial source	Strains (MLST)	Reference
Oral rinse-endodontic patient	EF1 (34),EF2 (283), EF3 (97) OS16 (173)	(Sedgley, 2004) (Sedgley *et al*., 2005)
Oral mucosal lesions	OMGS3919 (97)	(Dahlén, 2012)
Non oral human isolate	EF54 (381)	(Toledo-arana *et al*., 2001)
Oral Lab Strain	OG1RF (1)	(Bourgogne *et al*., 2008)
Non oral Lab strain	JH2-2 (8)	(Jacob & Hobbs, 1974)

Data were obtained from (Al-Zubidi et al., 2019)

Table 2.4 *E. coli* strains used in transformation.

Strains	Genotype	Source
DH5α	F – φ80dlacZΔM15 Δ(lacZYA-argF) U169 deoR recA1 endA1 hsdR17(rk–, mk+) phoA supE44 λ– thi-1 gyrA96 relA1	New England Biolabs (NEB)
BL21(DE3)	F – ompT, hsdSβ(rβ–mβ–) dcm gal λ(DE3 [lacI, lacUV5-T7, gene1, ind1, sam7, nin5])	Graham Stafford collection
C41(DE3)	F – ompT, hsdSβ(rβ–mβ–) dcm Ion	Graham Stafford collection

Table 2.5 Plasmid vectors used in this study

Plasmid	Description	Phenotype	source
pET21a	T7 promoter, His Tag, expression vector	Ampicillin resistant	Graham Stafford collection
pET21b	T7 promoter, His Tag, expression vector	Ampicillin resistant	Graham Stafford collection
pGEX4T3	tac promoter, glutathione S-transferase, expression vector	Ampicillin resistant	Graham Stafford collection

Table 2.6 Primers used in this study and their details

Name	Oligonucleotide Sequence
NLPC/P60 F	5’ CATATGGCTATAACAAAAGAAGATTTCGC
NLPC/P60 R	5’CTCGAGTGCACTAGGTGGAATACAAATACTTG
16s rRNA 27F	5′AGAGTTTGATCCTGGCTCAG
16s rRNA 1492R	5′CTACGGCTACCTTGTTACGA

2.3 Molecular biology techniques

2.3.1 Bacterial Genomic DNA extraction

DNA from bacterial cells was extracted using the Wizard Genomic DNA Purification kit. Briefly, 1 ml of overnight culture was centrifuged and the pellet was resuspended in 480 µl of EDTA (50 mM) and 120 µl of lysozyme (10 mg/ml). The sample was incubated at 37 ℃ for 45 min before centrifugation (16,000 x g, 3 min) and removal of the supernatant. Then, 600 µl of Nuclei Lysis Solution was added and the sample was incubated at 80 ℃ for 5 minutes. After the sample was cooled to room temperature, 3 µl of RNase (100 mg/ml) was added and the sample was incubated at 37 ℃ for 30 minutes. Then, 200 µl of Protein Precipitation Solution was added and the sample was vortexed for 20 seconds before it was incubated on ice for 5 min. The sample was then centrifuged (16,000 ×g, 3 min) and the supernatant was aspirated into a new 1.5 ml microcentrifuge tube containing 600µl of room temperature isopropanol. The sample was gently mixed and centrifuged (16,000 ×g, 2 min). The supernatant was removed and the DNA pellet was resuspended in 70% ethanol. The tube was centrifuged (16,000 ×g, 2 min) and ethanol was aspirated. The tube was then left open for 15 minutes to allow ethanol evaporation. The DNA pellet was then resuspended in 50 µl of autoclaved milli-Q-water and stored at -20 ℃.

2.3.2 Plasmid extraction

Plasmids of interest present in the *E.coli* stain DH5a were extracted by Monarch® Plasmid Miniprep Kit. Plasmid purification method according to the manufacturer's instructions.

2.3.3 Phage DNA extraction

To extract phage DNA, 1 µl of DNase I (1 U/µl) and 1µl of RNase A (100 mg/ml) were first added to 1 ml of concentrated phage lysate and incubated at 37 ℃ for 30 minutes. This step is important to ascertain the degradation of externally released bacterial nucleic acids. Then, 100

μg/ml of proteinase K was added to the mixture and incubated at 50 °C for 45 minutes. After that, DNA was separated from the denatured proteins by adding phenol: chloroform: isoamyl alcohol (25:24:1). After centrifugation at 14000 ×g for 5 min, the aqueous phase was aspirated into a new tube and this step was repeated once. Then, two volumes of ice-cold 100% ethanol and 1/10 volume of 3 M sodium acetate were added and left overnight at -20°C to precipitate DNA. Next day, the solution was centrifuged at 16,000 xg for 20 min to pellet DNA and the supernatant was discarded. 70% ethanol was added, and the tube was centrifuged at 14000 ×g for 5 min. The supernatant was discarded, and the tube was left open on the bench for 15 minutes to air-dry. The pellet was then suspended in 50 μl of autoclaved milli-Q-water and stored at -20 °C.

2.3.4 Cloning of putative lysins

The GenSmart™ Codon Optimization online tool was used to codon optimize *in silicon* the identified lysins (TAEP, PE, GDPD, TMP-LT) followed by the addition of suitable restriction sites in accordance with the plasmid vector. The gene synthesis was done by Genewiz in which the subcloning of the identified genes into the vectors pET21a, pET21b, and pGEX4T3 was also completed. Upon receiving the recombinant vectors, the lyophilised vector was suspended in nuclease-free H2O to obtain 100 ng/μL of DNA and stored at -20 ℃ to be used for transformation.

For the NLPC/P60 gene, this was amplified from the phiSHEF14 genome using primers designed with suitable restriction sites in accordance with the plasmid vector (pET21a). The amplified gene was cleaned up using Gel/PCR Extraction Kit (FastGene). To prepare the amplified gene for ligation, a restriction digestion step was first done using two restriction endonucleases enzymes to make the ends of both the insert and vector compatible. This step was explained in 2.3.8. The insert and vector were then ligated using the protocol described in 2.3.10. The vectors were then used in bacterial transformation as described in 2.3.5.

2.3.5 *E. coli* transformation by heat shock

E.coli Transformation was performed by adding 1 µl of a plasmid to 10 µl of competent cells in a 1.5 ml microcentrifuge tube which was followed by 30 min incubation on ice and then a heat shock step at 42 °C for 30 seconds. After 5 min incubation on ice, 900 µl of LB broth was added and the tube was left for 1 hour at 37 °C in a shaking incubator (180 rpm). After spinning at 13,000 rpm for 2 minutes, the pellet was resuspended in 100 µl LB and spread on LB agar containing ampicillin (50 µg/ml). After overnight incubation at 37 °C, one colony was picked and subcultured into a new plate.

2.3.6 Agarose gel electrophoresis

DNA samples were analysed by agarose gel electrophoresis. First, 1x TAE buffer was added to agarose powder to prepare 1% agarose solution which was dissolved by heating. To detect DNA in a gel, 1 µl of ethidium bromide (10 mg/ml) was added to the dissolved agarose before casting and a comb was placed to create wells. After solidification, 1x TAE was used to run gels in a BIORAD mini-sub-cell for small gels or a normal-sub-cell for large gels (BIO-RAD Laboratories, UK). A loading dye (New England Biolabs, UK) was added to DNA samples and a GeneRuler 1 kb DNA Ladder (Thermo fisher scientific, UK) was used to estimate the samples' molecular size.

2.3.7 Extraction of DNA fragments

Following agarose gel electrophoresis, DNA fragments of correct fragment sizes were cut from the gel using a scalpel, then solubilized and purified using the PCR and gel purification kit (BIOLINE) according to the manufacturer's instructions.

2.3.8 Restriction of DNA

The digestion of DNA was done via restriction endonucleases according to the manufacturer's instructions (New England Biolabs, UK). For double digestion, 50 µl reaction volume comprises 1 µg of DNA, 2 µl of each restriction enzyme, 5 µl of the recommended buffer by manufacturer's instruction and distilled nuclease-free water. The reaction was incubated at 37 °C for 2 h.

2.3.9 Dephosphorylation of DNA strand (5'end)

To prevent the re-circularization of plasmid DNA after digestion, calf intestine alkaline phosphatase (CIAP) (New England Biolabs, UK) was used. Briefly, 2.5 µl of CIAP and 5 µl of CIAP buffer were added to the restriction reaction and incubated for 1 h at 37 °C before stopping the reaction at 65 °C for 20 min. A purification step was followed using the PCR and gel purification kit (BIOLINE, UK) according to the manufacturer's instructions.

2.3.10 Ligation of DNA

To perform the ligation, both the insert and vector were first digested with appropriate restriction enzymes. After purification, a 20 µl reaction volume was set up according to the manufacturer's instructions. Briefly, 1 µl of T4 DNA Ligase, 2 µl of 10x Reaction Buffer, 1 µl of vector, 16 µl of insert and nuclease-free H_2O to reach the final reaction volume. This procedure was prepared on ice and the reaction was incubated at 16 °C overnight before performing the transformation experiment.

2.3.11 Preparation of *E. coli* electrocompetent cells

To prepare competent *E.coli* cells, 1 ml from an overnight culture was added to 50 ml LB broth and left at 37 °C shaking (180 rpm) till reach mid-exponential growth phase ($O.D._{600}$ = 0.6-0.8). Bacteria were then centrifuged and pelleted at 4 °C (6000 xg, 10 min) and then resuspended in 10 ml autoclaved ice-cold 0.1 M $CaCl_2$. After spinning at 4 °C for 10 min, the pellet was resuspended in 400 µl ice-cold autoclaved 0.1 M $CaCl_2$ with 100 µl of 50 % glycerol. This was followed by aliquoting into pre-chilled Eppendorf tubes (50 µl volume) and used immediately in the transformations experiment or stored at -80 °C for future use.

2.3.12 Protein expression

To perform the protein expression, an overnight culture of BL21(DE3) or C41(DE3) containing the desired vector was first prepared in 10 ml LB broth with ampicillin (50 µg/ml). This overnight culture was diluted 1:100 in 10 ml 2-TY medium and left in a shaking incubator (180 rpm) at 37 °C until reaching the mid-exponential growth phase ($O.D._{600}$= 0.6-0.8). At this point, 0.2 mM IPTG (induction) was added and the culture was left in the shaking incubator at either (37 °C for 3 h) or (30 °C for 12 h). 1 ml (before and after induction,) was taken and centrifuged at 13,000 xg for 2 min before keeping the pellet at -20 °C until testing protein expression. The pellet was resuspended in 2x SDS buffer, heated for 15 min at 95 °C and loaded (10 µl) into an SDS-PAGE to check protein expression. After overexpression was confirmed in a small volume (10 ml), a scale-up experiment was performed (500 ml). The Induction temperature and time was 25 °C for 16 h.

2.3.13 Protein solubility

After protein overexpression was confirmed, the protein solubility was checked to determine if the protein of interest is soluble or insoluble. After the protein induction was complete, bacterial cells (10 ml) were pelleted via centrifugation at 4 °C (6,000 xg, 10 min) and resuspended in 1 ml

50 mM sodium phosphate buffer (150 mM NaCl) containing protease inhibitor cocktail (cOmplete). To burst cells, a sonicator was used 3 times (20 seconds) with 30 seconds of incubation on ice between each time and the cells were kept on ice at the end. This was followed by a short centrifugation (1000 x g, 1 min, 4 °C) to pellet cell debris. The supernatant then was further centrifuged (10,000 xg, 10 min, 4 °C) to get the soluble fraction (supernatant) and the insoluble fraction (pellet). The cell debris pellet was resuspended in 200 µl phosphate buffer while the insoluble pellet was in 100 µl phosphate buffer. These fractions were then run on a gel to check protein solubility. For the 500 ml protein expression volume, 25 ml sodium phosphate buffer was used. After sonication, the soluble fraction (supernatant) was obtained by directly spinning the whole volume (25 ml) at (10,000 xg, 10 min, 4 °C).

2.3.14 His-tagged recombinant protein purification and dialysis

His-tagged proteins were purified from cell lysate via affinity chromatography using Ni-NTA metal chelate resin (QIAGEN). In these recombinant proteins, the 6x histidine tag placed at the C-terminal functions as electron donors on the surface of the protein and binds reversibly to the transition Ni2+ ion in the resin (Spriestersbach, Kubicek, Schäfer, Block, & Maertens, 2015). Therefore, this facilitates the protein of interest to be bound, washed and eluted. The resin (0.5-1 ml) was first treated with sodium phosphate buffer to remove ethanol (preservative) by centrifuging at 1,000 xg for 1 min. The resin was then loaded into the purification column and equilibrated with 10 ml of binding buffer (50 mM sodium phosphate + 5 mM imidazole). After the binding buffer is removed, the sample (25 ml) was then added to the column and the flow-through was collected. The washing buffer (50 mM sodium phosphate + 10 mM imidazole) was then added to remove nonspecific proteins. Next, the elution step was carried out by using a higher concentration of the competitive counter-ligand imidazole (50 mM sodium phosphate + 500 mM imidazole).

2.3.15 Polymerase chain reaction (PCR)

A PCR reaction was used to amplify a specific gene in either bacterial or phage DNA samples. Q5® High-Fidelity DNA Polymerase (New England Biolabs, UK) was used according to the manufacturer's instructions in 25 µl reaction volumes. These reactions consist of 5 µl of 5X Q5 Reaction Buffer, 0.25 µl of Q5® High-Fidelity DNA Polymerase, 1.25 µl of forward primer (10 µM), 1.25 µl of reverse primer (10 µM), 0.5 µl of dNTPs (10 mM), 1 µl of DNA and 15.75 µl of nuclease-free water. Q5® High-Fidelity 2X Master Mix was also occasionally used. The PCR reaction conditions are described in (Table 2.7).

Table 2.7 PCR thermal cycling conditions.

Step	Temperature °C	Time	Number of cycles
Initial denaturation	95	4min	1
Denaturation	95	30sec	40
Annealing	Primer temp -5	30sec	
Extension	72	30sec	
Final extension	72	2m	1

2.3.16 Colony PCR for the 16S ribosomal RNA gene

A bacterial colony was picked and suspended in PBS (50 µl) which was followed by heating at 95 °C for 10 min. The heated solution was centrifuged (11,000 xg, 1 min) and the supernatant was used as a DNA template in PCR reactions. To detect the amplification of the 16S rRNA gene, agarose gel electrophoresis was first performed using the PCR samples. DNA bands with the correct size were visualised by a UV transilluminator and excised using a scalpel. Then, the

excised bands were solubilised and purified using the PCR and gel purification kit (Bioline, UK) according to the manufacturer's instructions.

2.3.17 Sodium dodecyl sulphate Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse the molecular size of denatured proteins. Two gels with different acrylamide concentrations were used: 5% stacking and 12% resolving gels. Precast gels were also used occasionally. The gels were stained by Coomassie blue. The preparation of resolving gel and stacking gel was described in (Table 2.8).

Table 2.8 Preparation of resolving gel and stacking gel.

Components	**Resolving gel 12% (lower)**
Distilled water	4.3 ml
40% (w/v) acrylamide	3 ml
Upper resolving gel buffer: 18.17 g Tris Base, 0.4 g SDS dissolved in dH_2O, adjusted to pH 8.8 with NaOH, total volume 100 ml	2.5 ml
TEMED (Tetramethylene diamine)	5 µl
10% Ammonium persulfate (fresh)	350 µl
Components	**Stacking gel 5% (Upper)**
Distilled water (dH_2O)	4.7 ml
40% (w/v) acrylamide	0.975 ml
Lower resolving gel buffer: 6.06 g Tris Base, 0.4 g SDS dissolved in dH2O, adjusted to pH 6.8 with HCl, total volume 100 ml.	2.1 ml
TEMED (Tetramethylene diamine)	17 µl
10% Ammonium persulfate (fresh)	100 µl

The PROTEAN Tetra Cell (BIO-RAD Laboratories, UK) and 1x SDS-PAGE running buffer (25 mM Tris Base, 250 mM glycine, and 0.1% w/v SDS) were used to mount the gel after it had been set up. The protein samples were prepared by mixing an equivalent amount of the sample with 2x SDS lysis buffer (60 mM Tris HCl, pH 6.8, 2% SDS, 20% glycerol, 0.02% bromophenol blue mixed with 1 M DTT), followed by heating at 95 °C for 10 minutes. The EZ-Run™ Prestained Rec Protein Ladder (Fisher, UK) or Prime-Step™ Prestained Broad Range Protein Ladder (Fisher, UK) were used. Electrophoresis was then performed at a constant voltage of 140 V until the tracking dye had moved to the gel's bottom.

2.3.18 Bicinchoninic acid (BCA) assay

BCA assays were performed to calculate protein concentrations. Pierce BCA assay kit (Thermo-Fisher Scientific) used in accordance with the manufacturer's instructions. The Bovine Serum Albumin (BSA) was used as standard and a standard curve was drawn to determine the sample concentration. All steps were done in accordance with the manufacturer's instructions.

2.4 Phage techniques

2.4.1 Enrichment of bacteriophage

Phages in wastewater samples were enriched to facilitate phage isolation. Phage enrichment was performed using single or multiple hosts (Figure 2.2).

2.4.1.1 Single host

30 µl of concentrated wastewater samples were mixed with 10 ml of exponentially growing indicator bacteria and incubated overnight at 37 °C with shaking. The enriched sample was then

centrifuged (4,000 x g, 10 min, 4 °C) and the supernatant was filtered with a 0.45 µm syringe filter and stored at 4 °C.

2.4.1.2 Multiple hosts

In contrast to enriching phages from wastewater with a single bacterial host, concentrated wastewater samples were added to a bacterial culture consisting of 4 strains to increase the phage's likelihood of encountering their hosts (Hyman, 2019). All 4 strains were first individually cultured till reaching the mid-log phase. Then, 1 ml of each strain was added to a 26 ml BHI broth containing 1 ml of concentrated wastewater. The mixture was incubated overnight at 37 °C with shaking. Next day, the enriched sample was then centrifuged (4,000 xg, 10 min) and the supernatant was filtered with a 0.45 µm syringe filter and stored at 4 °C.

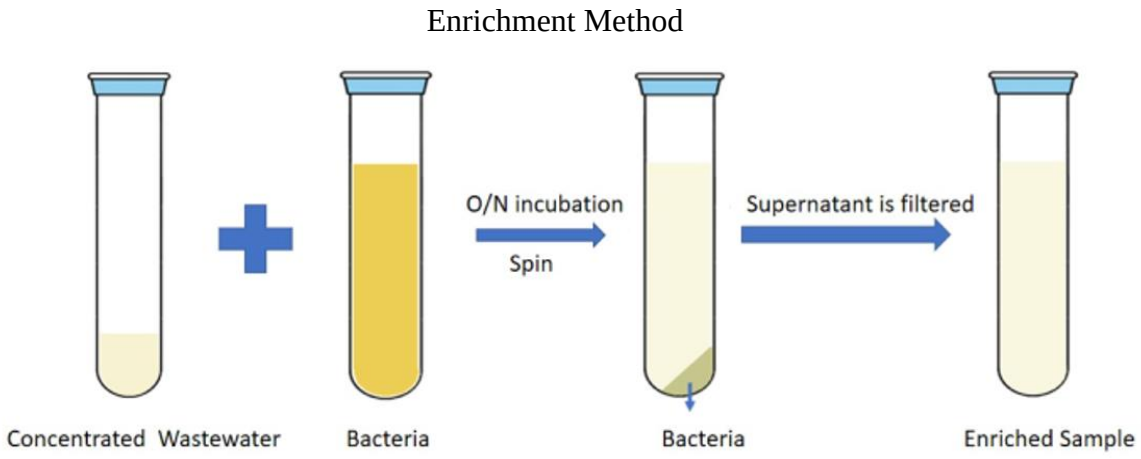


Figure 2.2 Schematic representation of the enrichment step.

2.4.2 Phage isolation

In order to isolate phages on agar plates, a double-layer agar was prepared. This involves two layers of agar: bottom solid agar (1.5% agar concentration) and top soft agar (0.4% agar). There are two main ways to isolate phages from enriched samples: the plaque assay and spot test.

2.4.2.1 Plaque assays

The plaque assay was performed by adding 5-50 µl of enriched samples or phage samples to 200 µl of overnight-grown indicator bacteria. Then, the mixture was added to a 4 ml molten soft agar (50 °C) and immediately poured onto a bottom solid agar (Figure 2.3). After overnight incubation at 37 °C, phage infection was observed as plaques (zones of bacterial lysis). The plaque assay can be used to purify and quantify phages.

2.4.2.2 Spot tests

For the spot test, 200 µl of indicator bacteria were mixed with 4 ml molten top soft agar and then poured onto a solid 1.5% bottom BHI agar (Figure 2.3). After the top agar solidifies, 5-10 µl of enriched samples or phage samples were spotted onto the upper agar and plates were incubated overnight at 37 °C.

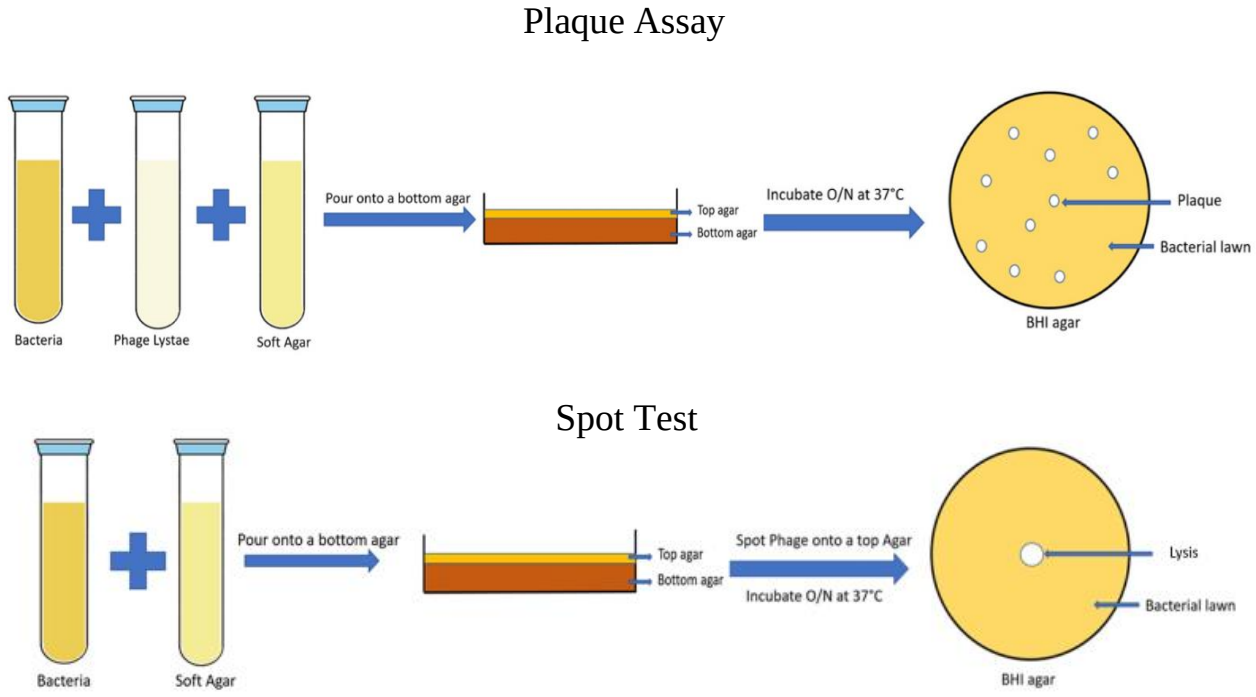


Figure 2.3 schematic diagrams of the Plaque assay and the Spot test.

2.4.3 Phage purification

To ensure isolating a single phage strain, a well-isolated plaque was picked with a sterile loop into a 500 µl BHI broth which was stored at 4 °C. The suspension was centrifuged, and the supernatant was filtered with a 0.45 µm syringe filter. 100 µl of the filtrate was mixed with 200 µl indicator bacteria. The phage-bacterial mixture was then mixed with 4 ml soft top agar and poured onto a solid bottom agar (plaque assay). The plates were incubated overnight at 37 °C to allow plaque formation. The phage purification step was accomplished by performing three successive pickings of a single plague. Then, the plaque from the last round was picked and suspended in 500 µl BHI broth or PBS and stored at 4 °C.

2.4.4 Preparation of Phage lysate

There are two main ways in which phages can be propagated to a high titer: liquid and plate lysates.

2.4.4.1 Liquid lysate

10 µl of phage suspension was added to 30 ml of early-log phase indicator bacteria (in BHI) and incubated overnight at 37 °C with shaking. The culture was then centrifuged at 5,000 xg for 15 min and the supernatant was filtered with a 0.45 µm syringe filter. The filtrate was stored at 4 °C.

2.4.4.2 Plate lysate

To prepare plate lysate, 4 ml PBS was poured onto the lysed bacterial lawn and left for 30 min at room temperature with shaking. Then, PBS was collected and centrifuged at 5000 xg for 5 min. The supernatant was filtered and stored at 4 °C.

2.4.5 Phage Host range

Various *E. faecalis* and *E. faecium* strains were tested for phage killing by using the double-layer agar technique. The strains were firstly tested by spot tests which 5 µl of phage (10^7 PFU/ml) were spotted on bacterial lawns and incubated at 37 °C. Next day, all lawns with lysis are further confirmed by spotting serially diluted phage 5µl (10^7,10^6, 10^5,10^4, 10^3 PFU/ml) and then incubating at 37 °C. On the following day, observing individual plaques indicates phage infection and lysis and rules out the lysis from without phenomenon. The phage titer is determined by counting the plaques and calculating the efficiency of platting (EOP) in percentage as follows: PFU/ml of phage on the test strain divided by the PFU/ml of phage on the isolating host.

2.4.6 Killing assays

The impact of phage on planktonic bacteria and their growth curve was tested. First, an overnight bacterial culture was diluted to $O.D._{600}$= 0.05 and 200ul (2 x 10^6 CFU) was loaded in a 96-wells plate. For the Phage, 20 µl were added to achieve a multiplicity of infections (MOI) of 10, 1 or 0.1 from phage titers 10^9, 10^8 and 10^7 PFU/ml, respectively. The plate was sealed and incubated overnight at 37 °C with shaking using the Tecan Sunrise Microplate Reader. The absorbance readings were then analysed using Excel and GraphPad Prism. For testing phage-resistant mutants, 15 µl of phage titers 10^9, 10^8 or 10^7 PFU/ml were added to 150 µl bacterial culture ($O.D._{600}$= 0.05). For phage-resistant mutants, MOI of 1 was used to generate the clones. After incubation, the culture was streaked on BHI agar which 20 colonies were selected and tested on bile esculin agar to further assess isolated clones.

2.4.7 Adaptation assays

To broaden the phage host range, adaptation assays were performed using different strains (E1071, dp9, V583 and 14RM0). After overnight incubation, all liquid cultures were diluted in

BHI broth to $O.D._{600}$= 0.1. For the experiment, 2.7 ml of either V583, dp9 or 14RM0 was mixed with 0.3 ml E1071. phiSHEF14 was added at MOI of 0.1 and the culture was left overnight in the shaking incubator at 37 °C. Next day, the culture (V1) was centrifuged and the supernatant was filtered. For the filtered supernatant, 50 µl was used for plaques assays to assess either phage titer on E1071 or phage isolation on the other strains. Another passage was carried out by adding 500 µl of the filtered supernatant to the 3 ml mixed culture. After overnight incubation, the culture (V2) was also treated as the V1 culture. This was continued for 7 passages.

2.4.8 Phage concentration by precipitation with Polyethylene Glycol (PEG 8000)

To precipitate phage particles, 10% PEG 8000/1M of NaCl was added to a phage lysate and left on ice for 1 hour. Then, phages were sedimented at 10,000 rpm for 30 minutes and the pellet was resuspended in 2 ml PBS. To remove the PEG, an equal volume of chloroform was added, and the solution was centrifuged at 5,000 xg for 10 min. The upper layer was aspirated into a new tube and used to determine phage titer.

2.4.9 Transmission Electron Microscopy (TEM)

4 µl of phage suspension was placed on carbon-coated copper grids for 5 min and then withdrawn using filter papers. The grids were then negatively stained with 2% (wt/vol) uranyl acetate (4 µl) pH 4 for 1 min. The stain on the grid was withdrawn with the help of filter paper. Particles were visualised using FEI Tecnai Transmission Electron Microscope at an accelerating voltage of 80 Kv. Electron micrographs were recorded using a digital camera and Digital Micrograph software.

2.4.10 Phage DNA sequencing

Extracted phage DNA was tested for quantity and quality by nanodrop (Nanodrop 2000; Thermo Scientific). Also, DNA samples were analyzed on agarose gels to ensure genome integrity. Then, phage genomes (50 ng/µl) were sent for sequencing at (MicrobesNG, Birmingham, UK) using Illumina. Some phage genomes were also sequenced using Oxford Nanopore Technology (ONT) at The Sheffield Bioinformatics Core Facility, University of Sheffield, UK. Annotation was done using Prokka 1.13 (Seemann, 2014). Multiple genome alignment was performed by BLASTn and visualised by EasyFig version 2.2.3 (Sullivan, Petty, & Beatson, 2011). PHASTER (PHAge Search Tool Enhanced Release) was used to confirm annotation (Arndt *et al*., 2016).

2.5 Bioinformatic analysis

2.5.1 Phage and prophage genomes

One hundred complete enterococcal phage DNA genomes available on the NCBI GenBank database were obtained as Genbank and Fasta sequences (up to 11/10/2020). The search for these genomes was done on the NCBI virus portal by using "bacteriophage" for virus choice, "Genbank" sequence type, "complete" for genome sequence and "Enterococcus" for host. For prophage genomes 203 complete *E. faecalis* and *E. faecium* bacterial genomes available on the NCBI GenBank database were obtained as Genbank sequences (up to 10-10-2020). The online web server PHASTER (Arndt et al., 2016) was used to identify putative intact prophages and their Fasta sequences were obtained for the annotation step. All phage and prophage genomes were re-annotated to ensure annotation consistency using RASTtk (new version of Phage Rapid Annotation using Subsystem Technology (RAST) pipeline (Brettin et al., 2015).

2.5.2 TAL identification and analysis

The tail module was identified between the head and lysis modules in most of the phage and prophage genomes based on RASTtk annotation. All tail proteins were checked for TALs using Pfam and NCBI conserved domains (CDD) databases. Structural analysis was also performed using the PHYRE2 webserver (Kelley et al., 2015). SnapGene software version 5.3.2 and Artemis (Carver et al., 2012) were used for genome scanning. The identified TAL proteins were aligned using ClustalW (genome.jp) and MultAlin webservers (Corpet, 1988). Phylogenetic trees were constructed using FastTree (genome.jp) and visualised using the ITOL online website (Letunic & Bork, 2021) (https://itol.embl.de/). To check putative peptidase classifications, the MEROPS database was employed (http://www.ebi.ac.uk/merops/) (Rawlings et al., 2018) while the CAZy (Carbohydrate Active Enzymes) database (CAZy; http://www.cazy.org) (Lombard et al., 2014) was used for predicted Glycoenzymes. Genome size analysis was performed using GraphPad Prism version 7 ,San Diego, California USA, www.graphpad.com.

2.6 Bacterial antagonistic assays

The cross-streaking assay was done as follows: an overnight culture was first streaked on BHI plates using a sterile swab. A perpendicular streaking was then followed from bacterial colonies using sterile loops. Plates were then incubated at 37 °C until the next day. For the spotting assay, an overnight culture was first streaked on BHI plates. A cell-free supernatant was obtained via centrifugation (5,000 xg, 5 min) and filtration (0.45 μm) and immediately spotted (3 μl) on each streaked strain. Plates are then incubated at 37 °C until the next day.

Chapter 3: The analysis of TALs in enterococcal phage and prophage genomes

3.1 Introduction

Phages are viruses that specifically infect bacteria which leads to cell lysis. This lysis is achieved via specific phage proteins that facilitate phage particles to be released from the inside out. There is also another type of phage lysins that are involved in the first steps of phage life cycles which facilitate phage adsorption and genome ejection (Rodríguez-Rubio et al., 2013). These lysins are mainly associated with the tail structure and therefore are commonly known as tail-associated lysins or TALs.

TALs have been investigated and analysed in several reviews that analyse phage targeting both Gram-negative and positive bacterial spp. (Drulis-Kawa et al., 2015; Oliveira et al., 2018; Pires et al., 2016). However, an investigation of TAL of phage targeting enterococcal species has still not been performed. Therefore, we aimed in this chapter to analyse enterococcal phage and prophage genomes and in particular focus on the TALs.

The phages' genomes are available on the NCBI database. Therefore, enterococcal phage genomes can be obtained as either FASTA or Genbank files. For prophages, these can be identified within bacterial genomes using the webserver PHASTER. After obtaining the genomes, a re-annotation step is performed to ensure genome annotation consistency which is done by the RASTtk pipeline.

A special characteristic of phages is that their genomes are organized into modules (Moura de Sousa et al., 2021). To clarify, this means that the genes with similar functions are located adjacent to each other. Therefore, scanning tail proteins for TALs can be achievable. Different analytical tools can be used in identifying TALs like Pfam, NCBI conserved domain and Phyre2.

The identified TALs are then analysed for domain architecture diversity and predicted activity. The abundance and location of TALs can also be assessed. TALs of phage and prophage can also be compared and analysed.

Specific objectives:

- Obtaining and reannotating enterococcal phage genomes
- Identifying and reannotating enterococcal prophage genomes
- Analysing the general features of phage and prophage genomes
- Identifying TALs in phage and prophage genomes

NB- The contents of this chapter have been published as of Feb 2023, and is appended at the end of this thesis:

Enterococcal bacteriophage: A survey of the tail associated lysin landscape

Alrafaie AM, Stafford GP. Virus Res. 2023 Apr 2;327:199073. doi: 10.1016/j.virusres.2023.199073. Epub 2023 Feb 22.

3.2 General analysis of enterococcal phage and prophage genomes

Using the NCBI virus portal, the genome sequences of 100 phage targeting either *Enterococcus faecalis* or *Enterococcus faecium* were collected for analysis. These comprised 86 phages isolated using *E. faecalis* strains while 10 phages were isolated using *E. faecium* strains. For predicted prophages, a total of 203 *E. faecalis* & *E. faecium* complete genomes were scanned using the PHASTER prediction tool with the default parameters set for classification of "intact" prophage set at (>90%), "questionable" (scoring 70-90%) and "incomplete" (scoring<70%). In this study we focused on the intact prophages as these have the highest confidence level to maintain a full set of functional modules and allow tail module identification. The PHASTER searches revealed 406 intact prophages in both *E. faecalis* and *E. faecium* bacterial genomes, meaning that in total this study examined 506 phage and prophage genomes.

3.2.1 Enterococcal phages

The 100 isolated phage genomes showed a variation in size from 16.9 to 156.5 kb (Figure 3.1). Our analysis showed that the phage genomes can be categorised into three main groups based on genome size and phage virion morphology. Phages with small genomes (<30.5 kb) are generally podoviruses (Rountreeviridae), medium-sized genomes (31-86.3 kb) siphoviruses (Efquatrovirus, Phifelvirus, Saphexavirus and Andrewesvirinae), and large genomes of over 130 kb myoviruses (Herelleviridae). There were some exceptions, for example, the smallest *Enterococcus* phage EFRM31 (16.9 kb) is a siphovirus (unclassified according to the current ICTV classification) with an isometric head and long non-contractile tail (206 nm tail length), whose genome has 35 predicted ORFs (Open Reading Frames) (Fard et al., 2010). Other examples of note within the podoviruses (Autographiviridae) are the *E. faecalis* phages EFA-1 (40.7 kb) and EFA-2 (39.9 kb) which have higher GC contents and number of ORFs (EFA-1 is 50.14%, 52 ORFs and EFA-2 is 48.55%, 49 ORFs) compared to the average value for enterococcal podoviruses (Autographiviridae and Sarlesvirinae) analysed in this study (35.1%, 30 ORFs).

Regarding morphology, all the 100 enterococcal phages are predicted morphologically to be either podoviruses (short-tailed), siphoviruses (long non-contractile tail) or myoviruses (contractile tail) based on database entries. While the morphological categorisation of podo-, myo- and sipho-virus has been widely used for many years, the recent increase in genomic information has identified a number of differences and allowed continual improvement of phage taxonomy (Turner et al., 2021). However, we will in some places use the commonly used morphological terms to simplify discussions. All of the 18 small genomed predicted podoviruses are classified as Copernicusvirus or Minhovirus within the Rountreeviridae or belong to Autographiviridae according to the new ICTV classifications and have a genome size of 17.9 to 40.7 kb (Figure 3.1) (Turner et al., 2021). The number of ORFs encoded in these genomes ranged from 22-52 with an average of 30.

Siphoviruses make up 64% of the isolated phage with genomes ranging from 16.9 to 86.3 kb. Based on the genome size and TAL analysis, siphoviruses can be classified into two groups: group-1 (21-43 kb, Efquatroviruses or Phifelviruses) and group-2 (55-86 kb, Saphexavirus or

Andrewesvirinae) (Figure 3.1). The average number of predicted ORFs in group-1 is 62 while this is 104 for Group-2. Lastly, we analysed 18 myovirus-type genomes (Herelleviridae, Schiekvirus or Kochikohdavirus) where the genomes varied from 130.9 to 156.5 kb (average 146.5 kb). The new classifications are further supported by these data since the genome sizes alone can indicate likely species membership according to our data.

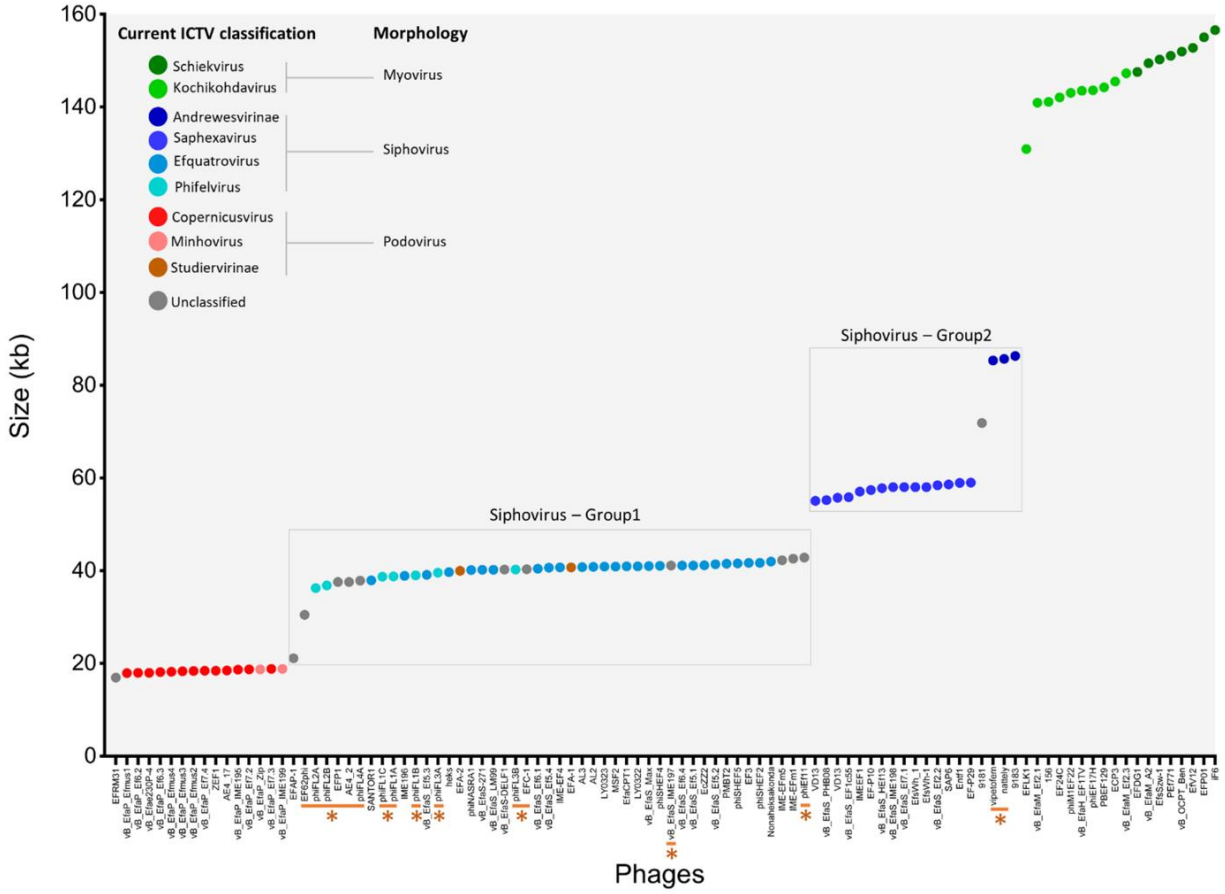


Figure 3.1 100 enterococcal phage genomes were plotted against genome size. The genomes are labelled in accordance with the new ICTV classification as follow: Schiekvirus (dark green), Kochikohdavirus (light green), Andrewesvirinae (dark blue), Saphexavirus (blue), Efquatrovirus (Azure), Phifelvirus (sky), Copernicusvirus (red), Minhovirus (orange), Studiervirinae (brown). The grey colour indicates unclassified genomes regarding the current ICTV classification. Phage morphologies are also included according to the ICTV classification. Temperate phages are underlined and labelled with asterisks.

Unsurprisingly, a positive correlation was seen between phage genome size and the number of ORFs with the small podoviruses having the lowest number of ORFs and the largest genomes (myoviruses) having the highest ORFs number (Figure 3.2A). The number of tRNAs also shows a positive correlation with the genome size, with podovirus genomes having no tRNAs while larger genomes of siphoviruses and myoviruses contain several putative tRNA genes (Figure 3.2B). In contrast, there is no clear correlation between the genome size and GC content (Figure 3.2C).

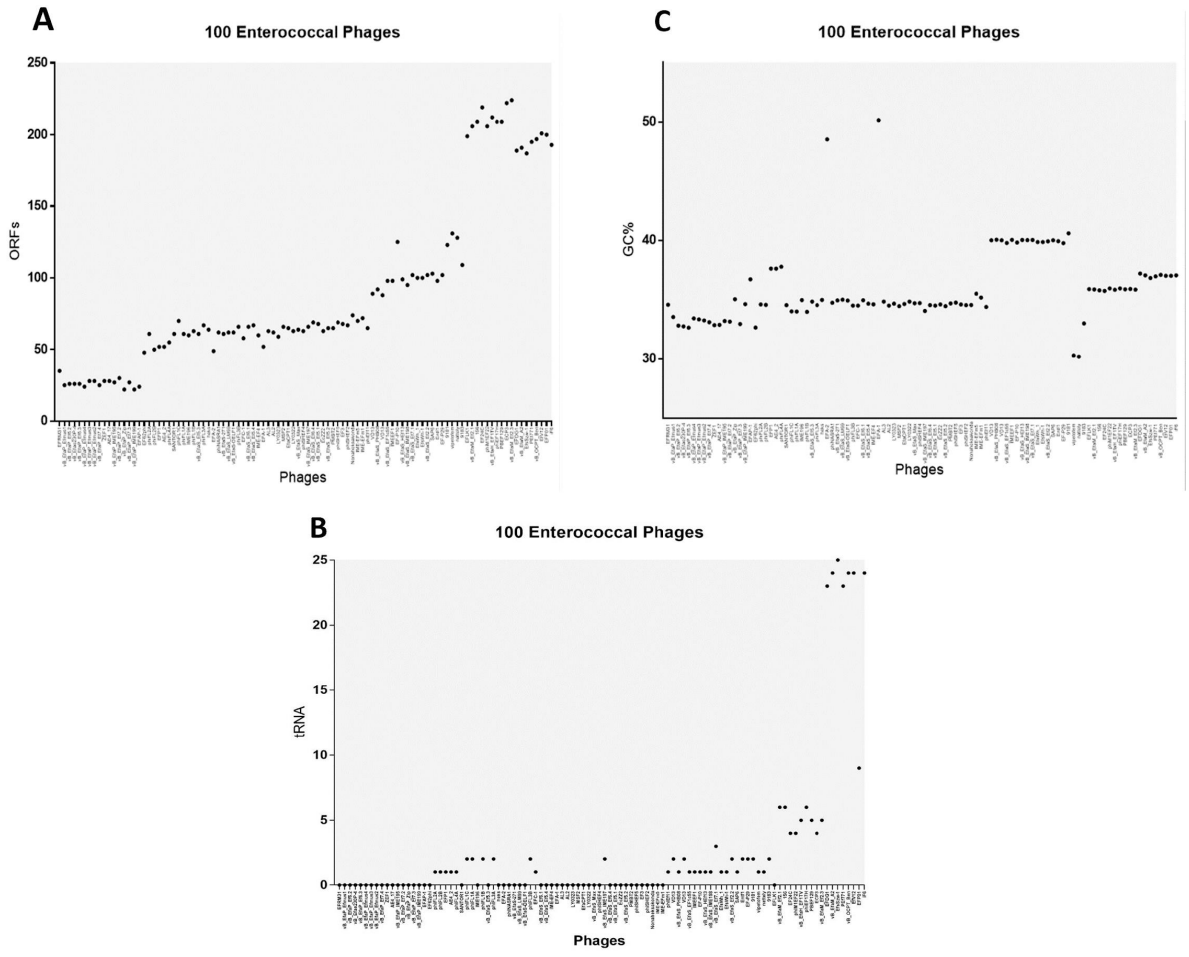


Figure 3.2 100 phage genomes were plotted against (A) ORFs, (B) tRNA and (C) GC%. The genomes are in ascending order in both Figures. The genomes are in ascending order.

Of the 100 enterococcal phage genomes we identified several temperate phages, based on the presence of integrase and repressor genes that are necessary for phage integration and maintenance during the lysogenic cycle into the bacterial genome. In our data, 16% of viruses are likely to be temperate, as they contain integrase and/or repressor genes (Figure 3.1).

3.2.2 Prophage genomes

The term "prophages" refers to phage genomes that are integrated into bacterial genomes. For prophages, only predicted intact prophage genomes were chosen and analysed. In our study, a total of 406 putative intact prophages (93 from *E. faecalis* & 313 from *E. faecium* genomes) were identified, with the most in one genome being five from 203 genomes that were scanned. These showed large variation in the predicted genome size (6.9 - 91.1 kb) (Figure 3.3A), with the smallest prophage containing 10 ORFs and the largest 121 ORFs with an average GC content of 35.9% (Figure 3.3D). It is worth mentioning here that not all identified intact prophage genomes possessed all necessary genes to complete the phage lifecycle indicating a limitation of the PHASTER webserver. Our analysis also showed a positive correlation between the number of ORFs and prophage genomes size (Figure 3.3B). The analysis of the number of tRNA genes showed no correlation (Figure 3.3C).

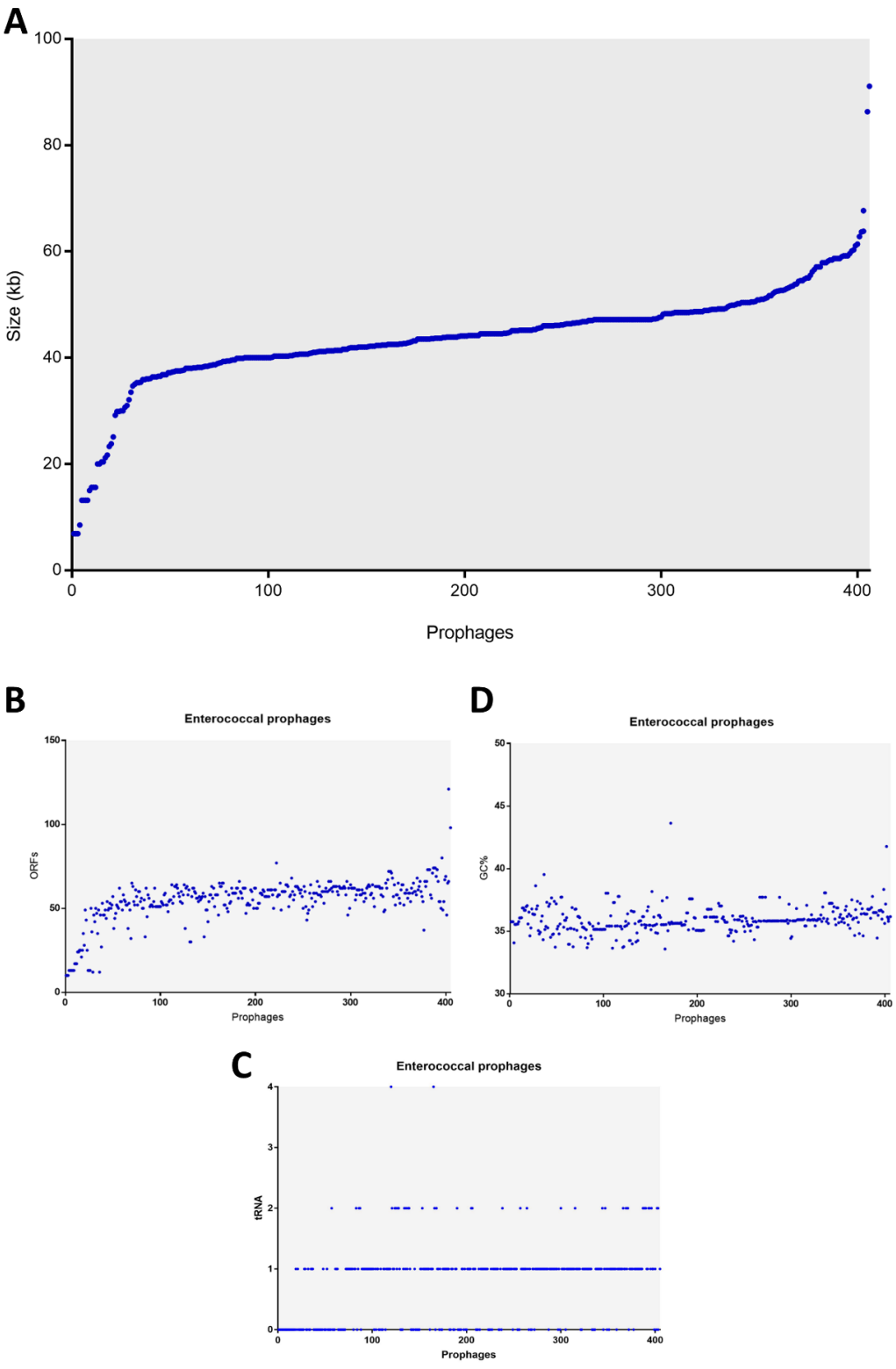


Figure 3.3 406 intact prophage genomes were plotted against (A) genome size, (B) ORFs, (C) tRNA and (D) GC%. The genomes are in ascending order.

3.3 Investigation of Tail-associated lysins

Phage genomes are generally organised in modules where related functional genes are grouped together such as packaging, head, tail and lysis functions. For example, the tail module of siphovirus type phage is considered to generally comprise of three main genes in the following order: “Tape measure protein”- TMP, “Distal tail protein”- Dit and “Tail associated Lysin”- Tal (Goulet et al., 2020) (Figure 3.4B). In this study, the term “TAL” means any lysin in the tail module while “Tal” is referring to the third gene in the siphovirus tail unit.

The TMP is usually one of the longest genes in phage genomes, and plays a role in controlling tail length, with the length of its translated protein approximately indicating the length of the phage tail (1 amino acid= 0.15nm) (Mahony et al., 2016). The TMP also helps facilitate genome ejection toward the bacterial cytoplasm, although mechanistic details are unclear (Mahony et al., 2016). This is evidenced by identifying domains with potential cell wall degrading function in TMPs as well as DNA-binding domains (Piuri & Hatfull, 2006; StockDale et al., 2013). TMPs are thought to be located in the lumen of the tail and interact with termination and initiation proteins as well as the polymeric Major Tail Protein (MTP)(Cornelissen et al., 2016; Kizziah et al., 2020).

The Dit is part of the baseplate and connects the tail with the tail tip as well as providing in some cases the site of attachment for a RBP “receptor-binding protein”- which may be housed on a fibrous protein (Kizziah et al., 2020). The RBPs are responsible for the specific recognition of bacterial receptors that may include outer membrane proteins, bacterial capsule, teichoic acids, pili and flagella (Bertozzi Silva et al., 2016; Letarov & Kulikov, 2017). In this study we assessed all genes in the putative tail modules (not only the putative Tal) for the presence of lysin-like domains, so as not to exclude any that might be associated directly with TMPs, RBP or tail fibres since many lysins used by phages in the first steps of phage infection are associated

with the tail and baseplate structure (Latka et al., 2017). After obtaining and reannotating the enterococcal phage and prophage genomes, all the predicted tail genes were scanned for the presence of predicted lysin domains using Pfam, the NCBI Domain database and the Phyre2 webserver. As a result, multiple types of lysins were identified in both phage and prophage genomes (Table 3.1)(Figure 3.4A).

Table 3.1 Summary of predicted lytic domains associated mainly with the tail module of our study set

	Domain	Activity	# Sequences (% total 544)
Lysins	Endopeptidase	Endopeptidase	383 (70.4%)
	Lytic transglycosylase	Lytic transglycosylase	98 (18.0%)
	NLPC/P60	Endopeptidase or Amidase	34 (6.2%)
	Glycerophosphodiester phosphodiesterase (GDPD)	Phosphodiesterase	22 (4.0%)
	Pectinesterase	Pectinesterase	7 (1.3%)

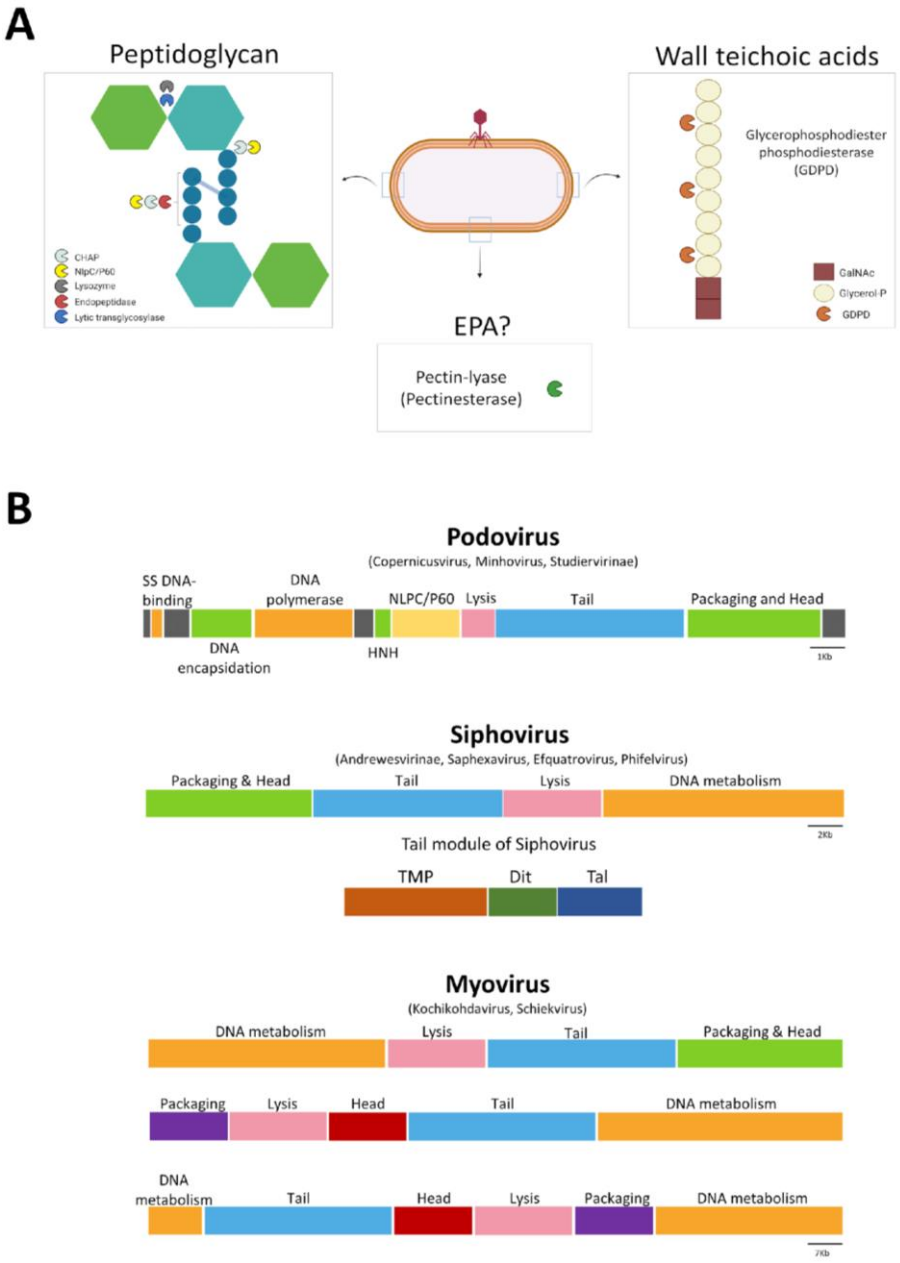


Figure 3.4 A) Schematic representation of the identified lysins and their targets in this study. B) The order of the Functional modules of the enterococcal Phage genomes. Modules and specific genes are coloured as follows: Green= DNA packaging and head, Red= Head alone, Purple= Packaging alone, Blue= Tail, Pink= lysis, and Orange= DNA metabolism. Of note, podovirus general genome organisation contains some labelled genes like yellow= NLPC/P60 gene. HNH stands for homing endonuclease. The general scheme of the siphovirus tail module is also drawn Brown=TMP, Dark green= Dit and Dark blue=Tal. The new ICTV classification is also indicated regarding each phage morphology.

Our analysis showed that the presence of a predicted endopeptidase is the most common lysin associated with tail proteins (70.4%), while lytic transglycosylase domains were present in 18.0% of the total identified lytic proteins. These two types of lysins are preferentially carried by phage-infecting Gram-positive bacteria (São-José, 2018). Other proteins were also observed to carry other potential lysins, namely peptidases of the NLPC/P60 family (6.2%), GDPD (4.0%) and lastly Pectinesterases (1.3%). Each one of these lysins is further discussed in the following sections.

3.3.1 Endopeptidases

The tail proteins associated with endopeptidase activity (TAEP) were identified in both phage and prophage genomes. We identified 383 TAEP proteins via homology with predicted phage endopeptidase domains (Pfam: PF06605). These TAEP proteins were then assessed for domain architectures (DA). This revealed 5 different groups (DA-EP), all containing a phage endopeptidase domain (Pfam: PF06605) located at the N-terminal end of the predicted protein (Figure 3.5). These domains are all found in the Tal position (i.e. TMP-Dit-Tal), although it is not clear if they have an endopeptidase activity themselves or are involved in forming active complexes or acting in a structural manner. Catalytically, endopeptidases target peptide bonds within peptidoglycan- either in the peptide stem or cross-bridge. Of our identified TAEP proteins 60.5% are within the DA1 architecture group and only contain an endopeptidase domain (Figure 3.5). This has also been noticed previously as most Tal proteins harboured a single lysin (Latka et al., 2017). To further analyse the endopeptidase domains, the MEROPS database was used to check the peptidase family of these sequences. To do this, three representative sequences from each DA were screened against the MEROPS_scan dataset, resulting in highlighting two types (M23B and C104) with high E-values ($<10^{-10}$).

The other DAs showed various lysin domains in addition to the endopeptidase domain. In DA2,3 and 4 a predicted lysozyme domain (Pfam: PF18013) was identified which is a structural homologue of a cell wall degrading enzyme in the bacteriophage phi29 tail (established using Phyre2 analysis) (Xiang et al., 2008). Besides the phage tail lysozyme, DA2 contains a CHAP domain (cysteine, histidine-dependent amidohydrolases/peptidase), while DA3 harbours a peptidase M23 domain (thought to target the peptide bonds in the peptidoglycan layer (Vermassen et al., 2019). DA4 also contains predicted amidase domains likely attacking the amide bond between MurNAc and the first amino acid L-alanine leading to the separation of the glycan and peptide units. Finally, DA5 contains a domain with homology to endosialidase chaperones. Of these domains, all have been associated with cell wall degradation or in the case of DA5- stabilisation of other catalytic domains. For example, CHAP domains (Pfam:PF05257) have been shown to act as endopeptidases (e.g. LysK CHAP) (Becker et al., 2009) or amidases Proença et al., 2012). The peptidase M23 domains (Pfam: PF01551) in DA3 are located at the C-terminal region as well as the predicted amidase domains (Pfam: PF05382) in DA4. The chaperone of endosialidase in DA5 has been shown to facilitate the folding and assembly of endosialidases and other phage proteins as well, as it is eventually cleaved off to ensure the stability of the native protein (Schwarzer et al., 2007).

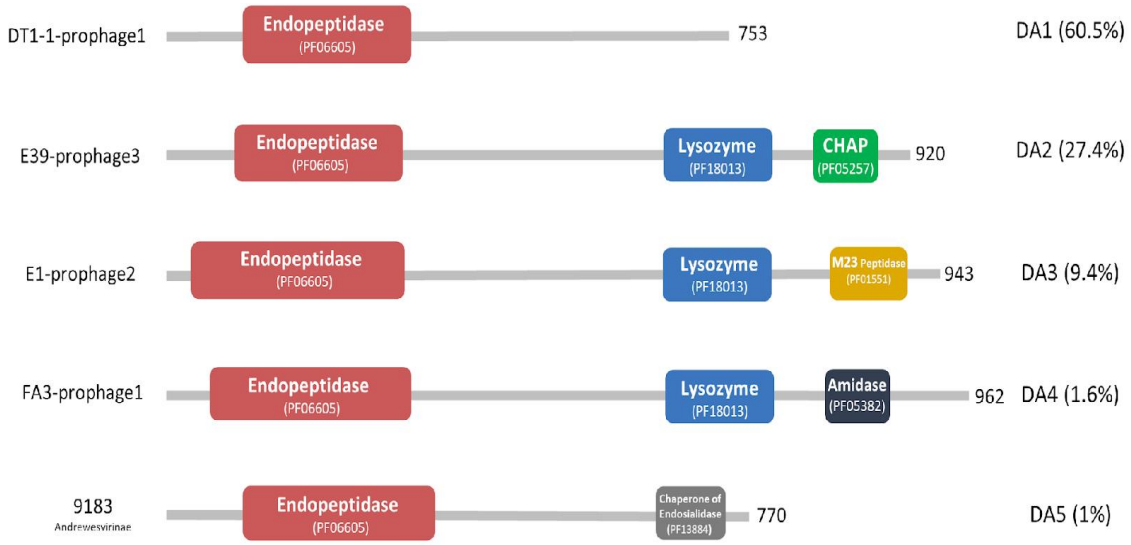


Figure 3.5 Domain architectures of TAEP proteins based on Pfam. Five DAs are shown with coloured domains. Red= endopeptidase, Blue= lysozyme, Green= CHAP, Yellow= M23 peptidase, Dark blue= amidase, Grey= chaperone of endosialidase. The left side contains the phage or prophage's name and ICTV classification while the right side contains the DA type and its abundance in percentage. The length of the protein is also indicated on the right side.

3.3.2 New Lipoprotein C/Protein of 60-kDa (NLPC/P60)

It is known that many phage proteins contain domains belonging to the NLPC/P60 family (New Lipoprotein C/Protein of 60-kDa). The NLPC/P60 family is a large group of papain-like cysteine proteases present in bacteria, like Escherichia coli (NLPC) and Listeria monocytogenes p60 (Anantharaman & Aravind, 2003). The members of the NLPC/P60 family can have (endo)peptidase as well as other activities such as amidase, transglutaminases and acetyltransferase and often contain a conserved catalytic N-terminal cysteine and C-terminal Histidine residue (Anantharaman & Aravind, 2003). In bacteria, the NLPC/P60 peptidases are likely to be involved in the bacterial cell cycle and morphogenesis by hydrolysing the peptidoglycan layer while in phage they likely aid in local peptidoglycan degradation and hence promoting genome injection (Fukushima et al., 2018; Griffin et al., 2022).

Our analysis identified 34 tail proteins that appear to be within the NLPC/P60 family (accession no.cl21534). Based on a phylogenetic tree made from amino acid alignments, these 34 sequences are classified into two main groups (Figure 3.6B). Group 1 includes sequences from podovirus-type genera while group 2 are from predicted myovirus subtypes.

Of these, Group 1 can be divided into three subgroups: 1A (Genus: Copernicusvirus) consists of 13 sequences which represent DA1 (Figure 3.6A). The 1B subgroup includes only one sequence (EF62phi) which is clearly diverged from the 1A sequences and contains an additional lysozyme domain as shown in DA2. The 1C subgroup (Genus: Minhovirus) consists of sequences from podoviruses that were isolated using *E. faecium* strains in contrast to subgroups 1A and 1B (host strains are *E. faecalis*) as well as the NLPC/P60 domain at the N-terminal region as represented in DA4 (Figure 3.6A). This may indicate differences in substrate specificity given the differing crosslinks between these spp. (Lys-Ala-Ala for *E. faecalis* and Lys-Asx for *E. faecium*) (Arbeloa et al., 2004) and is also the subject of current work in our lab.

For group 2, these NLPC/P60-containing proteins were all found in myoviruses (Herelleviridae). This group can also be further subdivided based on the phylogenetic tree and domain architecture into subgroups: 2A and 2B (Figure 3.6B). The 2A subgroup (Genus: Schiekvirus) contains only the NLPC/P60 domain, while the 2B subgroup (Genus: Kochikohdavirus) contains an additional M23 peptidase domain besides the NLPC/P60 domain.

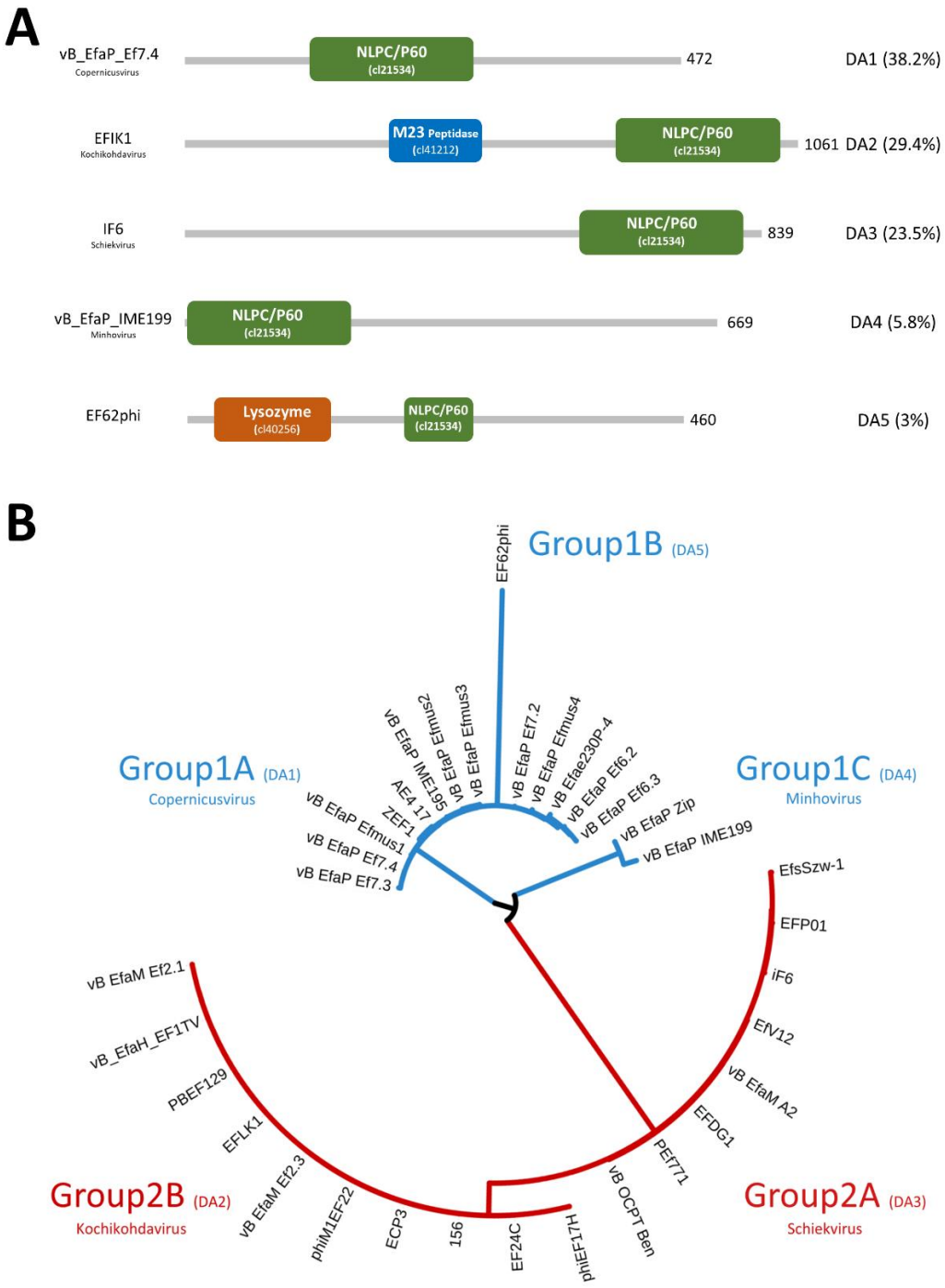


Figure 3.6 A) Domain architectures of NLPC/P60 containing proteins. Five DA are shown with coloured domains. Green= NLPC/P60, Blue= M23 peptidase, Orange= lysozyme. The domain type and abundance (%) are indicated on the right side while the phage or prophage name and the ICTV classification are on the left side. The length of the protein is also indicated. B) a phylogenetic tree of all identified NLPC/P60 containing proteins shows two main groups based on phage genomic classification and morphology: sequences from podoviruses are labelled blue while myoviruses are labelled red. The tree was constructed using FastTree and visualised using the ITOL online website.

3.3.3 Lytic transglycosylase

Our next step of analysis focuses on putative proteins containing lytic transglycosylases (LT) domains, enzymes that degrade the peptidoglycan layer by cleaving the β-1,4-glycosidic bond between N-Acetylmuramic acid (MurNAc) and N-Acetyl-D-glucosamine (GlcNAc) (Holtje et al., 1975). Of the 98 LT detected, 97 are contained within putative tail tape-measure proteins (TMPs) from the genomes of viruses with contractile or non-contractile tails. These contained a broad variety of predicted domain architectures but all had the LT domain at the C-terminal end (Figure 7A). The TMP proteins are usually the longest proteins in the phage genomes (Piuri & Hatfull, 2006) and the predicted length here varied from 1180 to 2254 aa (Figure 3.7A). Moreover, our analysis found that the location of the TMP-LT proteins in predicted enterococcal siphovirus genomes is always the same (i.e. TMP-Dit-Tal) (Goulet et al., 2020). Other studies from phage infecting other species have also identified LT within the TMPs (Piuri & Hatfull, 2006; StockDale et al., 2013). In our study, we do not see any other putative lysins than LT in TMP proteins.

Of note, the N-terminal region of the analysed TMPs often includes domains putatively involved in DNA binding or cleavage such as SCP-1, SMC, endonuclease and SbcC (Figure 3.7A). This coincides with the putative proposed function of TMP as facilitating DNA delivery and injection into bacterial cells (Mahony et al., 2016).

Since lytic transglycosylases are carbohydrate-targeting enzymes, the CAZy database was used to reveal that all the identified TMP-LT proteins belong to the specific glycosyl hydrolase (GH) family 23. The GH23 family includes lysozyme type G (EC 3.2.1.17), peptidoglycan lytic transglycosylase (EC 4.2.2.n1) and chitinase (EC 3.2.1.14). Amino acid sequence alignment and consensus analysis of our TMP-LTs revealed the presence of the GH23 conserved Glutamic acid (E) active site proton donor (Figure 7B, red arrow). Previous studies assigned LT enzymes into 8

families based on sequence motifs (Dik et al., 2017). Our identified TMP-LT sequences shared motifs with family 1A: motif I includes the catalytic residue E-S, motif II contains the G-L-M-Q residues, motif III consists of A/G-Y-N residues and motif IV is a conserved Y residue flanked by a hydrophobic residue (Figure 3.7B) (Dik et al., 2017). Indeed others have reviewed LTs and noted that the GXXQ of motif II is conserved amongst GH23 enzymes (Blackburn & Clarke, 2001; Dik et al., 2017; Scheurwater et al., 2008; Wohlkönig et al., 2010). Our data also revealed novel conserved motifs in the identified TMP-LT sequences that are not present in the family 1A i.e. T46, F47, G54, I59, L67, A68 (Figure 3.7B). Of note, the enterococcal phage LTs examined here contain extra conserved residues not present in the LTs in the literature- from either other Gram positives, Gram-negative bacteria or phage from Gram-negatives. Hence, we propose a new family that we label 1P (for phage) (Figure 3.8A).

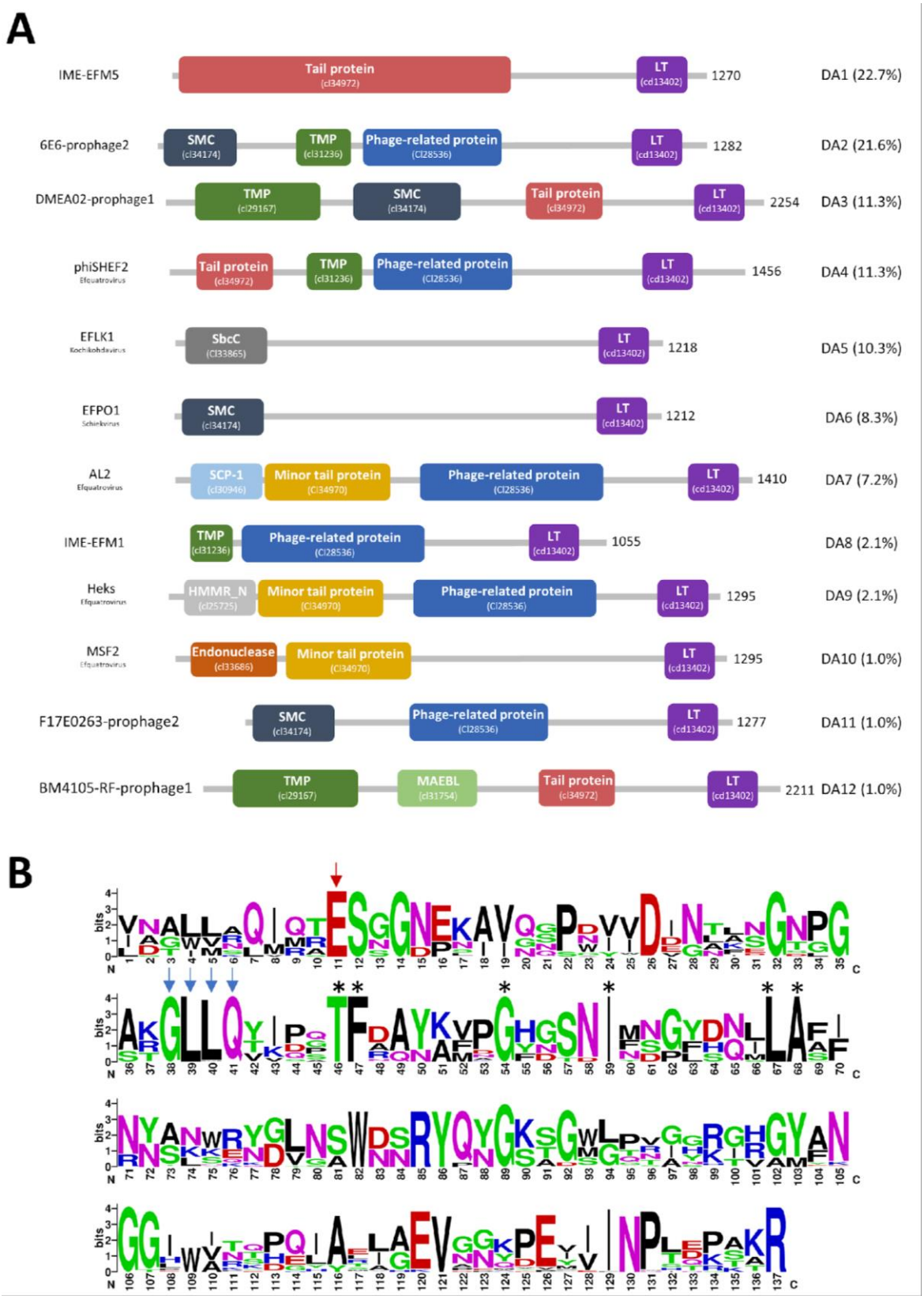

A
IME-EFM5
Tail protein (cl34972)
LT (cd13402)
1270
DA1 (22.7%)
6E6-prophage2
SMC (cl34174)
TMP (cl31236)
Phage-related protein (Cl28536)
LT (cd13402)
1282
DA2 (21.6%)
DMEA02-prophage1
TMP (cl29167)
SMC (cl34174)
Tail protein (cl34972)
LT (cd13402)
2254
DA3 (11.3%)
phiSHEF2
Efquatrovirus
Tail protein (cl34972)
TMP (cl31236)
Phage-related protein (Cl28536)
LT (cd13402)
1456
DA4 (11.3%)
EFLK1
Kochikohdavirus
SbcC (Cl33865)
LT (cd13402)
1218
DA5 (10.3%)
EFPO1
Schiekvirus
SMC (cl34174)
LT (cd13402)
1212
DA6 (8.3%)
AL2
Efquatrovirus
SCP-1 (cl30946)
Minor tail protein (Cl34970)
Phage-related protein (Cl28536)
LT (cd13402)
1410
DA7 (7.2%)
IME-EFM1
TMP (cl31236)
Phage-related protein (Cl28536)
LT (cd13402)
1055
DA8 (2.1%)
Heks
Efquatrovirus
HMMR_N (cl25725)
Minor tail protein (Cl34970)
Phage-related protein (Cl28536)
LT (cd13402)
1295
DA9 (2.1%)
MSF2
Efquatrovirus
Endonuclease (cl33686)
Minor tail protein (Cl34970)
LT (cd13402)
1295
DA10 (1.0%)
F17E0263-prophage2
SMC (cl34174)
Phage-related protein (Cl28536)
LT (cd13402)
1277
DA11 (1.0%)
BM4105-RF-prophage1
TMP (cl29167)
MAEBL (cl31754)
Tail protein (cl34972)
LT (cd13402)
2211
DA12 (1.0%)
B
bits
N
C

Figure 3.7 A) Domain architecture of TMP-LT proteins. Based on CDD, 12 DA are shown with coloured domains. Purple= lytic transglycosylase-like domain (LT), Green= tape measure protein domain (TMP), Dark blue= structural maintenance of chromosomes (SMC), Red= Tail protein, Sky blue= Synaptonemal complex protein (SCP-1), Orange= Minor tail protein, Dark Grey=SbcC, Blue= Phage-related protein, Light green= Merozoite Apical Erythrocyte Binding-ligand (MAEBL), Brown= Endonuclease, light grey= Hyaluronan mediated motility receptor N-terminal (HMMR_N). The domain type and abundance (%) are indicated on the right side while the phage or prophage name and the ICTV classification are on the left side. The asterisks indicate conserved residues specific for analysed sequences. The length of the protein is also mentioned. B) Weblogo of the TMP-LT domains showing highly conserved domains including the catalytic residue Glutamic acid (red arrow) and GH23 specific motif (Blue arrows). Sequence logos were created using Weblogo (http://weblogo.berkeley.edu/logo.cgi).

A

Family	Motif I	Motif II	Motif III	Motif IV	Motif V
1A	(E-S)	(G-L-M-Q)	(A/G-Y-N)	(x-Y)	
1E	(E-S)	(G-L/M-M-M/Q)	(A-Y-N)	(x-Y)	
1H	(E-S)	(G-W-F-Q)	(A-H-F)		
4A	(A-x_7-S-E)	(Y-x_4-G-x_5-D-x-H-P)	(S-T-x_4-G-R-Y-Q-x_5-W)	(W-x-S)	
TMP-LT	(E-S)	(G-L-L-Q)	(T-F-x_6-G-x_4-I-x_7-L-A)	(W-x_2-R-Y-x_2-G)	(G-x_{14}-V-x_9-N-x_6-R)

B

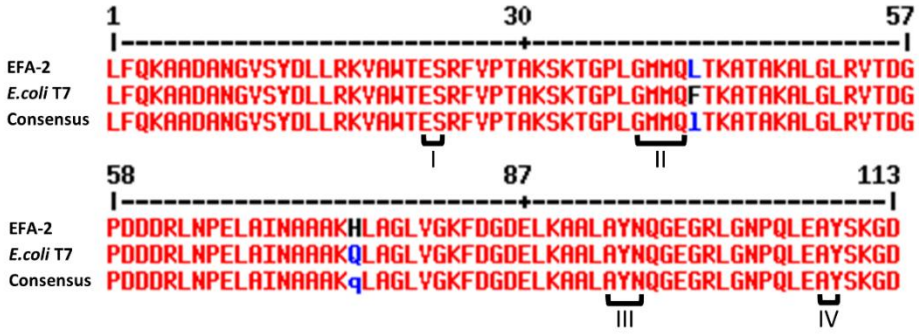


Figure 3.8 A) The conserved motifs of families 1A,1E,1H,4A and TMP-LT (blue) are shown. B) Sequences alignment of LT domain between EFA-2 and *E. coli* T7 phages whose motifs are numbered as I,II,III and IV.

Outside of the TMP-LTs discussed here, one outlier was found, this time in a predicted Studiervirinae genome EFA-2 (39.9 kb). EFA-2 has a large genome and an unusual genome organisation compared with the Sarlesvirinae podovirus genomes (Figure 3.1). However, the longest gene in this genome showed a predicted GH23-LT domain which is unusually located at the N-terminal end as opposed to the TMP-LTs which have the LT at the C-terminal end. CAZy database analysis showed that this LT has the closest homology with Gram-negative infecting phage lysins such as the *E. coli* T7 phage gp16 lytic transglycosylase protein (97% aa similarity) and is likely a member of family 1E as described by Dik et al (Dik et al., 2017) (Figure 3.8B).

3.3.4 Pectinesterase (PE)

After analysing the enterococcal phage and prophage genomes, seven predicted tail proteins from prophage, but none from lytic phage, were found to harbour a Tail-associated pectinesterase domain. Pectinesterases are enzymes that target pectin via the demethylation of galacturonosyl residues (Reid, 1950). Pectin is a main component of plant cell walls and is made up of three types: a homopolymer of galacturonic acid, and two forms of a rhamnogalacturonan (RG-I and RG-II) made up of repeating Gal-Rha disaccharides (Mohnen, 2008). Importantly, in enterococci, the cell wall contains a specialised polysaccharide called enterococcal polysaccharide antigen (EPA) that is made up of repeating rhamnose units, interspersed with other sugars and decorated with various modifications (Dale et al., 2017; Guerardel et al., 2020; Rigottier-Gois et al., 2015; Teng et al., 2009). Therefore, we hypothesise that the EPA structure in enterococci could be the target of these phage pectinesterases. The pectinesterase domains in the seven sequences were located approximately in the central region of the proteins and no other predicted domains were identified (Figure 3.9A). The pectinesterase genes were all located right after the common Tal position in siphovirus-type genomes (Efquatroviruses, Phifelviruses, Saphexavirus or Andrewesvirinae)(i.e. TMP-Dit-Tal) and of note were also present in concert with TAEP and TMP-LTs in 4 prophage genomes (Figure 3.9B). To further confirm our annotation, structural homology using Phyre2 was performed which identified structural

homologues in pectinesterase 1 or rhamnogalacturonan lyase families indicating that these putative genes may well be novel phage pectinesterases or EPA targeting enzymes. Similar pectinesterase/pectin lyase domains were also found in other phages targeting *Klebsiella pneumoniae* (J. Li et al., 2021; Pertics et al., 2021) and *Acinetobacter baumannii* (Shahed-Al-Mahmud et al., 2021) which all have showed a depolymerase activity upon expression and purification.

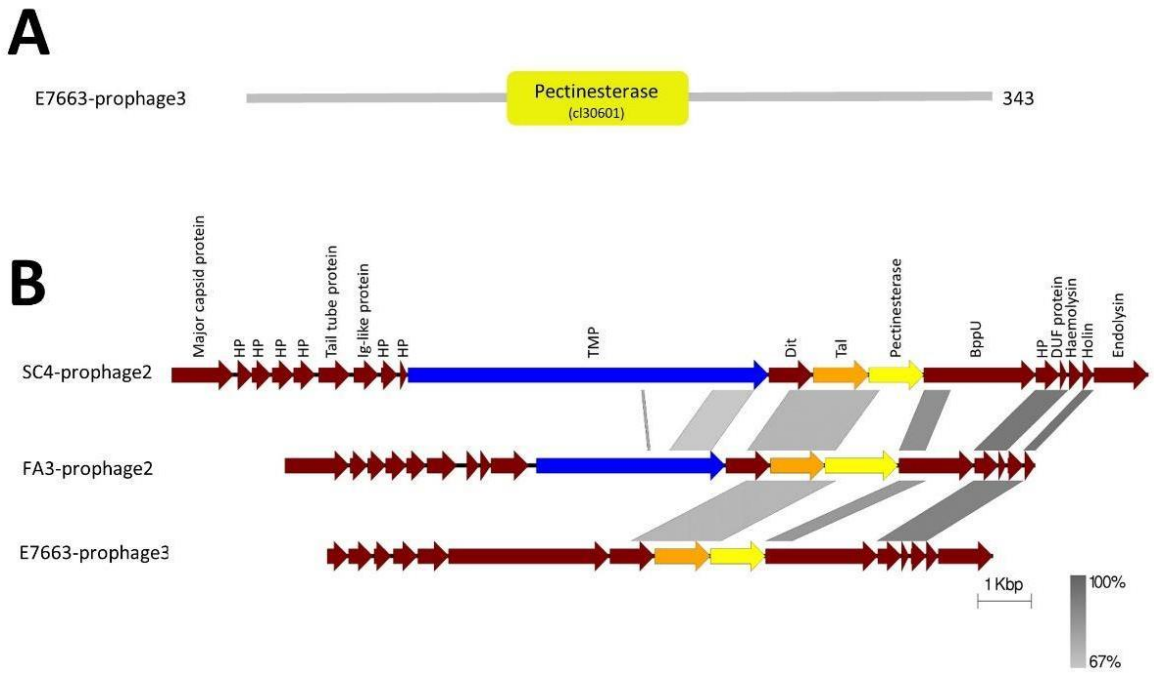


Figure 3.9 A) Domain architecture of a pectinesterase containing protein from E7663-prophage3 genome based on NCBI domain database which showed a pectinesterase domain. B) MSA of tail modules showing Pectinesterase protein (yellow), TMP-LT (Blue), TAEP (Orange). HP= Hypothetical proteins. The level of identity is indicated by the grey region between genomes.

3.3.5 Glycerophosphodiester phosphodiesterase (GDPD)

The final type of predicted lysin observed are glycerophosphodiester phosphodiesterases (GDPD), confirmed using Pfam, NCBI domain database and Phyre2. These GDPD enzymes can degrade the phosphodiester bonds holding wall teichoic acids to sn-glycerol 3-phosphate (Gro3P) and their corresponding alcohol (Cornelissen et al., 2016). Our analysis revealed 22 gene predictions carrying the GDPD domain in both phage and prophage genomes. The GDPD proteins display three domain architectures (DA-PD)(Figure 3.10A). The first and most common DA-PD (59.1%) contains only a GDPD domain (PF03009.17) with most sequences having a protein size of around 240 aa. The second DA harbours a GDPD domain and a predicted membrane domain (PF10110.9) (Figure 3.10A) homologues of which have been found in *Streptococcus* bacterial genomes (Chuang et al., 2015). The third DA contains the GDPD domain at the C-terminus with a predicted baseplate upper protein (BppU) located at the N-terminal end indicating that this is likely a multifunctional baseplate-lyase protein in phage 9183. Some of these proteins were found within the tail module (e.g., phage 9183 (Andrewesvirinae), vB_EfaS_IME197, SRCM103470-prophage2 (Figure 3.10B) while others were spotted throughout the genomes (e.g., BA17124, E39 and E745 prophages). It is also of note here that all the GDPD seen in the tail modules were in concert with a TAEP protein, suggesting potential synergy (Figure 3.10B).

The GDPD activity of other phage-encoded enzymes has been investigated and showed that five conserved residues are required: 2 catalytic Histidines that act as a general acid and general base in catalysing the hydrolysis of the 3'-5' phosphodiester bond (Rao et al., 2006; Shi et al., 2008) and 3 divalent metal-ion-binding residues (2 Glutamic acid residues and an Aspartic acid residue) (Cornelissen et al., 2016; Shi et al., 2008). In our analysis, the alignment of the enterococcal GDPD domains showed the presence of these highly conserved residues (Figure 3.10C).

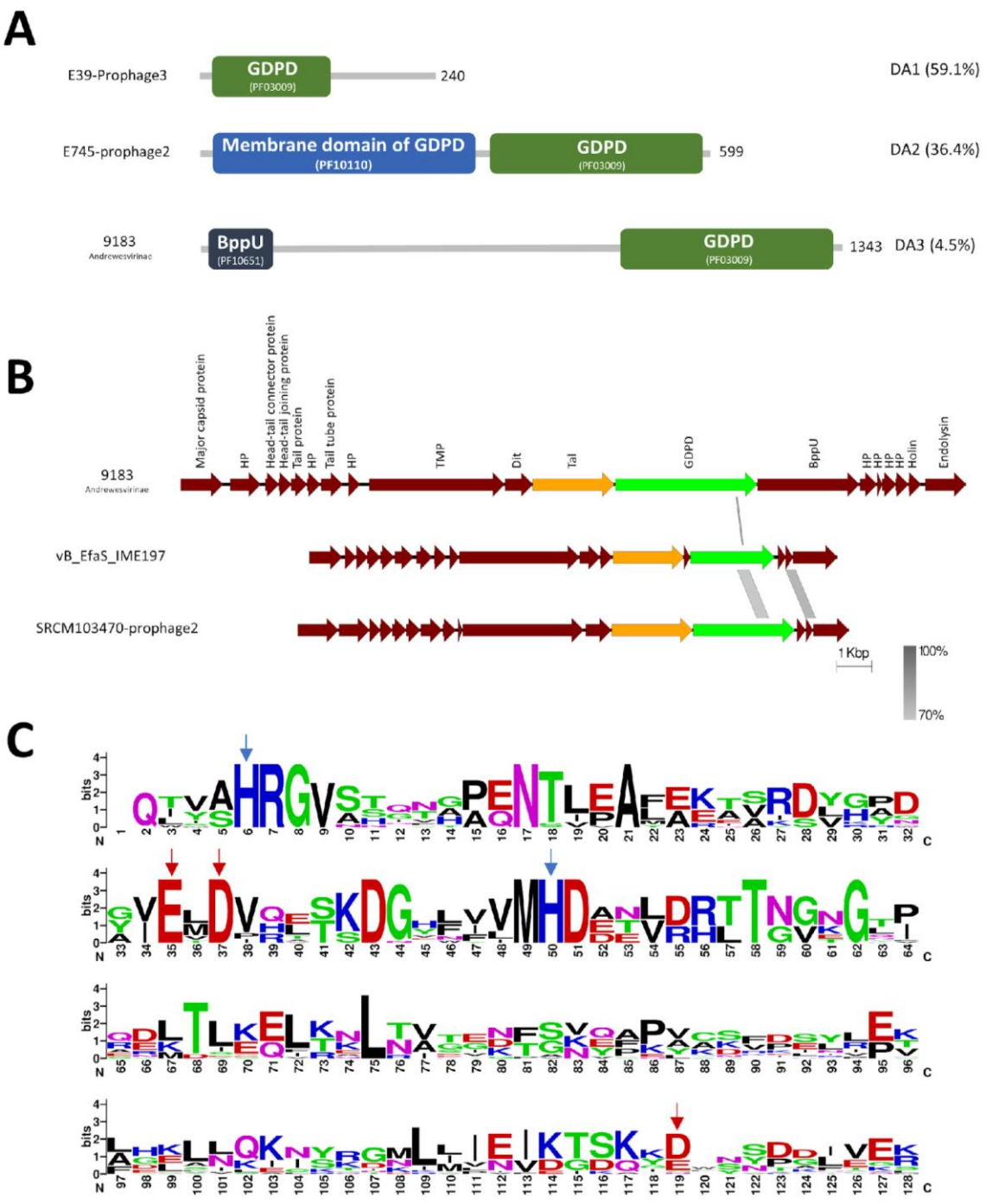


Figure 3.10 A) Domains architectures of the GDPD containing proteins based on Pfam. Three DA are shown with coloured domains. Green= GDPD, Blue= membrane and dark blue=baseplate upper protein. The domain type and abundance (%) are indicated on the right side while the phage or prophage name and the ICTV classification are on the left side. The length of the protein is also indicated. B) MSA of tail modules showing GDPD protein (Green), TAEP (Orange), HP= Hypothetical proteins. The level of identity is indicated by the grey region between genomes. C) weblogo of the aligned GDPD domains which shows catalytic residues (Blue arrows) and metal binding residues (Red arrows).

3.4 Patterns in the arrangement of lysins within phage and prophage genomes

During our analysis, we noted several patterns in the arrangement of potential lysins within phage genomes and specifically their tail modules. For the siphoviruses, the tail proteins usually follow the TMP-Dit-Tal tail order (Goulet et al., 2020), and this is also observed here with the analysed enterococcal phages. Notably, we observed that the type of lysin identified correlates with the phage genome size. Specifically, the smaller genome group (1, 21-43 kb, Efquatroviruses or Phifelviruses) contains TAEP proteins as well as Tape-measure LTs (TMP-LT) (Figure 3.11B) while the larger siphoviruses (Saphexavirus or Andrewesvirinae) (55-86 kb) have only a single predicted protein with a lytic domain (TAEP) (Figure 3.11C). Despite not identifying the Dit protein bioinformatically, in many cases we observed a small Hypothetical protein (HP) that we assume is the Dit protein in these phages. For the myoviruses (Schiekvirus or Kochikohdavirus), a TMP-LT protein and another adjacent tail protein containing an NLPC/P60 domain were identified in all analysed genomes. (Figure 3.11D).

Finally, enterococcal podoviruses (Copernicusvirus or Minhovirus), contain a morphology where the head seems to be connected to a baseplate with what one assumes is an infection system (and no tail measure protein is present). Hence, it is unsurprising that they do not display the TMP-Dit-Tal paradigm. Our data indicate that adjacent to the head and endolysin/holin pair most podoviruses genomes contain a predicted NLPC/P60 family protein (Figure 3.11A). The location of this protein is highly conserved among the analysed podovirus genomes and is likely part of a potential tailspike protein (unpublished data, personal communication, Professor Graham Stafford).

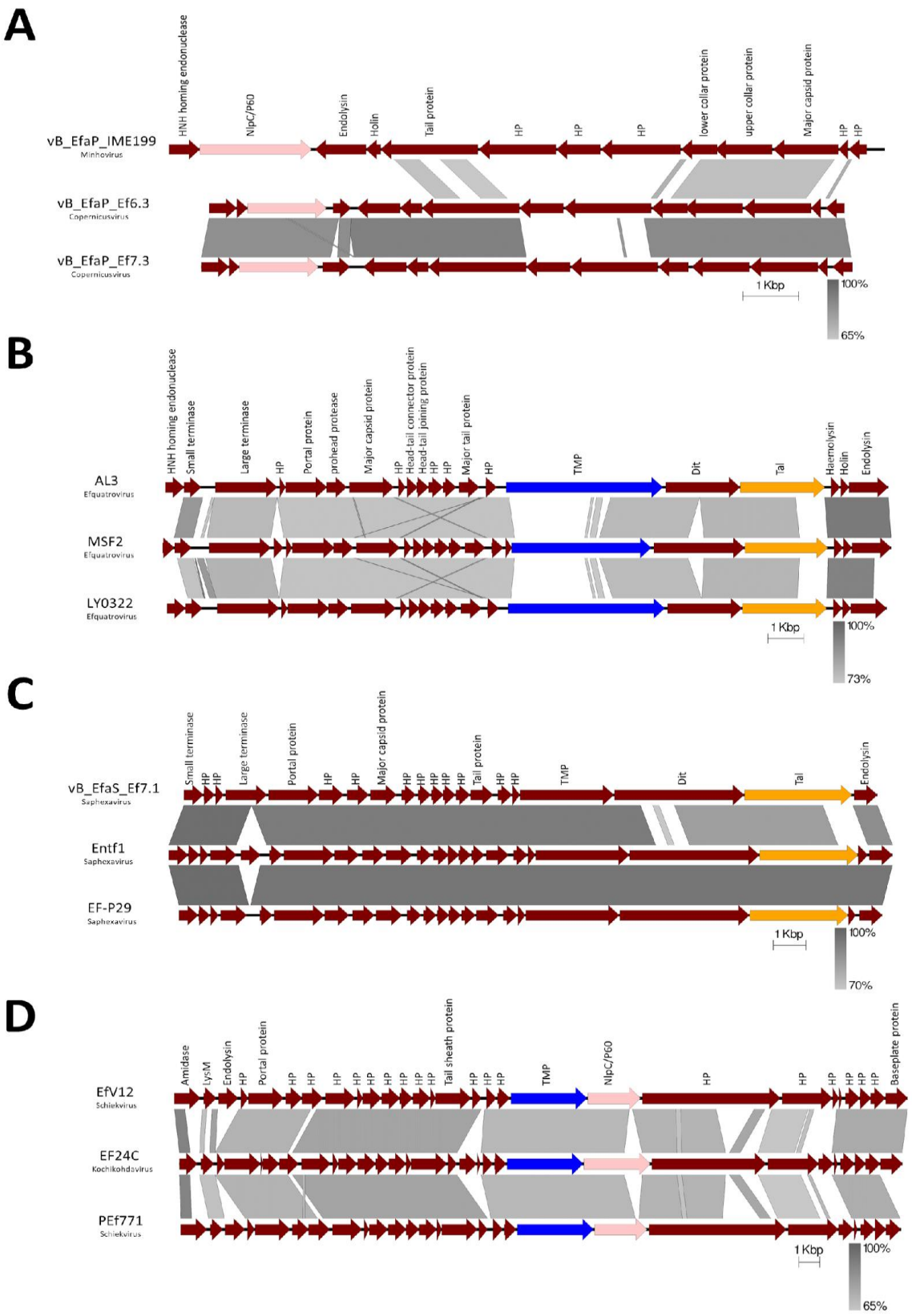

A
HNH homing endonuclease
NlpC/P60
Endolysin
Holin
Tail protein
HP
HP
HP
lower collar protein
upper collar protein
Major capsid protein
HP
HP
vB_EfaP_IME199
Minhovirus
vB_EfaP_Ef6.3
Copernicusvirus
vB_EfaP_Ef7.3
Copernicusvirus
1 Kbp
100%
65%
B
HNH homing endonuclease
Small terminase
Large terminase
HP
Portal protein
prohead protease
Major capsid protein
HP
Head-tail connector protein
Head-tail joining protein
HP
HP
Major tail protein
HP
TMP
Dit
Tal
Haemolysin
Holin
Endolysin
AL3
Efquatrovirus
MSF2
Efquatrovirus
LY0322
Efquatrovirus
1 Kbp
100%
73%
C
Small terminase
HP
HP
Large terminase
Portal protein
HP
HP
Major capsid protein
HP
HP
HP
HP
Tail protein
HP
HP
TMP
Dit
Tal
Endolysin
vB_EfaS_Ef7.1
Saphexavirus
Entf1
Saphexavirus
EF-P29
Saphexavirus
1 Kbp
100%
70%
D
Amidase
LysM
Endolysin
HP
Portal protein
HP
HP
HP
HP
HP
HP
HP
HP
Tail sheath protein
HP
HP
HP
TMP
NlpC/P60
HP
HP
HP
HP
HP
HP
Baseplate protein
EfV12
Schiekvirus
EF24C
Kochikohdavirus
PEf771
Schiekvirus
1 Kbp
100%
65%

Figure 3.11 The general organisation of Tail modules in enterococcal phage genomes. (A) examples of podovirus genomes harbour NLPC/P60 containing protein (pink colour). (B) Group 1 of siphovirus genomes contain both TAEP (orange) and TMP-LT (blue) proteins while Group 2 (C) contains only TAEP proteins (orange). (D) myovirus genomes harbour TMP-LT and NLPC/P60 proteins. The phage names and the ICTV classification are indicated on the left side. The level of identity is indicated by the grey region between genomes.

As discussed earlier, we have observed a correlation between phage genome size and phage morphology (i.e. small genomes are usually podoviruses while larger genomes are myoviruses). However, the exceptions to this correlation such as the siphovirus phage EFRM31 (16.9 kb) showed no TAL while the siphovirus phage EFAP-1 (21.1 kb) contains both the TAEP and TALT proteins. For podoviruses, phage EFA-2 (39.9 kb) contains an LT-containing protein (the largest protein in the genome), which could also act similarly to a TMP.

Enterococcal prophages genomes seem to conform to the pattern of siphovirus (Efquatroviruses, Phifelviruses, Saphexavirus or Andrewesvirinae) type tail modules (TMP-Dit-Tal) (Figure 3.12). For TAL, the majority (86.7%) of the prophage genomes have the endopeptidase TAEP in the Tal position, i.e. TMP-Dit-Tal(TAEP). The other genomic organisation observed is TMP with LT activity alone (6.9%). TMP and Tal with LT and TAEP activities, respectively, are observed in 6.3% of prophage genomes containing TAL (Figure 3.12). Lastly, 59 of the prophage genomes did not contain predicted lysins associated with tail proteins and the functional modules in some of these genomes were not conventionally organised albeit they are predicted to be intact prophages by PHASTER.

Furthermore, we found that genome organisation within the enterococcal prophages coincides with that within isolated phages in terms of module order (i.e. Packaging, Head, Tail, Lysis, DNA Metabolism) (Figure 3.4B). Additionally, the genome size of most analysed prophage was between 30-60 kb (Figure 3.3A) and the majority possess the typical tail module arrangement seen in siphoviruses (Efquatroviruses, Phifelviruses or Saphexavirus) (i.e. TMP-Dit-Tal). As

expected, we observed several lysogenic genes such as integrase, repressor and anti-repressor in these predicted prophages. Collectively, we propose that these prophages are likely to be Efquatroviruses, Phifelviruses or Saphexavirus.

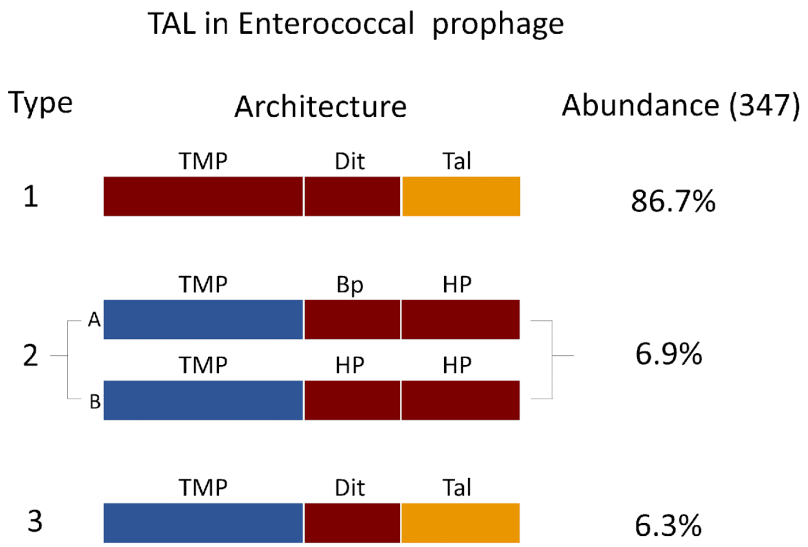


Figure 3.12 TAL in enterococcal prophage genomes. Three main types are shown regarding the location of TAL in the tail module. The abundance in percentage is calculated from the total prophage genomes with TAL. Gene annotation is shown as follows: Tape measure protein (TMP), Tail-associated lysin (Tal), Baseplate protein (Bp), Hypothetical protein (HP), predicted endopeptidase activity (orange), LT (blue), No lysin activity identified (dark brown).

3.5 Conclusion

In this chapter, the lysin landscape of enterococcal bacteriophage genomes was surveyed. The most commonly identified TAL domains were those targeting peptidoglycan, namely endopeptidases (TAEP), NLPC/P60 (endo)peptidases as well TMP located GH23 lytic transglycosylases- present within tail-tape measure proteins in all predicted tailed viruses surveyed. Lastly, other domains were also identified potentially targeting EPA (pectinesterases) and teichoic acids (GDPD). Overall, one predicts that these all target different parts of the cell

wall of these enterococci and that the differences in domain and sequence indicate differences in strain specificity that are not yet elucidated. Additionally, the finding that many phages contain multiple potential lysin domains suggests a layer of cooperation between these domains *in vivo* that we have yet to elucidate. Finally, our data reveal the extent and variety of enterococcal lytic domains as candidates for recombinant production as potential novel antimicrobials, either in isolation or in combination with each other or as potentiators of antibiotics. Finally, we have also laid a platform for the potential engineering of enterococcal phage akin to the recent refactoring study on T7 (Liang et al., 2022). Our group are also currently working on expressing examples of a range of these genes recombinantly with a view to the production of novel antimicrobials.

Chapter 4: Isolation and characterisation of phages targeting enterococci

4.1 Introduction

The crisis of antimicrobial resistance has caused about 700,000 deaths annually and therefore alternative therapies to antibiotics are needed (O'neill, 2016). One promising approach is phage therapy to treat human bacterial infections. Phages can be isolated from different environmental sources such as soil, river and wastewater. Phage isolation and characterisation facilitate phage therapy in several ways such as making phage available when clinically needed, providing material for phage engineering and providing a source for phage lysins.

Therefore, we first collected different *E. faecalis* and *faecium* strains, of which some are Vancomycin-resistant enterococci - VRE (Methods chapter, Table 2.1, Table 2.2 and Table 2.3). Some of our strains are also clinical isolates from patients with diabetic foot ulcers. We then aimed to isolate phages using wastewater samples against these various enterococcal strains.

Following isolation, we then aimed to characterise the isolated phages in terms of phage morphology, genome, host range and phage-host killing assay. We also attempted to further decipher the phage-host initial interaction by determining phage receptors on the bacterial host cell surface. Bacterial resistance to phages was also aimed to be investigated and analyzed.

Specific aims:

- Investigation of phages in wastewater samples
- Isolation of phages using different enterococcal strains
- Morphological, molecular and genomic characterisation of isolated phage
- Host range analysis of isolated phages
- Analysis of phage-resistant mutants

4.2 Investigation of bacteriophage within wastewater

In nature, bacteriophages are abundant and can be considered as one of the most abundant biological entities on Earth (Clokie et al., 2011). However, certain environments are known for their richness and diversity of phages such as wastewater which is frequently used for phage isolation. Therefore, wastewater samples were obtained from the Sheffield area namely Woodhouse mills wastewater treatment. Some wastewater samples were collected and processed by Dr Elizabeth Court (a postdoc in our lab) and I used the filtered samples in my isolation protocol. Other samples were collected by our collaborator Dr Henriette S. Jensen in the Department of Chemical and Biological Engineering and I personally processed the samples. These wastewater samples were received and handled at different periods throughout the PhD journey. The wastewater was sieved to eliminate any big particulates, but neither chemically nor biologically treated.

After processing and concentration of the wastewater samples, the presence of phages in these samples was first established before using them in any isolation experiment. Therefore, the samples were prepared for TEM examination. Upon examination, various phage morphologies were observed which are mainly characterised by a capsid with a tail. This shape encompasses three main morphology types: myoviruses, siphoviruses and podoviruses.

Myoviruses, phages with capsids and contractile tails, were identified in the concentrated wastewater samples (Figure 4.1). As myoviruses can have two forms under TEM (contracted and noncontracted), these were also observed in our samples as in (Figure 4.1 A-E) for the noncontracted form and (Figure 4.1 F) for the contracted form. A total of 12 myoviruses were seen under TEM. Of note, the variation in the capsids' shapes (mostly isometric) was seen in (Figure 4.1 A,B and G) as well as the tail lengths as in (Figure 4.1 B, C and D).

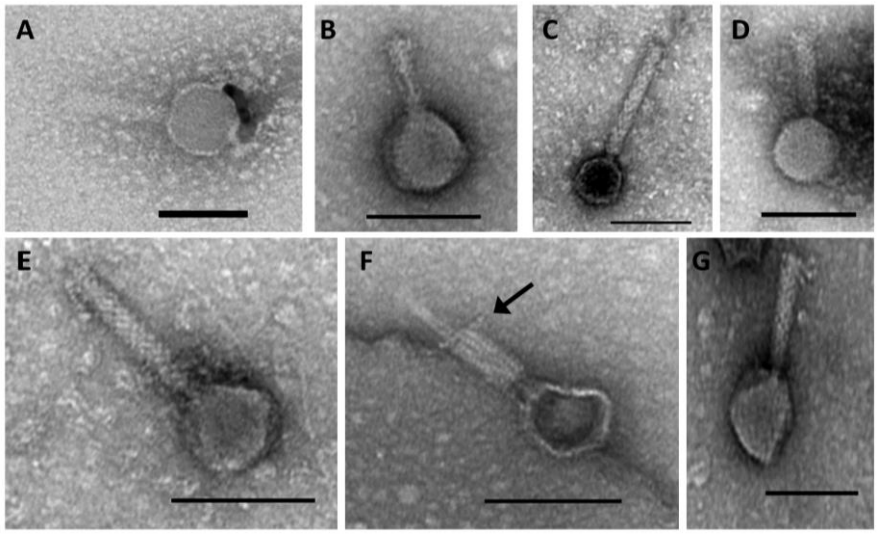


Figure 4.1. Electron micrographs of bacteriophage from wastewater that are myoviruses in morphology. Arrow indicates the baseplate structure. Scale bar 100 nm.

The other morphology that was observed under TEM was siphoviruses which are characterised by long noncontractile tails. As with myoviruses, the heads in these siphoviruses also varied from typical icosahedral capsids (Figure 4.2 E) to other structures (Figure 4.2 C and D). This variation was also seen with the tails as in (Figure 4.2 A and C). Nine siphoviruses were observed from the wastewater sample.

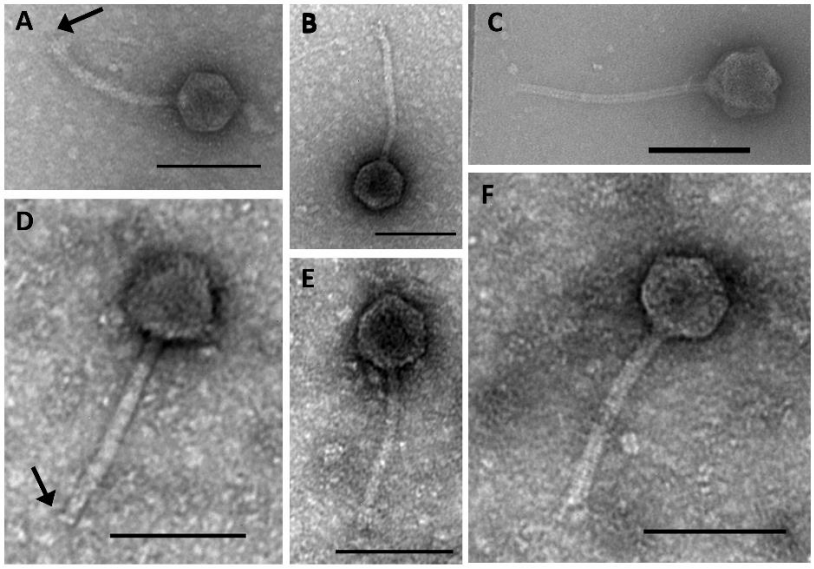


Figure 4.2. Electron micrographs of bacteriophage from wastewater that are siphoviruses in morphology. Baseplate structure is indicated by a black arrow. Scale bar 100nm.

For podoviruses, these are characterised by capsids and short tails. These phages showed varied head morphology as icosahedral capsids in (Figure 4.3 A,C and F) and irregular capsids (Figure 4.3 E and G). About 25 podoviruses were seen under TEM. Collectively, this shows the richness of wastewater with different phage morphologies and their suitability for phage isolation.

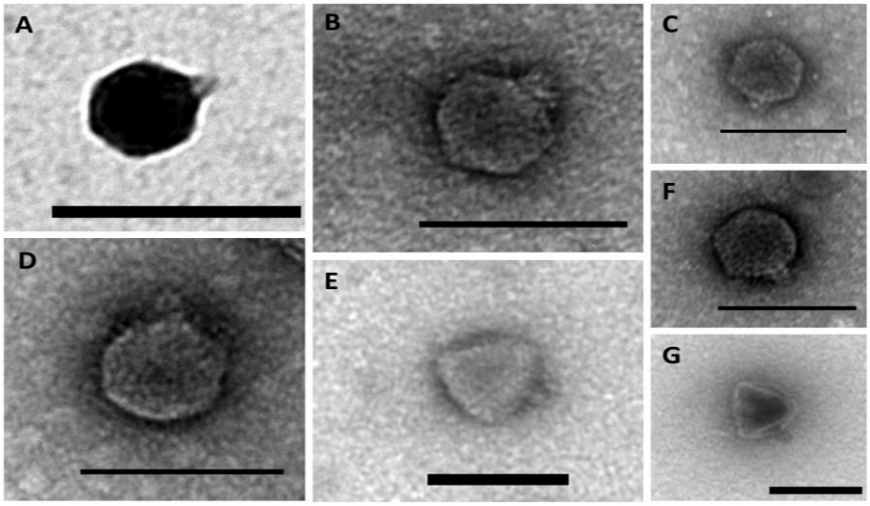


Figure 4.3. Electron micrographs of bacteriophage from wastewater that are podoviruses in morphology. Scale bar 100nm.

4.3 Isolation of phage

Since our wastewater samples were processed and concentrated, the isolation of phage targeting enterococci was started. Different strains were used in the isolation process ranging from lab strains as well as clinical isolates from patients with a diabetic foot ulcer. Initially, several attempts were performed to isolate phages using direct spotting of the concentrated wastewater onto the bacterial lawns which resulted in no observable lysis after overnight incubation. This direct spotting method was used on the *E. faecalis* strains OS16, JH2-2, EF2, EF54 and OMG3919 and the *E. faecium* strain E1162. Therefore, an enrichment method was used to increase the probability of phage isolation from wastewater. For this enrichment method, two approaches were conducted using either a single host or multiple hosts in the isolation process.

4.3.1 Single host isolation

In this approach, a concentrated wastewater sample was added to a single host bacterial culture and incubated overnight to allow phage infection and propagation. Next day, the culture was centrifuged, filtered and then spotted on the bacterial lawn. Using this method several positive results were obtained (lysis on lawns) for OS16, JH2-2, EF2, EF54 and OMGS3919 strains. To ensure that this lysis is actually a result of phage infection, plaque assays were performed from the enriched samples. This revealed that all the lysis on plates was a result of phage infection by observing discrete plaques. To purify phages, a single plaque was carefully picked with a sterile loop and mixed with its bacterial host and another plaque assay was performed. After three consecutive repeats, a plaque was picked and mixed with PBS and kept at 4 °C for further analysis. Using this method, a total of five phages were isolated (phiSHEF8-12) (Figure 4.5,Figure 4.6, Figure **4.7**, Figure 4.8Figure 4.9).

4.3.2 Multiple hosts isolation

In another approach, multiple hosts were used together in one culture for phage isolation/ enrichment. This was done by mixing four strains together and then the concentrated wastewater sample was added. After overnight incubation, the culture was centrifuged and filtered. The enriched sample was then spotted individually on each bacterial lawn. Next day, lysis on plates was checked for positive phage isolation which is further confirmed by plaque assays. The reason for choosing this method was to study the effect of mixed hosts on phage isolation and whether it may increase the likelihood of isolating phages as it was discussed in Hyman et al (Hyman, 2019). Additionally, using more than one host for phage isolation has been applied in previous studies (Betz & Anderson, 1964; Oliveira et al., 2009; Ross et al., 2016; Sillankorva et al., 2010). Regarding the possibility of prophage induction using this method, this is a poetical outcome and a reliable method to check this via phage genome sequencing.

Two groups of *E. faecium* strains were used in this approach, the first group was a set of 4 clinical isolates named (dp6-9) which were obtained from patients with diabetic foot ulcers at

The Northern general hospital. The second group was 4 strains (E1679, E1071, E1162, E980) which differ in their EPA classification based on Been et al (De Been et al., 2013). For each group, the abovementioned multiple hosts' technique was applied. As a result, a total of three phages targeting the E. faecium E1071 strain were successfully isolated (phiSHEF13, 14 and 16) (Figure 4.10, Figure **4.11**,Figure **4.12**).

4.4 Characterisation of isolated phage

A total of 8 phages (phiSHEF8-14,16) were successfully isolated from both *E. faecalis* and *E. faecium* strains. All these phages were stored at 4 °C either in BHI or PBS. Following this, a series of experiments were carried out to characterise these phages.

4.4.1 Plaque morphology

One of the first characterising information about phages is their morphology on plates meaning plaque morphology. This was assessed in all the isolated phages which showed varied results. Firstly, clear plaques with 3-4 mm diameter were seen for phiSHEF8 using OS16 strain, phiSHEF9 using JH2-2 strain, phiSHEF10 using EF54 strain, phiSHEF11 using EF2 strain and phiSHEF12 using OMG3919 (Figure 4.4 A). In contrast, tiny clear plaques with 0.5-1 mm diameter were observed for phiSHEF13,14,16 using E1071 (Figure 4.4 A). Of note, the soft agar concentration in all plates was constant at 0.4%. Interestingly, phiSHEF11 has shown a halo surrounding plaques and this was further investigated. A prolonged incubation (4 days) of these plaques has shown an increase in the halo diameter indicating a possible depolymerase effect on bacterial cells (Figure 4.4 B) (Pires et al., 2016). Due to time limits, no further investigation was carried out about this phenomenon. A thorough genomic analysis for depolymerase-containing proteins could shed light into this.

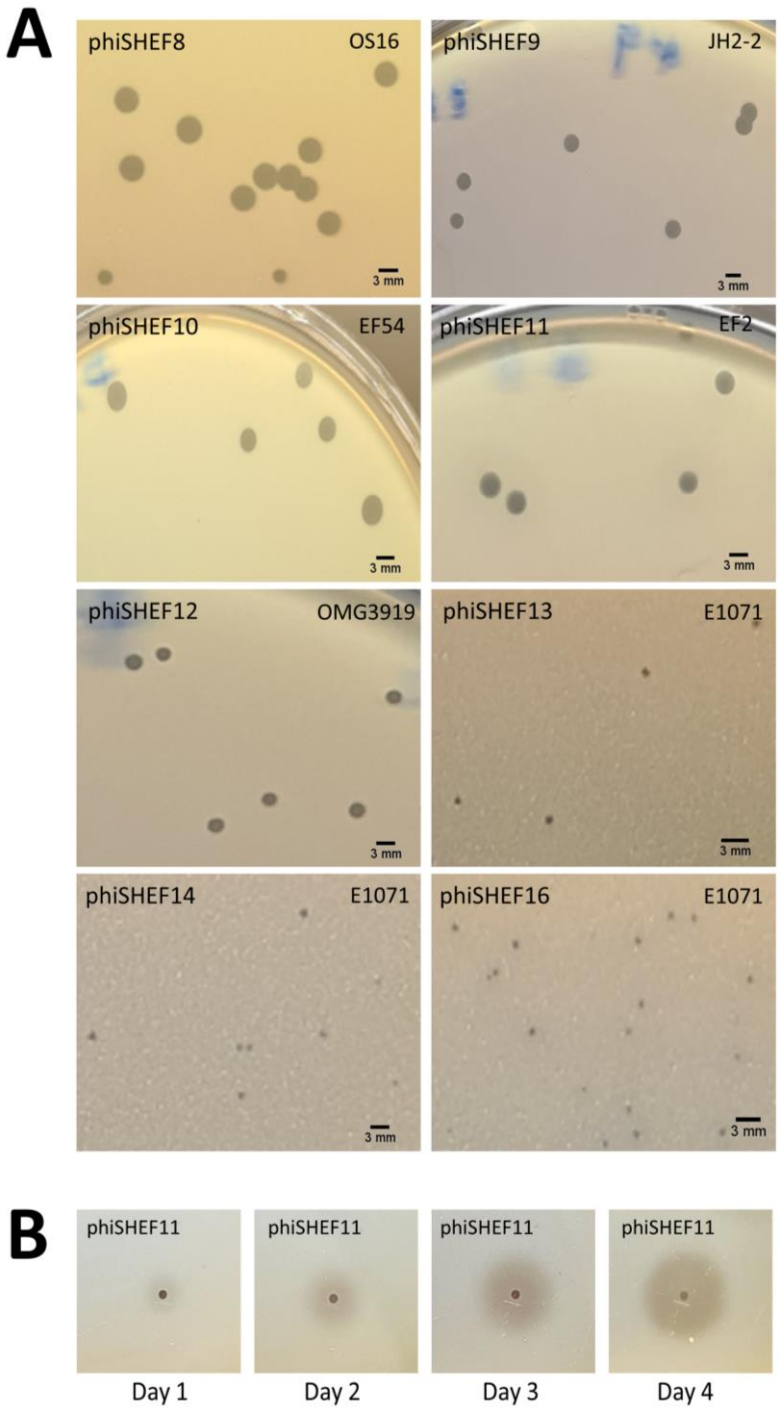


Figure 4.4. A) Plaque morphologies of isolated phages. All phages have clear plaques with various sizes. B) a phiSHEF11 plaque on ATCC19433 strain for 4 days which shows an increasing halo diameter.

4.4.2 Phage morphology

After plaque purification, phage morphology was determined. Upon TEM examination, the isolated phages showed icosahedral capsids and tails with varied lengths indicating that they belong to the Caudoviricetes class. For phiSHEF8-12, they possess long noncontractile tails which indicate that these are morphologically siphoviruses (Figure 4.5, 4.6, 4.7, 4.8, 4.9), respectively. These all approximately have similar head diameters, tail lengths and widths as described in Table 4.1. Among isolated phages, only phiSHEF14 was found to possess a short tail indicating this is a podovirus. The short tail of phiSHEF14 was about 24 nm in length while the head was about 54 nm in diameter. Tail fibres were also clearly observed in the phiSHEF14 structure (Figure 4.11 A and D). For phiSHEF13 and 16, the electron micrographs showed capsids with contractile tails indicating these are myoviruses. A clear baseplate structure was also observed in phiSHEF13 (Figure 4.10 B and C) and phiSHEF16 (Figure 4.12 B and D). The head diameter of these two phages was about 95 and 99 nm which are larger than the ones in sipho- and podoviruses Table 4.1. As myoviruses have contractile tails, it was of note that contracted and non-contracted virions were observed for phiSHEF13 and phiSHEF16 under TEM. The first is a contractile form which the phage uses to accomplish phage genome injection (Taylor et al., 2018). This form is recognised by the tail sheath being contracted along with the baseplate which makes the tail tube clearly observed (Figure 4.10 B) and (Figure 4.12 D). The other form is a noncontractile form in which the phage tail sheath is covering the tail tub indicating a potential phage for new host infection (Figure 4.10 C) and (Figure 4.12 B).

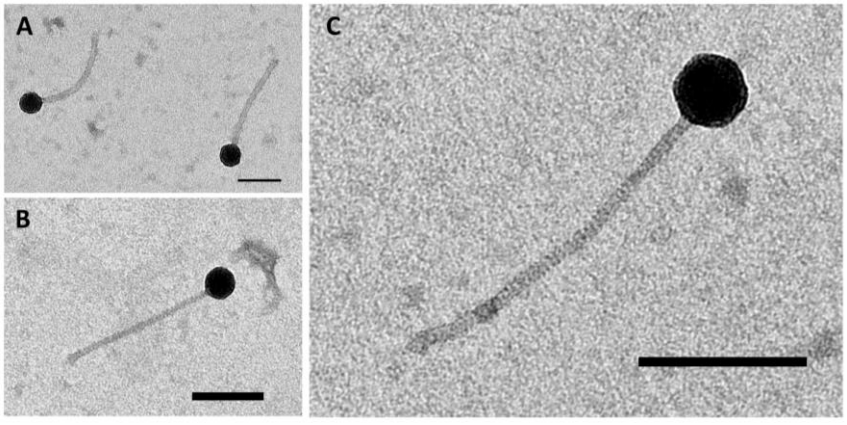


Figure 4.5. Electron micrographs of phiSHEF8.

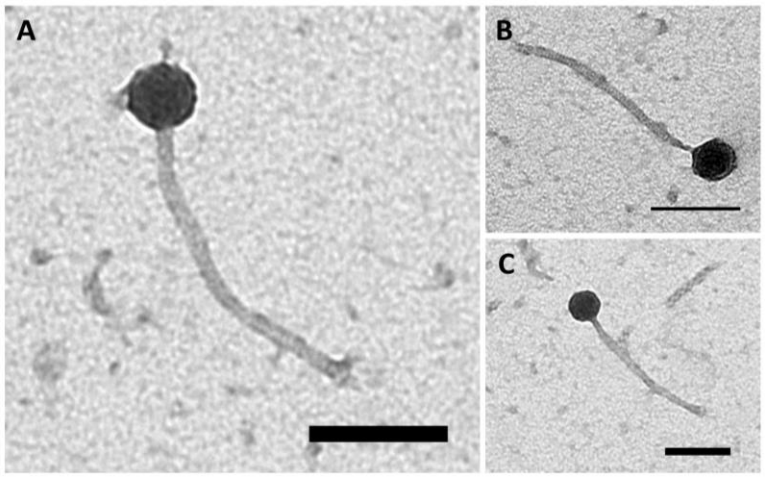


Figure 4.6. Electron micrographs of phiSHEF9

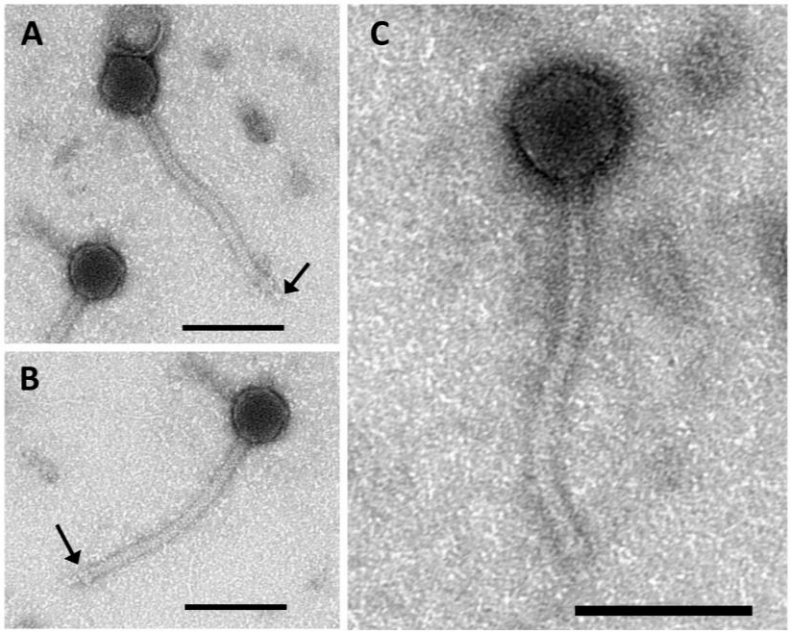


Figure 4.7. Electron micrographs of phiSHEF10. Black arrows in A indicate tail fibres while in B refer to baseplate structure.

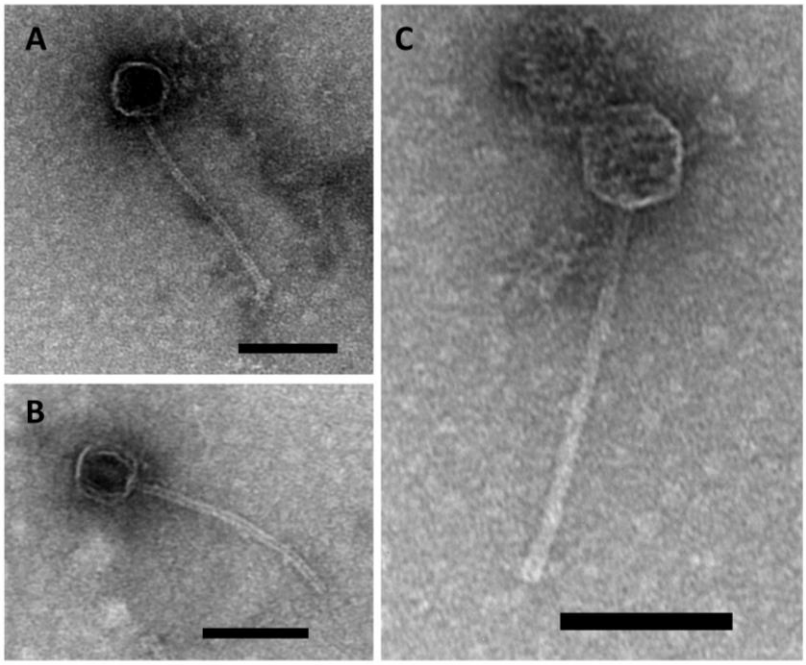


Figure 4.8. Electron micrographs of phiSHEF11.

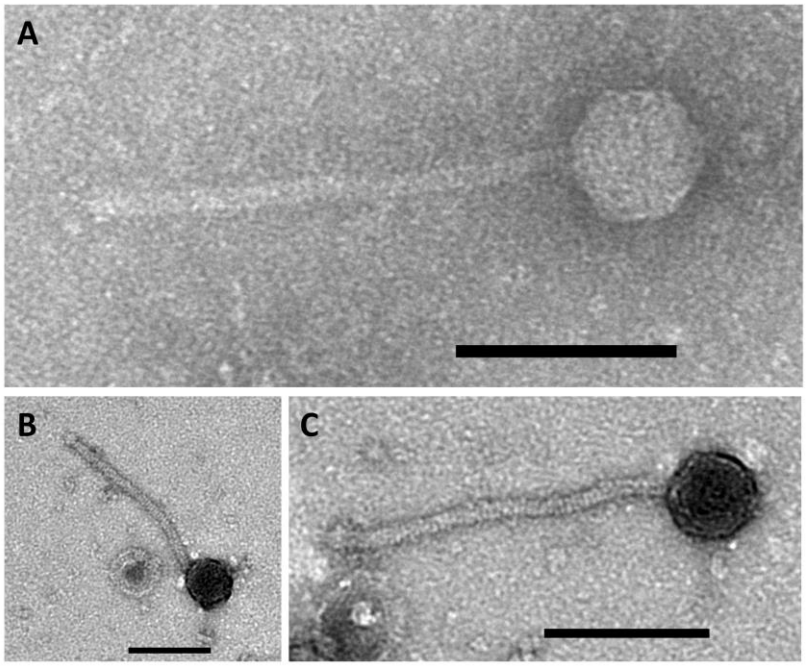


Figure 4.9. Electron micrographs of phiSHEF12.

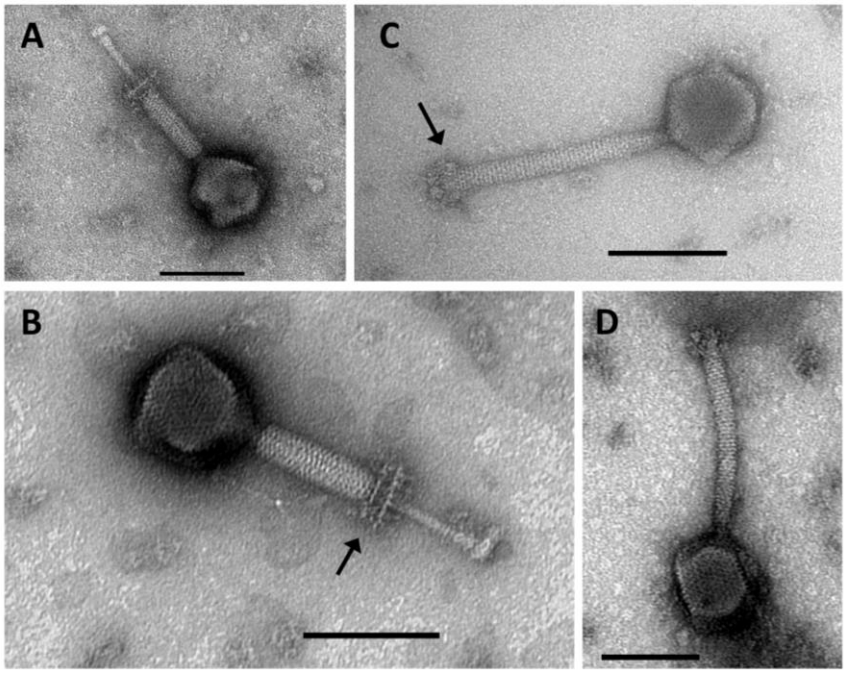


Figure 4.10. Electron micrographs of phiSHEF13. Baseplate is shown in B and C (black arrows)

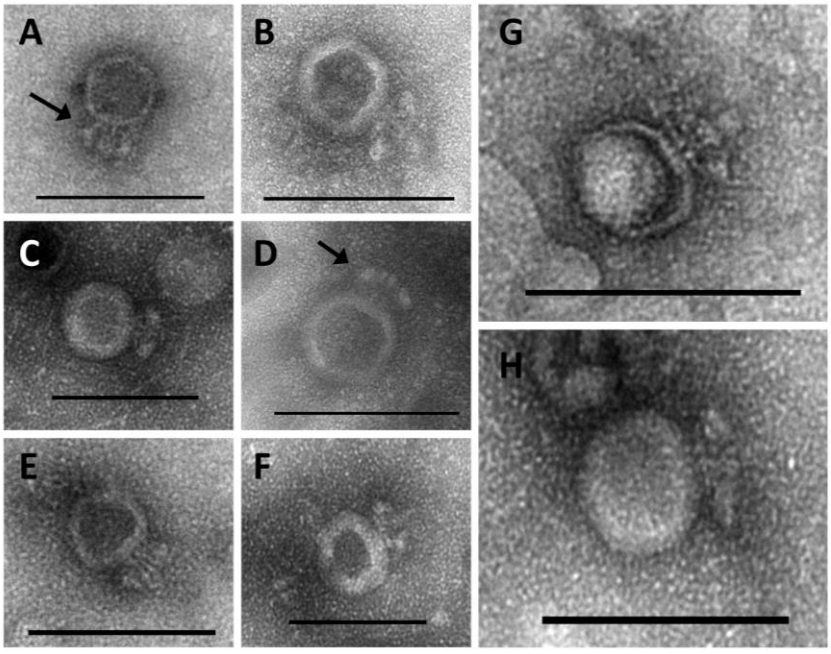


Figure 4.11. Electron micrographs of phiSHEF14. Arrows in A and D indicate tail fibres.

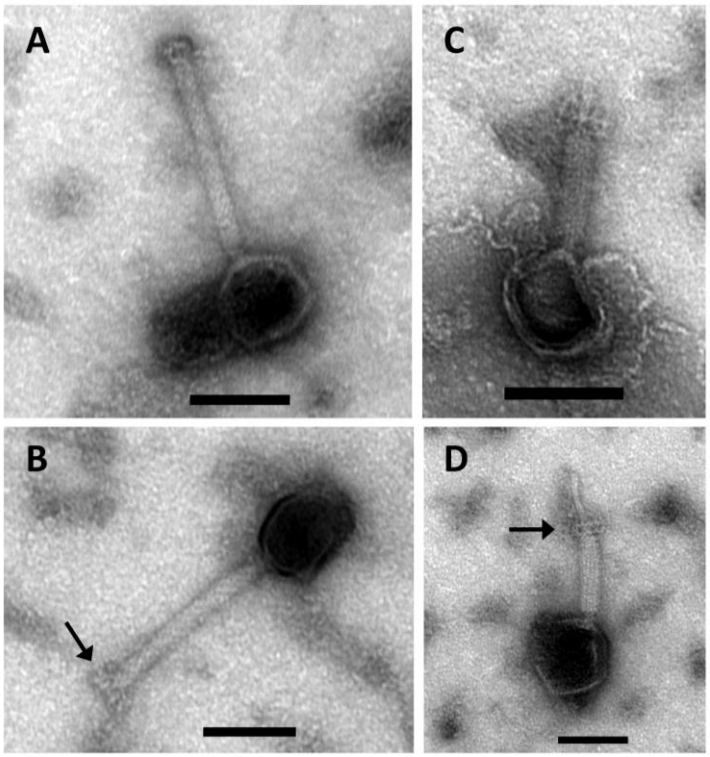


Figure 4.12. Electron micrographs of phiSHEF16. Baseplate is indicated by black arrows in B and D.

Table 4.1 Head and tail measurements. For each phage type, three phage particles were measured, and the mean value was used.

Phage	Head diameter nm (SD±)	Tail length nm (SD±)	Tail width nm (SD±)
PhiSHEF8	49 (3.6)	216.3 (2.1)	11 (1.0)
PhiSHEF9	48 (1)	232 (4.3)	12 (1.0)
PhiSHEF10	56 (2)	219 (1)	12 (0.57)
PhiSHEF11	57 (2)	219 (3.5)	12 (1)
PhiSHEF12	56 (2.6)	217 (4.3)	12.6 (0.57)
PhiSHEF13	95 (3)	199 (1.1)	22 (0.57)
PhiSHEF14	54 (1.5)	24 (2.1)	18 (5.5)
PhiSHEF16	99 (7.2)	220 (4.1)	25 (4.0)

4.4.3 Host range

As the isolated phages possess various morphologies and were isolated using different *E. faecalis* and *E. faecium* strains, the host range was determined. To enable this experiment, 36 *E. faecalis* and *faecium* strains were initially assessed for phage infection via spot tests using a defined phage titre at 10^7 PFU/ml. As a result, 20 strains showed lysis on plates which was further confirmed by serial dilution spot tests and plaque assays to determine the efficiency of plating (EOP) for each phage. EOP was calculated as follows: phage titer on the test strain divided by the phage titer on the isolating host. The interpretation of this calculation was classified as strong killing (EOP >0.01), weak killing (EOP <0.01) or no killing (no lysis on plates).

The host range analysis showed that all isolated siphoviruses (phiSHEF8-12) can infect only *E. faecalis* strains, with variable specificity, but no *E. faecium* strains. For example, both phiSHEF8 and 12 have a similar host range to phiSHEF10. The latter only differed in infecting the 8413 strain (Table 4.2). For phiSHEF9 and 11, the host range analysis of these showed completely different results indicating that these are possibly genomically diverged from each other. For phages infecting *E. faecium* strains, phiSHEF13 showed a broad host range infecting both *E.* faecalis and faecium strains while phiSHEF14 and 16 were limited to only E. faecium strains (Table 4.2). In fact, phiSHEF14 only infects the isolation host E1071 after testing the 36 different strains while phiSHEF16 infects 5 *E. faecium* strains with variable EOP.

The host range testing included several VRE strains, namely *E. faecalis* V583 and *E. faecium* dp9, E1679 and E1071. phiSHEF13 was able to effectively infect V583 and dp9 alongside with isolation host E1071. phiSHEF16 also showed a weak killing activity (EOP<0.01) against both dp9 and E1679 strains (Table 4.2).

The isolated phages showed no activity against 16 *E. faecalis* and *faecium* strains. The *E. faecalis* strains include dp2,dp4, R178, R197, R51, R5, K2756-02, R53, R70, ATCC 51575 while the *E. faecium* strains dp6, dp8, E980, E0317, E7345,E1162. Of note, the strain E1162 belongs to the EPA type 3 while the strains E980, E0317, and E7345 belong to the EPA type 4.

Table 4.2 Host range of the isolated phages. The enterococcal strains are on the left and VRE are labelled with a red colour. phiSHEF phages are on the top with different killing efficiency represented as shown below the table.

4.4.4 Genomic analysis

To further analyse our isolated phages, PEG-concentrated phage lysates were used for phage DNA extraction. The analysis of phage genomes is the most reliable test to ascertain that the isolated phages are novel and therefore can be added to phage databases. The extracted DNA samples were sent to (MicrobesNG, Birmingham, UK) for genomic sequencing using Illumina. A few phage samples were also sent to Sheffield Institute for Translational Neuroscience for genome sequencing using Nanopore technology. Comparing with Illumina, Nanopore technology allows fast sequencing and long reads. Therefore, the fast sequencing was exploited to quickly analyse phage genomes focusing on phage novelty and other genomic features like phage life cycle. The annotation of phage genomes was performed via the RASTtk pipeline (https://rast.nmpdr.org/rast.cgi) as well as manually by examining each coding sequence (CDS) for protein domains using Pfam and NCBI domain database.

From the host range analysis, two phages (phiSHEF8 and 12) showed identical host range results as well as phage morphology. A third phage (phiSHEF10) differed slightly from the other two in that it only weakly infects an additional *E. faecalis* strain 8413. To investigate this, the genomes of these three phage strains were sequenced (phiSHEF8 by Nanopore while phiSHEF10 and 12 by Illumina). After comparing the genomes of these phages with the viral database using BLASTn, phiSHEF8,10 and 12 showed 99.91, 99.61 and 100% identity, respectively, to another phage phiSHEF2 (**Table 4.3**). This was further investigated by aligning these genomes which showed complete alignment between phiSHEF2,8 and 12 (Figure 4.13A). For phiSHEF10, a complete alignment was found compared with the other three phages except for a specific region between the HNH endonuclease and DNA primase genes (Figure 4.13A). Upon analysing this region, only hypothetical genes were found after automatic and manual annotations. For phiSHEF2 and 10, a 97% coverage was found which can explain the difference in the alignment seen in (Figure 4.13A).

Almost an identical result was also found for phiSHEF9 (100% identity, 99% coverage) compared with another phage phiSHEF5 after BLASTn analysis. Aligning these two genomes

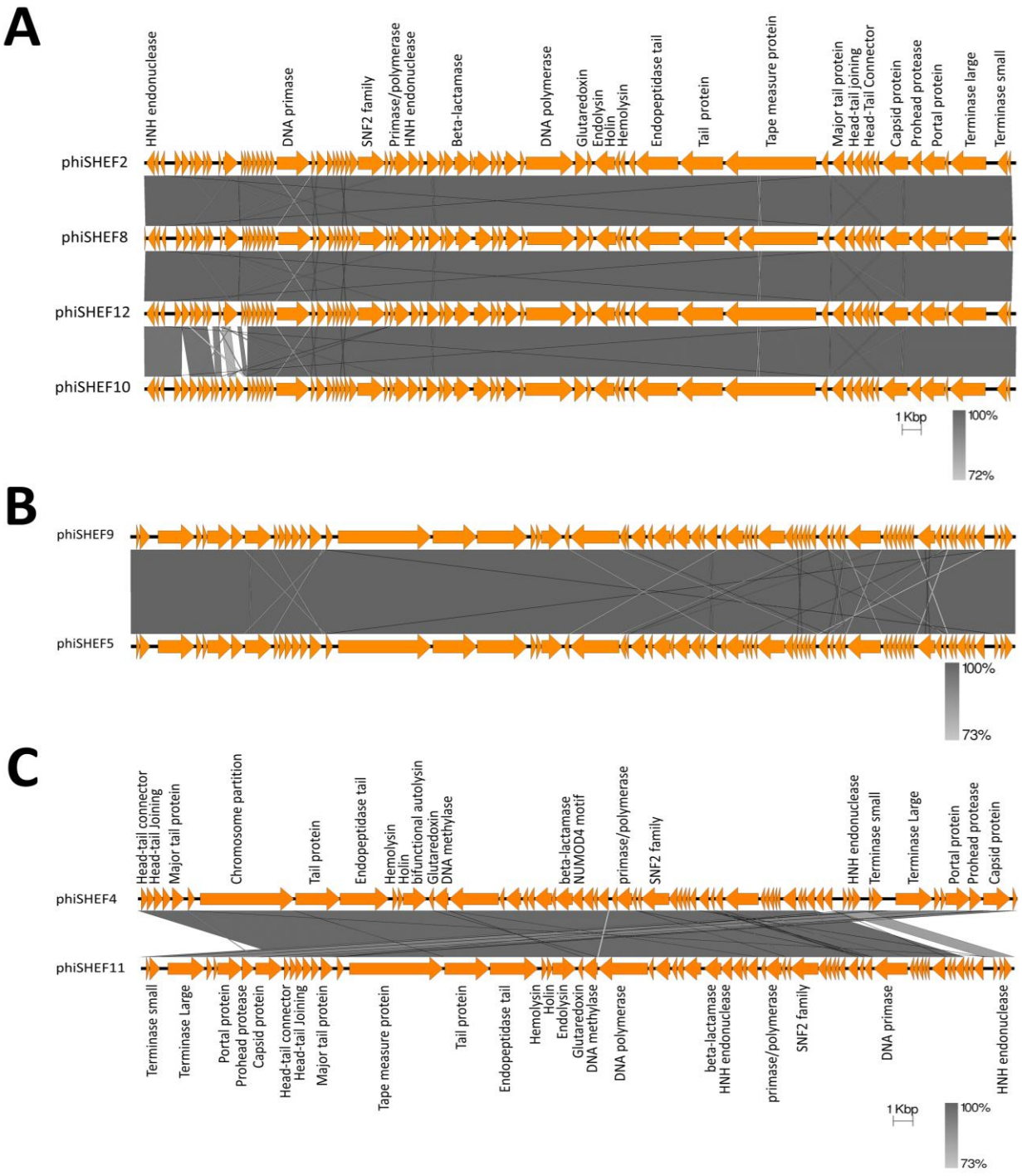
further confirmed these almost identical sequences (Figure 4.13 B). For the siphovirus phiSHEF11, the closest hit was phiSHEF4 phage (98 identity, 97% coverage) and the difference between these two genomes was concentrated on hypothetical genes (Figure 4.13 C).

For phiSHEF13, the closest hit was the EFDG1 phage with 92% identity (Figure 4.13 D). PhiSHEF14 also showed a similar result (92% identity) compared with the closest hit the vB_EfaP_Zip phage (Figure 4.13 F). Lastly, the closest hit for phiSHEF16 was Porthos phage with 97.1% identity (Figure 4.13 E).

Hence, the genomic analysis has shown that a total of 5 phages (phiSHEF10,11,13,14,16) are novel and therefore were deposited in the NCBI database.

Table 4.3. Genomic analysis of the isolated phages. *indicates Nanopore sequencing

phiSHEF	Size (kb)	Closest BLASTn hit	Identity %	Coverage %
8*	41.8	phiSHEF2	99.91	99
9	41.6	phiSHEF5	100	99
10	41.6	phiSHEF2	99.61	97
11	40.7	phiSHEF4	98.8	97%
12	41.7	phiSHEF2	100	100
13	151.3	EFDG1	92	89
14	19.3	vB_EfaP_Zip	92	91
16	152.9	Porthos	97.1	92


A
HNH endonuclease
DNA primase
SNF2 family
Primase/polymerase
HNH endonuclease
Beta-lactamase
DNA polymerase
Glutaredoxin
Endolysin
Holin
Hemolysin
Endopeptidase tail
Tail protein
Tape measure protein
Major tail protein
Head-tail joining
Head-Tail Connector
Capsid protein
Prohead protease
Portal protein
Terminase large
Terminase small
phiSHEF2
phiSHEF8
phiSHEF12
phiSHEF10
1 Kbp
100%
72%
B
phiSHEF9
phiSHEF5
100%
73%
C
Head-tail connector
Head-tail Joining
Major tail protein
Chromosome partition
Tail protein
Endopeptidase tail
Hemolysin
Holin
bifunctional autolysin
Glutaredoxin
DNA methylase
beta-lactamase
NUMOD4 motif
primase/polymerase
SNF2 family
HNH endonuclease
Terminase small
Terminase Large
Portal protein
Prohead protease
Capsid protein
phiSHEF4
phiSHEF11
Terminase small
Terminase Large
Portal protein
Prohead protease
Capsid protein
Head-tail connector
Head-tail Joining
Major tail protein
Tape measure protein
Tail protein
Endopeptidase tail
Hemolysin
Holin
Endolysin
Glutaredoxin
DNA methylase
DNA polymerase
beta-lactamase
HNH endonuclease
primase/polymerase
SNF2 family
DNA primase
HNH endonuclease
1 Kbp
100%
73%

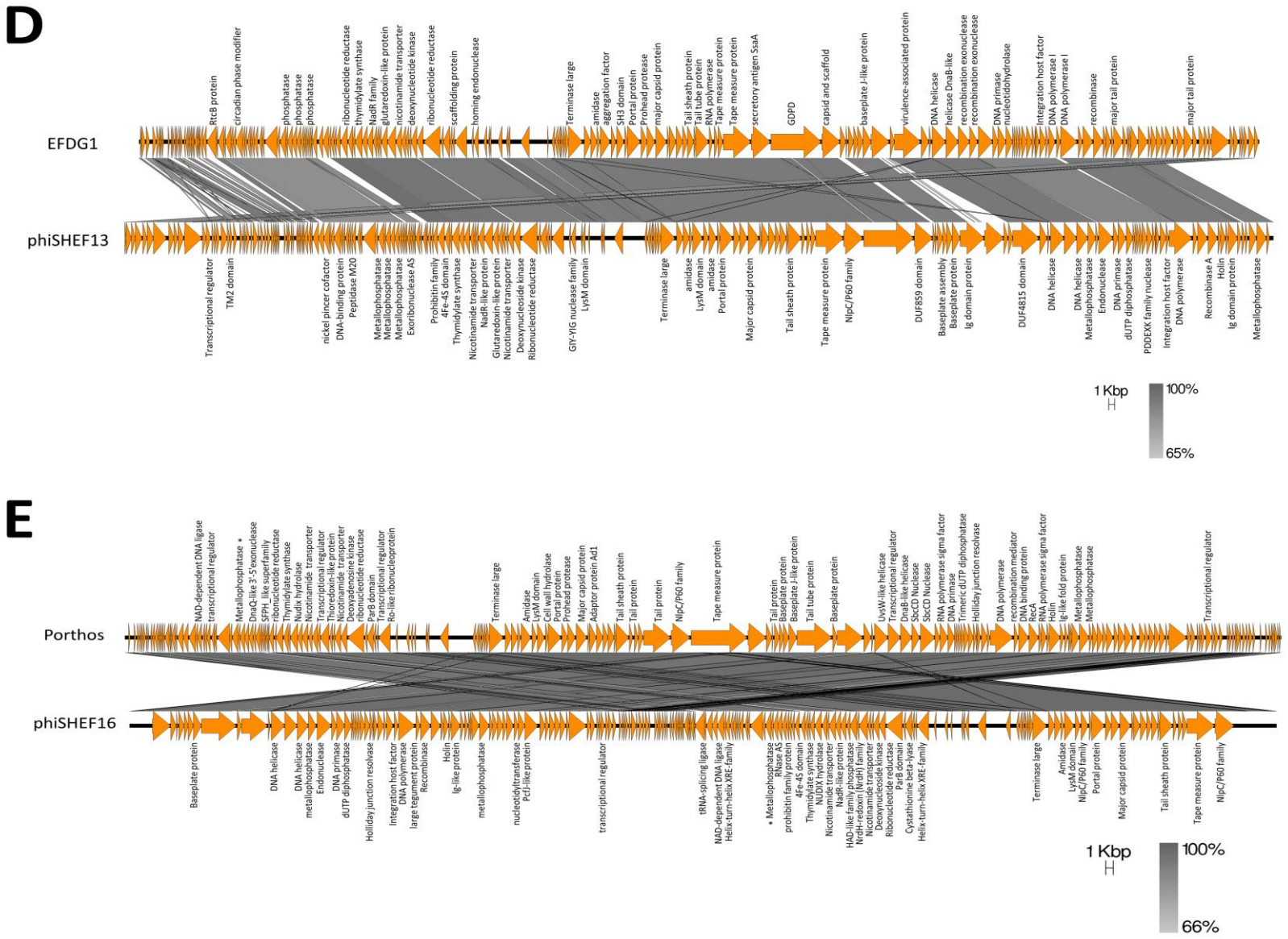

D
EFDG1
phiSHEF13
RtcB protein
circadian phase modifier
phosphatase
phosphatase
phosphatase
ribonucleotide reductase
thymidylate synthase
NadR family
glutaredoxin-like protein
nicotinamide transporter
deoxynucleotide kinase
ribonucleotide reductase
scaffolding protein
homing endonuclease
Terminase large
amidase
aggregation factor
SH3 domain
Portal protein
Prohead protease
major capsid protein
Tail sheath protein
Tail tube protein
RNA polymerase
Tape measure protein
Tape measure protein
secretory antigen SsaA
GDPD
capsid and scaffold
baseplate J-like protein
virulence-associated protein
DNA helicase
helicase DnaB-like
recombination exonuclease
recombination exonuclease
DNA primase
nucleotidohydrolase
Integration host factor
DNA polymerase I
DNA polymerase I
recombinase
major tail protein
major tail protein
Transcriptional regulator
TM2 domain
nickel pincer cofactor
DNA-binding protein
Peptidase M20
Metallophosphatase
Metallophosphatase
Metallophosphatase
Exoribonuclease AS
Prohibitin family
4Fe-4S domain
Thymidylate synthase
Nicotinamide transporter
NadR-like protein
Glutaredoxin-like protein
Nicotinamide transporter
Deoxynucleoside kinase
Ribonucleotide reductase
GIY-YIG nuclease family
LysM domain
Terminase large
amidase
LysM domain
amidase
Portal protein
Major capsid protein
Tail sheath protein
Tape measure protein
NlpC/P60 family
DUF859 domain
Baseplate assembly
Baseplate protein
Ig domain protein
DUF4815 domain
DNA helicase
DNA helicase
Metallophosphatase
Endonuclease
DNA primase
dUTP diphosphatase
PDDEXK family nuclease
Integration host factor
DNA polymerase
Recombinase A
Holin
Ig domain protein
Metallophosphatase
1 Kbp
100%
65%
E
Porthos
phiSHEF16
NAD-dependent DNA ligase
transcriptional regulator
Metallophosphatase *
DnaQ-like 3'-5'exonuclease
SPFH_like superfamily
ribonucleotide reductase
Thymidylate synthase
Nudix hydrolase
Nicotinamide transporter
Transcriptional regulator
Thioredoxin-like protein
Nicotinamide transporter
Deoxynucleoside kinase
ribonucleotide reductase
ParB domain
Transcriptional regulator
Ro-like ribonucleoprotein
Terminase large
Amidase
LysM domain
Cell wall hydrolase
Portal protein
Prohead protease
Major capsid protein
Adaptor protein Ad1
Tail sheath protein
Tail protein
Tail protein
NlpC/P60 family
Tape measure protein
Tail protein
Baseplate protein
Baseplate J-like protein
Tail tube protein
Baseplate protein
UvsW-like helicase
Transcriptional regulator
DnaB-like helicase
5'cCD Nuclease
3'cCD Nuclease
RNA polymerase sigma factor
DNA primase
Trimeric dUTP diphosphatase
Holliday junction resolvase
DNA polymerase
recombination mediator
DNA binding protein
RecA
RNA polymerase sigma factor
Holin
Ig-like fold protein
Metallophosphatase
Metallophosphatase
Transcriptional regulator
Baseplate protein
DNA helicase
DNA helicase
metallophosphatase
Endonuclease
DNA primase
dUTP diphosphatase
Holliday junction resolvase
Integration host factor
DNA polymerase
large tegument protein
Recombinase
Holin
Ig-like protein
metallophosphatase
nucleotidyltransferase
PcfJ-like protein
transcriptional regulator
tRNA-splicing ligase
NAD-dependent DNA ligase
Helix-turn-helix XRE-family
* Metallophosphatase
RNase AS
prohibitin family protein
4Fe-4S domain
Thymidylate synthase
NUDIX hydrolase
Nicotinamide transporter
NadR-like protein
HAD-like family phosphatase
NrdH-redoxin (NrdH) family
Nicotinamide transporter
Deoxynucleoside kinase
Ribonucleotide reductase
ParB domain
Cystathionine beta-lyase
Helix-turn-helix XRE-family
Terminase large
Amidase
LysM domain
NlpC/P60 family
Portal protein
Major capsid protein
Tail sheath protein
Tape measure protein
NlpC/P60 family
1 Kbp
100%
66%

Figure 4.13. Comparative analysis of the isolated phage genomes. A) phiSHEF2,8,12 and 10 genomes, B) phiSEHF5, C) phiSHEF11, D) phiSHEF13, E) phiSHEF16 F) phiSHEF14. The orange arrows indicate CDSs and the gene similarity profiles between phages are indicated in grayscale (and percentage). CDSs without a label indicate a hypothetical protein. The genomic comparison was made using BLASTn within EasyFig.

Following genome sequencing and annotation, the final five annotated genomes are shown in Table 6.1Table **6.2**Table **6.3**Table 6.4Table 6.5. As phage genomes are known for their modular organization (Moura de Sousa et al., 2021), this has also been found with the novel isolated phages here. This means that genes with related functions are generally located close to each other in the genome. For instance, the genomes of the siphoviruses phiSHEF10 and 11 have shown the following order of modules: Head-Tail-Lysis-Packaging-DNA metabolism. For the myoviruses phiSHEF13 and 16, the order of modules is as follows: Tail-Lysis-Head-Packaging-DNA metabolism. For the podovirus phiSHEF14, the modules (Packing-Lysis-Tail-Head) were identified.

Furthermore, the genomes of the novel isolated phages lack the main genes involved in the lysogenic life cycle (i.e. integrase and repressor genes) indicating that these phages are

obligatorily lytic. Additionally, hypothetical proteins (proteins with no known functions) were found to consist high percentage of the genomes. For example, 67.1 % of the phiSHEF13 genome belongs to hypothetical proteins (**Table 4.4**).

Table 4.4 Genomic characterisation of the novel isolated phages. HP refers to hypothetical proteins.

phiSHEF	ICTV classification	Size (kb)	CDS	GC%	HP %
10	Efquatrovirus (genus), SHEF2 (species)	41.6	67	35%	67.1%
11	Efquatrovirus (genus), SHEF4 (species)	40.7	63	35%	63.5%
13	Herelleviridae (family) Schiekvirus (genus)	151.3	219	37%	67.1%
14	Rountreeviridae (family); Sarlesvirinae (subfamily); Minhovirus (genus)	19.3	22	35%	45.5%
16	Herelleviridae (family) Schiekvirus (genus)	152.9	185	37%	65.4%

As phages mainly exploit their tail structure to accomplish the first steps in their life cycles (i.e. phage adsorption and genome ejection), certain phage lysins (called Tail associated lysins) facilitate this process. Therefore, an investigation for these lysins was performed on all the novel isolated phages. This revealed that the siphoviruses phiSHEF10 and 11 both possess proteins with predicted endopeptidase activities (phiSHEF10_17 and phiSHEF11_46). The analysis of the myoviruses (phiSHEF13 and 16) also revealed two proteins in the tail module with predicted lytic activities. These putative lysins harbour NLPC/P60 (phiSHEF13_50, phiSHEF16_1) as well as lytic transglycosylase (phiSHEF13_51, phISHEF16_2) domains. For phiSHEF14, a predicted NLPC/P60 domain (phiSHEF14_11) was identified in a protein located between the HNH homing endonuclease and the endolysin proteins. The identified NLPC/P60 proteins here were also compared with the enterococcal NLPC/P60 proteins (discussed in chapter 3). This showed

that phiSHEF14_11 belongs to group 1C (Minhovirus) while both phiSHEF13_50, and phiSHEF16_1 fit in group 2A (Schiekvirus) (Figure 4.14).

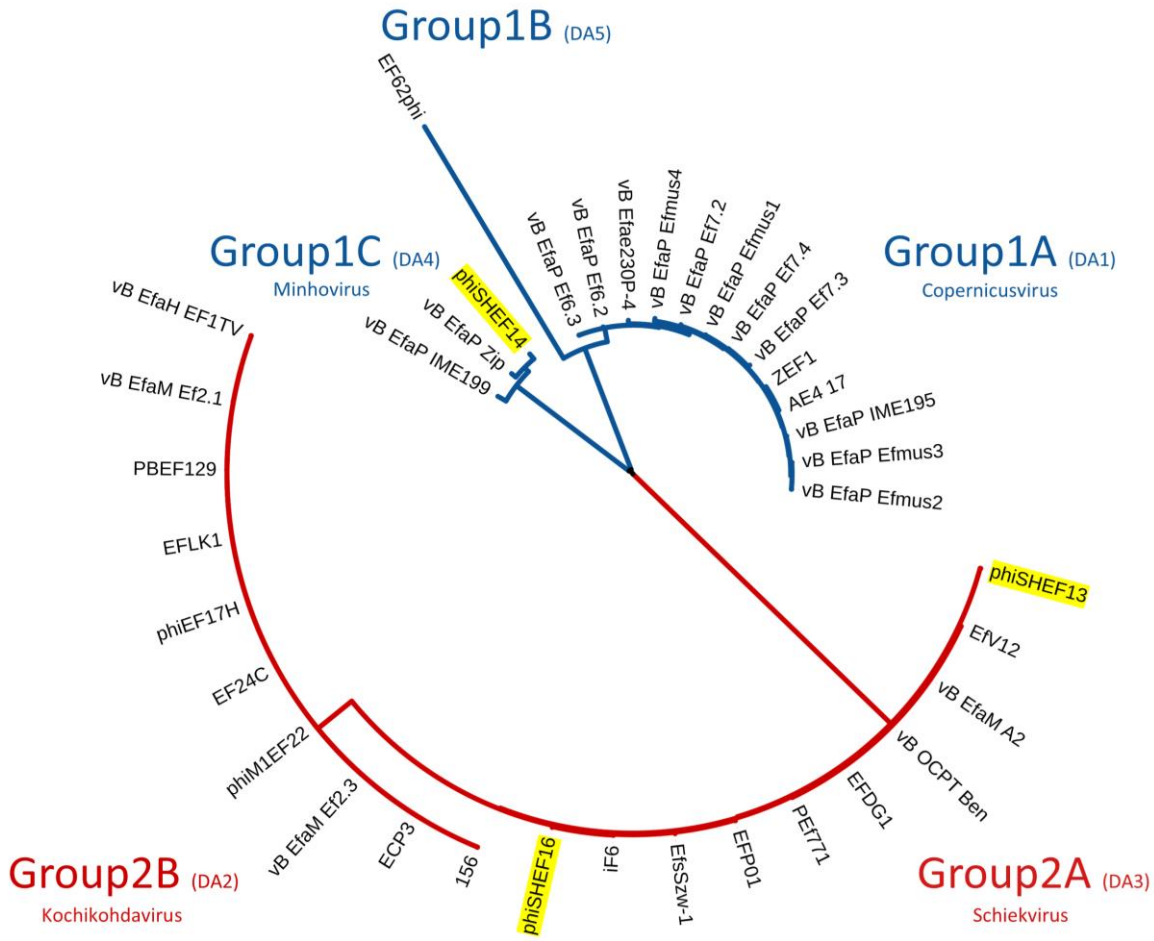


Figure 4.14. A phylogenetic tree of enterococcal NLPC/P60 containing proteins. This shows two main groups based on phage genomic classification and morphology: sequences from podoviruses labelled Blue and myoviruses labelled red. NLPC/P60 sequences from phiSHEF13, 14 and 16 were labelled with yellow. The tree was constructed using FastTree and visualised using the ITOL online website.

As three of the isolated phages were isolated using the *E. faecium* strain E1071, further investigation was carried out about any shared genomic features in these genomes. For this, the genomes of phiSHEF13, 14 and 16 were first aligned together to spot any shared genomic regions. This was resulted in finding a specific CDS that shows a level of similarity across the three genomes (Figure 4.15 A). These CDSs are phiSHEF13_41 (Ig-like domain-containing protein), phiSHEF14_15 (hypothetical protein) and phiSHEF16_201 (hypothetical protein) which are all located in the tail modules. These were then aligned together to have a closer look at the protein homology (Figure 4.15 B and C). The amino acid sequences of these proteins were also aligned using Multalin to further confirm the alignment results which revealed high similarity between the three proteins in the central region. The N and C- terminals of phishef13_41 shared homology with phishef16_201 while phiSHEF14_15 shared homology with the centre and C-terminal of phishef16_201 (Figure 4.16). Additionally, these proteins were further assessed using the structural analysis database (Phyre2) which showed that all three proteins have homology to a sugar-binding protein (structure of the n-terminal cbm22-1-cbm22-2 tandem domain from 2 *paenibacillus barcinonensis* xyn10) with over 98% confidence.

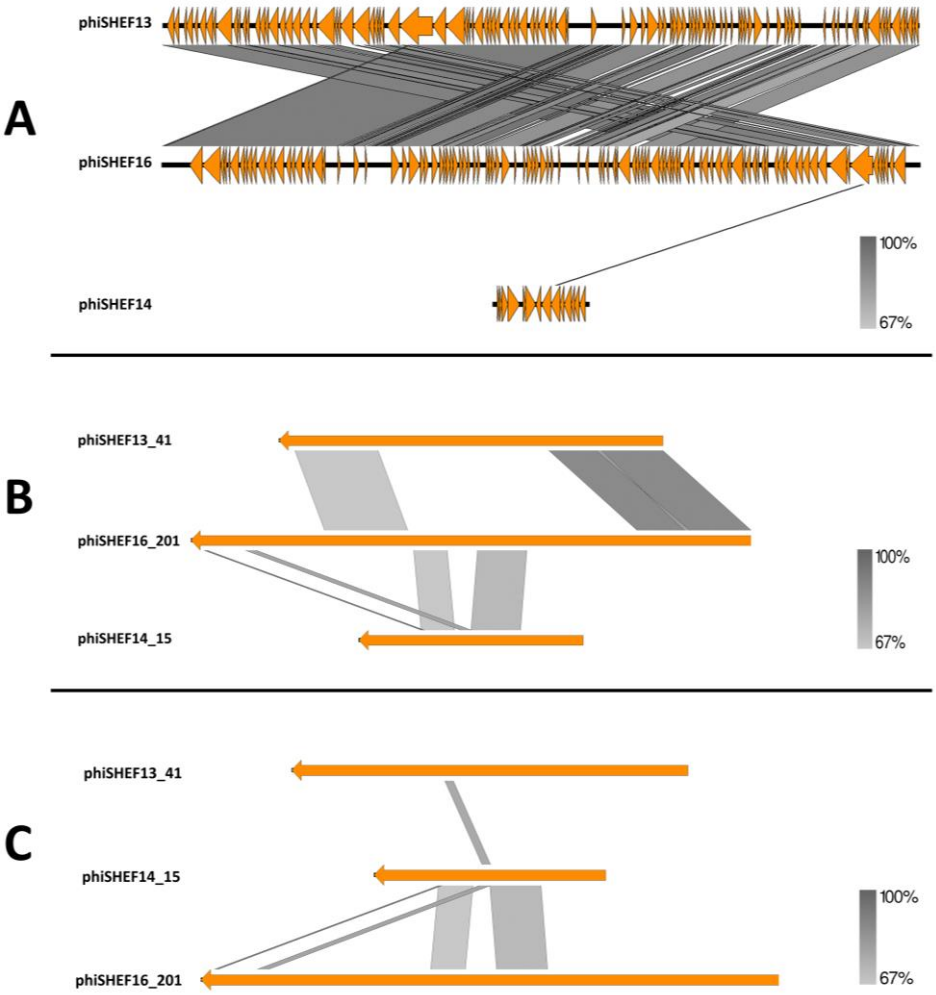


Figure 4.15. A) comparative genomic analysis of phiSHEF13, 14 and 16. B) and C) the alignment of the shared protein between the three phage genomes. The level of identity is indicated by the grey region between genomes.

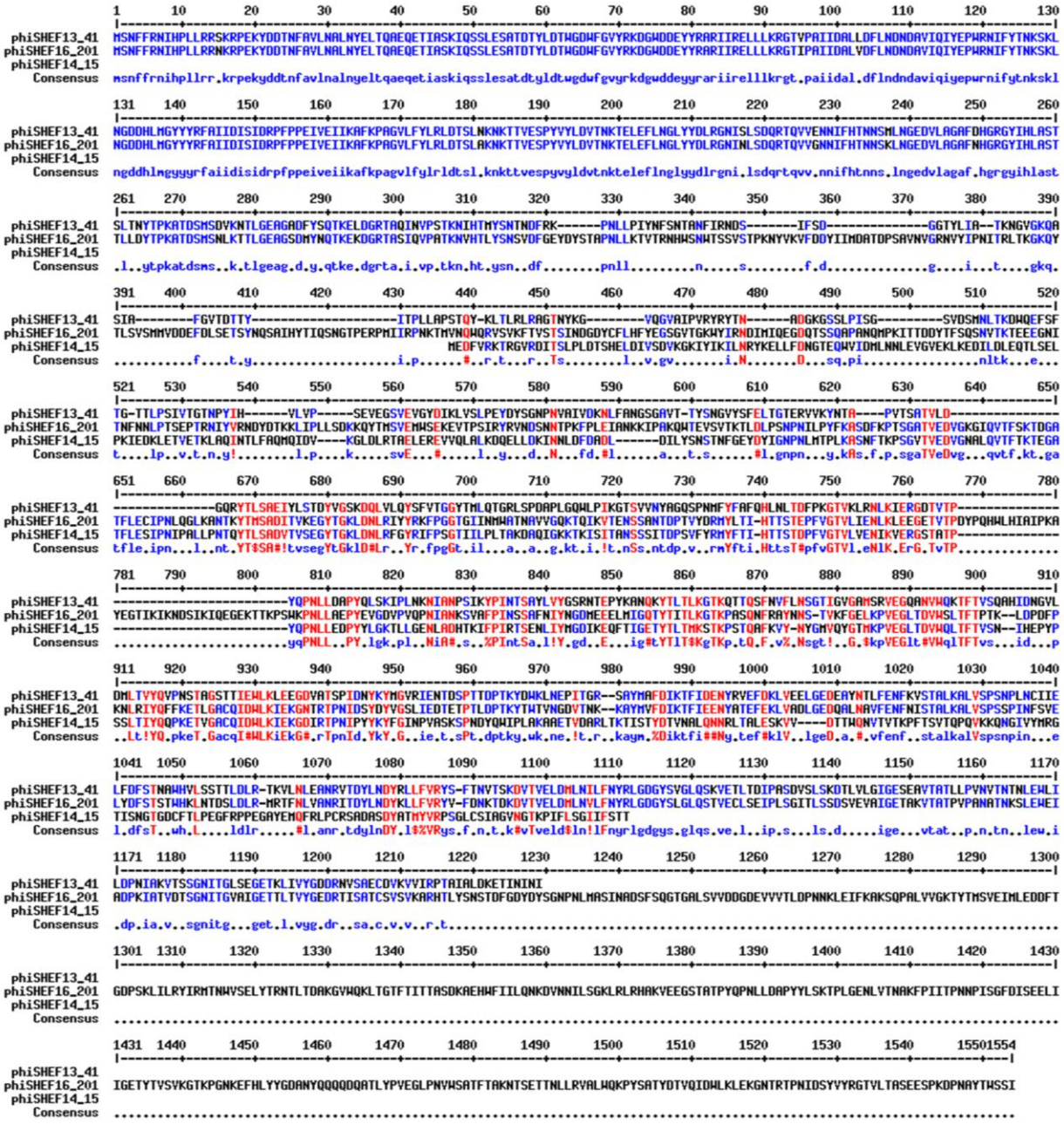


Figure 4.16. Multiple sequence alignment of phiSHEF13_41, phiSHEF16_201 and phiSHEF14_15 proteins. The alignments showed high similarity (red font) between the three proteins at the central region (440-1110aa). Both phiSHEF13_41 and phiSHEF16_201 show also high similarity (blue font) at the start (1-432aa) and end (1112-1215aa) of the sequences.

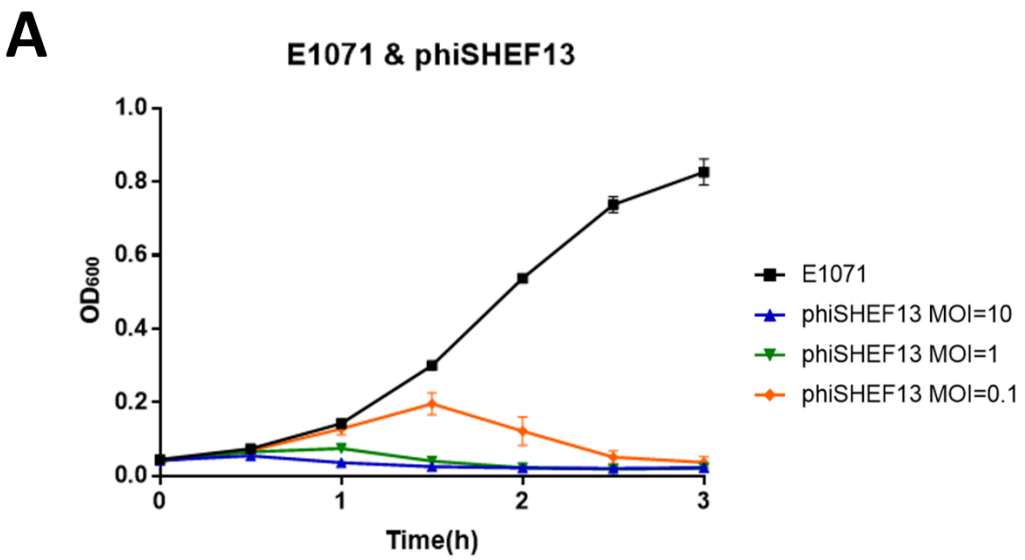
4.5 Further characterisation of phiSHEF13, 14 and 16:

Given that three phages were isolated using the VRE *E. faecium* E1071 strain, additional experiments were conducted to further investigate these phages and assess their efficiency.

4.5.1 Killing assays

As the three phages (phiSHEF13, 14 and 16) showed successful infection of the *E. faecium* E1071 strain, this infectivity was further analysed using killing assays. These assays show the effect of phages on planktonic bacteria using different multiplicity of infections MOI (10, 1 and 0.1) in a dose-dependent manner.

In all three phages, a rapid inhibition (less than 2h) of bacterial growth was observed at all MOIs applied (Figure 4.17). Of note, the lowest MOI (0.1) showed the slowest inhibitory effect on E1071 growth while the higher MOI (1 and 10) showed the quickest inhibition (less than 1h).



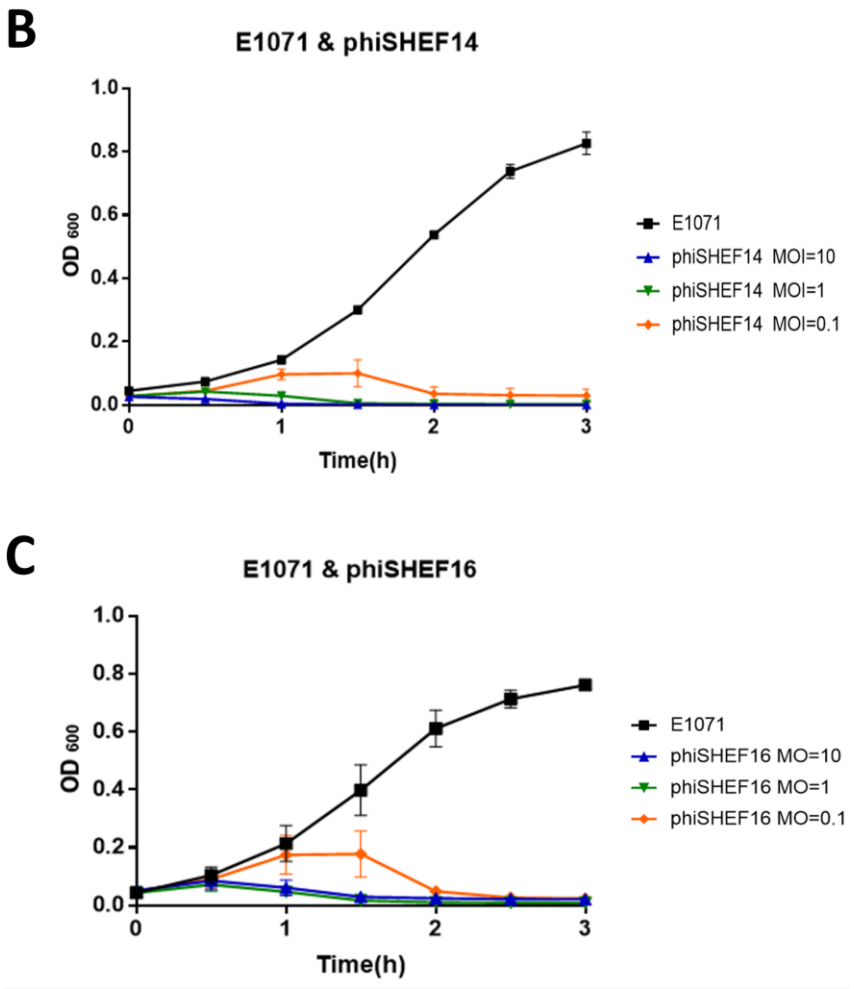


Figure 4.17. Killing assays of phiSHEF13, 14 and 16. The different MOI are indicated by colours: blue (MOI 10), green (MOI 1) and orange (MOI 0.1). This experiment was done in triplicate and Error bars represent SEM for three replicates.

Nevertheless, extended incubation of the killing assays showed the development of phage-resistant mutants (RM). For phiSHEF13, RM emergence was the quickest with MOI of 1 after 6 h incubation while MOI of 10 was the slowest, with RM developed after 9 h incubation (Figure 4.18 A). In contrast, phiSHEF14 showed the quickest RM development with an MOI of 10 while an MOI of 0.1 led to a slow RM emergence (Figure 4.18 B). For phiSHEF16, a varied development of RM results was seen and repeated experiments showed varied development of RM either after 6 h or 12 h in all MOIs (Figure 4.18 C).

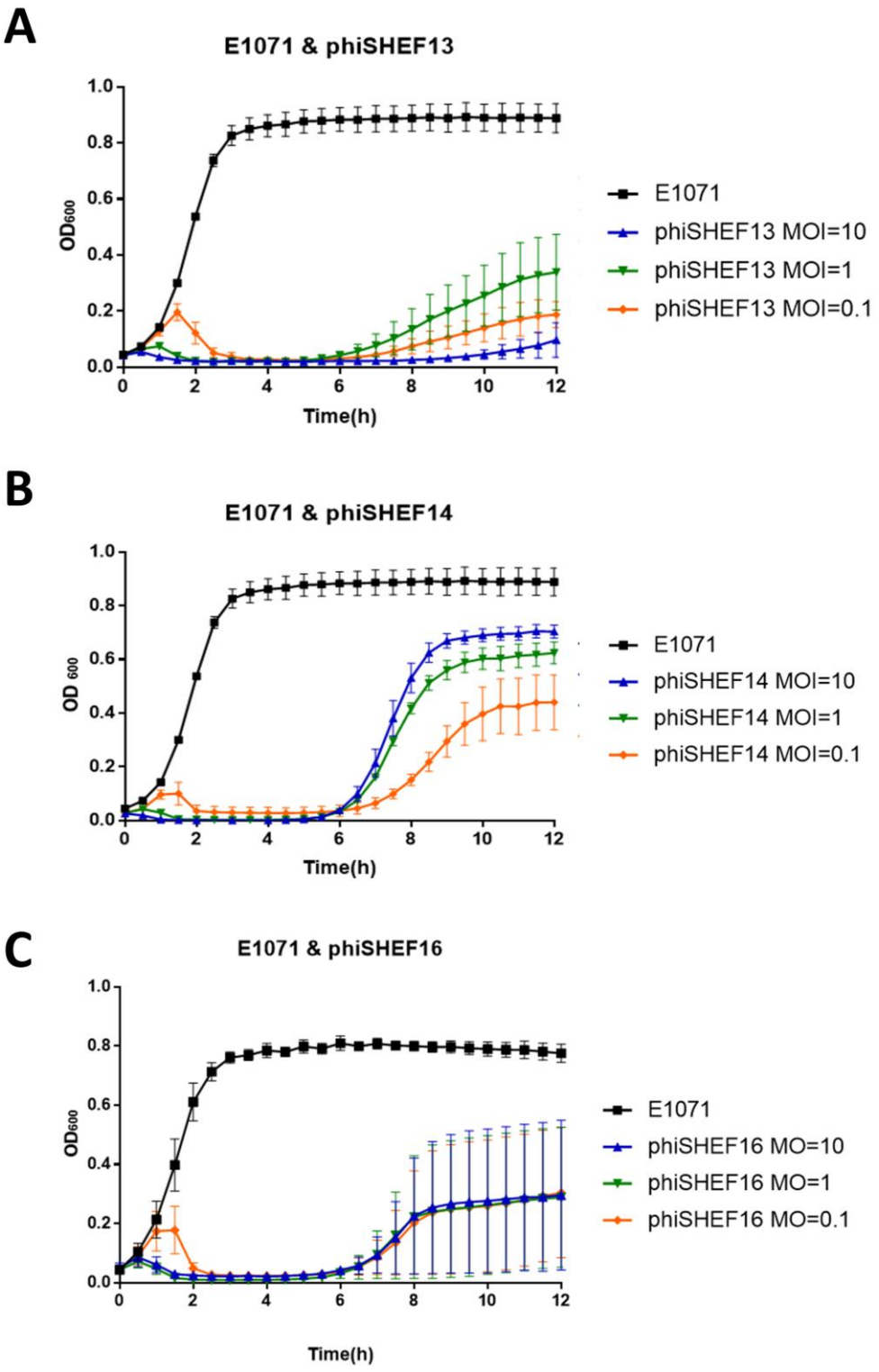


Figure 4.18. 12 hours Killing assays of phiSHEF13 ,14 and 16. The different MOI are indicated by colours: blue (MOI 10), green (MOI 1) and orange (MOI 0.1). This experiment was done in triplicate and Error bars represent SEM for three replicates.

As the E1071 strain can be infected with the three phages (phiSHEF13,14 and16), a cocktail-killing assay was also tested and analysed. This revealed a similar pattern in the first 3h of the infection in which the highest MOI showed the quickest inhibition of growth whereas the lowest MOI has the slowest inhibition (Figure 4.19 A). Prolonged incubation has resulted in the development of phage-resistant mutants in all MOIs. The timing for RM development varied from 6 to 12 hours after the killing assay has started. This variability in the biological repeats can be seen in (Figure 4.19 B).

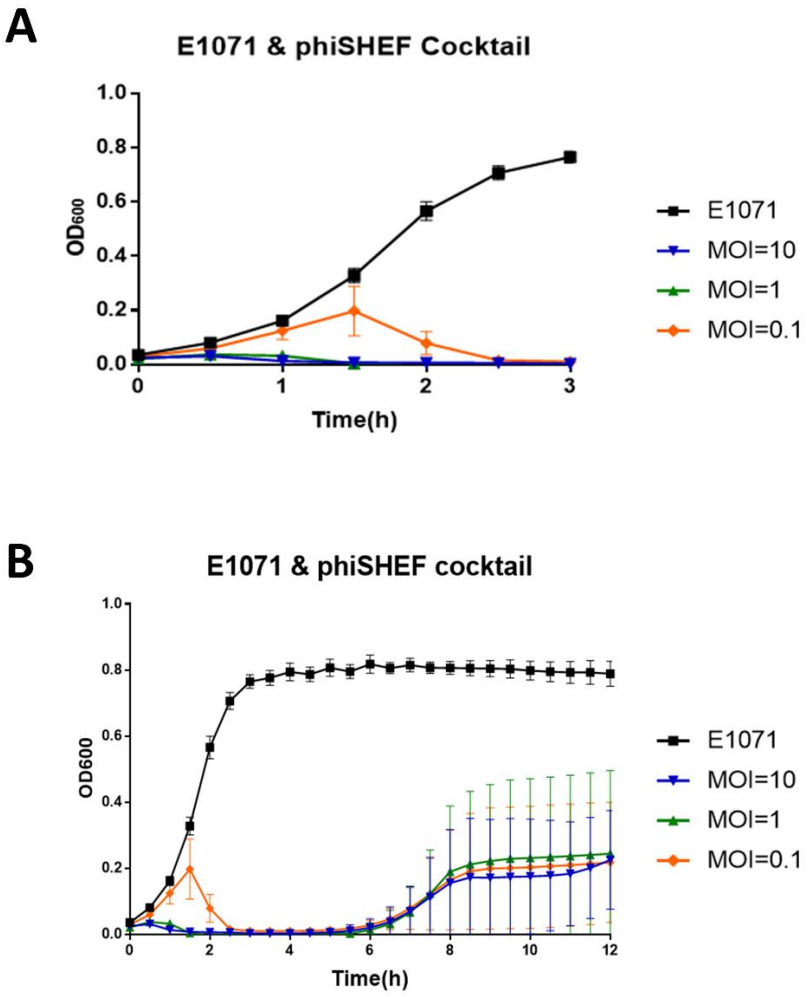


Figure 4.19. Phage cocktail killing assay using phiSHEF13, 14 and 16. A) killing assay of the first 3 hours. B) Killing assay of 12 hours showing the development of phage-resistant mutants after 6 h. The different MOI are indicated by colours: blue (MOI = 10), green (MOI = 1) and orange (MOI = 0.1). This experiment was done in triplicate and Error bars represent SEM for three replicates.

4.5.2 E1071 resistance mutants via phiSHEF13, 14 and 16

The killing assays that were performed using phiSHEF13, 14 and 16 have revealed the emergence of resistant mutants (RMs) using the E1071 strain. Therefore, further investigation of these RMs was carried out. Firstly, phage-resistant mutants were obtained by killing assays and streaked on BHI plates. 20 individual colonies were then selected from each tested phage (i.e. a total of 60 colonies from the three phages). These colonies were checked by Gram stain (Gram-positive diplococci) as well as on bile esculin agar (black pigmentation of colonies) to confirm that these are enterococcal bacteria (Chuard & Reller, 1998). Each RM colony was then tested for phage infectivity using the three phages (separately) at MOI of 1.

For RMs developed by phiSHEF13, both phages (phiSHEF14 and 16) inhibited the bacterial growth while no effect (Figure 4.20 A) or slight effect (Figure 4.20 B) was observed with phiSHEF13 (Table 4.10). All phiSHEF13 RM strains demonstrated susceptibility to inhibition by both phiSHEF14 and phiSHEF16 indicating no clear effect of phiSHEF13 resistance on their infectivity. As can be seen in (Figure 4.20 B), the growth of 13RM7 was slightly affected compared with the control. However, there is a clear difference between the effect of both phiSHEF14 and16 (complete inhibition of growth) and phiSHEF13 on 13RM7 (Figure 4.20 B) (Table 4.10).

For phiSHEF13 RMs, killing assays have revealed that both phiSHEF14 and 16 inhibited the bacterial growth while no effect (Figure 4.20 A) or slight effect (Figure 4.20 B) was observed with phiSHEF13 (Table 4.5). In all the tested phiSHEF13 RM strains, inhibition by both the phiSHEF14 and 16 was observed indicating no clear effect of phiSHEF13 resistance on their infectivity. As can be seen in (Figure 4.20 B), the growth of 13RM7 was slightly affected compared with the control. However, there is a clear difference between the effect of both phiSHEF14 and16 (complete inhibition of growth) and phiSHEF13 on 13RM7 (Figure 4.20 B) (Table 4.5)).

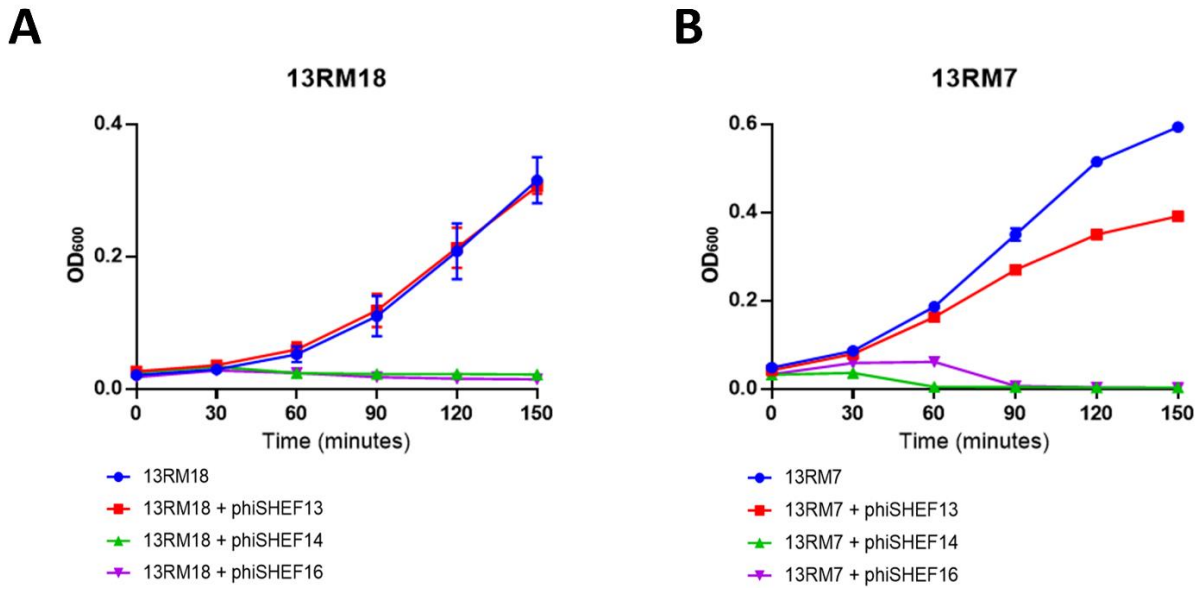


Figure 4.20. Killing assays of phiSHEF13 resistant mutants using phiSHEF13,14 and 16. A and B show the bacterial growth curve alone or with phage as indicated by colours: bacteria only (Blue), bacteria and phiSHEF13 (Red), bacteria and phiSHEF14 (Green), bacteria and phiSHEF16 (Purple). This experiment was done in duplicate and error bars represent SEM for two replicates.

For phiSHEF14 RMs, there were three main outcomes from testing phiSHEF13,14 and 16. The first outcome showed complete inhibition of strains by phiSHEF13 while no clear effect by both phiSHEF14 and 16 (Figure 4.21 A). This was seen in 14 of the 20 strains that were tested. The second outcome involved a complete inhibition by phiSHEF13 as well as phiSHEF16 (Figure 4.21 B) while phiSHEF14 showed no effect on bacterial growth as expected. The last outcome showed a quick and complete inhibition by phiSHEF13 while no effect (as expected) by phiSHEF14. For phiSHEF16, no effect on bacterial growth was observed till the growth reached OD600= 0.2, after which growth inhibition was observed (Figure 4.21 C). This can be explained as a possible expression of specific bacterial receptors at that time point (OD600= 0.2) allowed phiSHEF16 successfully infection and inhibition of bacterial growth.

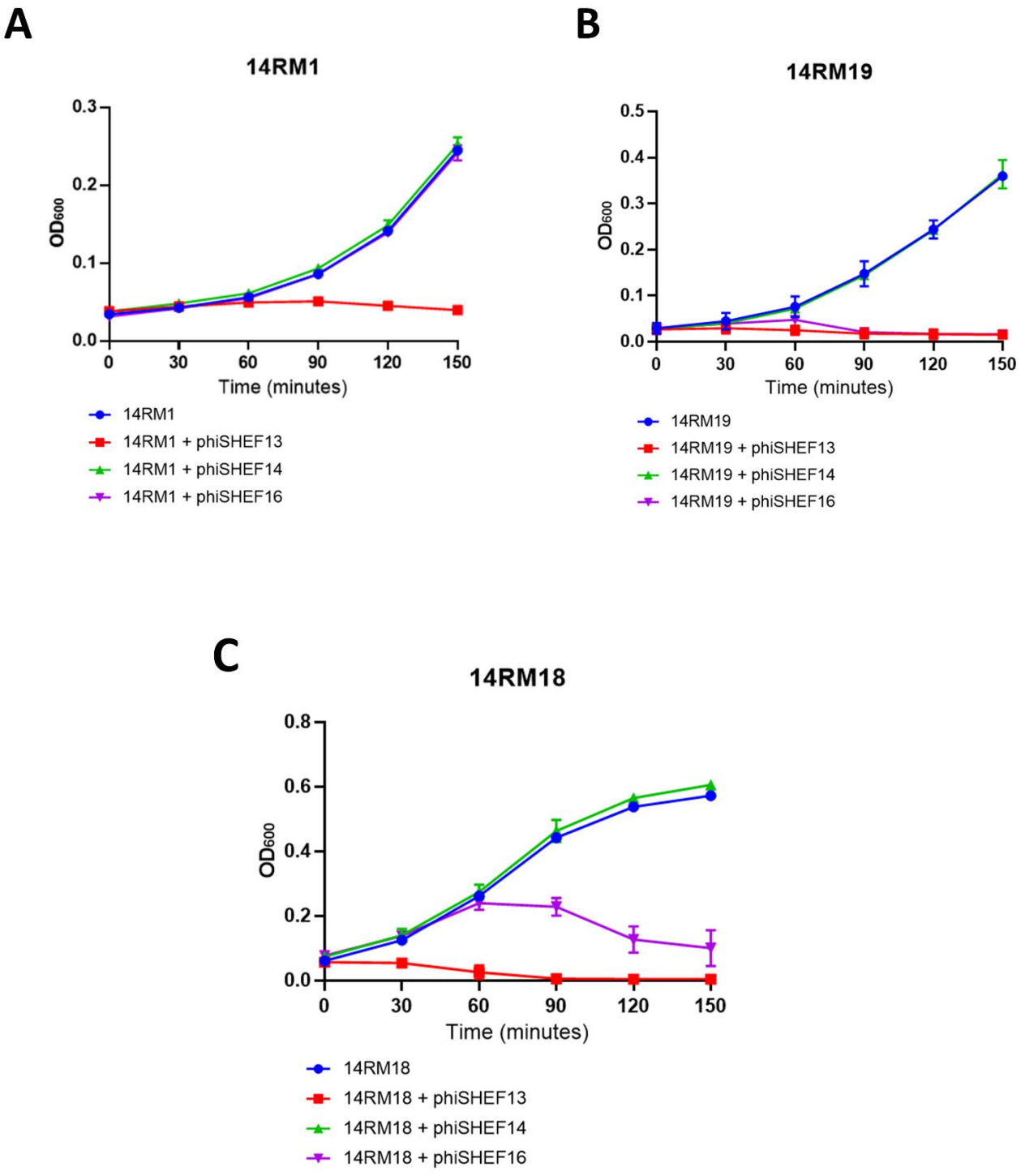


Figure 4.21. Killing assays of phiSHEF14 resistant mutants using phiSHEF13,14 and 16. A and B show the bacterial growth curve alone or with phage as indicated by colours: bacteria only (Blue), bacteria and phiSHEF13 (Red), bacteria and phiSHEF14 (Green), bacteria and phiSHEF16 (Purple). This experiment was done in duplicate and error bars represent SEM for two replicates.

For phiSHEF16, several RM strains (75% of tested strains) showed a complete inhibition by phiSHEF13 while no effect by both phiSHEF14 and 16 was observed (Figure 4.22 A). Interestingly, some other phiSHEF16 RM strains (e.g. 16RM16 and 16RM17) revealed complete resistance to the three phages 13,14 and 16 (Figure 4.22 B). This was observed in 25% of the tested 16RM strains.

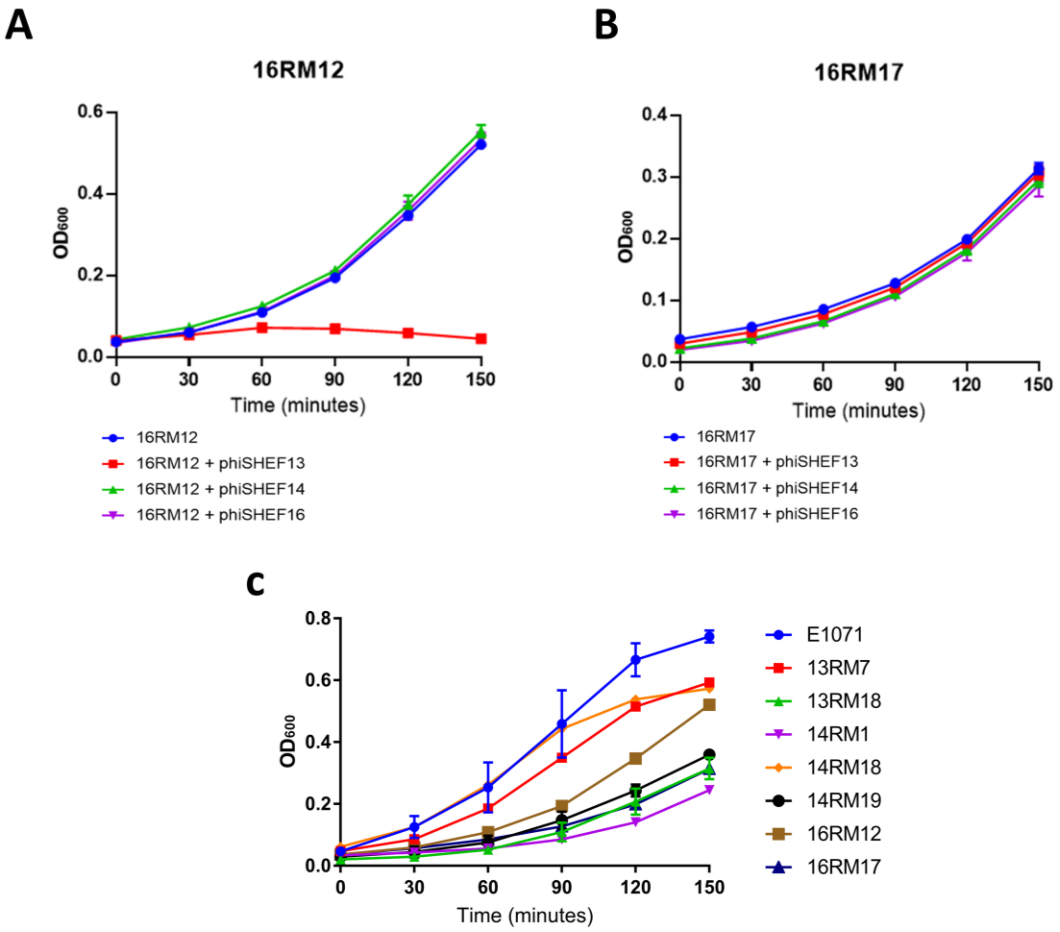


Figure 4.22. A and B) Killing assays of phiSHEF16 resistant mutants using phiSHEF13,14 and 16. The bacterial growth curve alone or with phage as indicated by colours: bacteria only (Blue), bacteria and phiSHEF13 (Red), bacteria and phiSHEF14 (Green), bacteria and phiSHEF16 (Purple). C) Growth curves of E1071 and RM strains as indicated. These experiments were done in duplicate and error bars represent SEM for two replicates.

The growth curves of RM strains were plotted against the wild type E1071 to assess any cost of fitness (Figure 4.22C). Regarding bacterial growth, the change in the values of $O.D._{600}$ between 150 min and 0 min was calculated which phage effect can be observed (Table 4.5). The bacterial doubling time was also calculated (Table 4.5) which showed varied results (i.e., decreased, no difference and higher growth rate compared with the wild type E1071). Of note, the inconsistency between the doubling time data and the bacterial growth curves (Figure 4.22C) is due to lower starting $O.D._{600}$ at 0 min for some strains (i.e., below $O.D._{600} = 0.05$). Additionally, some RM strains showed a cross-resistance involving phiSHEF14 and 16 (e.g. 14RM1 and 14RM10) or phiSHEF13 and 14 (e.g. 16RM16 and 16RM17) which could indicate a shared targeted receptor (Table 4.6).

Table 4.5. Bacterial growth comparison between E1071 and RM strains. This was done by calculating the difference in OD_{600} values between 150 min and 0 min. Bacterial doubling time was also calculated between 60 min and 90 min growth time points.

	Δ (OD_{600} 150min - 0min)				
Strain	Bacteria (B)	B + phiSHEF13	B+ phiSHEF14	B+ phiSHEF16	Doubling time (min)
E1071	0.69	-0.02	-0.03	-0.04	33
13RM1	0.53	0.33	-0.03	-0.03	31
13RM2	0.55	0.33	-0.03	-0.04	30
13RM3	0.16	0.15	-0.01	-0.01	30
13RM4	0.53	0.28	-0.02	-0.02	30
13RM5	0.59	0.28	-0.03	-0.04	35
13RM6	0.55	0.39	-0.03	-0.04	30
13RM7	0.55	0.35	-0.03	-0.03	33
13RM8	0.17	0.15	-0.01	-0.01	33
13RM9	0.56	0.41	-0.03	-0.04	31
13RM10	0.15	0.14	0.00	0.00	43
13RM11	0.53	0.40	-0.03	-0.04	35
13RM12	0.25	0.24	0.02	0.02	27
13RM13	0.47	0.30	-0.03	-0.03	35
13RM14	0.21	0.19	0.01	0.00	35
13RM15	0.56	0.30	-0.02	-0.02	30
13RM16	0.53	0.41	-0.02	-0.02	31
13RM17	0.55	0.38	-0.03	-0.03	33
13RM18	0.29	0.28	0.00	0.00	30
13RM19	0.26	0.24	0.01	0.01	23
13RM20	0.23	0.20	-0.01	-0.01	30
14RM1	0.21	0.00	0.21	0.21	38
14RM2	0.18	0.00	0.19	0.20	33
14RM3	0.19	0.00	0.20	0.19	43
14RM4	0.19	0.00	0.21	0.20	40
14RM5	0.36	0.00	0.38	0.19	27
14RM6	0.21	0.00	0.21	0.20	43
14RM7	0.23	0.00	0.24	0.23	33
14RM8	0.19	0.00	0.20	0.20	38
14RM9	0.21	0.00	0.21	0.21	35
14RM10	0.21	0.00	0.22	0.22	40
14RM11	0.44	0.03	0.47	0.46	31
14RM12	0.42	0.02	0.44	0.45	31

14RM13	0.41	-0.01	0.42	-0.01	24
14RM14	0.56	-0.03	0.44	-0.04	35
14RM15	0.48	0.00	0.43	0.33	30
14RM16	0.52	0.05	0.52	0.49	35
14RM17	0.50	0.03	0.52	0.52	35
14RM18	0.51	-0.05	0.53	0.02	35
14RM19	0.33	-0.01	0.34	-0.01	30
14RM20	0.49	-0.02	0.49	0.02	33
16RM1	0.18	0.00	0.19	0.19	38
16RM2	0.19	0.01	0.18	0.19	40
16RM3	0.18	0.00	0.18	0.18	40
16RM4	0.25	0.00	0.28	0.28	35
16RM5	0.11	0.01	0.11	0.11	46
16RM6	0.10	0.01	0.12	0.11	46
16RM7	0.10	0.01	0.11	0.11	43
16RM8	0.32	0.22	0.29	0.28	33
16RM9	0.11	0.00	0.12	0.11	46
16RM10	0.11	0.12	0.12	0.12	43
16RM11	0.51	0.01	0.50	0.50	35
16RM12	0.48	0.00	0.51	0.50	31
16RM13	0.50	0.01	0.52	0.53	28
16RM14	0.50	0.02	0.51	0.50	35
16RM15	0.50	0.01	0.50	0.49	31
16RM16	0.26	0.27	0.28	0.28	38
16RM17	0.28	0.28	0.27	0.27	40
16RM18	0.50	0.01	0.52	0.51	33
16RM19	0.46	0.36	0.46	0.48	38
16RM20	0.48	0.02	0.47	0.48	30

Table 4.6 Cross-resistance data of RM strains and phiSHEF13, 14 and 16. The percentage refers to the susceptibility of strains to the phage. 13RM refers to resistant mutants obtained after phiSHEF13 testing.

Strains/phages	phiSHEF13	phiSHEF14	phiSHEF16
13RM	0%	100%	100%
14RM	100%	0%	30%
16RM	75%	0%	0%

4.5.3 Investigating of phiSHEF13 host receptor

The data above suggest some changes that are shared in the resistance mechanism of these strains. To get an insight into this we investigated the phiSHEF13 host receptor. Phages begin their infection cycle by adsorbing on specific bacterial receptors and these receptors are involved in determining phage-host specificity (Sharma et al., 2017). The investigation of which receptors phages bind to facilitates the understanding of phage-host interaction. To enable this work, two genetically amenable *E. faecalis* strains namely V583 and OG1RF and their mutant strains were used in this investigation (obtained from Dr Stéphane Mesnage, school of biosciences, university of Sheffield) (Furlan et al., 2019). For all these strains, the best candidate phage from our collection and for this investigation is phiSHEF13 as it only infects V583 but not OG1RF. For V583, a mutant strain lacking the epa variable region (epaV) and another complemented strain (with the V583 epaV region) were used along with the wild type V583. For OG1RF, a mutant strain that lacks its own epaV region but was complemented with the V583 epaV region was used in addition to the wild-type OG1RF. Therefore, the impact of the V583 epaV region on the phiSHEF13 successful infection was assessed. This analysis was carried out using two approaches. The first was by testing the effect of phiSHEF13 on bacterial lawns via spot tests. The other approach was by using killing assays in which the inhibition of bacterial growth was analysed.

The effect of the epaV region was first investigated using the wild type V583 and its mutants (ΔepaV and complemented with epaV). For the V583 wild type, phiSHEF13 showed a rapid inhibition of bacterial growth as well as lysis on a bacterial lawn (Figure 4.23 A and F1). Testing the mutant strain (ΔepaV) revealed no effect of phiSHEF13 on both bacterial growth or on a bacterial lawn (Figure 4.23 B and F2). However, the sensitivity towards phiSHEF13 was restored upon complementation with epaV which was observed as inhibition of bacterial growth and lysis on a bacterial lawn (Figure 4.23 C and F3). Of note, the complemented strain showed less inhibitory effect by phiSHEF13 compared with the wild type and this could be due to the imperfect complementation. However, lysis on plates was observed which further confirms phage restoration of infectivity.

For OG1RF strains, both the wild type and the complemented (OG1RF-V583 epaV) strains were tested. No effect of phiSHEF13 on the OG1RF wild type was observed as expected on both bacterial growth and on a bacterial lawn (Figure 4.23 D and F4). However, the growth of the complemented strain was rapidly inhibited by phiSHEF13 and this was also confirmed by observing lysis on bacterial lawns (Figure 4.23 E and F5). Therefore, this indicates that the epaV region of V583 is necessary for phiSHEF13 infection.

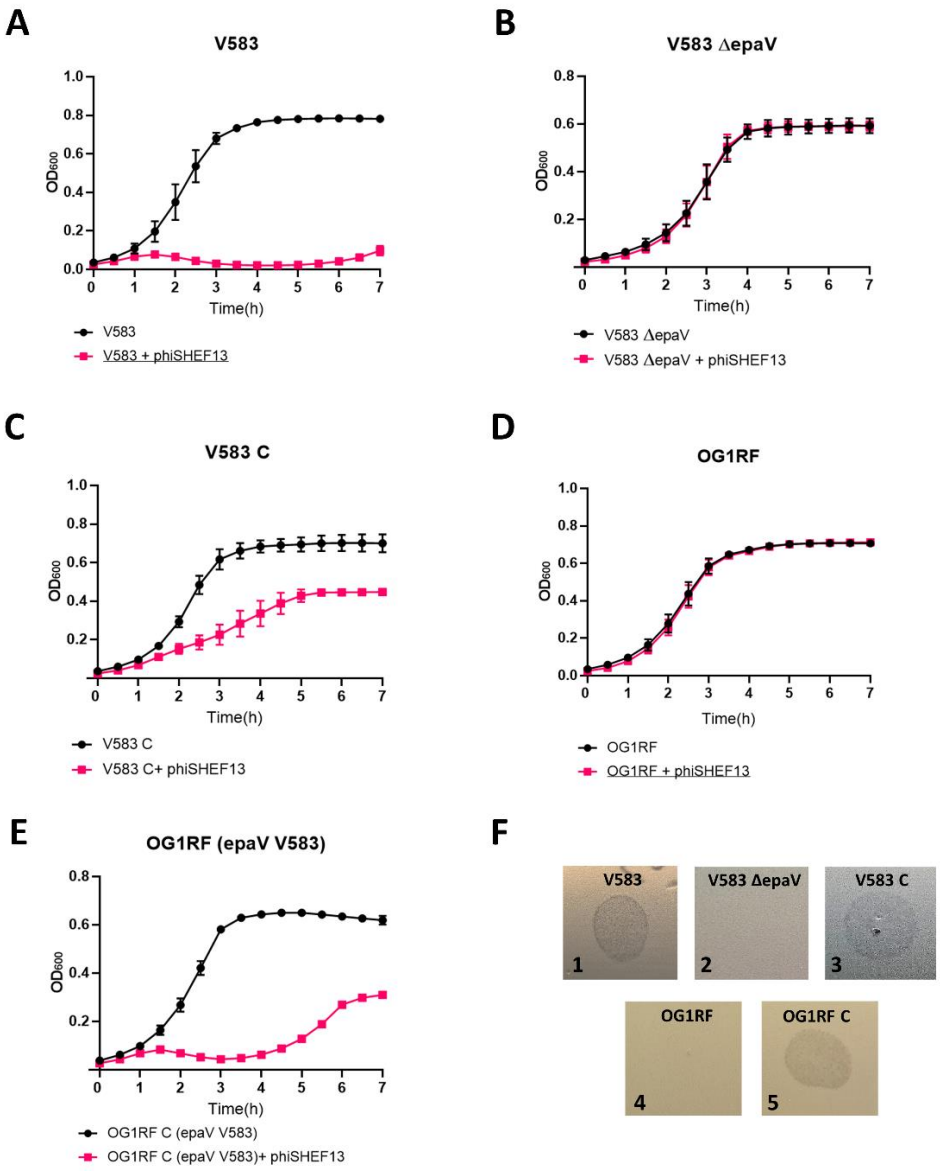


Figure 4.23. The effect of the V583 epaV region on phiSHEF13 infection. A-E) killing assays of strains with and without phiSHEF13. V583 C indicated a complemented V583 strain with the epa variable region. OG1RF (epaV V583) refers to the complemented OG1RF strain with the V583 epaV region. This experiment was done in triplicate and Error bars represent SEM for three replicates. F) Spots tests of phiSHEF13 on bacterial lawns in which lysis indicates a positive result.

4.6 Adaptation assays of phiSHEF14

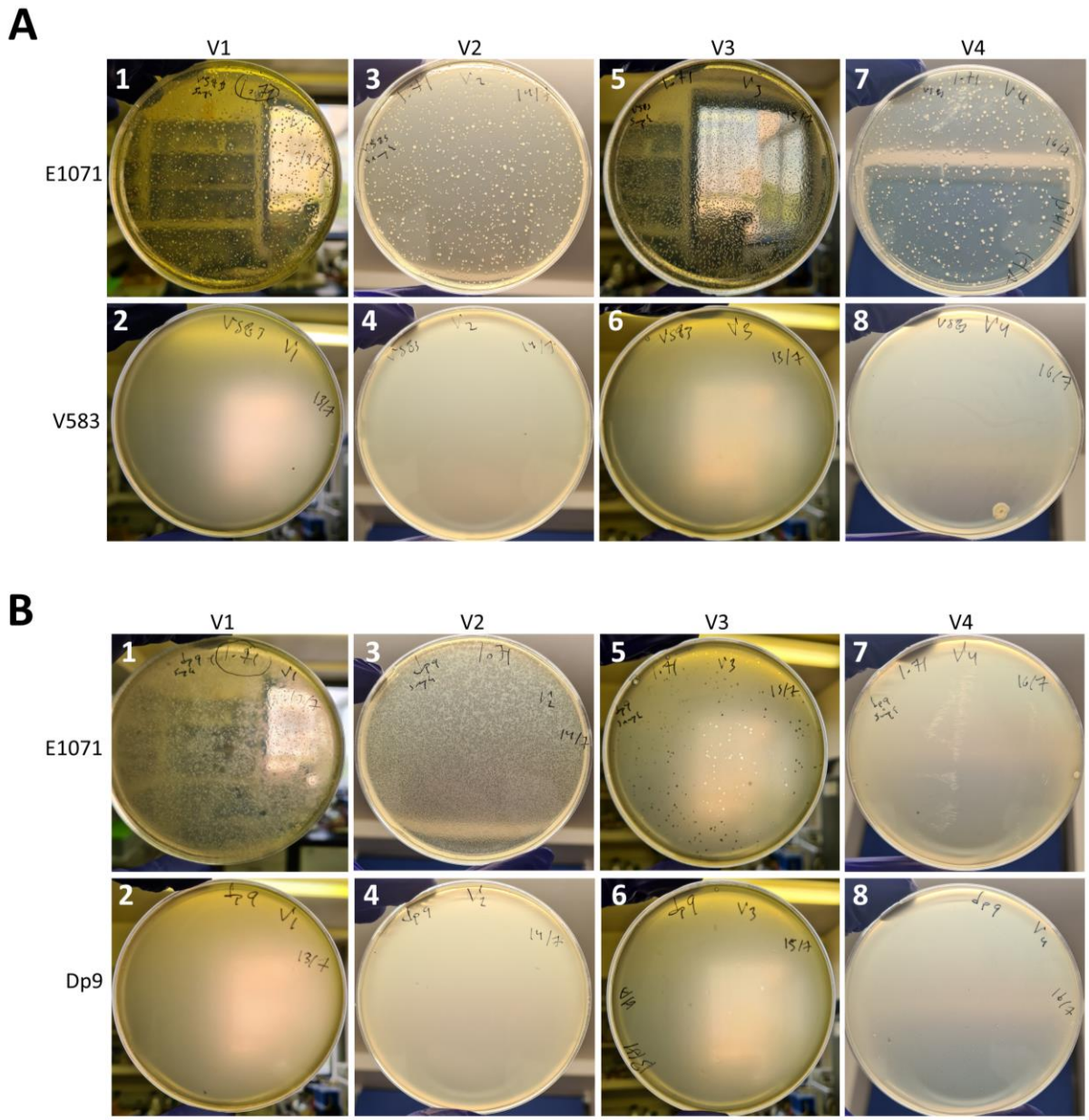
Phage specificity towards bacteria could be limited to a few strains or even a single host. These phages with narrow host range have been investigated to broaden their phage specificity to infect new hosts (Hall et al., 2013). One way to achieve this goal is by performing adaption assays that are focused on "training" phages for new hosts. This is done by the co-evolution of phage and bacteria which phage variants with natural mutations can be selected and propagated (Rohde et al., 2018).

To broaden the phage host range, phiSHEF14 was selected as it only infects a single host (*E. faecium* E1071). Three different strains were used in this investigation: the *E. faecalis* V583, the *E. faecium* dp9 and the phiSHEF14-resistant mutant of the E1071 strain (14RM0). All these strains are insusceptible to phiSHEF14. To perform this assay, the wild-type E1071 strain (susceptible to phiSHEF14) was co-cultured with the insusceptible strain (V583, Dp9 or 14RM0) along with phiSHEF14. This was done to allow phiSHEF14 to infect the E1071 strain and ensure phage high titre throughout the experiment. The phage titre was checked after every passage by plaque assays. The Methodology for this assay was described in the methods chapter 2.4.7.

For V583, testing phage infection after every passage showed no plaques on any V583 lawns indicating no isolation of any phiSHEF14 mutants targeting the V583 (Figure 4.24 A2,4,6,8). The E1071 plates showed complete lysis of the lawns indicating a high titre of phiSHEF14 (Figure 4.24 A1,3,5,7).

For dp9, there were also no plaques seen on the dp9 lawns after each passage (Figure 4.24 B2,4,6,8). However, our control E1071 plates showed a gradual reduction in phiSHEF14 titer as can be seen in (Figure 4.24 B1,3,5,7). This reduction of phage titer could be caused by the unavailability of the susceptible strain E1071 in the co-culture. To further investigate this, a cross streaking test was performed to identify any inhibitory effects between the dp9 and E1071 strains. As a result, a clear zone of inhibition was observed on E1071 by the dp9 strain (Figure 4.24 C). The control strain V583 showed no inhibition on the E1071 strain. Moreover, the

supernatant of dp9 liquid culture was spotted on the E1071 lawn which shows a lysis effect after overnight incubation indicating an inhibitory effect of dp9 on the E1071 strain (Figure 4.24 D).



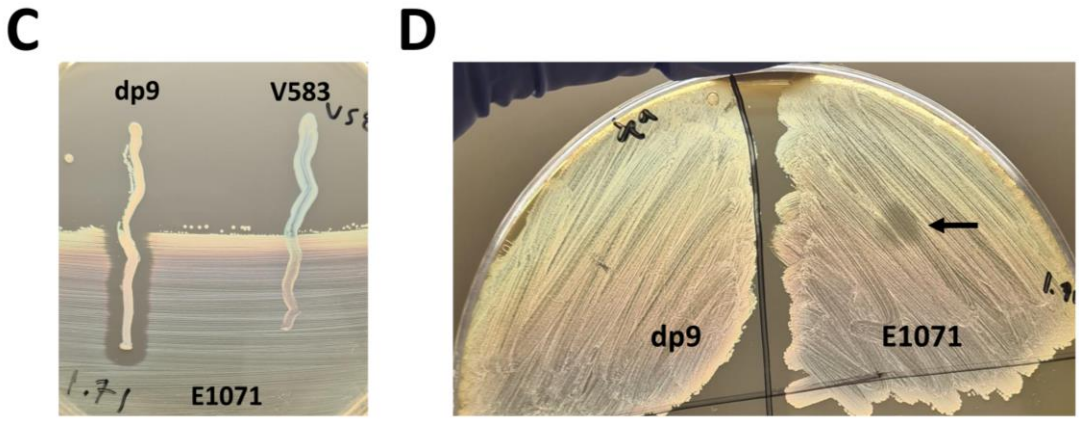


Figure 4.24. Adaptation assays of phiSHEF13 using V583 and dp9. Plaque assays on A) E1071 and V583 and B) E1071 and dp9. V1-4 indicates passages of phiSHEF14 with co-cultured strains. C) A cross-streaking test on E1071 by dp9 and V583 strains is shown which zone of inhibition is seen via dp9. D) spot tests using dp9 supernatant onto dp9 and E1071 strains are shown.

For the resistant mutant strain (14RM0), the adaptation protocol was followed as with the previous two strains (V583 and dp9). Similarly, this resulted in no plaques being seen on the 14RM0 lawns after passages while the control plates E1071 showed complete lawn lysis (Figure 4.25).

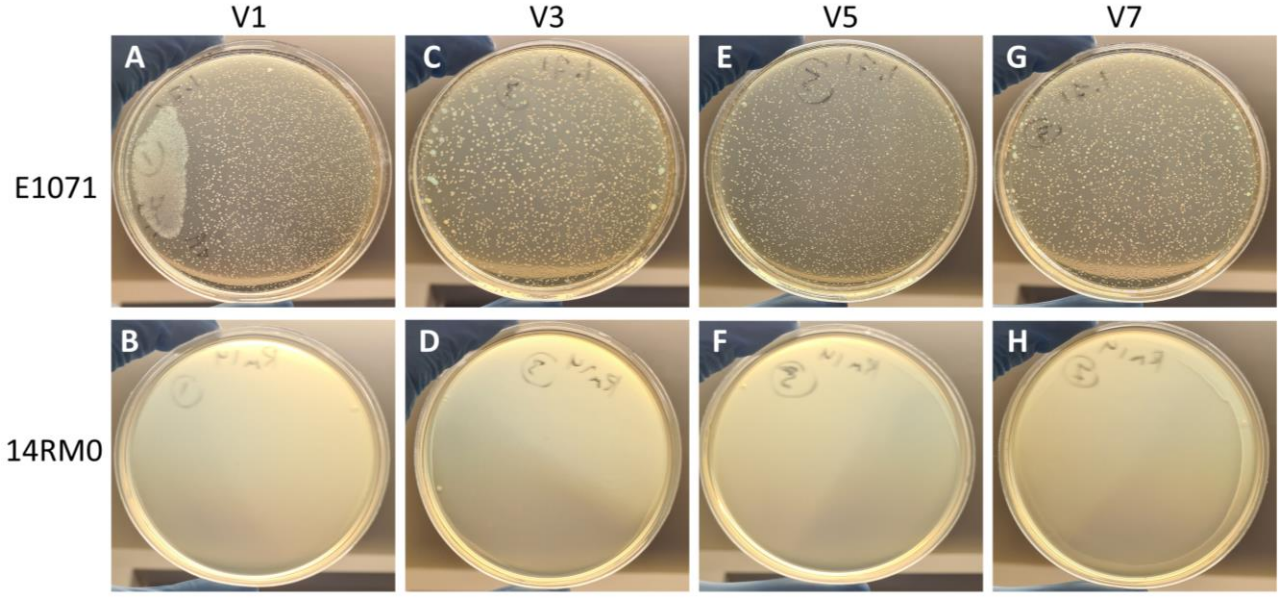


Figure 4.25. Adaptation assays of phiSHEF13 using E1071 and 14RM0. Plaque assays are shown for E1071 (A,C,E and G) and 14RM0 (B,D,F and H). V1-7 indicates passages of phiSHEF14 with co-cultured strains.

4.7 Assessment of antibacterial effect among enterococcal strains

As phages have been isolated using the multiple isolation technique and after observing the antibacterial effect of dp9 on E1071, the assessment of the antibacterial effect among enterococcal strains was then carried out. To investigate this, we assessed the antagonistic effect between the strains used in the multiple isolation technique via cross-streaking and spot assays. For group 1 (dp strains, dp6-9), the cross-streaking assay showed a zone of inhibition on dp8 streaked lawn from dp6, 7 and 9 (Figure 4.26 A3). This inhibition was also observed when the culture filtrate (cell-free supernatant) of dp6,7 and 9 was spotted on dp8 (**Figure 4.26** B1,2 and 4). This may suggest the involvement of either bacteriocin or spontaneously induced prophage effect that has led to dp8 inhibition (Vijayakumar & Muriana, 2015).

For group 2 (EPA strains), The cross-streaking assays were also applied and an inhibition zone was seen on E1679, E1071 and E1162 lawns via the E980 strain (Figure 4.27 A1,2 and 3). However, the cell-free supernatant of E980 showed no lysis upon spotting on E1679, E1071 and E1162 lawns (Figure 4.27B). A more sensitive assay to further assess the antagonistic effect is the agar-well diffusion assay to better estimate the inhibition effect.

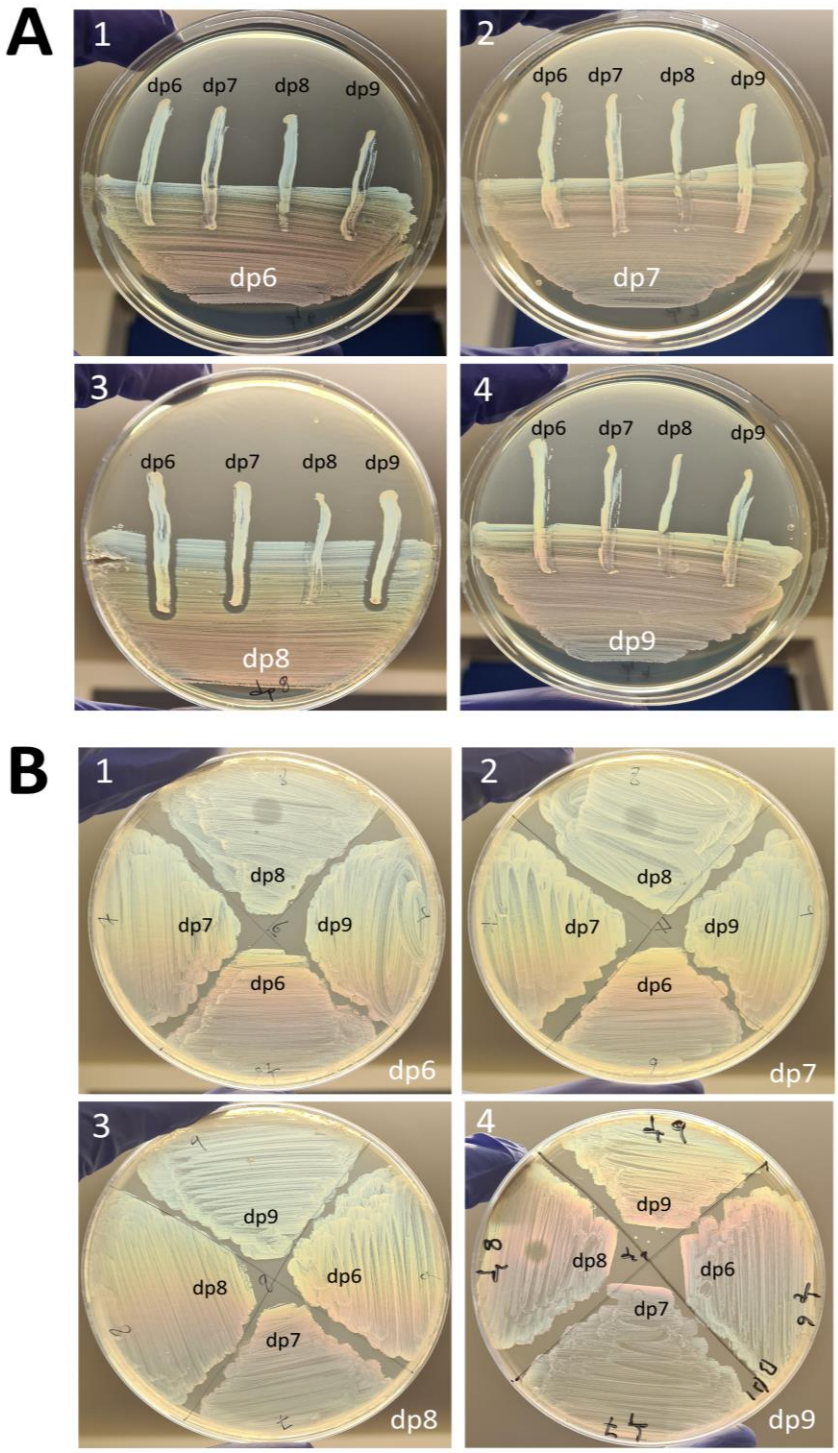


Figure 4.26. The assessment of antagonistic effect between dp *E. faecium* strains. A) cross-streaking assays of four strains (dp6-9). B) Spot tests using cell-free supernatant on bacterial lawns are shown.

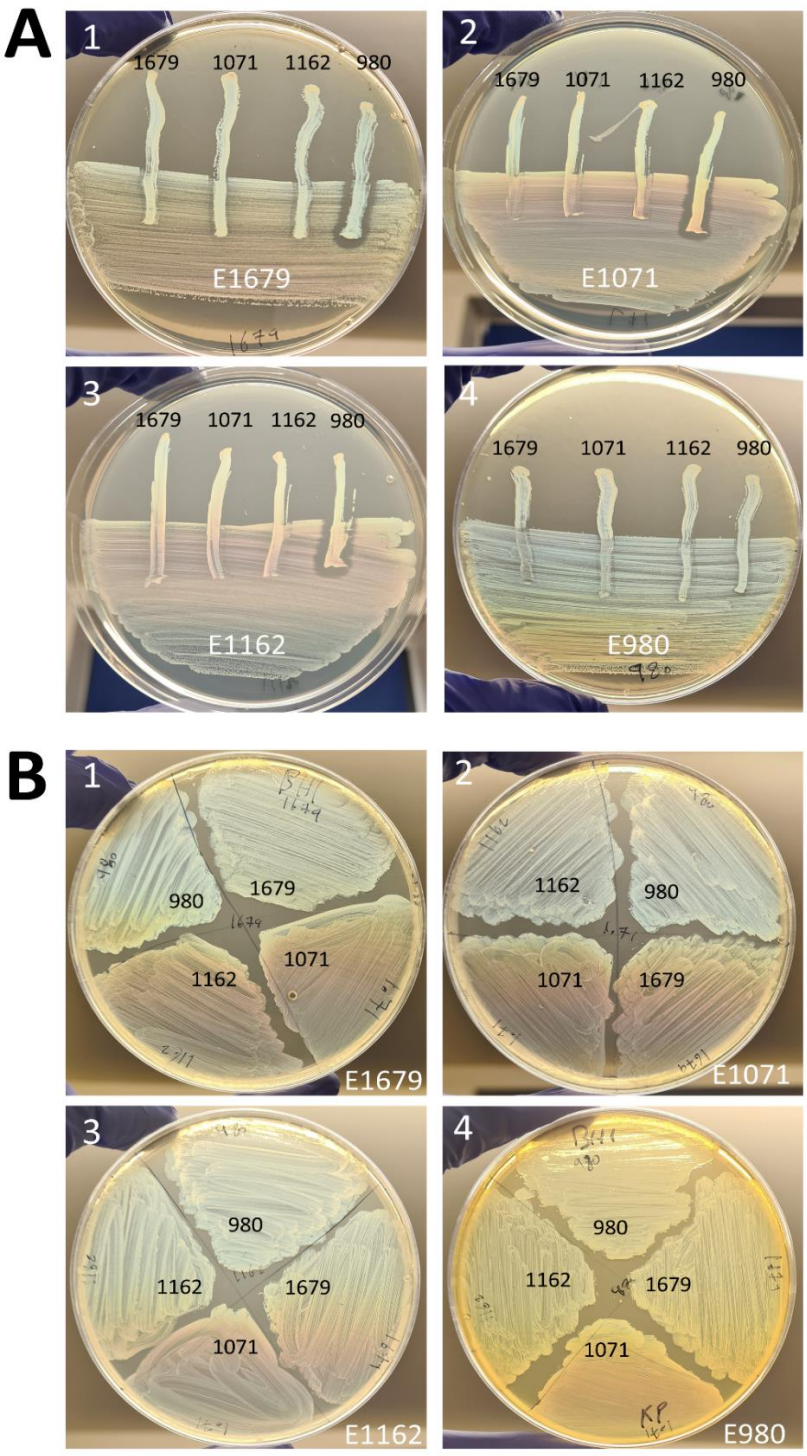


Figure 4.27 The assessment of the antagonistic effect between EPA *E. faecium* strains. A) cross-streaking assays of four strains (E1679, E1071, E1162 and E980). B) Spot tests using cell-free supernatant on bacterial lawns are shown.

Additionally, genomic analyses were also performed on these strains to further investigate the antagonistic effect seen. The BAGEL4 webserver (http://bagel4.molgenrug.nl/) was applied to identify bacteriocins while PHASTER (https://phaster.ca/) was used for prophage prediction (Table 4.7). It is important to mention here that this antagonistic assessment was performed after phages have already been isolated. Therefore, the antagonistic effect that is observed did not impact phages being isolated and hence the multiple hosts' isolation approach showed no apparent effect on phage isolation.

Table 4.7. Predicted bacteriocins and prophages in enterococcal strains. ND (not determined) due to PHASTER inaccessibility at the time of testing.

Strain	Bacteriocins	Prophages (intact)
dp7	Enterolysin_A	ND
	Bacteriocin_II; 81.2;Enterocin_A	
	Lactococcin; Bacteriocin_IIc;	
	Bacteriocin_II; 22.2;Bacteriocin_T8	
	Enterolysin_A	
dp9	Bacteriocin_II;Bacteriocin_T8	1
	Enterolysin_A	
	Bacteriocin_II; 81.2;Enterocin_A	
	Enterocin P	
	Lactococcin; Bacteriocin_IIc	
E1071	Bacteriocin_Iic	1
	Bacteriocin_IIc; Mersacidin; 7.2;Acidocin_LF221B(GassericinK7B)	
	Lactococcin; Bacteriocin_IIc;	
	Bacteriocin_II; 81.2;Enterocin_A	
	Enterolysin_A	
	Bacteriocin_IIc; 82.2;Enterocin_B	
E980	Enterolysin_A	2
	Enterocin P	
	Bacteriocin_II; 91.2;Enterocin_P	
	MR10B	
	Enterocin_L50a	
E1162	Bacteriocin_II; 81.2;Enterocin_A	1
	Enterolysin_A	
E1679	Bacteriocin_IId; 272.1;Enterocin_Nkr-5-3B	1

4.8 Genomic characterisation of the enterococcal clinical isolates

In our work here, we have used clinical isolates from patients with diabetic foot ulcers. These isolates are both *E. faecalis* and *E. faecium* strains and were labelled dp1-9. After the phage host range testing, the susceptible isolates (dp1,3,5,7,9) were selected for genomic analysis. The bacterial genomes were extracted and then sent for sequencing (MicrobesNG, Birmingham, UK) using Illumina. The FASTA files of each genome were also further analysed using the online tool "staramr" tool on usegalaxy.eu. This tool analyses the bacterial genome to mainly identify antimicrobial-resistant (AMR) genes (Table 4.8).

Table 4.8 Genomic characterisation of the enterococcal clinical isolates.

Species	Strains	Size (Mb)	Genotype	Predicted phenotype	MLST
E. faecalis	dp1	3.07	aac(6')-aph(2''), ant(6)-Ia, aph(3')-III, erm(B), lsa(A), tet(M)	amikacin, gentamicin, tobramycin, streptomycin, kanamycin, erythromycin, azithromycin, lincomycin, tetracycline	179
	dp3	3.04	aac(6')-aph(2''), ant(6)-Ia, aph(3')-III, erm(B), lsa(A), tet(M)	amikacin, gentamicin, tobramycin, streptomycin, kanamycin, erythromycin, azithromycin, lincomycin, tetracycline	179
	dp5	2.94	cat, dfrG, erm(B), lsa(A), str, tet(M)	chloramphenicol, trimethoprim, erythromycin, azithromycin, lincomycin, streptomycin, tetracycline	16
E. faecium	dp7	3.08	aac(6')-Ii, ant(6)-Ia, aph(3')-III, erm(B), msr(C), tet(M	amikacin, tobramycin, streptomycin, kanamycin, erythromycin, azithromycin, tetracycline	17
	dp9	3.11	aac(6')-aph(2''), aac(6')-Ii, ant(6)-Ia, ant(9)-Ia, aph(3')-III, dfrG, erm(A), erm(B), lnu(B), lsa(E), msr(C), tet(M), VanHAX	amikacin, gentamicin, tobramycin, streptomycin, spectinomycin, kanamycin, trimethoprim, erythromycin, azithromycin, lincomycin, unknown[lsa(E)_1_JX560992], tetracycline, vancomycin	787

4.9 Discussion

In this chapter, we set out to isolate and characterise phages targeting various *E. faecalis* and *faecium* strains. These strains were lab strains as well as clinical isolates from patients with diabetic foot ulcers. To increase the likelihood of isolating different phages, we obtained *E. faecium* strains that differ in the EPA structure as this antigen is commonly used by phages as a host receptor (Ho et al., 2018; Teng et al., 2009). EPA is a main component of the enterococcal cell envelope by binding to peptidoglycan (Guerardel et al., 2020) and its locus consists of two genetic loci: conserved and variable regions (Ramos et al., 2021). The variable region differs among enterococcal strains due to EPA decoration. Moreover, De Been et al have analysed the variable locus of *E. faecium* strains including VRE such as E1162 and E1071 which resulted in identifying 4 main variants based on genome alignment (De Been et al., 2013) Therefore, strains representing each variant have been obtained in our lab for phage isolation.

As a rich source for phages, wastewater was used in the isolation process. Several publications have shown the isolation of different phages from wastewater and specifically against *Enterococcus* spp. (Al-Zubidi et al., 2019; Chatterjee et al., 2019; D. Lee et al., 2019). The analysis of the obtained wastewater under TEM showed different phage morphologies (podoviruses, siphoviruses and myoviruses) which further confirms the richness of wastewater samples.

In this work, phage isolation using the enrichment method has successfully resulted in isolating 8 phages using both *E. faecalis* and *E. faecium* strains. In the literature, the enrichment method has also been used to isolate phages against other hosts (Alič et al., 2017; Viazis et al., 2011). Compared with the direct isolation method (no enrichment), the phage of interest in the enrichment method has the opportunity to propagate and increase in number which understandably makes the isolation process more prone to success. The 8 isolated phages were assessed under TEM which showed the three known tail phage morphologies: podovirus, siphovirus and myovirus. Enterococcal phages in the NCBI database as well as in the literature

are also dominantly tailed viruses (Bolocan et al., 2019). The isolation of only tailed phages can be explained as the tailed phage, belonging to the class Caudoviricetes, comprise the majority (96%) of all known phage morphologies (Fokine & Rossmann, 2014).

The investigation of the phage host range has shown different specificities among the isolated phages. PhiSHEF13 has a broad host range by infecting 11 *E. faecalis* and *E. faecium* strains while phiSHEF14 has a very narrow specificity as it only infects the *E. faecium* strain E1071. The infection of multiple strains suggests a shared receptor that is essential for phiSHEF13 infection or possessing unique phiSHEF13 proteins that helped infecting various strains. This varied specificity of phages is also observed in other enterococcal phages (D'Andrea et al., 2020; Del Rio et al., 2019) as well as other phages (Gibson et al., 2019). The different EPA *E. faecium* strains (i.e. the four variants of EPA) were also tested against the isolated phages. As the host range results show, phiSHEF13 and 14 were able to infect E1071 but not E4452 (both strains classified as variant 2) indicating dissimilarity between these strains (i.e. that cell wall different even if EPA locus is the same). The notion of both phiSHEF13 and 14 targeting the exact receptor on E1071 is unlikely as the phage RM experiments have revealed (discussed later). For phiSHEF16, however, it successfully infects strains in variants 1 (E1679 and E1636) and 2 (E1071 and E4452) which indicates that a possible shared receptor in variants 1 and 2 strains allowed phiSHEF16 infection. Other tested *E. faecium* strains like dp7 (infected by phiSHEF13) and dp9 (infected by phiSHEF13 and 16) show the difference between these two phages. For these clinical isolates (dp7 and dp9), the host range results show no clear conclusion which variants could these belong to. The genomic analysis of the *E. faecium* clinical isolates in relation to the EPA types is a future step which could not be done currently due to time limits. As the isolated phages have shown varied host range results, they are considered good candidates for phage cocktail strategy as varied enterococcal strains including VRE were effectively infected and lysed. Of note, some tested enterococcal strains in our stock are still resistant to our phage collection indicating a possible different surface receptor or defense mechanism which requires more attempts of phage isolation to broaden our enterococcal phage library.

The genomic analysis of the isolated phages revealed 5 novel types which undergo only a strict lytic phage life cycle. Three of the isolated phages showed high similarity and coverage (>99) as well as identical host range results. Therefore, these three phages (phiSHEF8, 9 and 12) were not considered novel. For phiSHEF10, this showed a 97% coverage result to its closest hit as well as a different host range result. Thus, this phage was added to the phage database as a novel one.

For phiSHEF13, BLASTn search showed the closest hit (92% identity) being the EFDG1 phage. The work of Khalifa et al showed that the EFDG1 phage has a broad host range infecting both *E. faecalis* and *E. faecium* strains including V583 and clinical isolates (Khalifa et al., 2015) and this coincides with phiSHEF13 host range analysis. For phiSHEF14, the closest hit was also a podovirus vB_EfaP_Zip that can infect both *E. faecalis* and *E. faecium* including the V583 strain (Melo et al., 2019). In the phage database, the number of isolated phages infecting *E. faecalis* is higher than *E. faecium* (Bolocan et al., 2019) indicating the need for more *E. faecium* phages. In our work, the three phages (phiSHEF13,14 and 16) were all isolated using the *E. faecium* E1071. The genomic alignment of these three phages has shown a shared section of a tail protein. Upon using Phyre2, the three proteins (phiSHEF13_41, phiSHEF14_15 and phiSHEF16_201) showed homology to a sugar-binding protein. Collectively, these proteins may be involved in phage adsorption or genome ejection as they are located within the tail module. Of note, the phiSHEF14_15 protein is also being further investigated by in the school of biosciences, University of Sheffield, to assess protein expression and purification as well as structure.

Further investigation of phiSHEF13,14 and 16 was then performed. Phage efficiency of killing planktonic bacteria was assessed via killing assays at different MOI. All three phages showed quick inhibition of bacterial growth which the highest MOI showing the fastest inhibition and vice versa. Phages can adsorb quickly onto bacterial cells (2-8 min) and complete their life cycle in about 30-50 min (Al-Zubidi et al., 2019; Imam et al., 2019; D. Lee et al., 2019). Therefore, the killing assays' results coincide with this general feature of phage-host interaction.

Phage-resistant mutants were observed upon analysing the killing assays of phiSHEF13,14 and 16 on the E1071 strain. A total of 60 mutant clones were selected for phage infectivity testing. For phiSHEF13 RM, all these were still sensitive to both phiSHEF14 and 16 indicating no effect of the bacterial mutations on the infectivity of these phages. Similar results were also seen with phage phiSHEF14 and 16. An interesting result with the phiSHEF14 RM was that some clones became sensitive to phiSHEF16 only at $O.D._{600} = 0.2$ suggesting an expression of phage receptor that allowed phage infection (Veselovsky et al., 2022). For phiSHEF16 RM, 5 clones showed resistance to the three phages indicating an evolved resistance mechanism like loss of receptors that impair phage infection (Denes et al., 2015). The different phage resistance profiles resulting in a cost of fitness have also been seen in previous publications (Bohannan & Lenski, 2000; Wright et al., 2018).

This RM data showed possible shared mechanisms of resistance between some phages (i.e. phiSHEF14 and 16). It is therefore worth assessing the phage cocktail of two phages like phiSHEF13 and 14 or 13 and 16. Moreover, an investigation of phage cocktail RM would also shed light on selecting an appropriate phage therapy regime. This would allow a better selection of phage cocktails to avoid phage competition on receptors and cross-resistance. Regarding the bacterial cost of fitness, one way to assess the effect of phage resistance can be through growth rates which our preliminary data showed a decreased, no difference or higher growth rate compared with wild type E1071. Previous publications on RM strains have also seen slow growth of some strains (Avrani & Lindell, 2015), no difference (Kortright et al., 2021) or higher growth rates (Nagarajan et al., 2019). These varied outcomes could be related to the mechanism of resistance which if receptors involved in nutrient uptake are affected a predicted consequence on growth rate is expected (Bohannan et al., 2002). Further investigation of these mutants (which was not done due to time limits) could include testing prolonged growth curves (24 h) compared with the wild-type E1071. As a possible trade-off of fitness, the antibiotic resistance profile can also be assessed for identifying any re-sensitisation results (Mangalea & Duerkop, 2020). In fact, a student in our lab (Elspeth Smith) has begun this work focusing on phiSHEF14 RM strains which presumptive data (compared with wild type E1071) showed ciprofloxacin susceptibility

but still vancomycin resistance upon using antibiotics disc assays. Interestingly, Elspeth has also isolated RM using the phage cocktail (phiSHEF13,14 and 16) which the preliminary data about antibiotic sensitivity profile showed vancomycin re-sensitisation along with other antibiotics. Of note, both the antibiotics and phage could have the same bacterial target in which bacterial cross-resistance may occur such as with phage T6, phage U115, and albicidin, a DNA gyrase inhibitor which a pleiotropic trade-up is often seen (Kortright et al., 2021). The bacterial virulence of these RM strains can also be assessed for such as biofilm formation or resistance to macrophages. Additional assessment can involve *Ex vivo* assays such as 3D human skin models or *In vivo* testing like the Zebrafish model. Additionally, the genetic differences between E1071 and the RM strains have also been considered and the DNA of the strains was extracted and sent for sequencing. The data after sequencing are being analysed to address any genomic variation between wild-type and mutants. This data is not included here due to unfinished analysis and time limits.

PhiSHEF13 was further investigated by assessing the effect of V583 epaV (variable region) on its infection. EPA is the enterococcal polysaccharide antigen and its coding genes are divided into conserved and variable regions. For the *E. faecalis* V583 strain, the variable region consists of 18 genes (Guerardel et al., 2020) and a knock-out mutant of this region was obtained and tested. This investigation has resulted in that the epaV of V583 is necessary for phiSHEF13 infection after testing wild types and mutants of both V583 and OG1RF strains. EPA is one of the commonly utilised receptors by phages to infect enterococcal cells (Ho et al., 2018; Teng et al., 2009b). The work here further confirms that EPA is a hotspot for phage infection. In the literature, the deletion of specific EPA genes affects phage adsorption and ultimately phage infection (Ho et al., 2018). Therefore, the effect of the epaV of V583 on phiSHEF13 is assumed to be also on phage adsorption (Chatterjee et al., 2019).

Given that phiSHEF14 only infects a single host, an attempt to broaden its host range was performed via adaptation assays. The ability of bacteriophages to quickly adapt to bacterial host populations is well recognised in the literature (Buckling et al., 2009; Buckling & Rainey, 2002). The co-evolution of phages and their hosts happens in nature consistently as a means of survivability and an "arms race" relationship (Betts et al., 2014; Stern & Sorek, 2011). The adaption or "training" of phages can be performed by serial passages of co-incubated phage/bacteria after a fresh medium is added (Betts et al., 2013; Friman et al., 2016). Therefore, our work here involved testing three insensitive strains to phiSHEF14: V583, dp9 and 14RM0. This investigation has resulted in no isolation of phiSHEF14 mutants that can infect the abovementioned strains. Although performing this experiment over 7 days, additional time could result in the development of phage mutants. Alternatively, phage mutants can be selected after challenging with sodium pyrophosphate which causes instability to phage particles and genome deletions (Gutiérrez et al., 2018b).

4.10 Conclusion

5 novel phages (phiSHEF10,11,13,14 and 16) were successfully isolated from wastewater samples in the Sheffield area. TEM examination has shown all three known tail morphologies: podovirus, siphovirus and myovirus. Upon testing 36 enterococcal strains, phiSHEF phages showed diverse host range results which phiHSEF13 infects 11 *E. faecalis* and *faecium* strains while phiSHEF14 only kills the isolation host E1071. All these phages undergo a strict lytic life cycle upon genomic analysis. Further investigation of phiSHEF13,14 and 16 showed quick inhibition of the VRE E1071 strain at different MOI upon killing assays analysis. The killing assays have also shown the development of phage RM which was investigated. Phage infectivity of the RM revealed different results which some clones were completely resistant to all three phages. The host receptor for phiSHEF13 was also assessed which showed the importance of the epaV of the V583 strain for successful phiSHEF13 infection. Lastly, phage adaptation assays were performed to broaden the host range of phiSHEF14 which resulted in no successful infection of dp9, V583 and 14RM0 strains.

Chapter 5: *In vitro* investigation of putative TAL proteins

5.1 Introduction

As the antibiotic resistance crisis has caused millions of deaths in the recent years, alternative therapies are needed to tackle antibiotic-resistant pathogens (Mobarki et al., 2019). One particular therapy that has shown promising results is phage-derived lysins which can cause bacterial death. These phage lysins have shown a potent effect on both *in vitro* (Larpin et al., 2018) and *in vivo* studies (Oliveira et al., 2018). In phage lifecycles, phage lysins are involved in the initial steps or late in the phage release. Phage lysins (in the late stage) are mainly holins, endolysins and spanins which help lyse bacteria from the inside out. In contrast, other lysins involve in facilitating phage infection by targeting various bacterial layers and these are mainly located within the tail structure. Therefore, these lysins are called tail-associated lysins (TAL) which were mentioned and discussed thoroughly in chapter 3.

In this chapter, the aim was to investigate a panel of five TALs for *in vitro* expression and purification. The TALs were selected based on the analysis of chapters 3 and 4. Different bacterial expression cells, induction temperatures and protein tags were assessed to determine the best expression conditions. Cloning of candidate TALs was performed by either gene amplification or gene synthesis.

Specific objectives :

- Select candidate TALs
- Gene cloning either by amplification or synthesis
- Assess TAL expression
- Purify and dialyse successfully expressed proteins

5.2 Determination of candidate TAL proteins

In chapter 3, several predicted proteins with lytic domains were identified which were located mostly within the tail module. As this was only conducted *in-silico*, we sought out to test some of these proteins *in vitro* to investigate their activity and efficiency. Therefore, we have selected five different predicted lytic proteins from the different categories for this investigation: TAEP, TMP-LT, PE, GDPD, NLPC/P60.

5.2.1 Tail protein associated with endopeptidase (TAEP)

TAEP is the most abundant type of predicted lytic proteins in the tail module of enterococcal phage and prophage genomes as our bioinformatic analysis (chapter 3) showed with about 70% of the total identified lytic proteins. TAEP proteins have shown various domain architectures ranging from only the endopeptidase domain to two additional lytic domains. In this work, a TAEP protein was selected that contains endopeptidase, lysozyme and amidase domains (Figure 5.1A). This selection was done as the protein harbours multiple lysins which might enhance or broaden the protein lytic effect. This protein was also identified in a predicted intact prophage in the *E. faecium* E1 strain genome. The full protein was aimed to be cloned and expressed.

5.2.2 Tape measure protein with lytic transglycosylase (TMP-LT)

Tape measure proteins are considered one of the largest proteins in phage genomes (Piuri & Hatfull, 2006) and they have a role in facilitating phage genome ejection inside bacterial cells (Mahony et al., 2016). Our analysis of these proteins in chapter 3 showed a Lytic transglycosylase (LT) domain positioned mostly at the C-terminal regions. LT targets the bacterial cell wall by cleaving the β-1,4-glycosidic bond between N-Acetylmuramic acid (MurNAc) and N-Acetyl-D-glucosamine (GlcNAc) (Holtje et al., 1975). Therefore, LT was included in our panel for testing protein expression and purification. In this work, the candidate TMP-LT protein was identified in the phage phiSHEF2 (phiSHEF2_12) which was also

previously isolated in our lab (Al-Zubidi et al., 2019). As TMP is a large protein (1456 aa), it was decided to only clone the predicted LT domain (125 aa) (Figure 5.1B).

5.2.3 Pectinesterase (PE)

Another type of lysin that was also identified in the work mentioned in chapter 3 was pectinesterases. Seven proteins with predicted pectinesterase domains have been analysed which only were found in prophage genomes. These PE proteins were located within the tail module and mostly after the TAEP protein. To our knowledge, the activity of pectinesterase on *Enterococcus* has not yet been investigated and enterococci have a unique antigen in their cell wall called enterococcal polysaccharide antigen (EPA) which is composed of repeating rhamnose units, interspersed with other sugars and decorated with various modifications. As pectinesterases have been shown to target pectin (rhamnose-containing substrates) (Reid, 1950), we hypothesise that the identified pectinesterase-containing proteins could target EPA in the enterococcal cell wall. Therefore, a pectinesterase protein from a prophage in the *E. faecium* Gr17 strain was selected for cloning and expression as these proteins were only identified in prophage genomes (Figure 5.1 C).

5.2.4 Glycerophosphodiester phosphodiesterase (GDPD)

GDPD is another type of lytic proteins that were identified in chapter 3. In that chapter, the genome scanning for GDPD-containing proteins showed that they are located within the tail module as well as throughout the genome. The target of GDPD is the wall teichoic acids via targeting the phosphodiester bonds (Cornelissen et al., 2016). Therefore, an example from the most commonly identified type (DA1) was selected: a GDPD protein from a prophage genome in the *E. faecium* E1334 strain which possesses only the GDPD domain located at the N-terminal region (Figure 5.1 D).

5.2.5 New Lipoprotein C/Protein of 60-kDa (NLPC/P60)

In chapter 4, the phiSHEF14 phage was isolated which infects the VRE strain E1071 and characterised which resulted in identifying an NLPC/P60 containing protein upon genome analysis. The NLPC/P60 has been shown to target the bacterial cell wall as peptidase or amidase (Anantharaman & Aravind, 2003). Therefore, we sought out to investigate the activity of the NLPC/P60 protein *in vitro*. The domain architecture analysis showed that this protein only contains an NLPC/P60 domain at the N-terminal region (Figure 5.1 E).

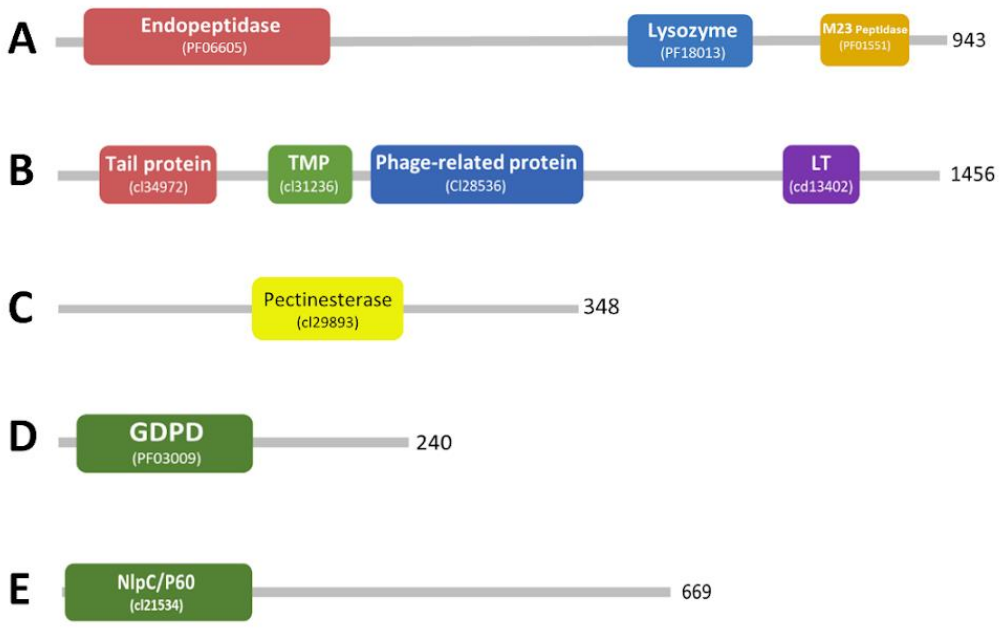
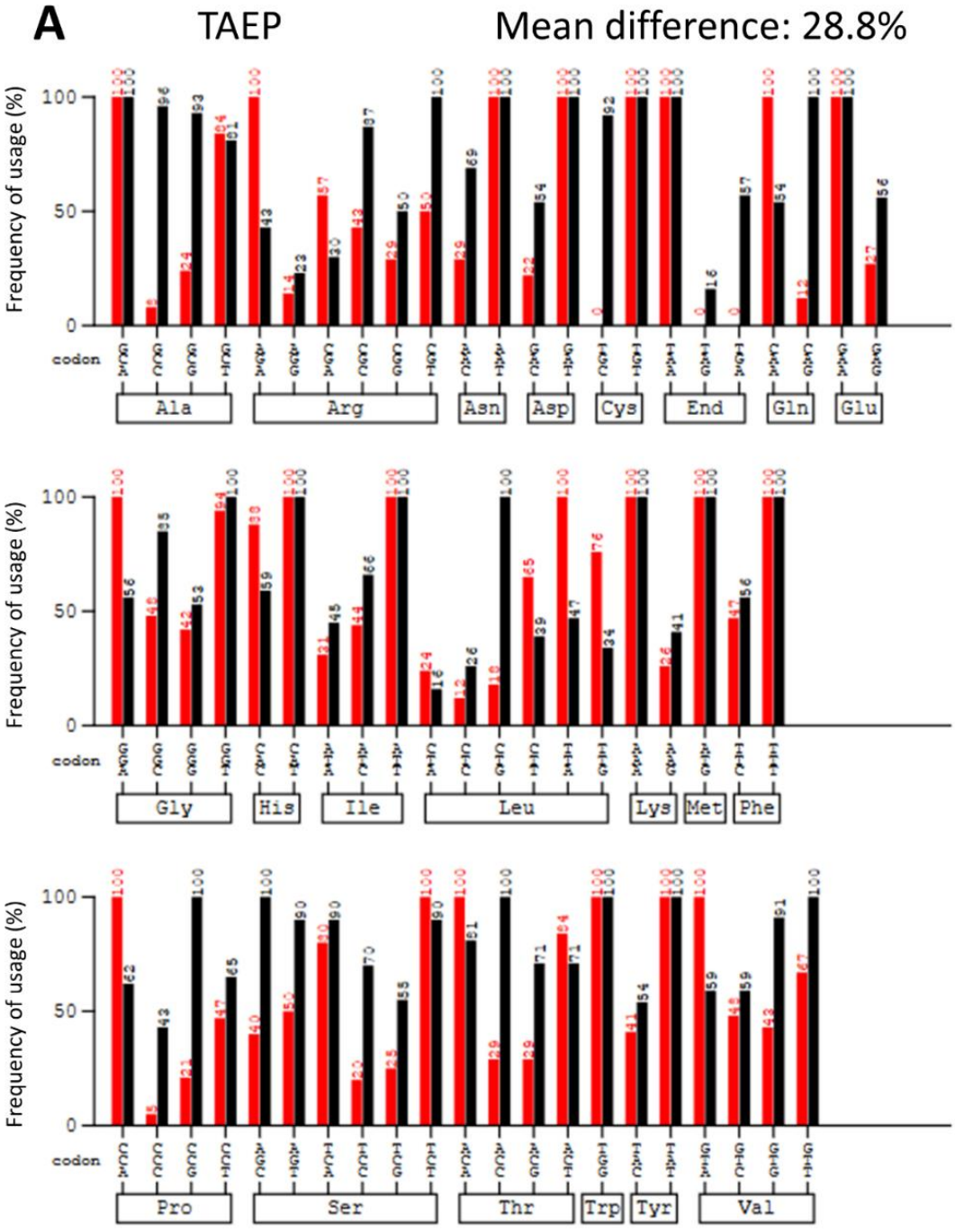


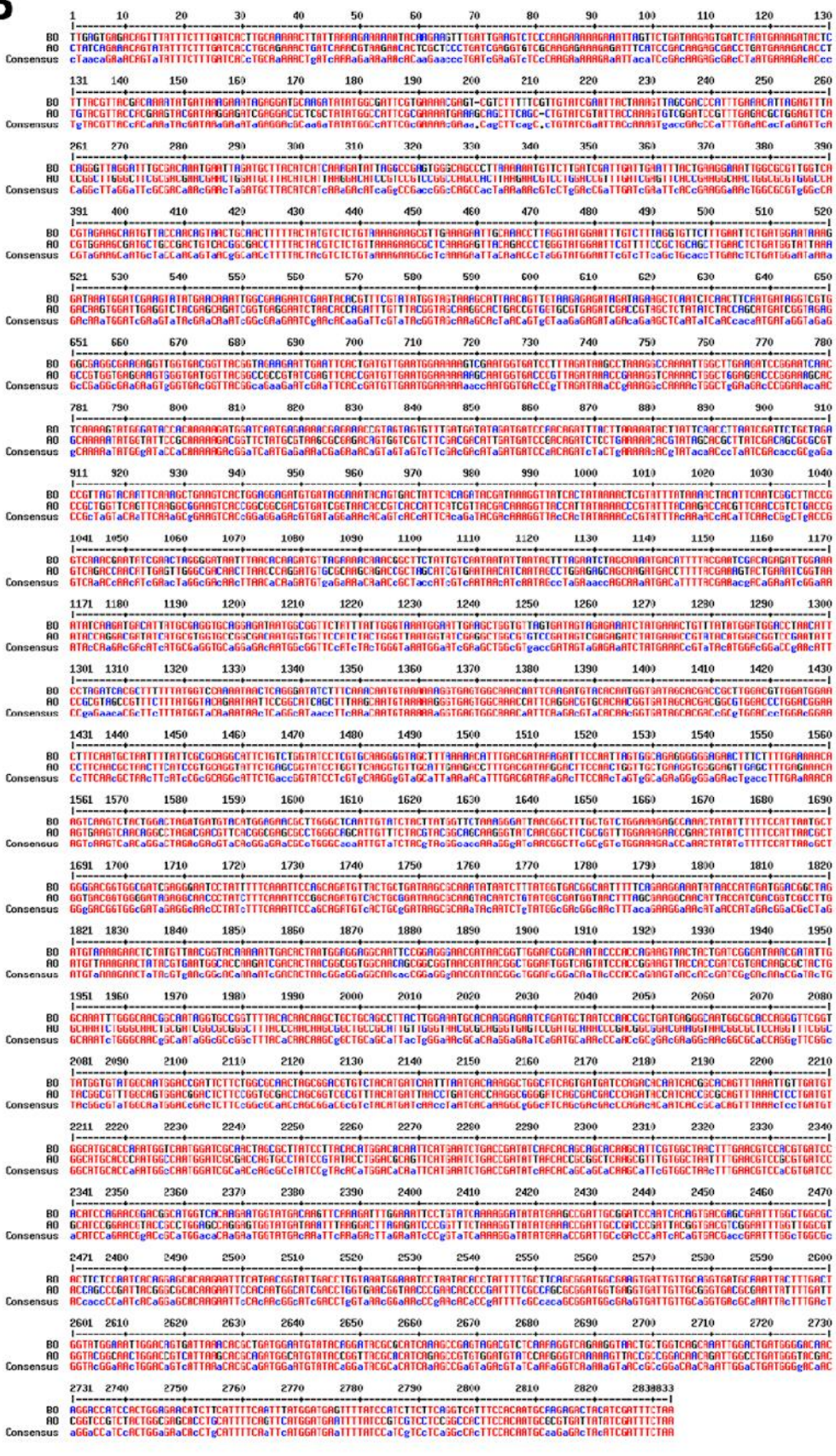
Figure 5.1 Domain architectures of the candidate TAL proteins. A) TAEP protein from intact prophage in the *E. faecium* E1 strain. B) TMP-LT protein (from phiSHEF2 phage) which the LT domain at the C-terminal is aimed for cloning. C) Pectinesterase protein from a prophage in the *E. faecium* Gr17 strain. D) GDPD protein from a prophage genome in the *E. faecium* E1334 strain. E) New Lipoprotein C/Protein of 60-kDa (NLPC/P60) from phiSHEF14 phage. Protein domains are shown with their accession numbers. The length of proteins (amino acids) is displayed on the right side.

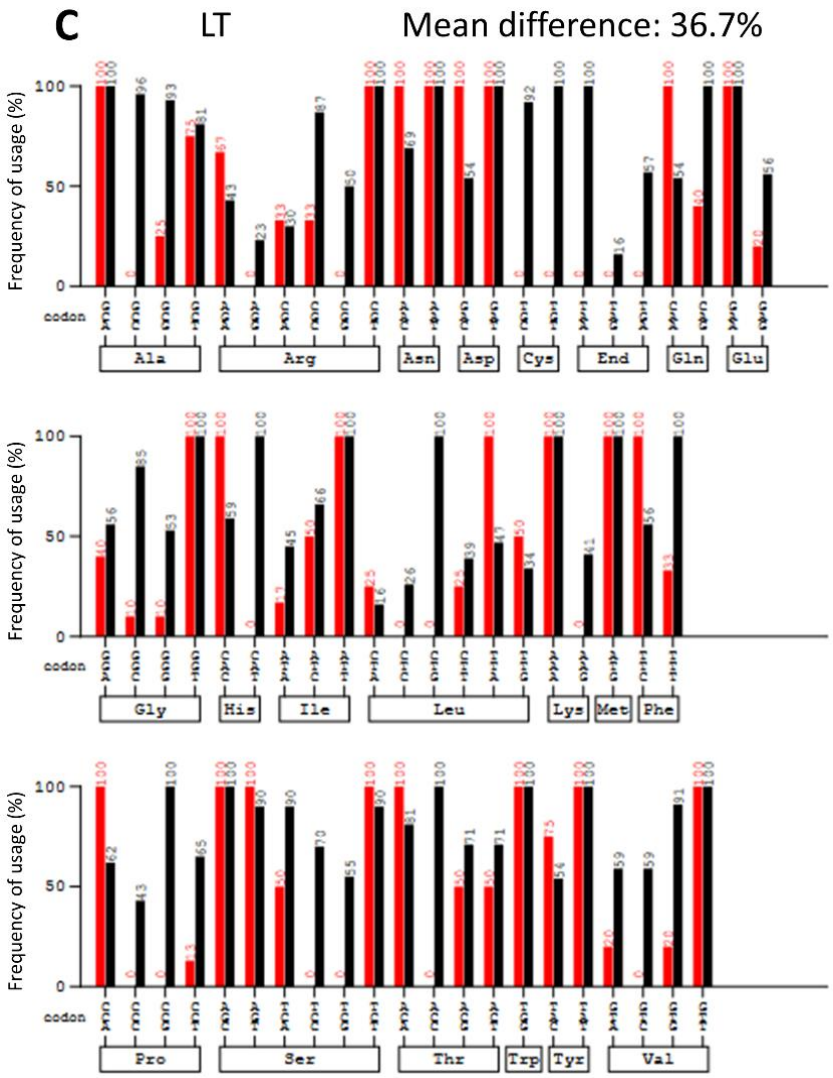
5.3 Codon optimisation

After selecting the proteins to be cloned, the difference in codon usage between the phage proteins and the expression host *E.coli* was first assessed. This was done via the website Graphical Codon Usage Analyser (GCUA) using the native sequence of the proteins. This resulted in a 28.8% mean difference in the frequency of codon usage between the TAEP and host *E.*coli (Figure 5.2 A). For LT, GDPD and PE, the differences were 36.9%, 25.9%, and 36.7%, respectively (Figure 5.2 C, E and G). This result shows how codon usage is different between E.coli and phage which emphasises the importance of codon optimization to increase the likelihood of successful protein expression. To enable codon optimisation, the GenSmart™ Codon Optimization online tool was used. Both sequences before and after codon optimisation were aligned to visualise the change in the nucleotide sequence (Figure 5.2 B, D, F and H). The codon optimisation step was done on TAEP, LT, PE and GDPD genes while NLPC/P60 was cloned via gene amplification.


A
TAEP
Mean difference: 28.8%
Frequency of usage (%)
codon
Ala
Arg
Asn
Asp
Cys
End
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val

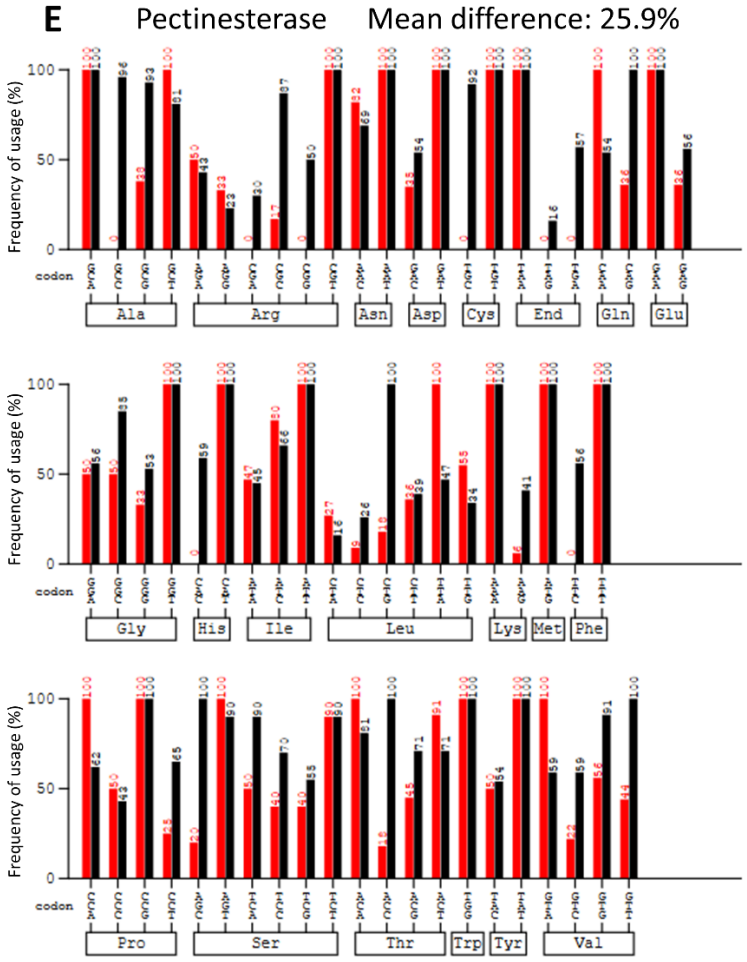
B

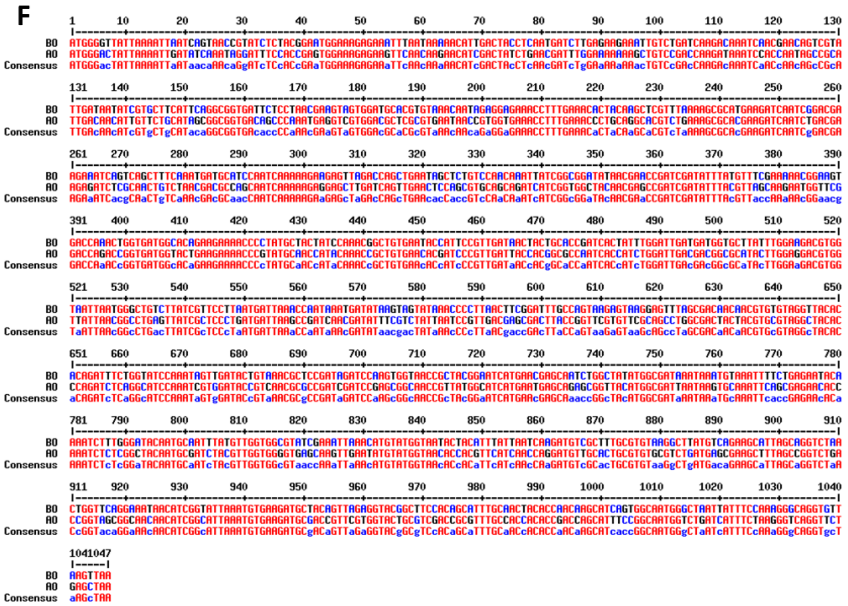


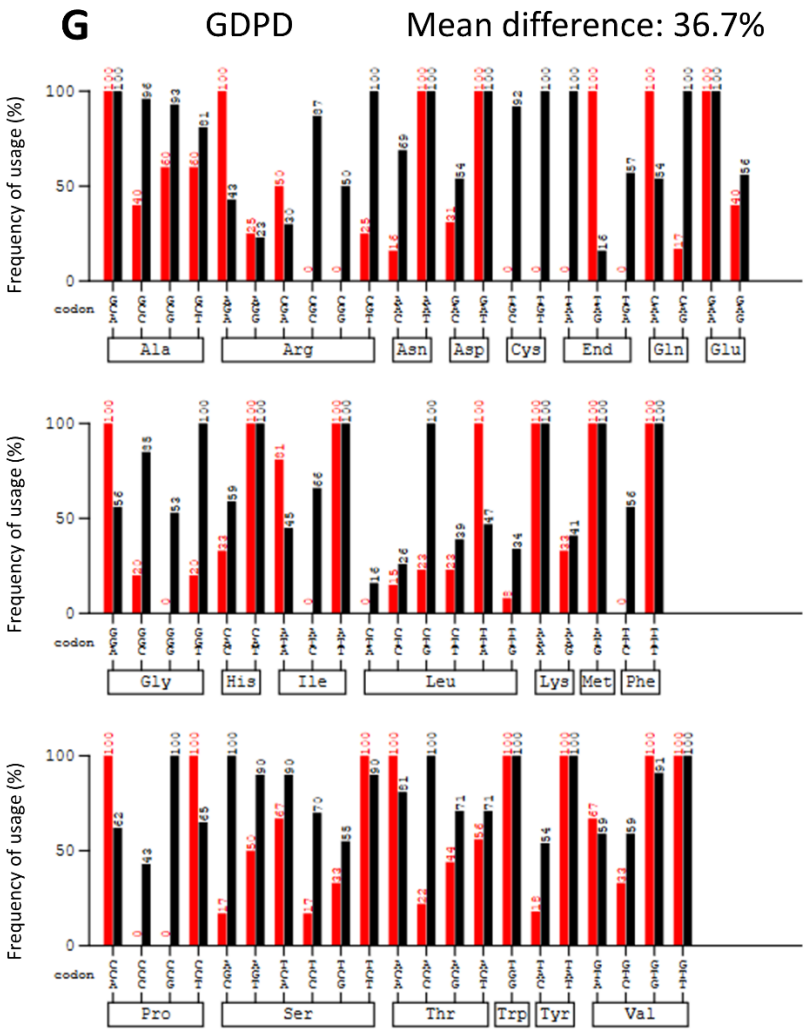


D

```
                1       10        20        30        40        50        60        70        80        90       100
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
BO              GTAGCTGGTTTAGTTGCTCAAATCATGCGCGAATCAGGTGGTAATGAAAAAATCGTTCAAAGTCCAGATGTGGTTGACATTAACACATTGAGTGGTAACC
AO              GTAGCTGGTCTTGTGGCTCAAATTATGCGTGAAAGCGGGGGTAACGAGAAAATCGTTCAGTCGCCGGATGTGGTTGACATCAACACCTTGTCCGGCAATC
Consensus       GTAGCTGGTcTaGTgGCTCAAATcATGCGcGAAacaGGgGGTAAcGAaAAAATCGTTCAaacgCCaGATGTGGTTGACATcAACACaTTGaccGGcAAcC

                101     110       120       130       140       150       160       170       180       190       200
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
BO              CAGCACGAGGATTATTACAATATATTCCACAGACGTTTGATGCGTACAAAGTTCCAGGTTATGGTAACATTAACAATGGATACCACCAATTGTTAGCATT
AO              CCGCACGTGGTTTGCTGCAATATATCCCGCAGACCTTCGACGCGTATAAAGTTCCGGGTTACGGAAACATTAATAACGGGTACCATCAGCTGCTGGCGTT
Consensus       CaGCACGaGGaTTacTaCAATATATcCCaCAGACcTTcGAcGCGTAcAAAGTTCCaGGTTAcGGaAACATTAAcAACGGaTACCAcCAacTGcTaGCaTT

                201     210       220       230       240       250       260       270       280       290       300
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
BO              CTTCAACAATAGCAACTGGGAAAATGACCTACAGTATGGTAAATCTGGTTGGGGTCCTCGTGGTCACCGTATCAGAGGGTACTTCAATGGAGGAATTGCA
AO              TTTTAATAACAGCAACTGGGAAAATGATCTACAATATGGTAAGTCTGGCTGGGGTCCGCGTGGTCACAGAATCCGTGGTTACTTTAATGGCGGTATTGCG
Consensus       cTTcAAcAACAGCAACTGGGAAAATGAcCTACAaTATGGTAAaTCTGGcTGGGGTCCgCGTGGTCACaGaATCaGaGGgTACTTcAATGGaGGaATTGCa

                301     310       320       330       340       350       360       370  375
                |--------+---------+---------+---------+---------+---------+---------+----|
BO              AGAACACCACAAATTGCAAGCCTTGCTGAAAATGGCTATCCAGAGGTTATAATTCCAACTGAACCATCTAAACGT
AO              CGCACCCGCAAATTGCAAGCCTGGCTGAAAACGGGTACCCGAGGTCATCATCCCGACTGAGCCGAGCAAACGT
Consensus       aGaACaCCaCAAATTGCAAGCCTgGCTGAAAAcGGcTAcCCaGAGGTcATaATcCCaACTGAaCCaaccAAACGT
```


E
Pectinesterase
Mean difference: 25.9%
Frequency of usage (%)
codon
Ala
Arg
Asn
Asp
Cys
End
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val


F
BO
AO
Consensus


G
GDPD
Mean difference: 36.7%
Frequency of usage (%)
codon
Ala
Arg
Asn
Asp
Cys
End
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val

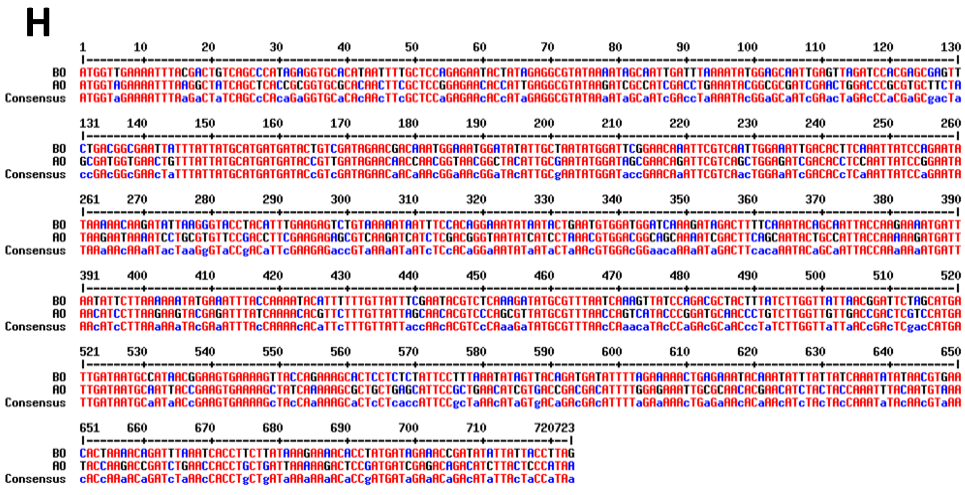

H
BO
AO
Consensus

Figure 5.2 Codon optimisation analysis. A ,C, E, G) the analysis of codon usage between TAEP, LT , GDP and PE and the expression cells *E. coli*, respectively. The percentage of similarity is shown on the top right side. Black indicates the frequency (%) of a given codon used in *E. coli*, whereas red indicates the frequency (%) of a given codon used in phage. B,D,F and H) The sequences (TAEP,LT,PE and GDPD), respectively, before and after codon optimisation were aligned using the online tool Multalin. High similarity between nucleotides is indicated by red colour while the difference in blue colour.

5.4 Prediction of protein solubility

One of the main issues with recombinant protein expression is protein insolubility (González-Montalbán et al., 2007). Therefore, an online tool called "protein-sol" (https://protein-sol.manchester.ac.uk/) was used to predict protein solubility which is based on amino acid sequences. The tool estimates protein solubility in comparison with the population average value for the experimental dataset which is 0.45. A greater value than 0.45 indicates higher protein solubility than the average soluble *E. coli* proteins (Niwa et al., 2009). The protein sequences (TAEP, LT, PE and GDPD) were run into this tool. The scores for these sequences were as follows: TAEP= 0.487, LT= 0.492, PE= 0.681 and GDPD= 0.661 (Figure 5.3 A, B, C and D). These values are all above the threshold value (0.45) indicating the probability of these proteins becoming in a soluble state upon expression and purification. In addition, the NLPC/P60 sequence (aimed to be cloned by gene amplification) was also tested which showed a lower value (0.277) than the threshold value predicting a possible insoluble state after protein expression (Figure 5.3 E). Additionally, signal peptides that facilitate the translocation of newly synthesised proteins to their secretory systems were also assessed in the selected proteins (Freudl, 2018). This was done using the webserver PSORTb (https://www.psort.org/psortb/) which resulted in no detection of signal peptides.

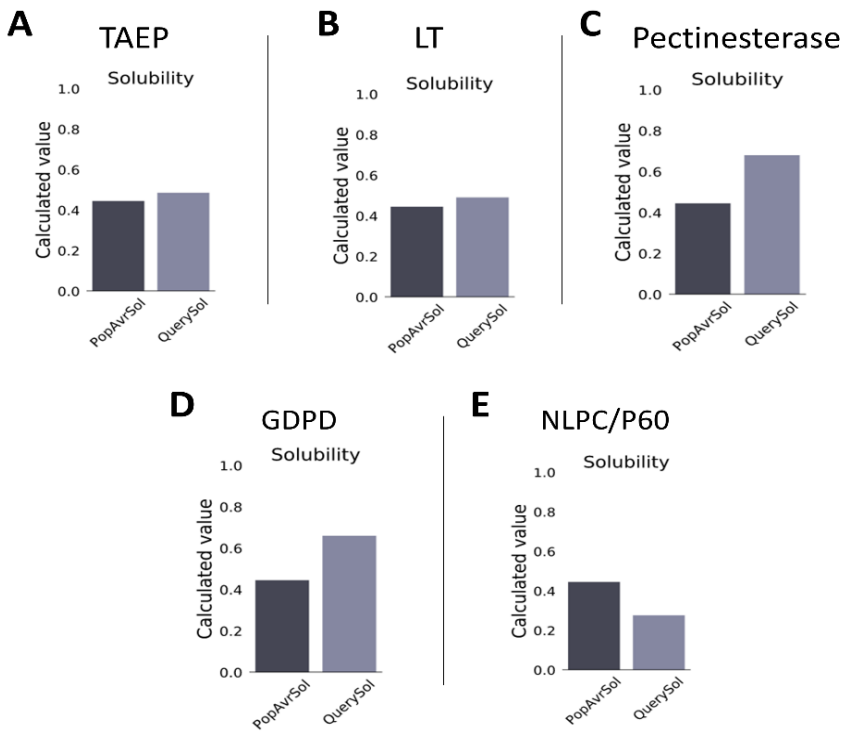


Figure 5.3 Prediction of protein solubility via the online tool "protein-sol". The results for TAEP (A) , LT (B) , Pectinesterase (C) , GDPD (D) and NLPC/P60 (E) are shown. PopAvrSol refers to the population average for the experimental dataset which is 0.45. protein sequences with a higher value (>0.45) indicate higher protein solubility.

5.5 Cloning by gene synthesis

Four sequences (TAEP, LT, PE & GDPD) were set out for cloning via gene synthesis. After codon optimisation, appropriate restriction sites were selected based on the protein sequences and the pET21b expression vector (Figure 5.4 A and B). The pET21b vector contains a T7 promoter and a His-tag sequence at the C-terminal region. For all the sequences, the *Nde*I and *Xho*I restriction sites were selected to be added into the sequences to allow cloning into the pET21b vector (Figure 5.5 A, C, E and G). After gene synthesis and receiving the samples from GENEWIZ (www.genewiz.com), the recombinant vectors were suspended in nuclease-free water and kept at -20 °C. Quality files were also received from GENEWIZ showing the successful gene synthesis (Figure 5.5 B, D, F and H).

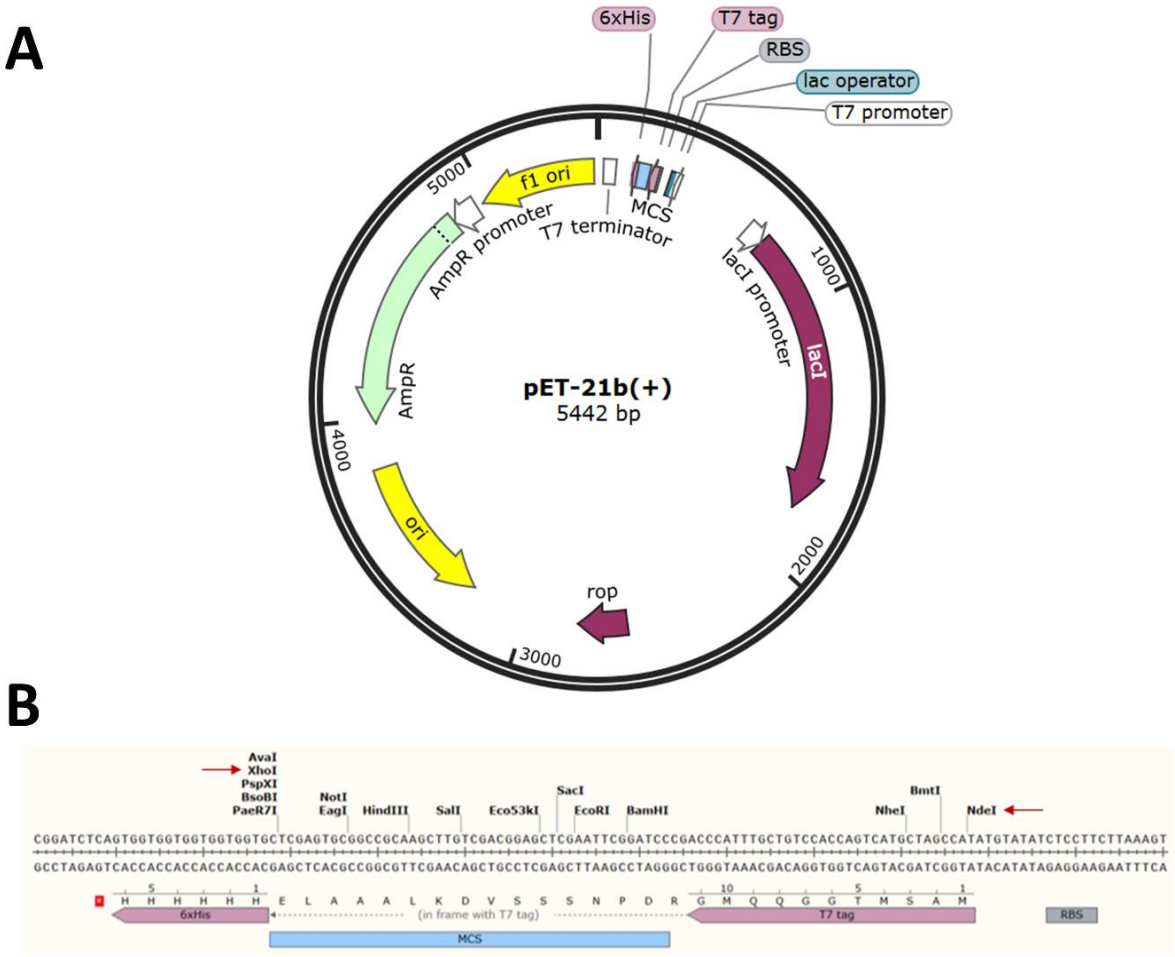
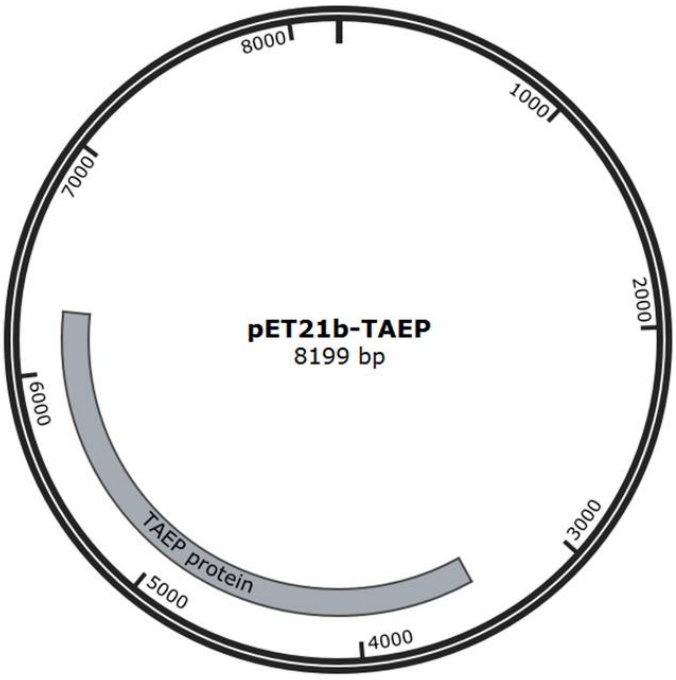


Figure 5.4 pET21b vector map and cloning details. A) the whole details of the pET21b vector which shows Ampicillin resistant gene (AmpR), multiple cloning site (MCS) and origin of replication (ori). B) the restriction sites selected for cloning are labelled (red arrows).

A

```
                10        20        30        40        50        60        70
CATATGCTATCAGAAACAGTATATTTCTTTGATCACCTGCAGAAACTGATCAAACGTAAGAACACTCGCTCCCTGAT    77
CGAGGTGTCGCAAGAGAAAGAGATTTCATCCGACAAGAGCGACCTGATGAAAGACACCTTGTACGTTACCACGAAGT   154
ACGATAAGGAGATCGAGGACGCTCGCTATATGGCCATTCGCGAAAATGAAAGCAGCTTCAGCCTGTATCGTATTACC   231
AAAGTGTCGGATCCGTTTGAGACGCTGGAGTTCACCGGCTTGGGCTTCGCGACGAACGAACTGGATGCTTACATCAT   308
TAAGGACATCCGTCCGTCCGGCCAGCCACTTAAGAACGTCCTGGACCGTTTGATCGAGTTCACCGAAGGCAACTGGC   385
GCGTGGGCCACGTGGAAGCGATGCTGCCGACTGTCACGGCGACCTTTTACTACGTCTCTGTTAAAGAAGCGCTCAAA   462
GAGTTACAGACCCTGGGTATGGAATTCGTTTTCCGCTGCAGCTTGAACTCTGATGGTATTAAAGACAAGTGGATTGA   539
GGTCTACGAGCAGATCGGTGAGGAATCTAACACCAGATTTGTTTACGGTAGCAAGGCACTGACCGTGGTGCGTGAGA   616
TCGACCGTAGCTCTATATCTACCAGCATGATCGGTAGAGGCCGTGGTGAGGAAGTGGGTGATGGTTACGGCCGCCGT   693
ATCGAGTTCACCGATGTTGAATGGAAAAAAAGCAATGGTGACCCGTTAGATAAACCGAAAGGTCAAAACTGGCTGGA   770
GGACCCGGAAAGCACGCAAAAATATGGTATTCCGCAAAAAGACGGTTCTATGCGTAAGCGCGAGACAGTGGTCGTCT   847
TCGACGACATTGATGATCCGACAGATCTCCTGAAAAACACGTATAGCACGCTTATCGACAGCGCGCGTCCGCTGGTT   924
CAGTTCAAGGCGGAAGTCACCGGCGGCGACGTGATCGGTAACACCGTCACCATTCATCGTTACGACAAAGGTTACCA  1001
TTATAAAACCCGTATTTACAAGACCACGTTCAACCGTCTGACCGGTCAGACCAACATTGAGTTGGGCGACAACTTAA  1078
CCCAGGATGTGCGCAAGCAGACCGCTAGCATCGTGAATAACATCAATAGCCTGGAGAGCAGCAAGATGACCTTTTAC  1155
GAAAGTACTGAAATCGGTAAATACCAGGACGATATCATGCGTGGTGCCGGCGACAATGGTGGTTCCATCTACTGGGT  1232
TAATGGTATCGAGGCTGGCGTGTCCGATAGTCGAGAGATCTATGAAACCGTATACATGGACGGTCCGAATATTCCGC  1309
GTAGCCGTTTCTTTATGGTACAGAATAATTCCGGCATCAGCTTTAAGCAATGTAAAAAGGGTGAGTGGCAAACCATT  1386
CAGGACGTGCACAACGGTGATAGCACGACGGCGTGGACCCTGGACGGAACCTTCAACGCTAACTTCATCCGTGCAGG  1463
TATTCTGAGCGGTATCCTGGTTCAAGGTGTTGCATTGAAGACCTTTGACGATAAGGACTTCCAACTGGTTGCTGAAG  1540
GTGGGGAGTTGAGCTTTGAGAAACAAGTGAAGTCAACAGGCCTAGACGACGTTCACGGCGAGCGCCTGGGCAGCATT  1617
GTTTCTACGTACGGCAGCAAGGGTATCAACGGCTTCGCGGTTTGGAAAGAACCGAACTATATCTTTTCCATTAACGC  1694
TGGTGACGGTGGGGATAGAGGCAACCCTATCTTTCAAATTCCGGCAGATGTCACTGCGGATAAGCGCAAGTACAATC  1771
TGTATGGCGATGGTAACTTTAGCGAAGGCAACATTACCATCGACGGTCGCCTTGATGTTAAAGAACTATACGTGAAT  1848
GGCACCAAGATCGACACTAACGGCGGTGGCAACAGCGGCGGTAACGATAACGGCTGGAATGGTCAGTATCCACCGGA  1925
AGTTACCACCGATCGTGACAAGCGCTACTGGCAAATCTGGGCAACTGCGATCGGCGCGGGCTTTACCCAACAAGCGG  2002
CTGCCGCATTGTTGGGTAACGCGCAGGGTGAGTCCGATGCAAACCCGACGGCGGACGAAGGTAACGGCGCTCCAGGT  2079
TTCGGCTACGGCGTTTGGCAGTGGACGGACTCTTCCGGTGCGACCAGCGGTCGCGTTTACATGATTAACCTGATGAC  2156
CAAGGCGGGGATCAGCGACGACCCAGATACCATCACCGCGCAGTTTAAACTCCTGATGTGGCATGCACCCAATGGCC  2233
AATGGATCGCGACCAGTGCCTATCCGTATACCTGGACGCAGTTCATGAATCTGACCGATATTAACACCGCGGCTCAA  2310
GCGTTTGTGGCTAATTTTGAACGTCCGCGTGATCCGCATCCGGAACGTACCGCCTGGAGCCAGGAGTGGTATGATAA  2387
ATTTAAGGACTTAGAGATCCCGGTTTCTAAAGGTTATATGAAACCGATTGCCGACCCGATTACGGTGACGTCGGAAT  2464
TTGGTTGGCGTACCAGCCCGATTACGGGCGCACAAGAATTCCACAATGGCATCGACCTGGTGAACGGTAACCCGAAC  2541
ACCCCGATTTTCGCCAGCGCGGATGGTGAGGTGATTGTTGCGGGTGACGCGAATTATTTTGATTGGTACGGCAACTG  2618
GACCGTCATTAAGCACGCAGATGGCATGTATACCGGTTACGCACATCAGAGCCGTGTGGATGTATCCAAGGGTCAAA  2695
AAGTTACCGCCGGACAACAGATTGGCCTGATGGGTACGACCGGTCCGTCTACTGGCGAGCACCTGCATTTTCAGTTC  2772
ATGGATGAATTTTATCCGTCGTCCTCCGGCCACTTCCACAATGCGCGTGATTATATCGATTTCCTCGAG           2841
```

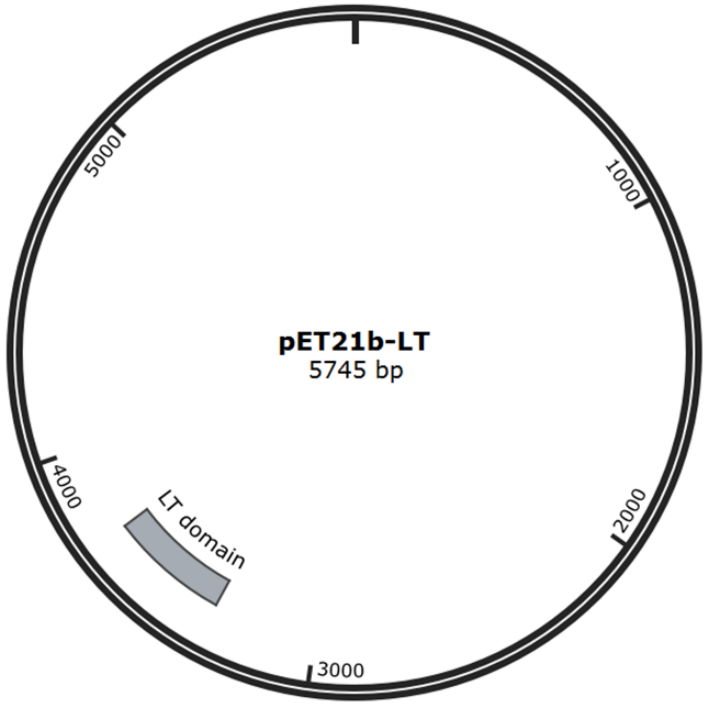
B



C

```
                10        20        30        40        50
CATATGGTAGCTGGTCTTGTGGCTCAAATTATGCGTGAAAGCGGGGGTAA      50
CGAGAAAATCGTTCAGTCGCCGGATGTGGTTGACATCAACACCTTGTCCG     100
GCAATCCCGCACGTGGTTTGCTGCAATATATCCCGCAGACCTTCGACGCG     150
TATAAAGTTCCGGGTTACGGAAACATTAATAACGGGTACCATCAGCTGCT     200
GGCGTTTTTTAATAACAGCAACTGGGAAAATGATCTACAATATGGTAAGT     250
CTGGCTGGGGTCCGCGTGGTCACAGAATCCGTGGTTACTTTAATGGCGGT     300
ATTGCGCGCACCCCGCAAATTGCAAGCCTGGCTGAAAACGGGTACCCGGA     350
GGTCATCATCCCGACTGAGCCGAGCAAACGTCTCGAG                  387
```

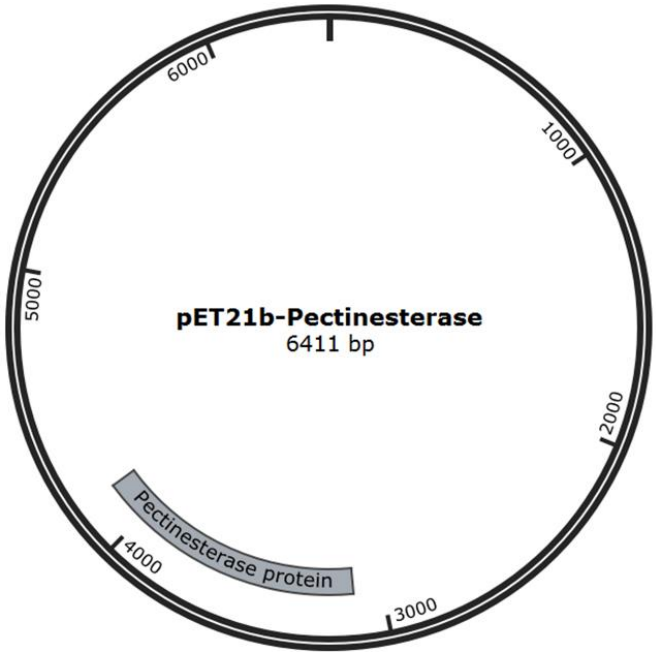
D



E

```
                10        20        30        40
CATATGGGACTATTAAAATTGATATCAAATAGGATTTCCACCGAG    45
TGGAAAGAGAAGTTCAACAAGAACATCGACTATCTGAACGATTTG    90
GAAAAAAAGCTGTCCGACCAAGATAAATCCACCAATAGCCGCATT   135
GACAACATTGTTCTGCATAGCGGCGGTGACAGCCCAAATGAGGTC   180
GTGGACGCTCGCGTGAATAACCGTGGTGAAACCTTTGAAACCCTG   225
CAGGCACGTCTGAAAGCGCACGAAGATCAATCTGACGAAGAGATC   270
TCGCAACTGTCTAACGACGCCAGCAATCAAAAAGAGGAGCTTGAT   315
CAGTTGAACTCCAGCGTGCAGCAGATCATCGGTGGCTACAACGAG   360
CCGATCGATATTTACGTTAGCAAGAATGGTTCGGACCAGACCGGT   405
GATGGTACTGAAGAAAACCCGTATGCAACCATACAAACCGCTGTG   450
AACACGATCCCGTTGATTACCACGGCGCCAATCACCATCTGGATT   495
GACGACGGCGCATACTTGGAGGACGTGGTTATTAACGGCCTGAGT   540
TATCGCTCCCTGATGATTAAGCCGATCAACGATATTTCGTCTATT   585
AATCCGTTGACGAGCGACTTACCGGTTCGTGTTCGCAGCCTGGCG   630
ACTACTACGTGCGTAGGCTACACCCAGATCTCAGGCATCCAAATC   675
GTGGATACCGTCAACGCGCCGATCGATCCGAGCGGCAACCGTTAT   720
GGCATCATGAATGAGCAGAGCGGTTACATGGCGATTAATAAGTGC   765
AAATTCAGCGAGAACACCAAATCTCTCGGCTACAATGCGATCTAC   810
GTTGGTGGGGTGAGCAAGTTGAATATGTATGGTAACACCACGTTC   855
ATCAACCAGGATGTTGCACTGCGTGTGCGTCTGATGAGCGAAGCT   900
TTAGCCGGTCTGACCGGTAGCGGCAACAACATCGGCATTAAATGT   945
GAAGATGCGACCGTTCGTGGTACTGCGTCGACCGCGTTTGCCACC   990
ACACCGACCAGCATTTCCGGCAATGGTCTGATCATTTCTAAGGGT  1035
CAGGTTCTGAGCCTCGAG                            1053
```

F



G

```
                10        20        30        40        50
CATATGGTAGAAAATTTAAGGCTATCAGCTCACCGCGGTGCGCACAACTT      50
CGCTCCGGAGAACACCATTGAGGCGTATAAGATCGCCATCGACCTGAAAT     100
ACGGCGCGATCGAACTGGACCCGCGTGCTTCTAGCGATGGTGAACTGTTT     150
ATTATGCATGATGATACCGTTGATAGAACAACCAACGGTAACGGCTACAT     200
TGCGAATATGGATAGCGAACAGATTCGTCAGCTGGAGATCGACACCTCCA     250
ATTATCCGGAATATAAGAATAAAATCCTGCGTGTTCCGACCTTCGAAGAG     300
AGCGTCAAGATCATCTCGACGGGTAATATCATCCTAAACGTGGACGGCAG     350
CAAAATCGACTTCAGCAATACTGCCATTACCAAAAAGATGATTAACATCC     400
TTAAGAAGTACGAGATTTATCAAAACACGTTCTTTGTTATTAGCAACACG     450
TCCCAGCGTTATGCGTTTAACCAGTCATACCCGGATGCAACCCTGTCTTG     500
GTTGTTGACCGACTCGTCCATGATTGATAATGCAATTACCGAAGTGAAAA     550
GCTATCAAAAAGCGCTGCTGAGCATTCCGCTGAACATCGTGACCGACGAC     600
ATTTTGGAGAAATTGCGCAACACGAACATCTACTACCAAATTTACAATGT     650
AAATACCAAGACCGATCTGAACCACCTGCTGATTAAAAAGACTCCGATGA     700
TCGAGACAGACATCTTACTCCCACTCGAG                          729
```

H

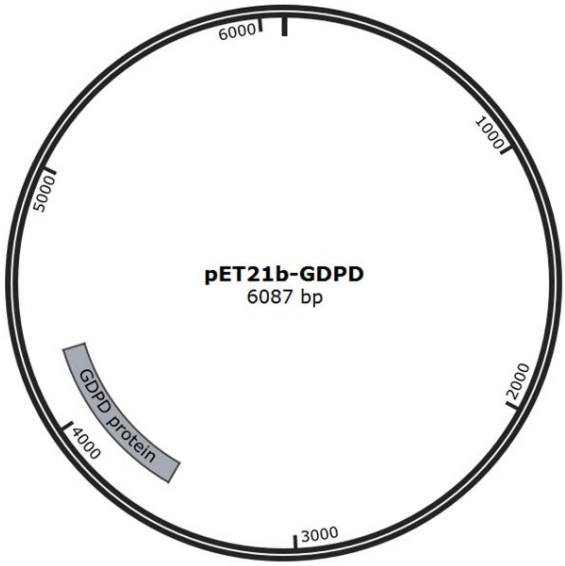


Figure 5.5 Cloning of TAEP, LT, PE and GDPD in the pET21b vector. A,C,E,G) show TAEP, LT, Pectinesterase and GDPD sequences, respectively, after adding the restriction sites *NdeI* (Green) and *XhoI* (Orange) while B,D,F,H show their successful subcloning into the vector after gene synthesis.

5.6 Cloning by gene amplification

In contrast to the above-mentioned proteins (TAEP, LT, PE and GDPD), the NLPC/P60 protein was decided to be cloned via gene amplification from phiSHEF14 that was isolated in this work and discussed in chapter 4. The phiSHEF14 phage is a podovirus which has a 19.3 kb genome size and its genome was sequenced and annotated in this work as well. Upon annotation, a protein with an NLPC/P60 domain was identified (phiSHEF14_11). To begin the cloning process, appropriate restriction sites were first selected in accordance with the vector pET21a. These restriction sites were *NdeI* and *XhoI* which the NLPC/P60 protein sequence lacks (Figure 5.6 A and B). Therefore, these restriction sites were included in the forward (*NdeI*) and reverse (*XhoI*) primers which were designed using the SnapGene software (Figure 5.7 B and C). Upon receiving the primers, PCR reactions were performed using the phiSHEF14 DNA as a template. Two annealing temperatures were first tested (58 °C & 60 °C) which both showed bands at the expected amplicon size (2010 bp) (Figure 5.7 D). Upon the successful amplification of the NLPC/P60 gene, the PCR clean-up kit was used. For the pET21a plasmid, the plasmid miniprep kit was used to extract the plasmid from the DH5α cells which is followed by a cleaning step using the PCR clean-up kit.

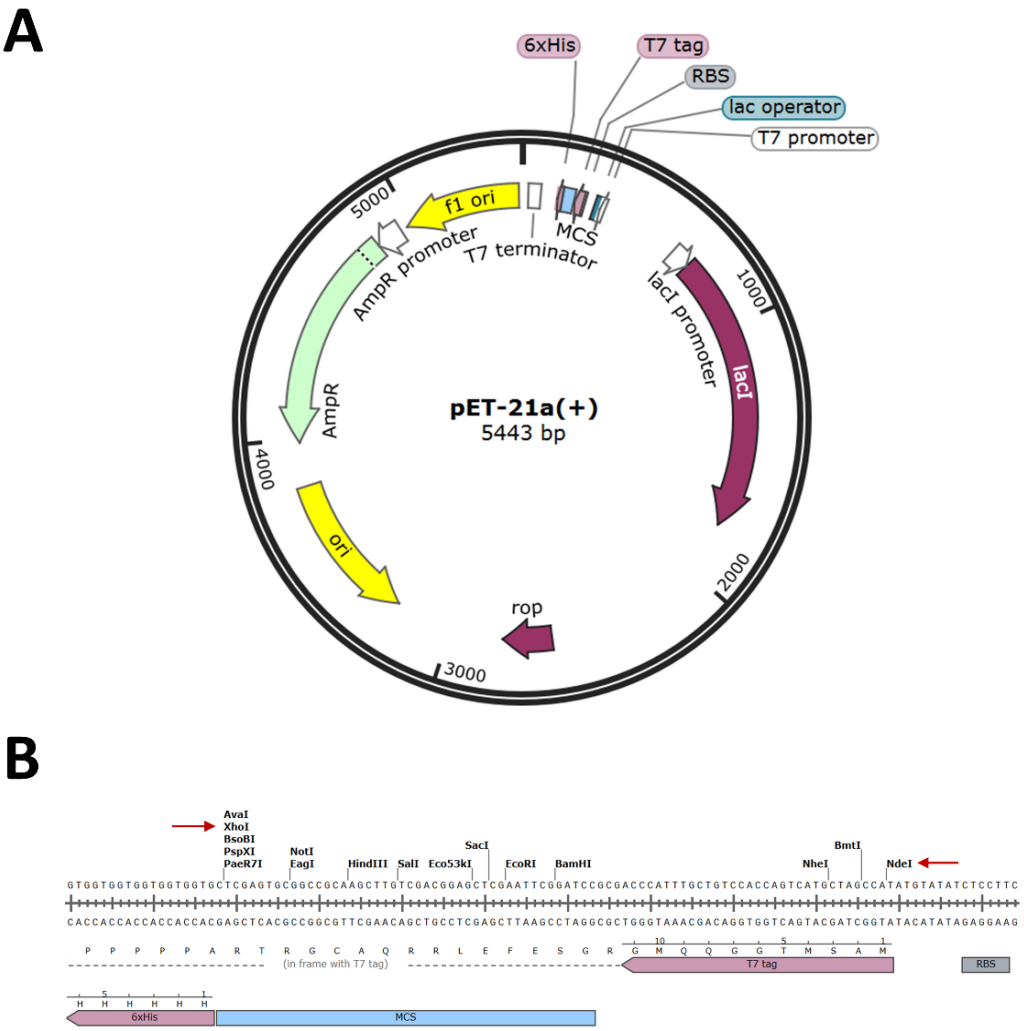


Figure 5.6 pET21a vector map and cloning details. A) the whole details of the pET21a vector which shows Ampicillin resistant gene (AmpR), multiple cloning site (MCS) and origin of replication (ori). B) the restriction sites chosen for cloning are labelled (red arrows).

Both the NLPC/P60 gene and pET21a vector were prepared for cloning separately via double digestion with *NdeI* and *XhoI* restriction enzymes. To avoid self-ligation in the vector, the alkaline phosphatase CIP (New England Biolabs) was used to dephosphorylate the vector's ends. Both the gene and vector were further checked on a gel after the digestion (Figure 5.7E). After this, a ligation step was performed in which the NLPC/P60 gene was ligated into the vector based on the compatibility of the insert and vector ends. This was done by using the T4 DNA-ligase enzyme.

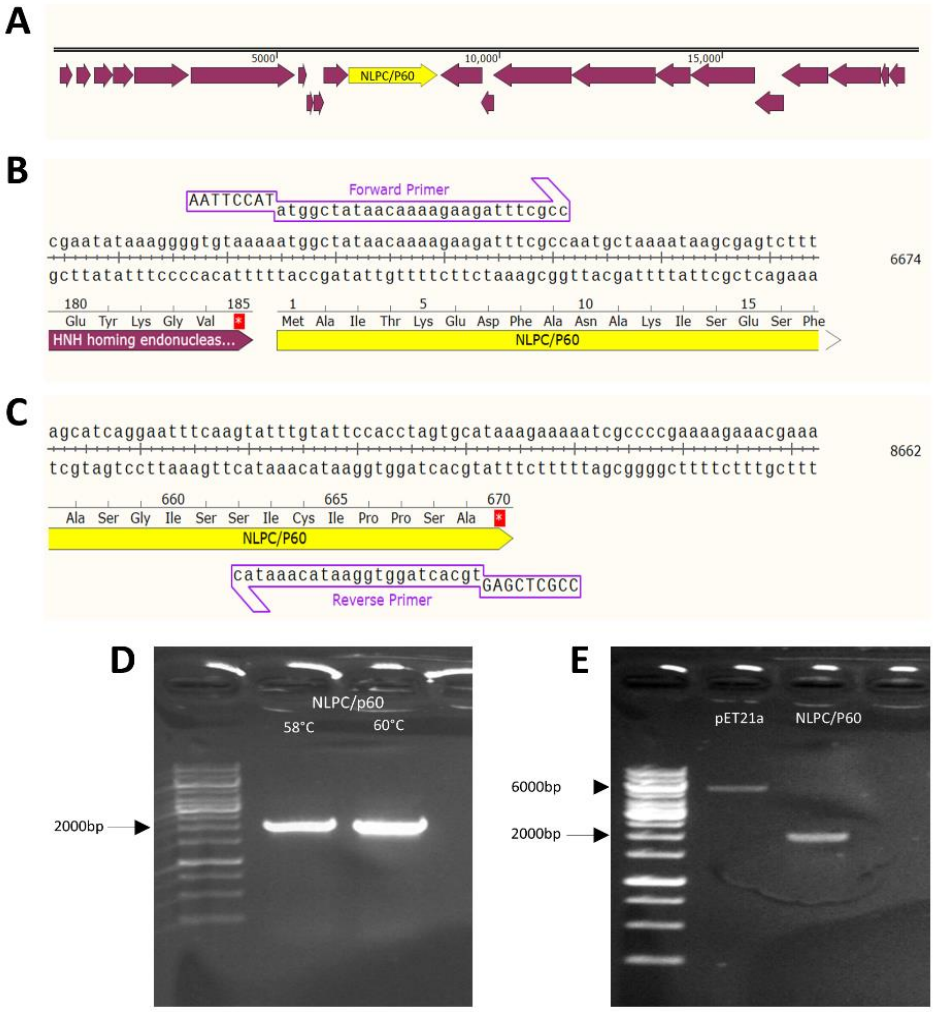


Figure 5.7 NLPC/P60 gene amplification. A) phiSHEF14 genome is shown and the NLPC/P60 gene is labelled (yellow). B) Forward and C) Reverse primers used for gene amplification are shown. D) the amplified gene is checked on a gel (around 2000 bp band) for both 58 °C and 60 °C annealing temperatures (PCR). E) the pET21a vector and NLPC/P60 gene after the double digestion with NdeI and XhoI.

After ligation, the pET21a-NLPC/P60 recombinant vector was inserted into the *E.coli* DH5α competent cells via the heat shock transformation technique. The *E.coli* cells were then grown overnight at 37 °C on LB containing ampicillin as a selective marker. Next day, the bacterial colonies were screened for the desired transformation via colony PCR. The primers used for cloning the NLPC/P60 gene were also used here to check the NLPC/P60 insert in the vector. This resulted in some colonies (C 2,3,4,10) showing a band at the expected size (2010bp) (Figure

5.8). Out of these positive colonies, the C3 colony was selected and a liquid culture in LB was prepared and stored at -80 °C.

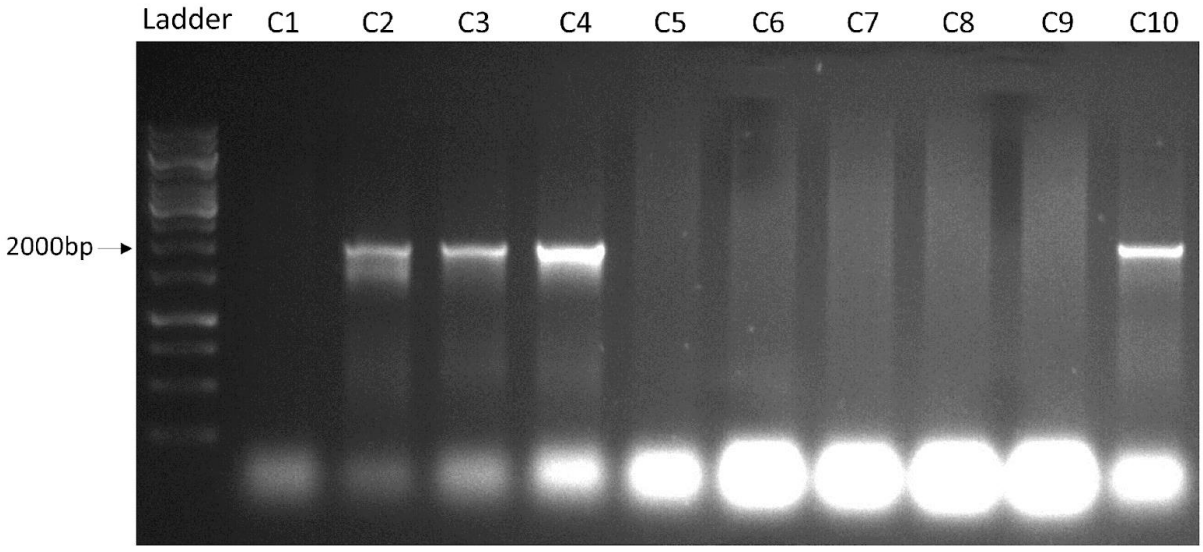


Figure 5.8 Colony PCR to assess the successful transformation of DH5α cells. Colonies (C) 3,4,5 and 10 show a band at approximately 2000 bp while other lanes show no bands at the expected size.

To further assess the recombinant vectors from both the gene amplified and synthesised techniques, each vector was digested with *NdeI* and *XhoI* restriction enzymes and ran on a gel to check sequence sizes. As a result, the gel showed expected band sizes that belong to the vector (approximately 6000 bp) and inserts (TAEP= 3000 bp band, LT= between 250 and 500 bp band, PE= 1000 bp band, GDPD= 750 bp band and NLPC=2000 bp band) (**Figure 5.9**). For the pET21b-TAEP gel bands, they look less clear compared with the other bands in other lanes. The preparation for the double digestion and gel loading was consistent among all samples but these faint bands could be due to pipetting errors.

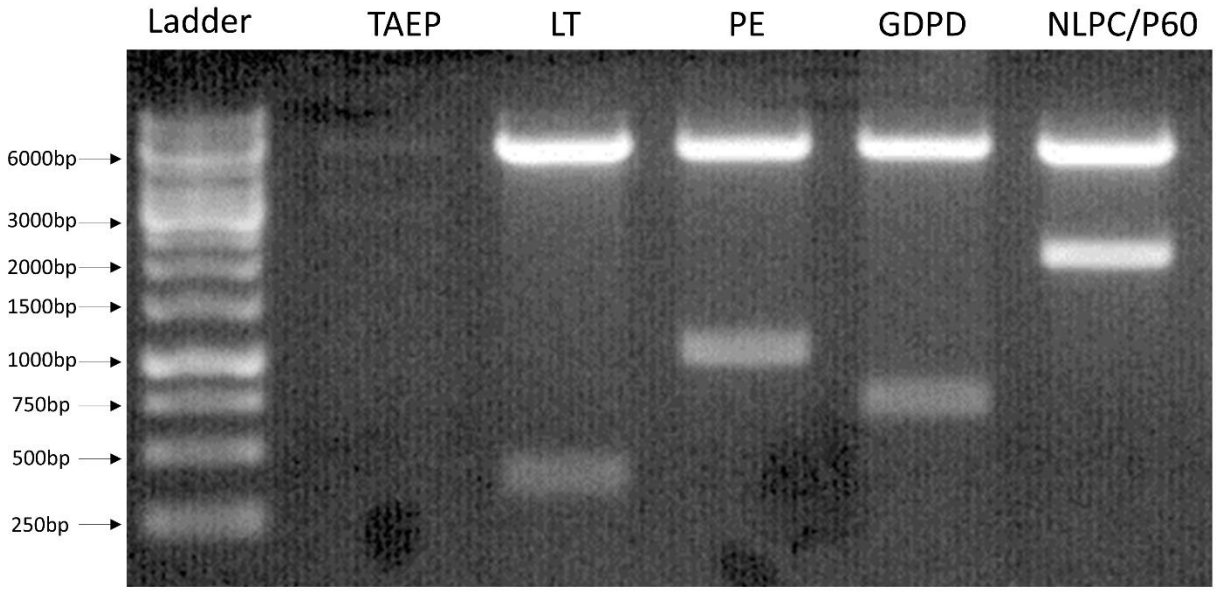


Figure 5.9 Double digestion of vectors carrying proteins of interest. The restriction enzymes used were NdeI and XhoI. Lanes are in the following order: ladder, pET21b -TAEP, pET21b -LT, pET21b -PE, pET21b -GDPD and pET21a-NLPC/P60.

5.7 Expression of His-tagged proteins

After the successful cloning and transformation of the abovementioned sequences, the assessment of protein expression was then carried out. To do this, the appropriate expression cells were transformed with a vector containing the sequence of interest. In this work, two *E.coli* expression cells were used: BL21(λDE3) and C41(λDE3).

5.7.1 BL21(λDE3)

The expression strain BL21(λDE3) has been used extensively for expressing recombinant proteins (Rosano & Ceccarelli, 2014). This strain contains a prophage λDE3 which is the source of the T7 polymerase needed for the protein expression. In addition, BL21(λDE3) is sensitive to ampicillin and this feature is exploited to assess successful transformation with pET21 vectors (carry an ampicillin resistance gene). The BL21(λDE3) also facilitates plasmid stability and

protein expression via lacking some genes that encode for proteases (Rosano & Ceccarelli, 2014).

Initially, a small culture size (10 ml) was first tested to verify protein expression and solubility at two different incubation times and temperatures (37 °C for 3 h and 30 °C for 12 h). Protein expression was induced with 0.25 mM IPTG for all proteins. Before adding IPTG, a sample was taken to compare uninduced and induced samples. After protein expression was complete, the bacterial cells were pelleted before mixing with 2x SDS lysis buffer and heating at 95 °C for 15 min. The samples were then loaded into polyacrylamide gels before staining with Coomassie blue. Using the ExPASy ProtParam web tool (http://web.expasy.org/protparam/), the five protein sequences showed an approximate protein size as follows: TAEP = 105 KDa, LT = 13.8 KDa, PE = 37 KDa, GDPD = 27 KDa and NLPC/P60 = 72 KDa. More details about the size of the analysed proteins are included in Table 5.1.

Table 5.1 TAL proteins properties

TALs	Gene length (bp)	Protein length (aa)	Protein weight (KDa) His tag	Protein weight (KDa) GST tag
TAEP	2832	943	105	131
LT (domain)	375	125	13.8	39.8
PE	1047	348	37	63
GDPD	723	240	27	53
NLPC/P60	2010	669	72	98

The expression experiment showed various results with the five sequences. For TAEP and PE, faint bands in the induced lane were observed around the expected size (105 and 37 KDa, respectively) (Figure 5.10). No overexpression bands were seen for LT and GDPD protein while the NLPC/P60 protein showed a thick band around the expected size in the induced lane (Figure 5.10).

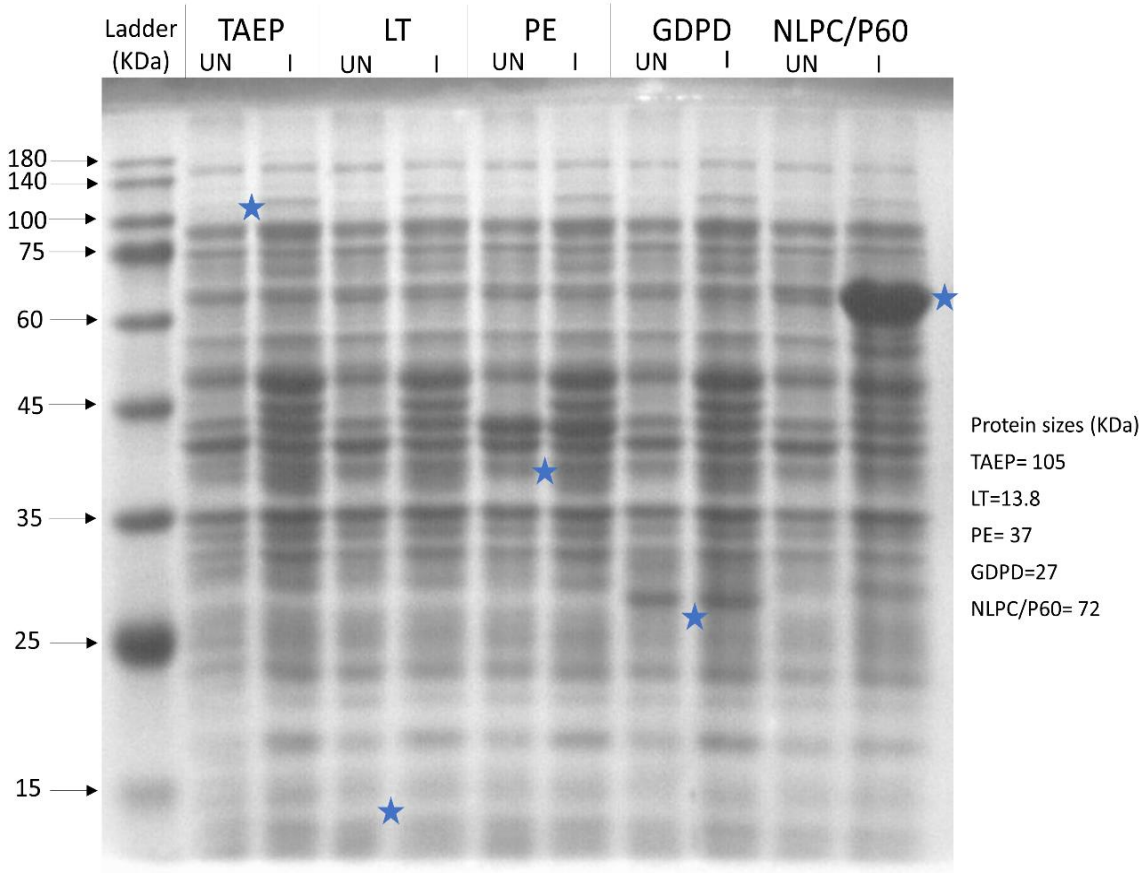


Figure 5.10 His-tag Protein overexpression using BL21 cells. Tested proteins (TAEP, LT, PE GDPD and NLPC/P60) were assessed using uninduced (UN) and induced (I) samples. The induction temperature was 30 °C. 12% polyacrylamide gel was used to run the SDS-PAGE. The stars indicate the predicated protein sizes.

5.7.2 C41(λDE3)

For the other expression strain C41(λDE3), it also contains some features, as BL21(λDE3), to help protein expression and vector stability and it has also been used for expressing toxic proteins (Miroux & Walker, 1996). As with BL21(λDE3), the C41(λDE3) cells containing pET21 vectors were induced with 0.25 mM IPTG and incubated at 37 °C for 3 h or 30 °C for 12 h. A sample was also collected before induction and labelled uninduced "UN" to compare the protein profile after induction. After Coomassie staining and comparing with the uninduced lanes, the gel showed thick bands in the induced lanes around the expected protein sizes for TAEP, PE, GDPD and NLPC/P60 (Figure 5.11). For the LT expression, the bands at the expected protein size were very faint in both uninduced and induced lanes (Figure 5.11). Therefore, it was decided to run it again at a higher gel concentration (15%) since the size of LT is relatively small (13.8 kDa). This was resulted in also no difference between the uninduced and induced samples and no overexpression was seen at the expected protein size (Figure 5.12). After the apparent expression of TAEP, PE, GDPD and NLPC/P60, a further investigation was carried out to test their solubility.

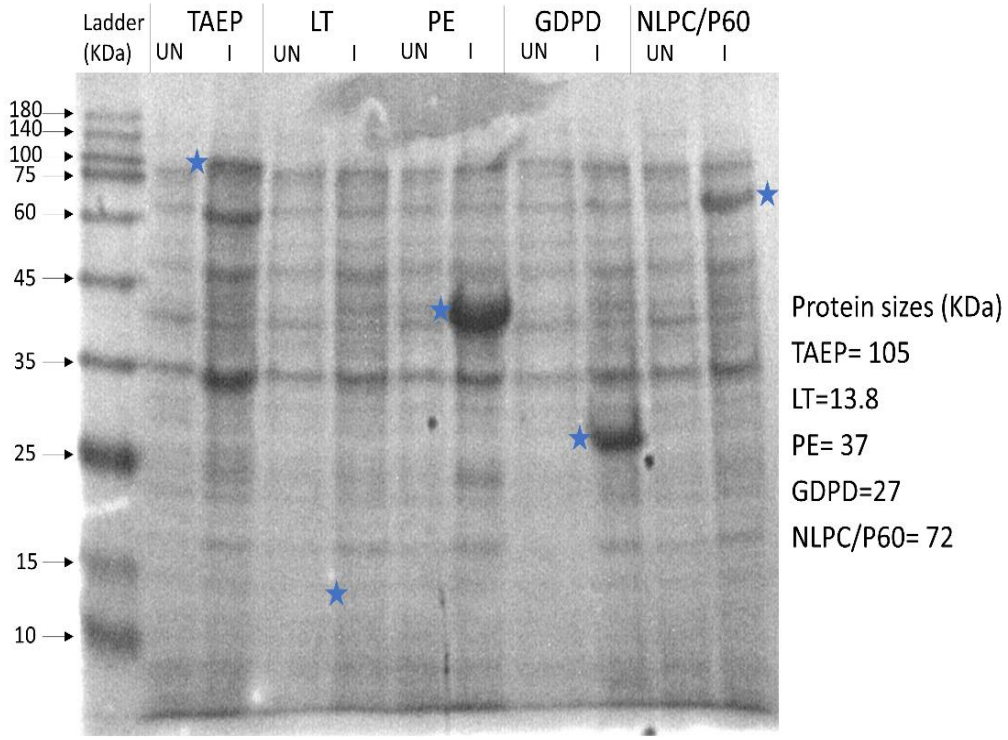


Figure 5.11 His-tag Protein overexpression using C41 cells. Tested proteins (TAEP, LT, PE GDPD and NLPC/P60) were checked from uninduced (UN) and induced (I) samples. The induction temperature was 30 °C. 12% polyacrylamide gel was used to run the SDS-PAGE. The stars indicate the predicted protein sizes.

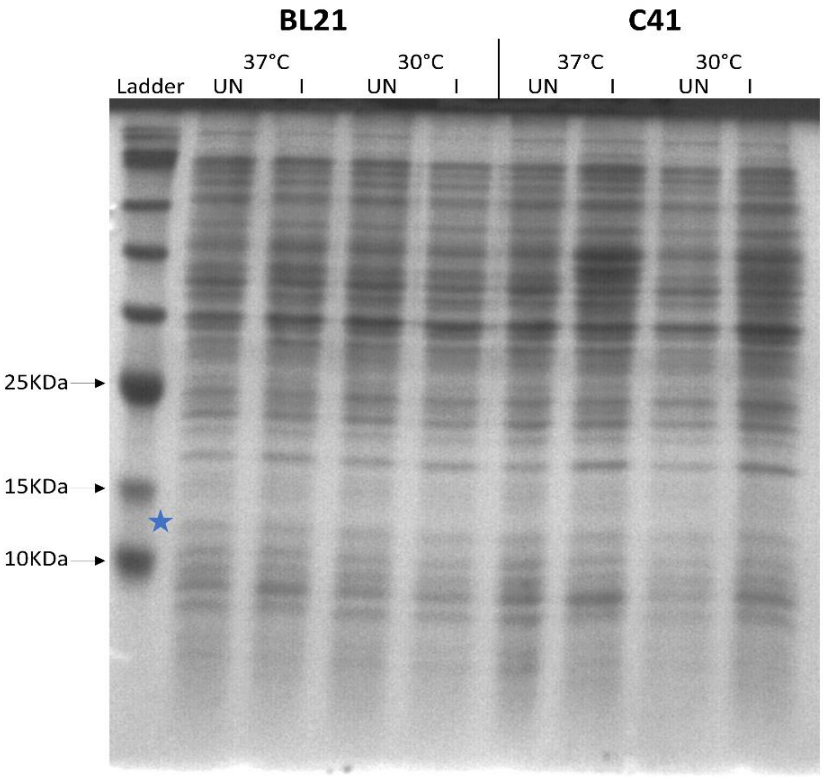


Figure 5.12 LT protein overexpression using BL21(λDE3) and C41(λDE3) cells. The protein was tested under two temperatures (30 °C and 37 °C) and both uninduced (UN) and induced (I) samples were run on a 15% polyacrylamide gel was used to run the SDS-PAGE. The star indicates the predicted protein size.

5.8 Protein solubility

The protein expression showed several promising bands at the expected protein size for TAEP, PE, GDPD and NLPC/P60. To test protein solubility, C41(λDE3) cells containing recombinant vectors were first pelleted after the protein induction was complete at 30 °C for 12 h. The bacterial cells were then resuspended in a sodium phosphate buffer and sonicated to lyse the

cells. This was followed by a quick centrifugation (1000 xg, 1 min) to pellet cell debris "CD" and the supernatant was further centrifuged at high speed (13,000 xg, 10 min) to get the insoluble fraction "IS" in the pellet and the soluble fraction "S" in the supernatant. All three samples (CD, IS and S) were treated with SDS loading dye and ran on polyacrylamide gels. After staining, the gels showed different solubility results (Figure 5.13). For PE, heavy bands were shown in all CD, IS and S lanes at the protein expected size (37 kDa). The GDPD sample showed heavy bands at the expected size (27 KDa) in the CD and IS lanes while no band was seen in the S lane indicating that this protein is insoluble. The NLPC/P60 showed a band in the soluble lane at the expected protein size (72 KDa) although it is less heavy than the bands in both CD and IS lanes. For TAEP, no heavy bands were shown in the three lanes (CD, IS and S) at the protein expected size (105 kDa). This can be explained as the protein expression result of TAEP was misinterpreted and this solubility test result confirms the lack of expression at the conditions applied (temperature and time). As a result of this protein solubility experiment, both PE and NLPC/P60 were selected for protein purification.

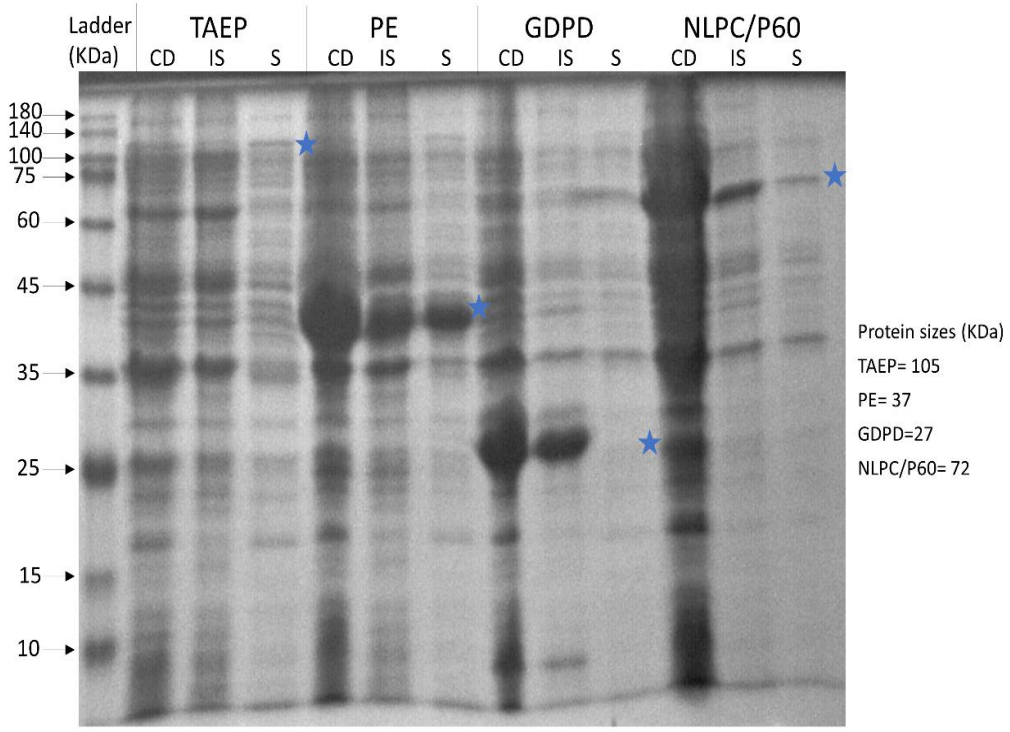


Figure 5.13 Protein solubility tests of TAEP, PE GDPD and NLPC/P60 proteins. Three samples were run for each protein: cell debris (CD), insoluble (IS) and soluble (S). The stars indicate the predicated protein sizes.

5.9 Protein purification and dialysis

Upon successful protein expression and solubility, both PE and NLPC/P60 were further selected for protein purification. To assess this, a large protein lysate (500 mL) was prepared and the protein expression conditions were changed to 25 °C for 16 h to likely enhance protein solubility especially for the NLPC/P60 protein as the solubility test showed a faint band at the soluble fraction.

Both the PE and NLPC/P60 proteins are His-tagged as they were cloned in the pET21b and pET21a vectors, respectively. To purify the proteins, a nickel resin was added to the soluble fraction of the samples to allow the binding between the Histidine residues (His-tag) and the nickel resin. Untagged proteins were removed through a washing step. For protein elution, a sodium phosphate buffer with 500 mM imidazole (competitive binding molecule) was applied which allowed the protein of interest (PE or NLPC/P60) to be eluted and collected.

Following the elution step, the PE protein samples (uninduced, induced, lysate, flow-through-washing, and elution) were prepared with SDS loading dye and loaded into an SDS-PA gel. The gel showed heavy bands for PE protein in all the elution lanes (E1-4) and less intense bands in the wash lane indicating successful protein elution (Figure 5.14 A). Due to a practical error, the ladder and uninduced lanes got contaminated with either induced or lysate samples which can be seen in the gel (Figure 5.14 A). The gel was repeated with only the elution sample (less volume loaded) to better analyse the bands. This showed heavy bands at a protein size between 43-55 KDa which is slightly higher than the expected PE protein size (37K Da) (Figure 5.15 A).

For the NLPC/P60 protein, the gel showed heavy bands in all elution lanes at a protein size between 72 and 95 KDa which is also slightly higher than the expected protein size (72 KDa) (Figure 5.14 B). The NLPC/P60 gel had the same practical error as the PE gel (they were run together) and that is why some lanes (ladder and uninduced) in the NLPC/P60 gel also looks contaminated. To make the elution bands clearer, another gel was run with only the elution fractions (less volume loaded) which shows heavy bands in the E1-3 (Figure 5.15 A).

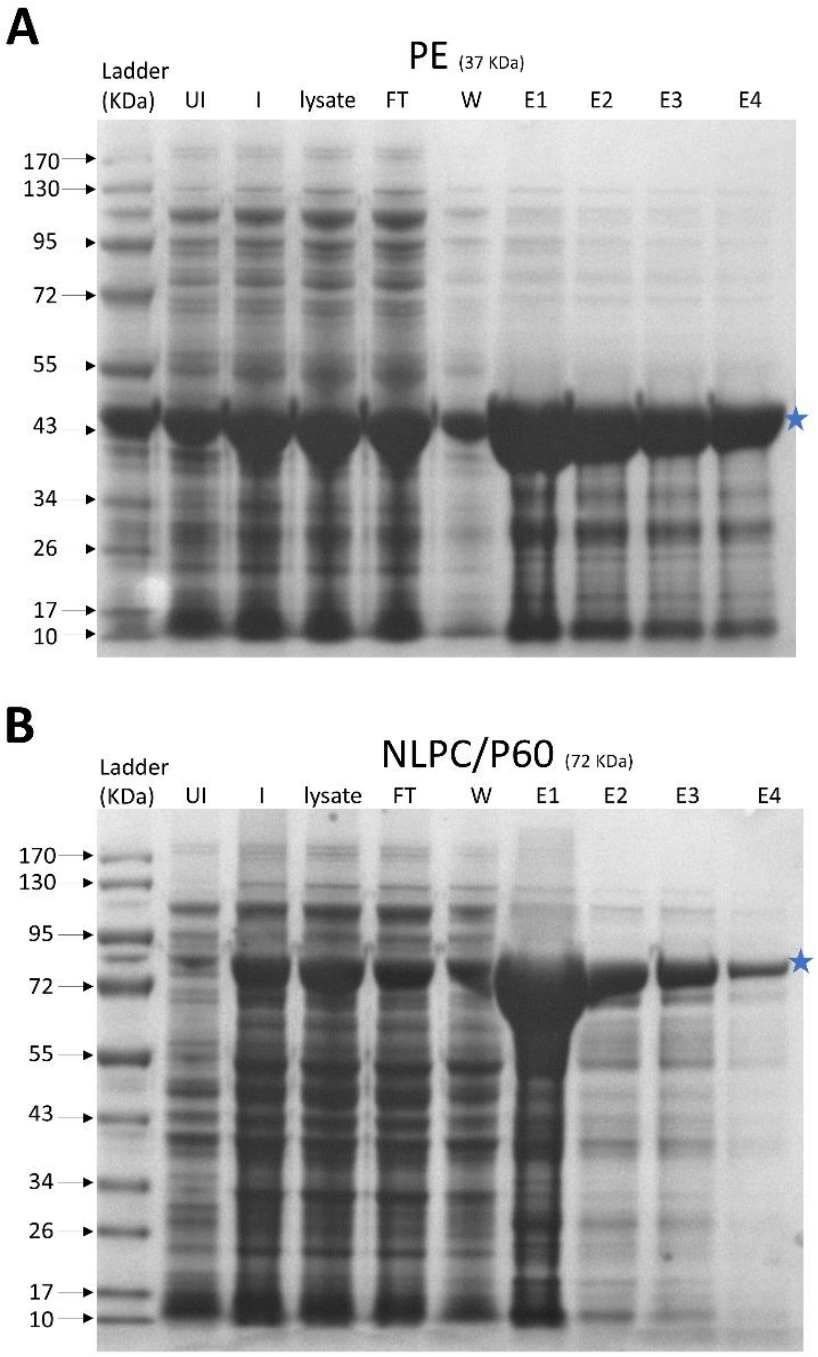


Figure 5.14 Purification of A) PE and B) NLPC/P60 proteins. The samples ran on the gel were as follow: Uninduced "UN", induced "I", lysate "soluble fraction", Flow-through "FT, Wash "W", Elution"E1-4". The stars indicate the predicated protein sizes.

Upon the successful purification of both PE and NLPC/P60 proteins, a dialysis experiment was performed to remove imidazole. The dialysed PE and NLPC/P60 samples were then checked on a gel showing clear protein bands like the elution samples (Figure 15 B).

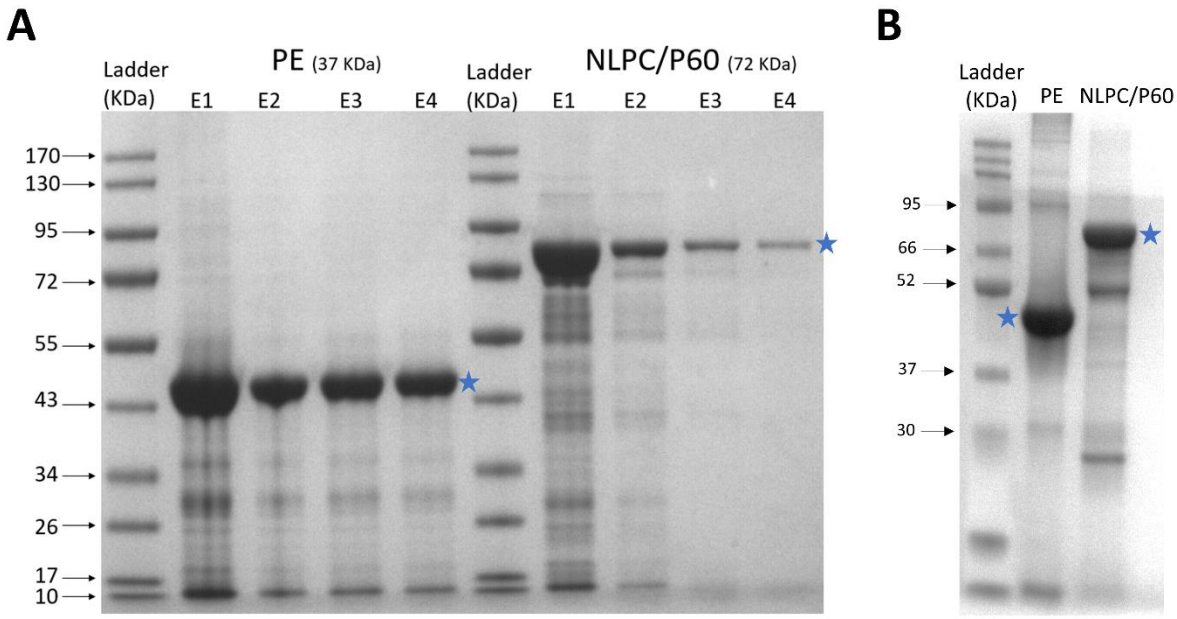


Figure 5.15 A) PE and NLPC/P60 elution samples (E1-4). B) PE and NLPC/P60 samples after protein dialysis. The stars indicate the predicated protein sizes.

Upon dialysis, the concentration of the purified proteins was determined via BCA assays which showed that the PE was measured to be 3.7 mg/ml while the NLPC/P60 was 4.8 mg/ml.

5.10 Antibacterial assessment of His-tagged purified proteins

Following purification, the purified NLPC/P60 and PE proteins were then assessed for lytic activities using spot testing. 3 µl of undiluted filtered proteins were separately spotted on the *E. faecium* E1071 lawn. Next day, no effect (no lysis) on the bacterial lawn was observed. Additionally, the NLPC/P60 protein is being further analysed by our collaborator Dr Stephane

Mesnage (school of biosciences, University of Sheffield) using extracted peptidoglycan material from the E1071 strain. So far, the work has revealed that the NLPC/P60 protein has two enzymatic activities: acting as an endopeptidase and glucosyl hydrolase (glucosaminidase) based on LC-MS/MS analysis. Structural predictions (done by our collaborators Dr John B Rafferty and Mr Maz Robertson, school of biosciences, university of Sheffield) also revealed predicted endopeptidase and glucosyl hydrolase domains using AlphaFold modelling.

5.11 Expression of GST-tagged proteins

As some proteins (TAEP and LT) showed no successful expression, a different protein tag "Glutathione S-transferase" (GST) was then assessed. The GDPD protein was also tested as its expression resulted in insoluble protein. GST-tag fusion is commonly used in protein expression and purification assays (Walls & Loughran, 2011). To enable this, the pGEX4T3 vector was used which contains a GST protein and an ampicillin resistance gene (Figure 5.16 A). The coding sequences of TAEP, LT and GDPD were cloned into pGEX4T3 via gene synthesis. The restriction sites *BamHI* and *XhoI* were used for LT and GDPD proteins while *TspMI* and *XhoI* were for TAEP (Figure 5.16 B).

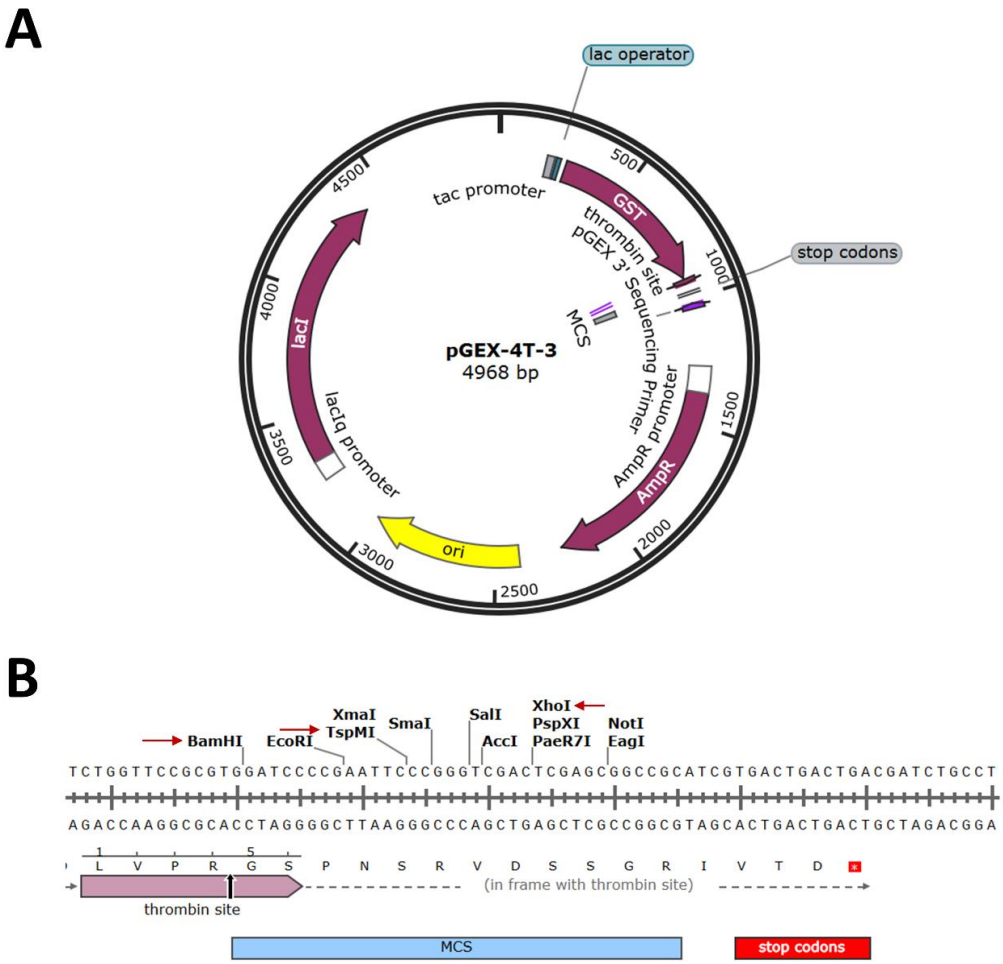


Figure 5.16 pGEX4T3 vector map and cloning details. A) the details of the pGEX4T3 vector which shows Ampicillin resistant gene (AmpR), multiple cloning site (MCS) and origin of replication (ori). B) the restriction sites chosen for cloning are labelled (red arrows).

For TAEP, the initial testing of TAEP subcloning into the pGEX4T3 vector (using SnapGene software) did not show in-frame translation with the GST protein. To solve this, two base pairs namely Cytosine (C) and Thymine (T) were added to the beginning of the TAEP sequence resulting in adding an alanine (Ala) amino acid (Figure 5.17 C). This resulted in having the TAEP in-frame with the GST protein.

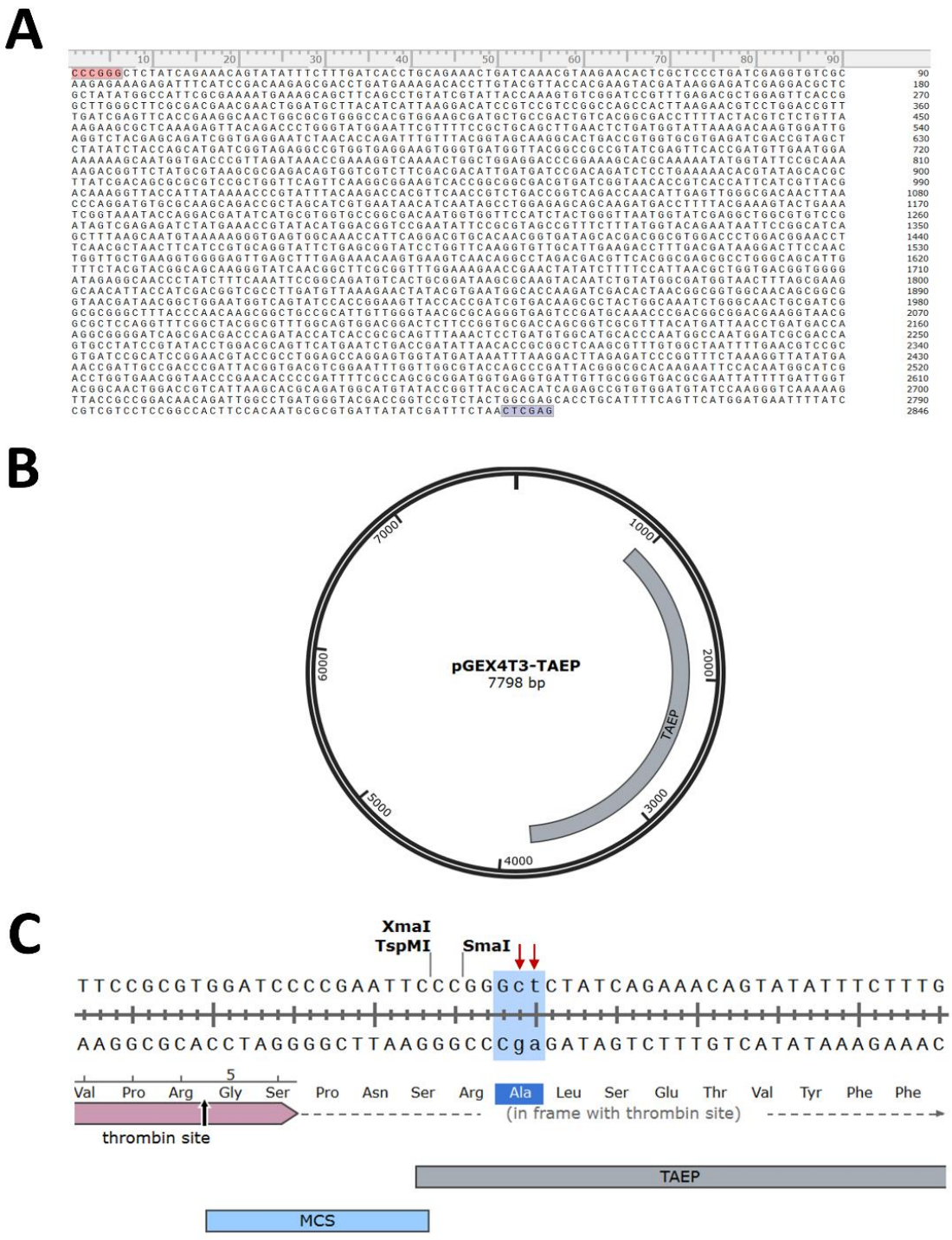
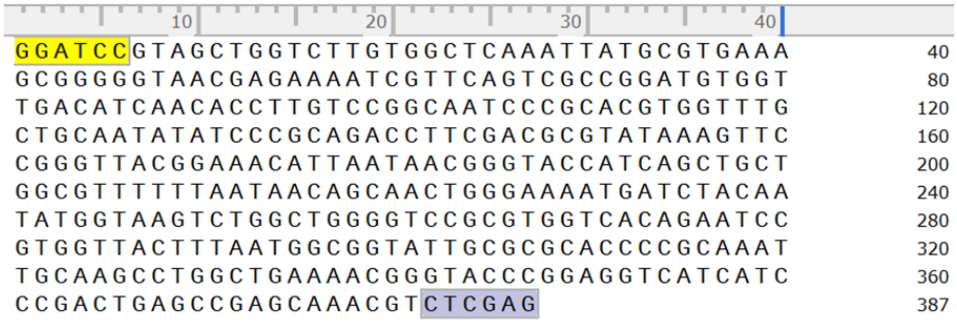


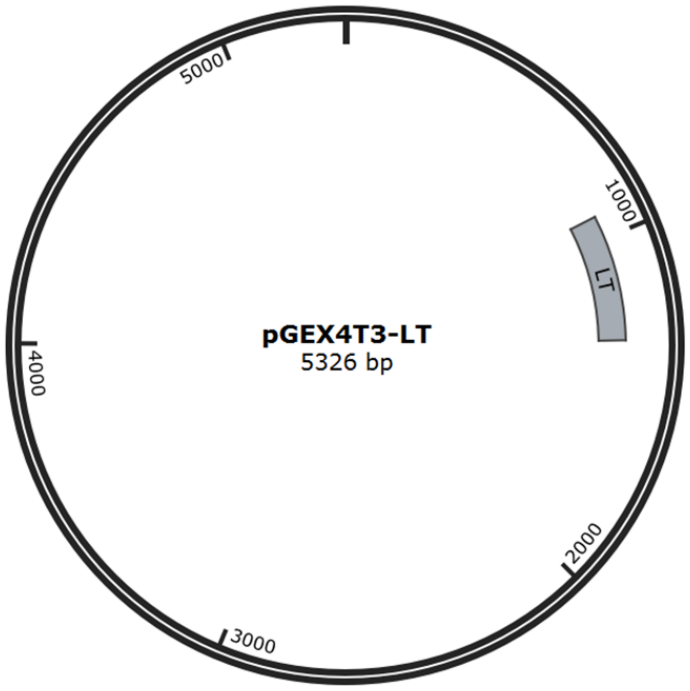
Figure 5.17 Cloning of TAEP into the pGEX4T3 vector. A) the TAEP sequence after adding the restriction sites *TspMI* (Red) and *XhoI* (Blue). B) subcloning of the TAEP sequence into the vector. C) the TAEP sequence after adding cytosine (C) and thymine (T) (red arrows) which results in introducing an alanine (Ala) amino acid.

For GDPD and LT, both showed in-frame translation with GST which needed no further modification (Figure 5.18). The cloning of the three proteins (TAEP, LT and GDPD) was done via gene synthesis by GENEWIZ.

A



B



C

```
GGATCCGTAGAAAATTTAAGGCTATCAGCTCACCGCGGTGCGCACAACTT   50
CGCTCCGGAGAACACCATTGAGGCGTATAAGATCGCCATCGACCTGAAAT   100
ACGGCGCGATCGAACTGGACCCGCGTGCTTCTAGCGATGGTGAACTGTTT   150
ATTATGCATGATGATACCGTTGATAGAACAACCAACGGTAACGGCTACAT   200
TGCGAATATGGATAGCGAACAGATTCGTCAGCTGGAGATCGACACCTCCA   250
ATTATCCGGAATATAAGAATAAAATCCTGCGTGTTCCGACCTTCGAAGAG   300
AGCGTCAAGATCATCTCGACGGGTAATATCATCCTAAACGTGGACGGCAG   350
CAAAATCGACTTCAGCAATACTGCCATTACCAAAAAGATGATTAACATCC   400
TTAAGAAGTACGAGATTTATCAAAACACGTTCTTTGTTATTAGCAACACG   450
TCCCAGCGTTATGCGTTTAACCAGTCATACCCGGATGCAACCCTGTCTTG   500
GTTGTTGACCGACTCGTCCATGATTGATAATGCAATTACCGAAGTGAAAA   550
GCTATCAAAAAGCGCTGCTGAGCATTCCGCTGAACATCGTGACCGACGAC   600
ATTTTGGAGAAATTGCGCAACACGAACATCTACTACCAAATTTACAATGT   650
AAATACCAAGACCGATCTGAACCACCTGCTGATTAAAAAGACTCCGATGA   700
TCGAGACAGACATCTTACTCCCATAACTCGAG                     732
```

D

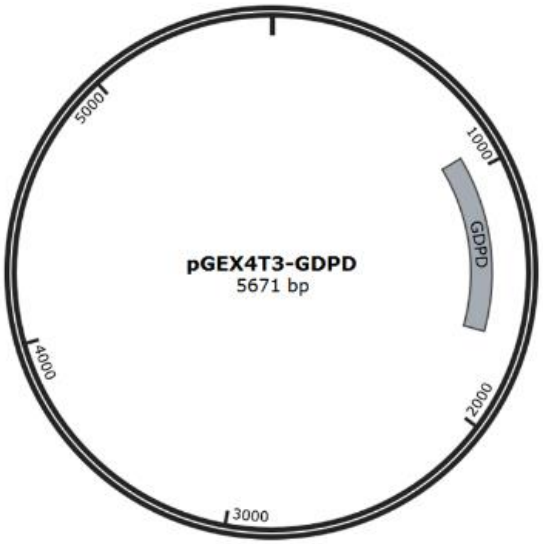


Figure 5.18 Cloning of LT and GDPD into the pGEX4T3 vector. A) the LT sequence after adding the restriction sites *BamHI* (yellow) and *XhoI* (Blue). B) subcloning of LT sequence into the vector. C) the GDPD sequence after adding the restriction sites *BamHI* (yellow) and *XhoI* (Blue). D) subcloning of GDPD sequence into the vector.

5.12 Protein expression and solubility tests

As with the pET21 vectors, BL21(λDE3) and C41(λDE3) cells were transformed with the pGEX4T3 vectors containing TAEP, LT or GDPD sequences. The protein expression was done at 25 °C for 16 h and both uninduced and induced samples (at 0.25 mM IPTG) were run on a gel to assess protein expression. Of note, the protein size of GST is 26 KDa which was added to the size of the protein of interest upon expression Table 5.1. The TAEP protein showed no overexpression at the expected protein band size of 131 KDa in both cells (Figure 5.19 A). In contrast, thick bands at the expected protein size (53 KDa) indicating overexpression were observed for GDPD in both BL21(λDE3) and C41(λDE3) cells (Figure 5.19 A). For LT, C41(λDE3) cells showed overexpression at around the expected band size (39.8 KDa) while less overexpression is seen with BL21(λDE3) (Figure 5.19 B).

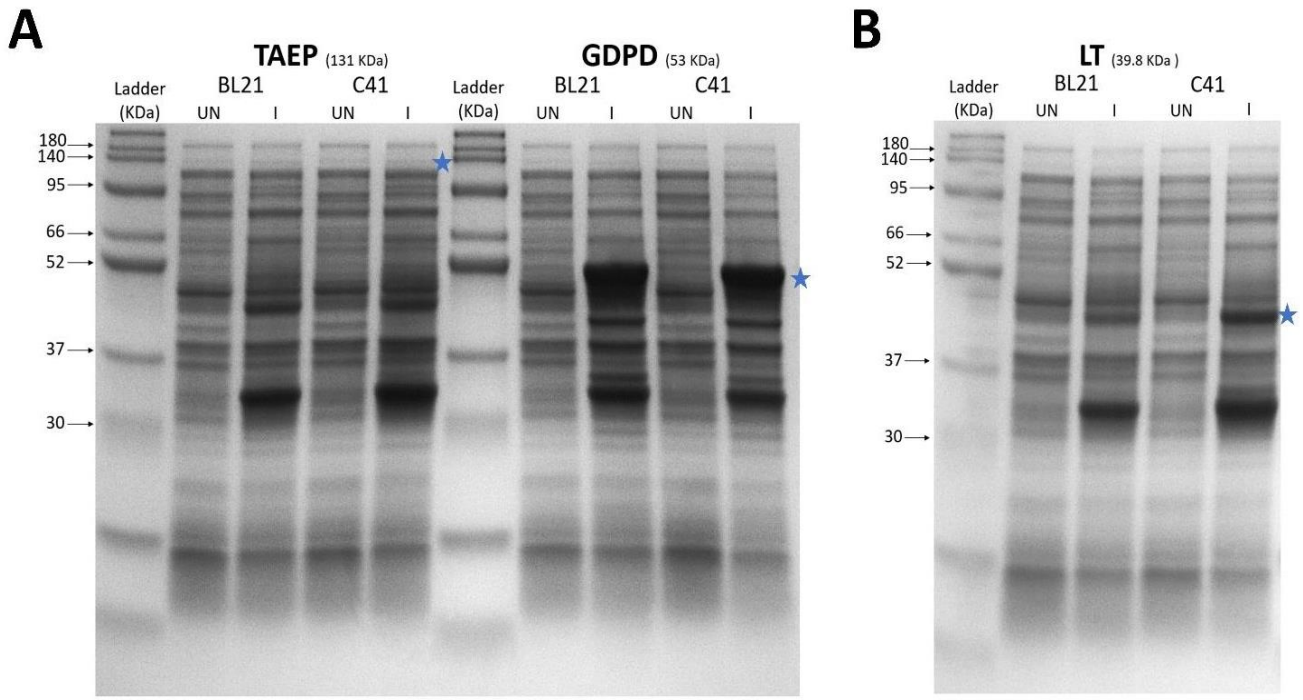


Figure 5.19 TAEP, GDPD and LT protein overexpression using BL21 and C41 cells. A) TAEP and GDPD uninduced (UN) and induced (I) samples were run on a 12% polyacrylamide gel. B) the LT uninduced (UN) and induced (I) samples are shown.

Given the results of the protein overexpression, GDPD and LT were further assessed for protein solubility. For this, GDPD from both BL21(λDE3) and C41(λDE3) cells and LT from C41(λDE3) cells were tested which insoluble and soluble samples were run on a gel. This resulted in very faint bands at the expected size (53 KDa) in the soluble fraction for the GDPD in both cells. For LT, no visible bands were seen at the expected band size (39.8 KDa) in the soluble fraction indicating that LT is in an insoluble form. No further work was done with the insoluble protein due to time limits.

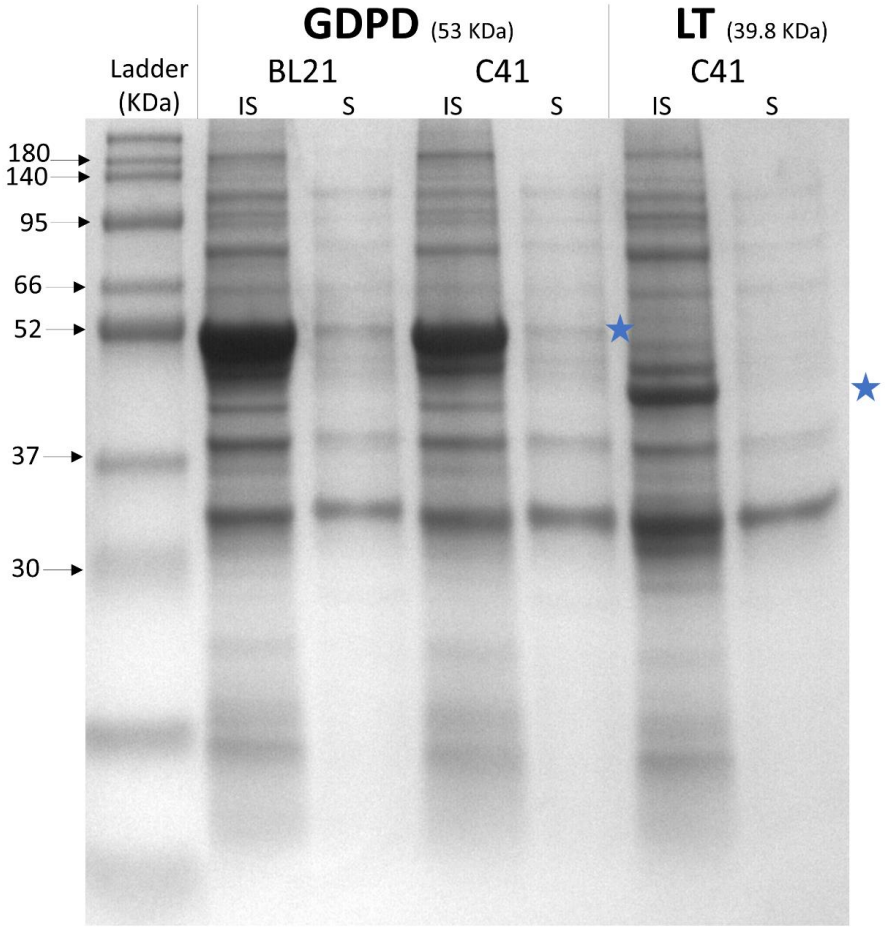


Figure 5.20 Solubility assays of the GDPD and LT proteins. Insoluble (IS) and soluble (S) fractions were run on a 12% polyacrylamide gel. BL21(λDE3) and C41(λDE3) cells are indicated at the top of the gel.

5.13 Discussion

The treatment of bacterial infections needs further investigation and development as antibiotics have become ineffective against various bacterial strains. Various antibacterial approaches have been developed to target antibiotic-resistant pathogens. One way that has shown promising results is to exploit phages and their lytic proteins. In the lytic phage life-cycle, phages lyse their host cells from inside using lytic proteins to allow virions release. Phages also utilise lytic-type proteins to locally degrade bacterial layers to accomplish the initial steps in the phage life cycle. These latter lytic proteins are mainly associated with the phage tail structure and therefore commonly known as tail-associated lysins TALs.

As whole phage particles can be used in targeting and lysing bacteria, purified phage lysins have also shown a potent effect on various bacterial species (Pertics et al., 2021; Rodríguez et al., 2011; Shahed-Al-Mahmud et al., 2021). Among phage lysins, endolysins are widely investigated and tested on different bacterial species. Endolysins facilitate phage release from their host by targeting the bacterial cell wall. As opposed to endolysins, TALs are involved in facilitating phage to begin their life cycle. These TALs are essential for phages to effectively adsorb on their hosts as well as to accomplish phage genome ejection. To our knowledge, TALs in enterococcal phage and prophage genomes have not been investigated.

Therefore, an investigation was carried out about identifying and analysing TALs in 506 enterococcal phage and prophage genomes which was discussed in chapter 3 (Alrafaie & Stafford, 2022). Several predicted lytic proteins were identified such as endopeptidase, lytic transglycosylases, pectinesterases and GDPD. Thus, one candidate sequence from each type was selected for protein expression and purification. Additionally, the analysis of the phiSHEF14 genome in chapter 4 has shown a predicted lytic protein harbouring an NLPC/P60 domain (phiSHEF14_11). Therefore, this protein was also selected to test its expression and purification. Of note, phiSHEF14 is also being further investigated by our collaborators especially to analyse its structure and associated proteins.

To select proteins for expression, a web server (protein-sol) that predicts protein solubility was used to increase the likelihood of expressing soluble proteins. All the proteins tested using gene synthesis showed predicted soluble status based on the web server. This web server has been used in previous publications (Benítez-Chao et al., 2022; Hernández-Ramírez et al., 2020). As opposed to the protein solubility prediction, GDPD showed an insoluble form while the NLPC/P60 was a soluble protein upon protein expression. The PE expression showed soluble protein which coincided with the protein solubility prediction. This indicates the inconsistency between the protein solubility prediction and upon expression. Following this, a codon optimisation step was performed to facilitate protein translation and expression in the *E.coli* cells. Codon optimisation is a critical step as this could determinately affect protein expression (Elena et al., 2014).

As the type of expression cells is critical in protein expression, two commonly used cells were tested: BL21(λDE3) and C41(λDE3) (Rosano & Ceccarelli, 2014). They both contain a prophage (DE3) that facilitates protein expression upon induction with IPTG. Phage proteins have been expressed and purified using these cells (Legotsky et al., 2014; Li et al., 2021). Moreover, both His and GST tags were also used to assess protein expression and solubility. Additionally, protein expression was tested under different temperature and time conditions (37 °C for 3 h, 30 °C for 12 h and 25 °C for 16 h). This was done to assess the best condition to obtain the highest soluble protein yield.

The selected proteins were tested at the different above-mentioned conditions which showed various results. For TAEP, the protein expression was not successful as no band was seen at the expected protein size. One possible solution for this issue is to use BL21 with *rne* mutation which can cause more mRNA stability and therefore better protein expression (Heyde & Nørholm, 2021). The other tested protein was LT which the His-tagged protein resulted in no expression

using both BL21 and C41 cells. However, a successful expression but only in the insoluble form was observed with a GST tag using both BL21 and C41 cells. This result shows that the GST tag can support protein expression (Walls & Loughran, 2011). For GDPD, a successful protein expression was achieved but as insoluble protein for the His tagged samples. One way to fix this issue is by using denaturant agents for refolding and solubilisation such as urea and guanidinium chloride (Gn-HCl) (Alibolandi & Mirzahoseini, 2011; Burgess, 2009; Leal et al., 2006). Moreover, protein expression at low temperatures can help improve protein solubility and reduce inclusion bodies formation (Rosano & Ceccarelli, 2014). The GST tag of GDPD showed faint bands at the soluble fractions upon using the different cells (BL21 and C41) indicating a better solubility result with the GST tag (Hammarström et al., 2009).

The pectinesterase protein showed a successful expression using the His tag and it was in a soluble form when the solubility test was done. Regarding its size on the gel, the solubility assay showed a protein size between 35 and 45 KDa bands which coincides with the predicted protein size of 37 kDa. However, the protein size on the gel became higher after protein purification and dialysis to be between 43 and 55 KDa and about 47 kDa. This could be due to misfolded protein that has led to higher protein size on the gel. The other successfully expressed protein was the NLPC/P60 protein which was cloned via gene amplification from phiSHEF14. Protein purification and dialysis were accomplished and a thick band around the expected size of 72 KDa was seen. The assessment of the antibacterial effect using spot testing on the E1071 lawn resulted in no effect (no lysis). This could be explained as these tested proteins may require additional proteins to show the lytic effect as they are part of the tail complex structure. Due to time limits, further investigation of the proteins activities was not performed. However, Dr Stephane Mesnage (our collaborator from School of Biosciences) has analysed the effect of the NLPC/P60 purified protein on extracted peptidoglycan and showed both endopeptidase and glucosaminidase activities on peptidoglycan extracted material from the *E. faecium* strain E1071. Further analysis would include testing the effect on planktonic bacteria (different *E. faecalis* and *E. faecium*) using serially diluted protein.

5.14 Conclusion

Using phage-derived proteins can be beneficial to tackle the issue of antibiotic-resistant bacteria. Therefore, we investigated here the expression and purification of five different phage-predicted lytic proteins. Different expression cells BL21(λDE3) and C41(λDE3), temperature (37 °C, 30 °C and 25 °C), and protein tags (His and GST) were assessed. The proteins LT and GDPD were only expressed in the insoluble form under specific conditions while the TAEP protein showed no successful expression However, two proteins namely pectinesterase and NLPC/P60 showed successful expression and purification. Due to time limits, a preliminary analysis regarding the antibacterial effect of the purified proteins was performed. An extensive analysis is being carried out in the lab as well as with our collaborators.

Chapter 6: Final discussion and future work

6.1 Major findings

We have investigated in this work different aspects of bacteriophages ranging from bioinformatic analysis of TAL proteins (Chapter 3), isolation and characterisation of bacteriophages (chapter 4), and lastly TAL protein cloning and expression (chapter 5). All this work has revealed some new findings which are summarized in the following sections.

Chapter 3: The analysis of TAL in enterococcal phage and prophage genomes

- Phages utilise lysins associated with tail proteins in order to effectively infect their host (Latka et al., 2017). To our knowledge, the TALs from enterococcal phages have not yet been analysed in any previous publications. Therefore, a total of 506 enterococcal phage and prophage genomes were obtained to allow genomes scanning of tail modules. Prophage genomes were identified using the PHASTER web server. The analysed phage genomes in this study have shown a correlation between size and phage morphology and current ICTV genomic classification: podoviruses (Rountreeviridae) are phages with genomes under 30.5 kb, siphoviruses (Efquatrovirus, Phifelvirus, Saphexavirus, and Andrewesvirinae) with genomes between 31 and 86.3 kb, and myoviruses with genomes exceeding 130 kb (Herelleviridae). The analysis of phage lifecycle showed that 86 genomes (of 100) have strict lytic lifecycle as no lysogenic genes such as integrase or repressor genes were identified. For prophages, 406 genomes were obtained and scanned after reannotation using RASTtk. The size of most analysed prophage genomes was between 20-60 kb.

- As phage genomes are organised into modules (Moura de Sousa et al., 2021), the tail module was identified and scanned for TAL in both enterococcal phage and prophage genomes. This has revealed various lysins with different abundance: endopeptidase-

containing proteins (70.4%), lytic transglycosylase proteins (18.0%), NLPC/P60 family (6.2%), GDPD (4.0%), and lastly pectinesterases (1.3%). For the endopeptidase-containing proteins (TAEP), these are the most abundant lytic type and are found in five different domain architectures which some contain additional lytic domains such as lysozyme and amidase domains. For NLPC/P60 proteins, these were found in myo- and podoviruses and possess various domain architectures. The identified LT domains were found associated with TMP and belong to the GH23 based on the CAZy database. Aligning the LT domains has revealed some shared motifs with published LT families while other motifs were different. Therefore, a new LT family for phage-infecting Gram-positive bacteria is proposed. The pectinesterase proteins were found in 7 prophage genomes and we hypothesise the EPA surface antigen is their target. For GDPD, these were found in both phage and prophage genomes and target the cell wall teichoic acids (Shi et al., 2008).

- The identified lysins and the size of phage genomes were found to be correlated as myoviruses (largest size, Schiekvirus or Kochikohdavirus) possess TMP-LT and NLPC/P60 lysins while podoviruses (smallest size, Copernicusvirus or Minhovirus) contain NLPC/P60 proteins. For siphoviruses, two groups based on size were assigned which Group 1 (21.1-42.8 kb, Efquatroviruses or Phifelviruses) has TAEP and TMP-LT lysins whereas Group 2 (55-86 kb, Saphexavirus or Andrewesvirinae) contains only the TMP-LT lysin. For prophages, the tail module proteins resemble the siphovirus tail module organisation (i.e. TMP-Dit-Tal) as well as the genome sizes were between 20-60 Kb which is similar to the analysed siphovirus genomes.

Chapter 4: Isolation and characterisation of bacteriophages targeting enterococci

- Wastewater samples were used as an environmental source of phages and the TEM analysis has shown various phage morphologies (myoviruses, siphoviruses and

podoviruses). Phage isolation also resulted in obtaining the three different tail morphologies as phiSHEF8-12 are siphoviruses, phiSHEF13 and 16 are myoviruses while phiSHEF14 is a podovirus upon TEM examination. Host range analysis showed that the phiSHF13 can infect both *E. faecalis* and *E. faecium* strains which some are VRE such as E1071 and V583. The podovirus phiSHEF14 has a very narrow host range as it only infects the isolation host E1071 after testing 36 strains. Genomic analysis of the isolated phages has revealed 5 novel phages (phiSHEF10,11,13,14 and 16) with strict lytic lifecycles as they lack lysogenic genes. In addition, these novel phages have been shown to harbour tail-associated lysins such as endopeptidase activities (phiSHEF10_17 and phiSHEF11_46), lytic transglycosylases (phishef13_51 and phishef16_2) and NLPC/P60 (phiSHEF13_50, phiSHEF16_1 and phiSHEF14_11). Upon aligning the phiSHEF13,14 and 16 genomes which all infect the E1071, a shared CDS has been found (phiSHEF13_41, phiSHEF14_15 and phiSHEF16_201) which has a predicted sugar-binding ability based on Phyre2 analysis.

- Further characterisation was carried out focusing on phages infecting the *E. faecium* E1071 namely phiSHEF13,14 and 16. The assessment of the killing kinetics of planktonic bacteria has revealed quick inhibition of bacterial growth which the highest MOI showing the fastest inhibition and vice versa. A phage cocktail of the three phages has also shown a similar result of bacterial inhibition. The development of phage-resistant mutants (RM) was also observed during the killing assays in all phages as well as the cocktail. Therefore, an investigation of these resistant clones was carried out and 20 clones from each phage (60 clones in total) were selected and assessed for phage infection. This revealed that the phiSHEF13 RMs were still susceptible to phiSHEF14 and 16 while resistant to phiSHEF13. For phiSHEF14 RMs, all clones were susceptible to phiSHEF13 and resistant to phiSHEF14 whereas a mixed result (susceptible and resistant) for phiSHEF16. For phiSHEF16 RMs, all tested clones were resistant to phiSHEF14 and 16 while susceptible and resistant results were observed with phiSHEF13.

- The phage host receptor was also investigated with phiSHEF13 using the *E. faecalis* OG1RF and V583 wild-type and mutant strains. EPA is an essential antigen within the enterococcal cell envelope and it is also a hotspot for phage adsorption and infection (Chatterjee et al., 2019; Ho et al., 2018). The variable region of EPA (epaV) from the V583 was tested for its role in phiSHEF13 infectivity. Of note, the wild-type OG1RF is resistant to phiSHEF13 infection while V583 is susceptible. Upon analysis, the V583 wild type and complemented strains were susceptible to phiSHEF13 while the ΔepaV mutant showed resistance. For OG1RF, the wild type was resistant to infection while the complemented strain (with V583 epaV) showed a susceptible result. Collectively, this indicates that the V583 epaV region is important for phiSHEF13 successful infection.

- Phage adaptation or "training" was also assessed as phiSHEF14 showed a very narrow host range infecting only the isolation host *E. faecium* E1071. Three strains were used in this experiment: *E. faecalis* V583, *E. faecium* dp9 and the resistant mutant 14RM0. This was done via serial passages of phage for up to 7 days and the *E. faecium* E1071 strain was added to ensure high phage titer throughout the experiment. Ultimately, plaque assays showed no isolation of mutant phiSHEF14 that can infect the tested hosts.

Chapter 5: Cloning and expression of TAL proteins

- After the *in silico* analysis of TAL proteins was done in chapter 3, an investigation of protein expression in the lab was carried out. Candidate proteins were selected to cover all identified types of lysins: TAEP, NLPC/P60, LT, pectinesterase and GDPD. Cloning by gene synthesis was done for all proteins except NLPC/P60 as it was amplified from the phiSHEF14 genome. Codon optimisation was performed to increase the likelihood of successful protein expression. The *E.coli* expression cells BL21(λDE3) and C41(λDE3) were used as well as the His- and GST tags.

- After experiments were carried out, TAEP has shown no expression upon testing the abovementioned different conditions. For GDPD, a successful expression was observed but only as an insoluble protein. This was also the case for LT which the GST-tagged protein only showed a band in the insoluble fraction. For pectinesterase and NLPC/P60, both were successfully expressed and purified. The expression cell used was C41(λDE3) while His-tag was for purification. The work of our collaborator (Dr Stephane Mesnage) showed a dual activity of the NLPC/P60 protein: an endopeptidase and glucosyl hydrolase (glucosaminidase).

6.2 Future work and limitations

Chapter 3: The analysis of TAL in enterococcal phage and prophage genomes

- Our analysis of the TAL involved a total of 100 enterococcal phage genomes which were obtained from the NCBI database. To date, there are over 160 enterococcal phages deposited at the NCBI virus portal. Therefore, our investigation here focused on about 62% of the available enterococcal phages and future work may involve covering the unanalysed phages. For prophages, we focused only on intact prophages as their genomes are highly likely to contain intact phage modules. Future work can involve assessing incomplete prophage genomes for TAL if the tail module is still intact.

- Phages mainly utilise lysins associated with tail proteins (Latka et al., 2017) and therefore our analysis was focused on these TAL proteins. However, lytic proteins can also be found in other phage parts such as capsid (Moak & Molineux, 2004) or neck (Gutiérrez et

al., 2015) which investigating these regions could also result in identifying novel lytic proteins or domains.

- The data in this study can be exploited in phage engineering as phages with various TAL could broaden phage host specificity by attacking different bacterial targets. Moreover, the different identified lytic domains can also be used in a shuffling and fusion approach to generate chimeric lytic enzymes (São-José, 2018).

Chapter 4: Isolation and characterisation of bacteriophages targeting enterococci

- Bacteriophages have been considered a promising therapy to tackle the issue of antibiotic resistance (Ghosh et al., 2019). One way to achieve this is by isolating potent phages from the environment targeting resistant bacteria. In our work here, five novel phages targeting *E. faecalis* and *E. faecium* strains were isolated. The isolation methodology that was used involved single and multiple hosts. Multiple hosts isolation technique increases the likelihood of phage isolation (Hyman, 2019) which testing the antagonistic effect between hosts is needed to ensure bacterial availability. In addition, our *E. faecium* strains collection was based on the EPA different variants (De Been et al., 2013). Some strains in variant 1 (E1636 and E1679) and 2 (E1071 and E4452) showed susceptible results to some isolated phages namely phiSHEF13,14 and 16. Future work can involve phage isolation focusing on strains in variants 3 and 4 to potentially enhance phage isolation diversity.

- The five novel phages isolated in this work were characterised via phage and plaque morphologies, host range and genomic analysis. Although these data are very informative, further experiments on these phages would also facilitate better understanding. One important experiment to be done is the one-step growth which can explain phage kinetics especially phage burst size, eclipse and latent periods. Another

assay that can also be done is the adsorption assay which entails determining the time needed for phage to adsorb on their host cells.

- Our isolated phages showed a varied host range result targeting different *E. faecalis* and *E. faecium* strains. phiSHEF18-12 exclusively infect *E. faecalis* while phiSHEF14 and 16 target *E. faecium*. PhiSHEF13 has a broad host range covering 11 *E. faecalis* and *E. faecium* strains. Therefore, this data is encouraging to design an efficient phage cocktail to tackle both *E. faecalis* and *E. faecium* strains. Moreover, host range results also showed no activity on some enterococcal strains using the isolated phages. Further phage isolation could be done by targeting the still resistant strains to broaden the enterococcal phage library and eventually design a better phage cocktail to cover more hosts. Additionally, the epaV region of V583 showed its essential involvement in phiSHEF13 successful infection which further investigation could involve specifically determining the involved gene/s using gene knockout or silencing.

- The activity of the phiSHEF13, 14 and 16 was assessed on planktonic bacteria using killing assays. Although quick inhibition was observed, assessing phage activities on bacterial biofilms is important to further confirm phage potency. For this, future work may involve assessing phage activities (individual or cocktail) on susceptible strains such as the VRE V583 for phiSHEF13 or E1071 for phage cocktail. *Ex vivo* studies can also be assessed using 3D human skin models and a good bacterial candidate would be the VRE clinical isolate dp9 (isolated from a patient with a diabetic foot ulcer) which is susceptible to phiSHEF13.

- Phage-resistant mutants (RM) were also isolated and investigated in this work. The assessment of RM resistance or susceptibility to phages was done using killing assays. A period of 2.5 hours was set to assess the effect of phages on bacterial clones. This timing

was selected as phiSHEF13,14 and 16 have shown quick bacterial inhibition (less than 2 hours) using the E1071 strain. Future work can involve longer incubation time (20 h) to better assess the development of phage RM and the effect on growth rate. In this vein, 10 phage RM clones were selected and sent for sequencing to be compared with the wild type (E1071) to identify any genomic variations. The data for this are being analysed at the moment and are not included here due to time limits. Furthermore, the cost of fitness concept can also be investigated with these phage RM (Mangalea & Duerkop, 2020). Antibiotic susceptibility tests can be performed to assess changes in MIC (minimum inhibitory concentration). As E1071 is a VRE strain, vancomycin as well as other antibiotics can be tested (Arias & Murray, 2012). This antibiotic susceptibility testing is currently being performed in the lab by one of the students (Ms Elspeth Smith) and the preliminary data supports the antibiotic re-sensitisation phenomenon.

- The phage adaptation methodology did not result in the successful isolation of broadened phage strains (phiSHEF14). Therefore, a further modified approach involving a cocktail of phages instead of a single phage could be assessed (Appelman's approach) (Burrowes et al., 2019). The phage cocktail approach provides an opportunity for phage genetic recombination upon co-infection in addition to the spontaneous mutation which both could generate new phage strains with broad host range features (Burrowes et al., 2019).

<u>Chapter 5: Cloning and expression of TAL proteins</u>

- Protein expression has been carried out in *E.coli* hosts especially the strains BL21(λDE3) and C41(λDE3) (Rosano & Ceccarelli, 2014). The candidate proteins selected for protein expressions showed varied results. TAEP showed no expression when the cells, tags and incubation times were applied. Future work may involve attempting a different TAEP or using different expression cells such as BL21(DE3) pLysS Competent Cells. These cells contain a plasmid that encodes for T7 lysozyme which lowers the basal expression of the

gene of interest if protein toxicity is the issue (Rosano & Ceccarelli, 2014). Additionally, using BL21 with the *rne* mutation would increase mRNA stability and consequently protein expression (Heyde & Nørholm, 2021). The other assessed proteins (GDPD and LT) showed insoluble protein expression. This result can be improved to solubilise the proteins by using urea or guanidinium chloride for protein refolding (Alibolandi & Mirzahoseini, 2011; Burgess, 2009; Leal et al., 2006) or by lowering the temperature of protein expression (Bhatwa et al., 2021).

- As both Pectinesterase and NLPC/P60 proteins were successfully expressed and purified, the assessment of their activity can be further investigated. An experiment to be done is the killing assays which decipher the effect of the lytic proteins on the bacterial growth pattern. Additionally, enterococcal phage-derived pectinesterases potentially target EPA while NLPC/P60 attack the peptidoglycan layer (Anantharaman & Aravind, 2003). Therefore, a combination of both purified proteins can also be assessed as well as in addition to whole phage particles or antibiotics.

6.3 Conclusion

Phages are bacterial viruses that have been exploited for over a century (S. T. Abedon et al., 2011) and their abundance in nature has helped to isolate diverse phage strains. Phage isolation and characterisation are essential steps to broaden our knowledge about these viruses. In addition, targeting highly antibiotic-resistant bacteria is needed to make phage therapy a possible alternative treatment. In this work, five novel phages have been isolated and characterised which some target VRE strains. Moreover, phages require local degradation of the bacterial layer to accomplish the first steps in the phage lifecycle which involve specific lysins mainly named TAL. Therefore, these lytic proteins have been investigated bioinformatically in this work via scanning a total of 506 enterococcal phage and prophage genomes. This has resulted in identifying a panel of lysins that can target bacterial peptidoglycan, teichoic acid and potentially EPA. Finally and to complete the project story, the *in vitro* analysis of protein cloning and expression was carried out focusing on TAL proteins. Of the candidate TAL proteins tested, two proteins (pectinesterase and NLPC/P60) were successfully expressed and purified. Nowadays, phage therapy has gained more recognition worldwide and phage manufacturing is considered to follow the Good Manufacturing Practice (GMP) guidelines to ensure safety and efficacy (Bretaudeau et al., 2020). For the treatment of diabetic foot infections, phage therapy is also considered a promising option and more assessment of phage efficiency on single/multispecies biofilms are needed (Pouget et al., 2021). The overall data in this work adds to the knowledge about phages and hopefully would be useful for eventually applied to tackle the issue of antibiotic resistance.

Covid-19 impact statement

My PhD study started in March 2019 and the first year was mainly focused on mastering main phage techniques such as plaque assays, enrichment protocols and isolation methods. This took a while as I had no previous experience with phage protocols. At the end of the first year, I was successful in isolating different phages and ready to start phage characterisation. However, the Covid-19 lockdown took place causing lab experiments to be postponed. During that time, we started a bioinformatic project focusing on phage tail-associated lysins. The labs were opened again after about 6 months with limited working hours. This restricted access to labs lasted for about another 4 months before it was relaxed to its current no-restriction access. This lost period has affected my work in doing more in-depth analysis regarding the isolated phages such as the one-step growth experiment and adsorption assay. Additionally, some of the performed experiments could be modified to improve or confirm the results such as with the RM data (chapter 4) and the purified proteins (chapter 5).

References

Abdelrahman, F., Easwaran, M., Daramola, O. I., Ragab, S., Lynch, S., Oduselu, T. J., Khan, F. M., Ayobami, A., Adnan, F., Torrents, E., Sanmukh, S., & El-Shibiny, A. (2021). Phage-encoded endolysins. In *Antibiotics* (Vol. 10, Issue 2, pp. 1–31). https://doi.org/10.3390/antibiotics10020124

Abedon, S. T., Kuhl, S. J., Blasdel, B. G., & Kutter, E. M. (2011). Phage treatment of human infections. *Bacteriophage*, *1*(2), 66–85. https://doi.org/10.4161/bact.1.2.15845

Ackermann, H. W. (2007). 5500 Phages examined in the electron microscope. Archives of Virology, 152(2), 227–243. https://doi.org/10.1007/s00705-006-0849-1

Ackermann, H. W., & Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. Archives of Virology, 157(10), 1843–1849. https://doi.org/10.1007/s00705-012-1383-y

Ackermann, H.-W. (2011). The first phage electron micrographs. Bacteriophage, 1(4), 225–227. https://doi.org/10.4161/bact.1.4.17280

Aggarwala, V., Liang, G., & Bushman, F. D. (2017). Viral communities of the human gut: Metagenomic analysis of composition and dynamics. Mobile DNA, 8(1), 1–10. https://doi.org/10.1186/s13100-017-0095-y

Alibolandi, M., & Mirzahoseini, H. (2011). Chemical assistance in refolding of bacterial inclusion bodies. Biochemistry Research International, 2011. https://doi.org/10.1155/2011/631607

Alič, Š., Naglič, T., Tušek-Žnidarič, M., Ravnikar, M., Rački, N., Peterka, M., & Dreo, T. (2017). Newly isolated bacteriophages from the Podoviridae, Siphoviridae, and Myoviridae families have variable effects on putative novel Dickeya spp. Frontiers in Microbiology, 8(SEP), 1–14. https://doi.org/10.3389/fmicb.2017.01870

Almohamad, S., Somarajan, S. R., Singh, K. V., Nallapareddy, S. R., & Murray, B. E. (2014). Influence of isolate origin and presence of various genes on biofilm formation by Enterococcus faecium. FEMS Microbiology Letters, 353(2), 151–156. https://doi.org/10.1111/1574-6968.12418

Al-Zubidi, M., Widziolek, M., Court, E. K., Gains, A. F., Smith, R. E., Ansbro, K., Alrafaie, A., Evans, C., Murdoch, C., Mesnage, S., Douglas, C. W. I., Rawlinson, A., & Stafford, G. P. (2019). Identification of novel bacteriophages with therapeutic potential that target enterococcus faecalis. Infection and Immunity, 87(11), 1–19. https://doi.org/10.1128/IAI.00512-19

Anantharaman, V., & Aravind, L. (2003). Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biology, 4(2), 1–12. https://doi.org/10.1186/gb-2003-4-2-r11

Arbeloa, A., Hugonnet, J. E., Sentilhes, A. C., Josseaume, N., Dubost, L., Monsempes, C., Blanot, D., Brouard, J. P., & Arthur, M. (2004). Synthesis of mosaic peptidoglycan cross-bridges by hybrid peptidoglycan assembly pathways in gram-positive bacteria. Journal of Biological Chemistry, 279(40), 41546–41556. https://doi.org/10.1074/jbc.M407149200

Archimbaud, C., Shankar, N., Forestier, C., Baghdayan, A., Gilmore, M. S., Charbonné, F., & Joly, B. (2002). In vitro adhesive properties and virulence factors of Enterococcus faecalis strains. Research in Microbiology, 153(2), 75–80. https://doi.org/10.1016/S0923-2508(01)01291-8

Arias, C. A., & Murray, B. E. (2012). The rise of the Enterococcus: Beyond vancomycin resistance. Nature Reviews Microbiology, 10(4), 266–278. https://doi.org/10.1038/nrmicro2761

Armstrong, D. G., Boulton, A. J. M., & Bus, S. A. (2017). Diabetic foot ulcers and their recurrence. New England Journal of Medicine, 376(24), 2367–2375. https://doi.org/10.1056/NEJMra1615439

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Arsène, S., & Leclercq, R. (2007). Role of a qnr-like, gene in the intrinsic resistance of Enterococcus faecalis to fluoroquinolones. Antimicrobial Agents and Chemotherapy, 51(9), 3254–3258. https://doi.org/10.1128/AAC.00274-07

Avrani, S., & Lindell, D. (2015). Convergent evolution toward an improved growth rate and a reduced resistance range in Prochlorococcus strains resistant to phage. Proceedings of the National Academy of Sciences of the United States of America, 112(17), E2191–E2200. https://doi.org/10.1073/pnas.1420347112

Bakhshinejad, B., & Ghiasvand, S. (2017). Bacteriophages in the human gut: Our fellow travelers throughout life and potential biomarkers of heath or disease. In Virus Research (Vol. 240, Issue March, pp. 47–55). Elsevier. https://doi.org/10.1016/j.virusres.2017.07.013

Banla, L. I., Salzman, N. H., & Kristich, C. J. (2019). Colonization of the mammalian intestinal tract by enterococci. *Current Opinion in Microbiology*, *47*, 26–31. https://doi.org/10.1016/j.mib.2018.10.005

Becker, S. C., Dong, S., Baker, J. R., Foster-frey, J., Pritchard, D. G., & Donovan, D. M. (2009). *LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells*. *294*, 52–60. https://doi.org/10.1111/j.1574-6968.2009.01541.x

Belkaid, Y., & Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. Cell, 157(1), 121–141. https://doi.org/10.1016/j.cell.2014.03.011

Bertin, A., de Frutos, M., & Letellier, L. (2011). Bacteriophage-host interactions leading to genome internalization. Current Opinion in Microbiology, 14(4), 492–496. https://doi.org/10.1016/j.mib.2011.07.010 Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. In FEMS Microbiology Letters (Vol. 363, Issue 4, pp. 1–11). https://doi.org/10.1093/femsle/fnw002

Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. In *FEMS Microbiology Letters* (Vol. 363, Issue 4, pp. 1–11). https://doi.org/10.1093/femsle/fnw002

Betts, A., Kaltz, O., & Hochberg, M. E. (2014). Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. Proceedings of the National Academy of Sciences of the United States of America, 111(30), 11109–11114. https://doi.org/10.1073/pnas.1406763111

Betts, A., Vasse, M., Kaltz, O., & Hochberg, M. E. (2013). Back to the future: Evolving bacteriophages to increase their effectiveness against the pathogen Pseudomonas aeruginosa PAO1. Evolutionary Applications, 6(7), 1054–1063. https://doi.org/10.1111/eva.12085

Betz, J. V, & Anderson, K. E. (1964). ISOLATION AND CHARACTERIZATION OF BACTERIOPHAGES ACTIVE ON CLOSTRIDIUM SPOROGENES. Journal of Bacteriology, 87(2), 408–415. http://www.ncbi.nlm.nih.gov/pubmed/14151064

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. Frontiers in Bioengineering and Biotechnology, 9(February), 1–18. https://doi.org/10.3389/fbioe.2021.630551

Bhavsar, A. P., & Brown, E. D. (2006). Cell wall assembly in Bacillus subtilis: How spirals and spaces challenge paradigms. Molecular Microbiology, 60(5), 1077–1090. https://doi.org/10.1111/j.1365-2958.2006.05169.x

Bhaya, D., Davison, M., & Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. Annual Review of Genetics, 45(1), 273–297. https://doi.org/10.1146/annurev-genet-110410-132430

Blackburn, N. T., & Clarke, A. J. (2001). Identification of four families of peptidoglycan lytic transglycosylases. *Journal of Molecular Evolution*, *52*(1), 78–84. https://doi.org/10.1007/s002390010136

Blower, T. R., Chai, R., Przybilski, R., Chindhy, S., Fang, X., Kidman, S. E., Tan, H., Luisi, B. F., Fineran, P. C., & Salmond, G. P. C. (2017). Evolution of Pectobacterium bacteriophage ΦM1 to escape two bifunctional type III toxin-antitoxin and abortive infection systems through mutations in a single viral gene. Applied and Environmental Microbiology, 83(8), 1–17. https://doi.org/10.1128/AEM.03229-16

Blower, T. R., Evans, T. J., Przybilski, R., Fineran, P. C., & Salmond, G. P. C. (2012). Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. PLoS Genetics, 8(10), e1003023. https://doi.org/10.1371/journal.pgen.1003023

Bohannan, B. J. M., & Lenski, R. E. (2000). Linking genetic change to community evolution: Insights from studies of bacteria and bacteriophage. Ecology Letters, 3(4), 362–377. https://doi.org/10.1046/j.1461-0248.2000.00161.x

Bohannan, B. J. M., Kerr, B., Jessup, C. M., Hughes, J. B., & Sandvik, G. (2002). Trade-offs and coexistence in microbial microcosms. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 81(1–4), 107–115. https://doi.org/10.1023/A:1020585711378

Bolocan, A. S., Upadrasta, A., De Almeida Bettio, P. H., Clooney, A. G., Draper, L. A., Ross, R. P., & Hill, C. (2019). Evaluation of phage therapy in the context of enterococcus faecalis and its associated diseases. Viruses, 11(4), 1–18. https://doi.org/10.3390/v11040366

Bondi, M., Laukova, A., De Niederhausern, S., Messi, P., Papadopoulou, C., & Economou, V. (2020). Controversial Aspects Displayed by Enterococci: Probiotics or Pathogens? *BioMed Research International*, *2020*. https://doi.org/10.1155/2020/9816185

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature, 493(7432), 429–432. https://doi.org/10.1038/nature11723

Bonilla, N., Rojas, M. I., Netto Flores Cruz, G., Hung, S.-H., Rohwer, F., & Barr, J. J. (2016). Phage on tap-a quick and efficient protocol for the preparation of bacteriophage laboratory stocks. PeerJ, 4, e2261. https://doi.org/10.7717/peerj.2261

Born, Y., Fieseler, L., Klumpp, J., Eugster, M. R., Zurfluh, K., Duffy, B., & Loessner, M. J. (2014). The tail-associated depolymerase of Erwinia amylovora phage L1 mediates host cell adsorption and enzymatic capsule removal, which can enhance infection by other phage. Environmental Microbiology, 16(7), 2168–2180. https://doi.org/10.1111/1462-2920.12212

Bouhss, A., Josseaume, N., Severin, A., Tabei, K., Hugonnet, J. E., Shlaes, D., Mengin-Lecreulx, D., Van Heijenoort, J., & Arthur, M. (2002). Synthesis of the L-alanyl-L-alanine cross-bridge of Enterococcus faecalis peptidoglycan. Journal of Biological Chemistry, 277(48), 45935–45941. https://doi.org/10.1074/jbc.M207449200

Bowler, P. G., Duerden, B. I., & Armstrong, D. G. (2001). Wound microbiology and associated approaches to wound management. Clinical Microbiology Reviews, 14(2), 244–269. https://doi.org/10.1128/CMR.14.2.244-269.2001

Brede, D. A., Snipen, L. G., Ussery, D. W., Nederbragt, A. J., & Nes, I. F. (2011). Complete genome sequence of the commensal enterococcus faecalis 62, isolated from a healthy norwegian infant. *Journal of Bacteriology*, *193*(9), 2377–2378. https://doi.org/10.1128/JB.00183-11

Bretaudeau, L., Tremblais, K., Aubrit, F., Meichenin, M., & Arnaud, I. (2020). Good Manufacturing Practice (GMP) Compliance for Phage Therapy Medicinal Products. Frontiers in Microbiology, 11(June). https://doi.org/10.3389/fmicb.2020.01161

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., & Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, *5*. https://doi.org/10.1038/srep08365

Broeker, N. K., Roske, Y., Valleriani, A., Stephan, M. S., Andres, D., Koetz, J., Heinemann, U., & Barbirz, S. (2019). Time-resolved DNA release from an O-antigen–specific Salmonella bacteriophage with a contractile tail. Journal of Biological Chemistry, 294(31), 11751–11761. https://doi.org/10.1074/jbc.RA119.008133

Brown, S., Santa Maria, J. P., & Walker, S. (2013). Wall Teichoic Acids of Gram-Positive Bacteria. Annual Review of Microbiology, 67(1), 313–336. https://doi.org/10.1146/annurev-micro-092412-155620

Brum, J. R., & Sullivan, M. B. (2015). Rising to the challenge: Accelerated pace of discovery transforms marine virology. Nature Reviews Microbiology, 13(3), 147–159. https://doi.org/10.1038/nrmicro3404

Bruttin, A., & Bru, H. (2005). VOLUME 1: A RAPID INJECTION MOLDING REFERENCE GUIDE FOR PRODUCT DESIGNERS AND ENGINEERS Designing for Moldability 3D PRINTING CNC MACHINING INJECTION MOLDING. 49(7), 2874–2878. https://doi.org/10.1128/AAC.49.7.2874

Buckling, A., & Rainey, P. B. (2002). Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings of the Royal Society B: Biological Sciences, 269(1494), 931–936. https://doi.org/10.1098/rspb.2001.1945

Buckling, A., Craig MacLean, R., Brockhurst, M. A., & Colegrave, N. (2009). The Beagle in a bottle. Nature, 457(7231), 824–829. https://doi.org/10.1038/nature07892

Burger, M. M. (1966). Teichoic acids: antigenic determinants, chain separation, and their location in the cell wall. Proceedings of the National Academy of Sciences, 56(3), 910–917. https://doi.org/10.1073/pnas.56.3.910

Burgess, R. R. (2009). Chapter 17 Refolding Solubilized Inclusion Body Proteins. In Methods in Enzymology (1st ed., Vol. 463, Issue C). Elsevier Inc. https://doi.org/10.1016/S0076-6879(09)63017-2

Burrowes, B. H., Molineux, I. J., & Fralick, J. A. (2019). Directed in vitro evolution of therapeutic bacteriophages: The appelmans protocol. Viruses, 11(3). https://doi.org/10.3390/v11030241

Bychowska, A., Theilacker, C., Czerwicka, M., Marszewska, K., Huebner, J., Holst, O., Stepnowski, P., & Kaczyński, Z. (2011). Chemical structure of wall teichoic acid isolated from Enterococcus faecium strain U0317. Carbohydrate Research, 346(17), 2816–2819. https://doi.org/10.1016/j.carres.2011.09.026

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., & McQuillan, J. A. (2012). Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, *28*(4), 464–469. https://doi.org/10.1093/bioinformatics/btr703

Cattoir, V. (2022). The multifaceted lifestyle of enterococci: genetic diversity, ecology and risks for public health. Current Opinion in Microbiology, 65, 73–80. https://doi.org/10.1016/j.mib.2021.10.013

CDC. (2018). Timeline of Antibiotic Resistance Compared to Antibiotic Development.

https://www.cdc.gov/drugresistance/about.html
Chajęcka-Wierzchowska, W., Zadernowska, A., & Łaniewska-Trokenheim, Ł. (2017). Virulence factors of Enterococcus spp. presented in food. LWT - Food Science and Technology, 75, 670–676. https://doi.org/10.1016/j.lwt.2016.10.026
Chan, B. K., Sistrom, M., Wertz, J. E., Kortright, K. E., Narayan, D., & Turner, P. E. (2016). Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa. Scientific Reports, 6(March), 1–8. https://doi.org/10.1038/srep26717
Chang, C., & Lin, H. (2016). Dysbiosis in gastrointestinal disorders. Best Practice and Research: Clinical Gastroenterology, 30(1), 3–15. https://doi.org/10.1016/j.bpg.2016.02.001
Chapot-Chartier, M. P., & Kulakauskas, S. (2014). Cell wall structure and function in lactic acid bacteria. Microbial Cell Factories, 13(Suppl 1), 1–23. https://doi.org/10.1186/1475-2859-13-S1-S9
Chatterjee, A., Johnson, C. N., Luong, P., Hullahalli, K., McBride, S. W., Schubert, A. M., Palmer, K. L., Carlson, P. E., & Duerkop, B. A. (2019). Bacteriophage resistance alters antibiotic-mediated intestinal expansion of enterococci. Infection and Immunity, 87(6), 1–14. https://doi.org/10.1128/IAI.00085-19
Chiang, Y. N., Penadés, J. R., & Chen, J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. PLOS Pathogens, 15(8), e1007878. https://doi.org/10.1371/journal.ppat.1007878
Chuang, Y., Peng, Z., Tseng, S., Lin, Y., Sytwu, H., & Hsieh, Y. (2015). *Impact of the glpQ2 Gene on Virulence in a Streptococcus pneumoniae Serotype 19A Sequence Type 320 Strain*. *83*(2), 682–692. https://doi.org/10.1128/IAI.02357-14
Chuard, C., & Reller, L. B. (1998). Bile-esculin test for presumptive identification of enterococci and streptococci: Effects of bile concentration, inoculation technique, and incubation time. Journal of Clinical Microbiology, 36(4), 1135–1136. https://doi.org/10.1128/jcm.36.4.1135-1136.1998
Clemente, J. C., Ursell, L. K., Parfrey, L. W., & Knight, R. (2012). The impact of the gut microbiota on human health: An integrative view. In Cell (Vol. 148, Issue 6, pp. 1258–1270). NIH Public Access. https://doi.org/10.1016/j.cell.2012.01.035
Clewell, D. B. (1993). Bacterial sex pheromone-induced plasmid transfer. Cell, 73(1), 9–12. https://doi.org/10.1016/0092-8674(93)90153-H
Clokie, M. R. J., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. Bacteriophage, 1(1), 31–45. https://doi.org/10.4161/bact.1.1.14942
Coburn, P. S., & Gilmore, M. S. (2003). The Enterococcus faecalis cytolysin: A novel toxin active against eukaryotic and prokaryotic cells. Cellular Microbiology, 5(10), 661–669. https://doi.org/10.1046/j.1462-5822.2003.00310.x
Cochran, P. K., Kellogg, C. A., & Paul, J. H. (1998). Prophage induction of indigenous marine lysogenic bacteria by environmental pollutants. Marine Ecology Progress Series, 164, 125–133. https://doi.org/10.3354/meps164125
Coia, J., & Cubie, H. (1995). Enterococcus species (Streptococcus beta-haemolytic, group D). In The Immunoassay Kit Directory (Vol. 1, pp. 710–712). Springer Netherlands. https://doi.org/10.1007/978-94-009-0359-3_14
Cornelissen, A., Sadovskaya, I., Vinogradov, E., Blangy, S., Spinelli, S., Casey, E., Mahony, J., Noben, J. P., Dal Bello, F., Cambillau, C., & Van Sinderen, D. (2016). The baseplate of Lactobacillus delbrueckii bacteriophage Ld17 harbors a glycerophosphodiesterase. *Journal of Biological Chemistry*, *291*(32), 16816–16827. https://doi.org/10.1074/jbc.M116.728279
Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, *16*(22), 10881–10890. https://doi.org/10.1093/nar/16.22.10881
D'Andrea, M. M., Frezza, D., Romano, E., Marmo, P., Henrici De Angelis, L., Perini, N., Thaller, M. C., & Di Lallo, G. (2020). The lytic bacteriophage vB_EfaH_EF1TV, a new member of the Herelleviridae family, disrupts biofilm produced by Enterococcus faecalis clinical strains. Journal of Global Antimicrobial Resistance, 21, 68–75. https://doi.org/10.1016/j.jgar.2019.10.019
Dale, J. L., Nilson, J. L., Barnes, A. M. T., & Dunny, G. M. (2017). Restructuring of Enterococcus faecalis biofilm architecture in response to antibiotic-induced stress. *Npj Biofilms and Microbiomes*, *3*(1), 1–9. https://doi.org/10.1038/s41522-017-0023-4
Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nature Communications, 3(May), 945–947. https://doi.org/10.1038/ncomms1937
Davies, J. (1996). Origins and evolution of antibiotic resistance. Microbiología (Madrid, Spain), 12(1), 9–16. https://doi.org/10.1128/mmbr.00016-10

De Been, M., Van Schaik, W., Cheng, L., Corander, J., & Willems, R. J. (2013). Recent recombination events in the core genome are associated with adaptive evolution in Enterococcus faecium. Genome Biology and Evolution, 5(8), 1524–1535. https://doi.org/10.1093/gbe/evt111

Debarbieux, L., Pirnay, J.-P., Verbeken, G., De Vos, D., Merabishvili, M., Huys, I., Patey, O., Schoonjans, D., Vaneechoutte, M., Zizi, M., & Rohde, C. (2016). A bacteriophage journey at the European Medicines Agency. FEMS Microbiology Letters, 363(2), fnv225. https://doi.org/10.1093/femsle/fnv225

Del Rio, B., Sánchez-Llana, E., Redruello, B., Magadan, A. H., Fernández, M., Martin, M. C., Ladero, V., & Alvarez, M. A. (2019). Enterococcus faecalisbacteriophage 156 is an effective biotechnological tool for reducing the presence of tyramine and putrescine in an experimental cheese model. Frontiers in Microbiology, 10(MAR), 1–10. https://doi.org/10.3389/fmicb.2019.00566

Denes, T., Den Bakker, H. C., Tokman, J. I., Guldimann, C., & Wiedmann, M. (2015). Selection and characterization of phage-resistant mutant strains of Listeria monocytogenes reveal host genes linked to phage adsorption. Applied and Environmental Microbiology, 81(13), 4295–4305. https://doi.org/10.1128/AEM.00087-15

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of Bacteriology, 190(4), 1390–1400. https://doi.org/10.1128/JB.01412-07

Devoto, A. E., Santini, J. M., Olm, M. R., Anantharaman, K., Munk, P., Tung, J., Archie, E. A., Turnbaugh, P. J., Seed, K. D., Blekhman, R., Aarestrup, F. M., Thomas, B. C., & Banfield, J. F. (2019). Megaphages infect Prevotella and variants are widespread in gut microbiomes. Nature Microbiology, 4(4), 693–700. https://doi.org/10.1038/s41564-018-0338-9

Dik, D. A., Marous, D. R., Fisher, J. F., & Mobashery, S. (2017). Lytic transglycosylases: concinnity in concision of the bacterial cell wall. *Critical Reviews in Biochemistry and Molecular Biology*, *52*(5), 503–542. https://doi.org/10.1080/10409238.2017.1337705

Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. Nature Reviews Microbiology, 18(3), 125–138. https://doi.org/10.1038/s41579-019-0311-5

Djurkovic, S., Loeffler, J. M., & Fischetti, V. A. (2005). Synergistic killing of Streptococcus pneumoniae with the bacteriophage lytic enzyme Cpl-1 and penicillin or gentamicin depends on the level of penicillin resistance. Antimicrobial Agents and Chemotherapy, 49(3), 1225–1228. https://doi.org/10.1128/AAC.49.3.1225-1228.2005

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A.-S., & Lavigne, R. (2013). Learning from Bacteriophages - Advantages and Limitations of Phage and Phage-Encoded Protein Applications. Current Protein and Peptide Science, 13(8), 699–722. https://doi.org/10.2174/138920312804871193

Duckworth, D. H. (1976). "Who discovered bacteriophage?". Bacteriological Reviews, 40(4), 793–802. http://www.ncbi.nlm.nih.gov/pubmed/795414%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC413985

Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. M., & Hooper, L. V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. Proceedings of the National Academy of Sciences, 109(43), 17621–17626. https://doi.org/10.1073/pnas.1206136109

Duplessis, C., Biswas, B., Hanisch, B., Perkins, M., Henry, M., Quinones, J., Wolfe, D., Estrella, L., & Hamilton, T. (2018). Refractory Pseudomonas bacteremia in a 2-year-old sterilized by bacteriophage therapy. Journal of the Pediatric Infectious Diseases Society, 7(3), 253–256. https://doi.org/10.1093/jpids/pix056

Dutka-Malen, S., & Courvalin, P. (1990). Update on glycopeptide resistance in enterococci. Antimicrobic Newsletter, 7(11–12), 81–86. https://doi.org/10.1016/0738-1751(90)90027-A

Dy, R. L., Richter, C., Salmond, G. P. C., & Fineran, P. C. (2014). Remarkable Mechanisms in Microbes to Resist Phage Infections. Annual Review of Virology, 1(1), 307–331. https://doi.org/10.1146/annurev-virology-031413-085500

Dykhuizen, D. (2005). Species Numbers in Bacteria. Proceedings. California Academy of Sciences, 56(6 Suppl 1), 62–71. http://www.ncbi.nlm.nih.gov/pub med/21874075%0Ahttp://www.pubmedcentr al.nih.gov/articl erender.fcgi?artid=PMC3160642

Eaton, T. J., & Gasson, M. J. (2001). Molecular Screening of Enterococcus Virulence Determinants and Potential for Genetic Exchange between Food and Medical Isolates. Applied and Environmental Microbiology, 67(4), 1628–1635. https://doi.org/10.1128/AEM.67.4.1628-1635.2001

Edwards, R. a. (2005). Viral metagenomics. Reviews in Medical Virology, 17(2), 115–131.

https://doi.org/10.1002/rmv.532
Elbreki, M., Ross, R. P., Hill, C., O'Mahony, J., McAuliffe, O., & Coffey, A. (2014). Bacteriophages and Their Derivatives as Biotherapeutic Agents in Disease Prevention and Treatment. Journal of Viruses, 2014, 1–20. https://doi.org/10.1155/2014/382539
Emori, T. G., & Gaynes, R. P. (1993). An overview of nosocomial infections, including the role of the microbiology laboratory. Clinical Microbiology Reviews, 6(4), 428–442. http://www.ncbi.nlm.nih.gov/pubmed/8269394%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC358296
Fard, R. M. N., Barton, M. D., Arthur, J. L., & Heuzenroeder, M. W. (2010). Whole-genome sequencing and gene mapping of a newly isolated lytic enterococcal bacteriophage EFRM31. *Archives of Virology*, *155*(11), 1887–1891. https://doi.org/10.1007/s00705-010-0800-3
Fiore, E., Van Tyne, D., & Gilmore, M. S. (2019). Pathogenicity of Enterococci. Microbiology Spectrum, 7(4), 189–211. https://doi.org/10.1128/microbiolspec.GPP3-0053-2018
Fischetti, V. A. (2008). Bacteriophage Lysins as Effective Antimicrobials. Current Opinion in Microbiology, 11(5), 393–400. https://doi.org/10.1016/j.mib.2008.09.012.Bacteriophage
Fischetti, V. A. (2010). Bacteriophage endolysins: A novel anti-infective to control Gram-positive pathogens. International Journal of Medical Microbiology, 300(6), 357–362. https://doi.org/10.1016/j.ijmm.2010.04.002
Fleischmann, R. D., and O. White, Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., & ... et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science, 269(5223), 496–512.
Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. In The British Journal of Experimental Pathology (Vol. 10, pp. 226–236). https://doi.org/11545337
Fokine, A., & Rossmann, M. G. (2014). Molecular architecture of tailed double-stranded DNA phages. Bacteriophage, 4(2), e28281. https://doi.org/10.4161/bact.28281
Friman, V. P., Soanes-Brown, D., Sierocinski, P., Molin, S., Johansen, H. K., Merabishvili, M., Pirnay, J. P., De Vos, D., & Buckling, A. (2016). Pre-adapting parasitic phages to a pathogen leads to increased pathogen clearance and lowered resistance evolution with Pseudomonas aeruginosa cystic fibrosis bacterial isolates. Journal of Evolutionary Biology, 29(1), 188–198. https://doi.org/10.1111/jeb.12774
Fukushima, T., Uchida, N., Ide, M., Kodama, T., & Sekiguchi, J. (2018). DL-endopeptidases function as both cell wall hydrolases and poly-γ-glutamic acid hydrolases. Microbiology (United Kingdom), 164(3), 277–286. https://doi.org/10.1099/mic.0.000609
Furlan, S., Matos, R. C., Kennedy, S. P., Doublet, B., Serror, P., & Rigottier-Gois, L. (2019). Fitness Restoration of a Genetically Tractable Enterococcus faecalis V583 Derivative To Study Decoration-Related Phenotypes of the Enterococcal Polysaccharide Antigen. MSphere, 4(4). https://doi.org/10.1128/msphere.00310-19
Galloway-Peña, J. R., Nallapareddy, S. R., Arias, C. A., Eliopoulos, G. M., & Murray, B. E. (2009). Analysis of Clonality and Antibiotic Resistance among Early Clinical Isolates of Enterococcus faecium in the United States. The Journal of Infectious Diseases, 200(10), 1566–1573. https://doi.org/10.1086/644790
Garen, A., & Puck, T. T. (1951). The first two steps of the invasion of host cells by bacterial viruses. II. The Journal of Experimental Medicine, 94(3), 177–189. https://doi.org/10.1084/jem.94.3.177
Ghosh, C., Sarkar, P., Issa, R., & Haldar, J. (2019). Alternatives to Conventional Antibiotics in the Era of Antimicrobial Resistance. Trends in Microbiology, 27(4), 323–338. https://doi.org/10.1016/j.tim.2018.12.010
Gibson, S. B., Green, S. I., Liu, C. G., Salazar, K. C., Clark, J. R., Terwilliger, A. L., Kaplan, H. B., Maresso, A. W., Trautner, B. W., & Ramig, R. F. (2019). Constructing and Characterizing Bacteriophage Libraries for Phage Therapy of Human Infections. Frontiers in Microbiology, 10(November), 1–17. https://doi.org/10.3389/fmicb.2019.02537
Golding, I. (2011). Decision Making in Living Cells: Lessons from a Simple System. Annual Review of Biophysics, 40(1), 63–80. https://doi.org/10.1146/annurev-biophys-042910-155227
Gordillo Altamirano, F. L., & Barr, J. J. (2019). Phage Therapy in the Postantibiotic Era. In Clinical microbiology reviews (Vol. 32, Issue 2, pp. e00066-18). American Society for Microbiology Journals. https://doi.org/10.1128/CMR.00066-18
Górski, A., Międzybrodzki, R., Łobocka, M., Głowacka-Rutkowska, A., Bednarek, A., Borysowski, J., Jończyk-Matysiak, E., Łusiak-Szelachowska, M., Weber-Dabrowska, B., Bagińska, N., Letkiewicz, S., Dabrowska, K., & Scheres, J. (2018). Phage therapy: What have we learned? In Viruses (Vol. 10, Issue 6). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/v10060288

Goulet, A., Spinelli, S., Mahony, J., & Cambillau, C. (2020). Conserved and diverse traits of adhesion devices from siphoviridae recognizing proteinaceous or saccharidic receptors. *Viruses*, *12*(5), 1–21. https://doi.org/10.3390/v12050512

Gouliouris, T., Raven, K. E., Moradigaravand, D., Ludden, C., Coll, F., Blane, B., Naydenova, P., Horner, C., Brown, N. M., Corander, J., Limmathurotsakul, D., Parkhill, J., & Peacock, S. J. (2019). Detection of vancomycin-resistant Enterococcus faecium hospital-adapted lineages in municipal wastewater treatment plants indicates widespread distribution and release into the environment. Genome Research, 29(4), 626–634. https://doi.org/10.1101/gr.232629.117

Griffin, M. E., Klupt, S., Espinosa, J., & Hang, H. C. (2022). Peptidoglycan NlpC/P60 peptidases in bacterial physiology and host interactions. Cell Chemical Biology, 2022. https://doi.org/10.1016/j.chembiol.2022.11.001

Guerardel, Y., Sadovskaya, I., Maes, E., Furlan, S., Chapot-Chartier, M. P., Mesnage, S., Rigottier-Gois, L., & Serror, P. (2020). Complete structure of the enterococcal polysaccharide antigen (EPA) of vancomycin-resistant enterococcus faecalis v583 reveals that EPA decorations are teichoic acids covalently linked to a rhamnopolysaccharide backbone. MBio, 11(2). https://doi.org/10.1128/mBio.00277-20

Gulhan, T., Boynukara, B., Ciftci, A., Sogut, M. U., & Findik, A. (2015). Characterization of Enterococcus faecalis isolates originating from different sources for their virulence factors and genes, antibiotic resistance patterns, genotypes and biofilm production. Iranian Journal of Veterinary Research, 16(3), 261–266. https://doi.org/10.22099/ijvr.2015.3191

Gurk-Turner, C. (2000). Quinupristin/Dalfopristin: The First Available Macrolidelincosamide-Streptogramin Antibiotic. Baylor University Medical Center Proceedings, 13(1), 83–86. https://doi.org/10.1080/08998280.2000.11927646

Gutiérrez, D., Briers, Y., Rodríguez-Rubio, L., Martínez, B., Rodríguez, A., Lavigne, R., & García, P. (2015). Role of the pre-neck appendage protein (Dpo7) from phage vB_SepiS-phiIPLA7 as an anti-biofilm agent in staphylococcal species. Frontiers in Microbiology, 6(NOV), 1–10. https://doi.org/10.3389/fmicb.2015.01315

Gutiérrez, D., Fernández, L., Rodríguez, A., & García, P. (2018a). Are phage lytic proteins the secret weapon to kill staphylococcus aureus? MBio, 9(1), 1–17. https://doi.org/10.1128/mBio.01923-17

Gutiérrez, D., Fernández, L., Rodríguez, A., & García, P. (2018b). Practical method for isolation of phage deletion mutants. Methods and Protocols, 1(1), 1–14. https://doi.org/10.3390/mps1010006

Hall, J. P. J., Harrison, E., & Brockhurst, M. A. (2013). Viral host-adaptation: Insights from evolution experiments with phages. Current Opinion in Virology, 3(5), 572–577. https://doi.org/10.1016/j.coviro.2013.07.001

Hancock, L. E., & Gilmore, M. S. (2002). The capsular polysaccharide of Enterococcus faecalis and its relationship to other polysaccharides in the cell wall. Proceedings of the National Academy of Sciences of the United States of America, 99(3), 1574–1579. https://doi.org/10.1073/pnas.032448299

Hancock, L. E., Murray, B. E., & Sillanpää, J. (2014a). Enterococcal Cell Wall Components and Structures. Enterococci: From Commensals to Leading Causes of Drug Resistant Infection, 1–44. http://www.ncbi.nlm.nih.gov/pubmed/24649506

Hancock, L. E., Murray, B. E., & Sillanpää, J. (2014b). Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. Enterococcal Cell Wall Components and Structures, 1–35. https://doi.org/10.1016/j.jinteco.2006.12.001

Harms, A., Brodersen, D. E., Mitarai, N., & Gerdes, K. (2018). Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. Molecular Cell, 70(5), 768–784. https://doi.org/10.1016/j.molcel.2018.01.003

Health Canada. (2013). Final Special Access Programme (SAP) for drugs Guidance Document. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/dhp-mps/alt_formats/hpfb-dgpsa/pdf/acces/sapg3_pasg3-eng.pdf

Henein, A. (2013). What are the limitations on the wider therapeutic use of phage? Bacteriophage, 3(2), e24872. https://doi.org/10.4161/bact.24872

Herskowitz, I., & Hagen, D. (1980). THE LYSIS-LYSOGENY DECISION OF PHAGE λ: EXPLICIT PROGRAMMING AND RESPONSIVENESS. 399–445.

Heyde, S. A. H., & Nørholm, M. H. H. (2021). Tailoring the evolution of BL21(DE3) uncovers a key role for RNA stability in gene expression toxicity. Communications Biology, 4(1), 1–9. https://doi.org/10.1038/s42003-021-02493-4

HHS. (2017). National Action Plan for Combating Antibiotic-Resistant Bacteria: Progress Report for Years 1 and 2 Prepared by the United States Task Force for Combating Antibiotic-Resistant Bacteria. October.

https://aspe.hhs.go v/system/files/pdf/258516/ProgressYears1and2CARBNationalActionPlan.pdf
Ho, K., Huo, W., Pas, S., Dao, R., & Palmer, K. L. (2018). Loss-of-Function Mutations in epaR Confer Resistance to ϕNPV1 Infection in Enterococcus faecalis OG1RF. Antimicrobial Agents and Chemotherapy, 62(10), 1–11. https://doi.org/10.1128/AAC.00758-18
Holtje, J. V, Mirelman, D., Sharon, N., & Schwarz, U. (1975). Novel Type of Murein Transglycosylase in Escherichia coli. *Journal of Bacteriology*, *124*(3), 1067–1076. https://doi.org/https://doi.org/10.1128/jb.124.3.1067-1076.1975
Hoskisson, P. A., & Smith, M. C. (2007). Hypervariation and phase variation in the bacteriophage "resistome." Current Opinion in Microbiology, 10(4), 396–400. https://doi.org/10.1016/j.mib.2007.04.003
Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. ISME Journal, 11(7), 1511–1520. https://doi.org/10.1038/ismej.2017.16
Hu, B., Margolin, W., Molineux, I. J., & Liu, J. (2013). The Bacteriophage T7 Virion Undergoes Extensive Structural Remodeling During Infection. Science, 339(6119), 576 LP – 579. https://doi.org/10.1126/science.1231887
Hu, B., Margolin, W., Molineux, I. J., & Liu, J. (2015). Structural remodeling of bacteriophage T4 and host membranes during infection initiation. Proceedings of the National Academy of Sciences of the United States of America, 112(35), E4919–E4928. https://doi.org/10.1073/pnas.1501064112
Huycke, M. M., Sahm, D. F., & Gilmore, M. S. (1998). Multiple-Drug Resistant Enterococci: The Nature of the Problem and an Agenda for the Future - Volume 4, Number 2—June 1998 - Emerging Infectious Disease journal - CDC. Emerg. Infect. Dis., 4(2), 239–249. https://doi.org/10.3201/eid0402.980211
Hyman, P. (2019). Phages for Phage Therapy: Isolation, Characterization, and Host Range Breadth. Pharmaceuticals (Basel, Switzerland), 12(1). https://doi.org/10.3390/ph12010035
Hynes, W. L., & Walton, S. L. (2000). Hyaluronidases of Gram-positive bacteria. FEMS Microbiology Letters, 183(2), 201–207. https://doi.org/10.1016/S0378-1097(99)00669-2
Iacumin, L., Manzano, M., & Comi, G. (2016). Phage Inactivation of Listeria monocytogenes on San Daniele Dry-Cured Ham and Elimination of Biofilms from Equipment and Working Environments. Microorganisms, 4(1). https://doi.org/10.3390/microorganisms4010004
Imam, M., Alrashid, B., Patel, F., Dowah, A. S. A., Brown, N., Millard, A., Clokie, M. R. J., & Galyov, E. E. (2019). vB_PaeM_MIJ3, a Novel Jumbo Phage Infecting Pseudomonas aeruginosa, Possesses Unusual Genomic Features. Frontiers in Microbiology, 10(January 2018), 1–15. https://doi.org/10.3389/fmicb.2019.02772
Isnard, C., Malbruny, B., Leclercq, R., & Cattoira, V. (2013). Genetic basis for in vitro and in vivo resistance to lincosamides, streptogramins a, and pleuromutilins (LSAP Phenotype) in Enterococcus faecium. Antimicrobial Agents and Chemotherapy, 57(9), 4463–4469. https://doi.org/10.1128/AAC.01030-13
Johnson, A. P. (1994). The pathogenicity of enterococci. Journal of Antimicrobial Chemotherapy, 33(6), 1083–1089. https://doi.org/10.1093/jac/33.6.1083
Kakasis, A., & Panitsa, G. (2019). Bacteriophage therapy as an alternative treatment for human infections. A comprehensive review. In International Journal of Antimicrobial Agents (Vol. 53, Issue 1, pp. 16–21). https://doi.org/10.1016/j.ijantimicag.2018.09.004
Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, *10*(6), 845–858. https://doi.org/10.1038/nprot.2015.053
Khalifa, L., Brosh, Y., Gelman, D., Coppenhagen-Glazer, S., Beyth, S., Poradosu-Cohen, R., Que, Y. A., Beyth, N., & Hazan, R. (2015). Targeting Enterococcus faecalis biofilms with phage therapy. Applied and Environmental Microbiology, 81(8), 2696–2705. https://doi.org/10.1128/AEM.00096-15
Kizziah, J. L., Manning, K. A., Dearborn, A. D., & Dokland, T. (2020). Structure of the host cell recognition and penetration machinery of a Staphylococcus aureus bacteriophage. *PLoS Pathogens*, *16*(2). https://doi.org/10.1371/journal.ppat.1008314
Klare, I., Konstabel, C., Mueller-Bertling, S., Werner, G., Strommenger, B., Kettlitz, C., Borgmann, S., Schulte, B., Jonas, D., Serr, A., Fahr, A. M., Eigner, U., & Witte, W. (2005). Spread of ampicillin/vancomycin-resistant Enterococcus faecium of the epidemic-virulent clonal complex-17 carrying the genes esp and hyl in German hospitals. European Journal of Clinical Microbiology and Infectious Diseases, 24(12), 815–825. https://doi.org/10.1007/s10096-005-0056-0

Korczynska, M., Mukhtar, T. A., Wright, G. D., & Berghuis, A. M. (2007). Structural basis for streptogramin B resistance in Staphylococcus aureus by virginiamycin B lyase. Proceedings of the National Academy of Sciences of the United States of America, 104(25), 10388–10393. https://doi.org/10.1073/pnas.0701809104

Kortright, K. E., Chan, B. K., Koff, J. L., & Turner, P. E. (2019). Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. In Cell Host and Microbe (Vol. 25, Issue 2, pp. 219–232). https://doi.org/10.1016/j.chom.2019.01.014

Kortright, K. E., Doss-Gollin, S., Chan, B. K., & Turner, P. E. (2021). Evolution of Bacterial Cross-Resistance to Lytic Phages and Albicidin Antibiotic. Frontiers in Microbiology, 12(June), 1–10. https://doi.org/10.3389/fmicb.2021.658374

Krause, K. M., Serio, A. W., Kane, T. R., & Connolly, L. E. (2016). Aminoglycosides: An overview. Cold Spring Harbor Perspectives in Medicine, 6(6), 1–18. https://doi.org/10.1101/cshperspect.a027029

Kristich, C. J., & Little, J. L. (2012). Mutations in the β subunit of RNA polymerase alter intrinsic cephalosporin resistance in enterococci. Antimicrobial Agents and Chemotherapy, 56(4), 2022–2027. https://doi.org/10.1128/AAC.06077-11

Krüger, D. H., & Bickle, T. A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiological Reviews, 47(3), 345–360.http://www.ncbi.nlm.nih.gov/pubmed/6314109%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC281580

Krupovic, M., Turner, D., Morozova, V., Dyall-Smith, M., Oksanen, H. M., Edwards, R., Dutilh, B. E., Lehman, S. M., Reyes, A., Baquero, D. P., Sullivan, M. B., Uchiyama, J., Nakavuma, J., Barylski, J., Young, M. J., Du, S., Alfenas-Zerbini, P., Kushkina, A., Kropinski, A. M., … Adriaenssens, E. M. (2021). Bacterial Viruses Subcommittee and Archaeal Viruses Subcommittee of the ICTV: update of taxonomy changes in 2021. Archives of Virology, 166(11), 3239–3244. https://doi.org/10.1007/s00705-021-05205-9

Kwon, K. T., & Armstrong, D. G. (2018). Microbiology and Antimicrobial Therapy for Diabetic Foot Infections. Infection & Chemotherapy, 50(1), 11–20. https://doi.org/10.3947/ic.2018.50.1.11

LaFee, S., & Buschman, H. (2019). With OK From FDA, UC San Diego Researchers Prepare to Launch Novel Phage Study. https://health.ucsd.edu/news/releases/Pages/2019-01-08-FDA-okays-uc-san-diego-to-launch-novel-phage-study.aspx

Lancefield, R. C. (1933). A SEROLOGICAL DIFFERENTIATION OF HUMAN AND OTHER GROUPS OF HEMOLYTIC STREPTOCOCCI. The Journal of Experimental Medicine, 57(4), 571–595. https://doi.org/10.1084/jem.57.4.571

Latka, A., Maciejewska, B., Majkowska-Skrobek, G., Briers, Y., & Drulis-Kawa, Z. (2017). Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. Applied Microbiology and Biotechnology, 101(8), 3103–3119. https://doi.org/10.1007/s00253-017-8224-6

Le Marrec, C., Van Sinderen, D., Walsh, L., Stanley, E., Vlegels, E., Moineau, S., Heinze, P., Fitzgerald, G., & Fayard, B. (1997). Two groups of bacteriophages infecting Streptococcus thermophilus can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. Applied and Environmental Microbiology, 63(8), 3246–3253.

Leach, K. L., Swaney, S. M., Colca, J. R., McDonald, W. G., Blinn, J. R., Thomasco, L. M. M., Gadwood, R. C., Shinabarger, D., Xiong, L., & Mankin, A. S. (2007). The Site of Action of Oxazolidinone Antibiotics in Living Bacteria and in Human Mitochondria. Molecular Cell, 26(3), 393–402. https://doi.org/10.1016/j.molcel.2007.04.005

Leal, A. T., Pohl, P. C., Ferreira, C. A. S., Nascimento-Silva, M. C. L., Sorgine, M. H. F., Logullo, C., Oliveira, P. L., Farias, S. E., Da Silva Vaz, I., & Masuda, A. (2006). Purification and antigenicity of two recombinant forms of Boophilus microplus yolk pro-cathepsin expressed in inclusion bodies. Protein Expression and Purification, 45(1), 107–114. https://doi.org/10.1016/j.pep.2005.07.009

Lee, D., Im, J., Na, H., Ryu, S., Yun, C. H., & Han, S. H. (2019). The Novel Enterococcus Phage vB_EfaS_HEf13 Has Broad Lytic Activity Against Clinical Isolates of Enterococcus faecalis. Frontiers in Microbiology, 10(December), 1–14. https://doi.org/10.3389/fmicb.2019.02877

Lee, T., Pang, S., Abraham, S., & Coombs, G. W. (2019). Antimicrobial-resistant CC17 Enterococcus faecium: The past, the present and the future. Journal of Global Antimicrobial Resistance, 16, 36–47. https://doi.org/10.1016/j.jgar.2018.08.016

Lepage, P., Leclerc, M. C., Joossens, M., Mondot, S., Blottière, H. M., Raes, J., Ehrlich, D., & Doré, J. (2013). A metagenomic insight into our gut's microbiome. Gut, 62(1), 146–158. https://doi.org/10.1136/gutjnl-2011-

301805
Letarov, A. V, & Kulikov, E. E. (2017). *Adsorption of Bacteriophages on Bacterial Cells*. *82*(13), 1632–1658. doi:10.1134/S0006297917130053
Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301
Li, J., Sheng, Y., Ma, R., Xu, M., Liu, F., Qin, R., Zhu, M., Zhu, X., & He, P. (2021). Identification of a depolymerase specific for k64-serotype klebsiella pneumoniae: Potential applications in capsular typing and treatment. *Antibiotics*, *10*(2), 1–18. https://doi.org/10.3390/antibiotics10020144
Liang, J., Zhang, H., Tan, Y. L., Zhao, H., & Ang, E. L. (2022). Directed Evolution of Replication-Competent Double-Stranded DNA Bacteriophage toward New Host Specificity. *ACS Synthetic Biology*, *11*(2), 634–643. https://doi.org/10.1021/acssynbio.1c00319
Loeffler, J. M., Nelson, D., & Fischetti, V. A. (2001). Rapid killing of Streptococcus pneumoniae with a bacteriophage cell wall hydrolase. Science, 294(5549), 2170–2172. https://doi.org/10.1126/science.1066869
Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M., & Henrissat, B. (2014). *The carbohydrate-active enzymes database (CAZy) in 2013*. *42*(November 2013), 490–495. https://doi.org/10.1093/nar/gkt1178
Lourenço, M., De Sordi, L., & Debarbieux, L. (2018). The Diversity of Bacterial Lifestyles Hampers Bacteriophage Tenacity. Viruses, 10(6). https://doi.org/10.3390/v10060327
Love, R. M. (2001). Enterococcus faecalis - A mechanism for its role in endodontic failure. *International Endodontic Journal*, *34*(5), 399–405. https://doi.org/10.1046/j.1365-2591.2001.00437.x
Lowe, A. M., Lambert, P. A., & Smith, A. W. (1995). Cloning of an Enterococcus faecalis endocarditis antigen: Homology with adhesins from some oral streptococci. Infection and Immunity, 63(2), 703–706. https://doi.org/10.1128/iai.63.2.703-706.1995
Lu, M.-J., & Henning, U. (1994). Superinfection exclusion by T-even-type coliphages. Trends in Microbiology, 2(4), 137–139. https://doi.org/10.1016/0966-842X(94)90601-7
Łusiak-Szelachowska, M., Zaczek, M., Weber-Dabrowska, B., Międzybrodzki, R., Kłak, M., Fortuna, W., Letkiewicz, S., Rogóz, P., Szufnarowski, K., Jończyk-Matysiak, E., Owczarek, B., & Górski, A. (2014). Phage neutralization by sera of patients receiving phage therapy. Viral Immunology, 27(6), 295–304. https://doi.org/10.1089/vim.2013.0128
Mahony, J., Alqarni, M., StockDale, S., & Spinelli, S. (2016). Functional and structural dissection of the tape measure protein of lactococcal phage TP901-1. *Nature Publishing Group*, *August*, 1–10. https://doi.org/10.1038/srep36667
Majewska, J., Beta, W., Lecion, D., Hodyra-Stefaniak, K., Kłopot, A., Kazmierćzak, Z., Miernikiewicz, P., Piotrowicz, A., Ciekot, J., Owczarek, B., Kopciuch, A., Wojtyna, K., Harhala, M., Mąkosa, M., & Dąbrowska, K. (2015). Oral application of T4 phage induces weak antibody production in the gut and in the blood. Viruses, 7(8), 4783–4799. https://doi.org/10.3390/v7082845
Mangalea, M. R., & Duerkop, B. A. (2020). Fitness Trade-Offs Resulting from Bacteriophage Resistance. American Society for Microbiology, June, 1–15.
Manrique, P., Bolduc, B., Walk, S. T., van der Oost, J., de Vos, W. M., & Young, M. J. (2016). Healthy human gut phageome. Proceedings of the National Academy of Sciences, 113(37), 10400–10405. https://doi.org/10.1073/pnas.1601060113
Marles-Wright, J., & Lewis, R. J. (2007). Stress responses of bacteria. Current Opinion in Structural Biology, 17(6), 755–760. https://doi.org/10.1016/j.sbi.2007.08.004
Marshall, S. H., Donskey, C. J., Hutton-Thomas, R., Salata, R. A., & Rice, L. B. (2002). Gene dosage and linezolid resistance in Enterococcus faecium and Enterococcus faecalis. Antimicrobial Agents and Chemotherapy, 46(10), 3334–3336. https://doi.org/10.1128/AAC.46.10.3334-3336.2002
Maura, D., & Debarbieux, L. (2012). On the interactions between virulent bacteriophages and bacteria in the gut. Bacteriophage, 2(4), 229–233. https://doi.org/10.4161/bact.23557
Melo, L. D. R., Ferreira, R., Costa, A. R., Oliveira, H., & Azeredo, J. (2019). Efficacy and safety assessment of two enterococci phages in an in vitro biofilm wound model. Scientific Reports, 9(1), 1–12. https://doi.org/10.1038/s41598-019-43115-8
Meuric, V., Le Gall-David, S., Boyer, E., Acuña-Amador, L., Martin, B., Fong, S. B., Barloy-Hubler, F., & Bonnaure-Mallet, M. (2017). Signature of microbial dysbiosis in periodontitis. Applied and Environmental Microbiology, 83(14), 1–13. https://doi.org/10.1128/AEM.00462-17
Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., & Lenski, R. E. (2012). Repeatability and

Contingency in the Evolution of a Key Innovation in Phage Lambda. Science (New York, N.Y.), 335(6067), 428. https://doi.org/10.1126/SCIENCE.1214449

Międzybrodzki, R., Hoyle, N., Zhvaniya, F., Łusiak-Szelachowska, M., Weber-Dąbrowska, B., Łobocka, M., Borysowski, J., Alavidze, Z., Kutter, E., Górski, A., & Gogokhia, L. (2018). Current Updates from the Long-Standing Phage Research Centers in Georgia, Poland, and Russia. In Bacteriophages. https://doi.org/10.1007/978-3-319-40598-8_31-1

Miller, W. R., Munita, J. M., & Arias, C. A. (2014). Mechanisms of antibiotic resistance in enterococci. Expert Review of Anti-Infective Therapy, 12(10), 1221–1236. https://doi.org/10.1586/14787210.2014.956092

Mirzaei, M. K., & Maurice, C. F. (2017). Ménage à trois in the human gut: Interactions between host, bacteria and phages. Nature Reviews Microbiology, 15(7), 397–408. https://doi.org/10.1038/nrmicro.2017.30

mith, R. E., Salamaga, B., Szkuta, P., Hajdamowicz, N., Prajsnar, T. K., Bulmer, G. S., Fontaine, T., Kołodziejczyk, J., Herry, J. M., Hounslow, A. M., Williamson, M. P., Serror, P., & Mesnage, S. (2019). Decoration of the enterococcal polysaccharide antigen EPA is essential for virulence, cell surface charge and interaction with effectors of the innate immune system. PLoS Pathogens, 15(5), 1–24. https://doi.org/10.1371/journal.ppat.1007730

Moak, M., & Molineux, I. J. (2004). Peptidoglycan hydrolytic activities associated with bacteriophage virions. Molecular Microbiology, 51(4), 1169–1183. https://doi.org/10.1046/j.1365-2958.2003.03894.x

Mobarki, N., Almerabi, B., & Hattan, A. (2019). Antibiotic Resistance Crisis. International Journal of Medicine in Developing Countries, 40(4), 561–564. https://doi.org/10.24911/ijmdc.51-1549060699

Moellering, R. C., & Weinberg, A. N. (1971). Studies on antibiotic synergism against enterococci. Journal of Clinical Investigation, 50(12), 2580–2584. https://doi.org/10.1172/jci106758

Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology*, *11*(3), 266–277. https://doi.org/10.1016/j.pbi.2008.03.006

Molineux, I. J., & Panja, D. (2013). Popping the cork: mechanisms of phage genome ejection. Nature Reviews Microbiology, 11(3), 194–204. https://doi.org/10.1038/nrmicro2988

Moon, T. M., D'Andréa, É. D., Lee, C. W., Soares, A., Jakoncic, J., Desbonnet, C., Garcia-Solache, M., Rice, L. B., Page, R., & Peti, W. (2018). The structures of penicillin-binding protein 4 (PBP4) and PBP5 from Enterococci provide structural insights into -lactam resistance. Journal of Biological Chemistry, 293(48), 18574–18585. https://doi.org/10.1074/jbc.RA118.006052

Moura de Sousa, J. A., Pfeifer, E., Touchon, M., & Rocha, E. P. C. (2021). Causes and Consequences of Bacteriophage Diversification via Genetic Exchanges across Lifestyles and Bacterial Taxa. Molecular Biology and Evolution, 38(6), 2497–2512. https://doi.org/10.1093/molbev/msab044

Munita, J. M., Tran, T. T., Diaz, L., Panesso, D., Reyes, J., Murray, B. E., & Arias, C. A. (2013). A liaF Codon Deletion Abolishes Daptomycin Bactericidal Activity against Vancomycin-Resistant Enterococcus faecalis . Antimicrobial Agents and Chemotherapy, 57(6), 2831–2833. https://doi.org/10.1128/aac.00021-13

Nagarajan, V., Peng, M., Tabashsum, Z., Salaheen, S., Padilla, J., & Biswas, D. (2019). Antimicrobial Effect and Probiotic Potential of Phage Resistant Lactobacillus plantarum and its Interactions with Zoonotic Bacterial Pathogens. Foods, 8(6), 1–13. https://doi.org/10.3390/foods8060194

Nallapareddy, S. R., Qin, X., Weinstock, G. M., Hook, M., & Murray, B. E. (2000). Enterococcus faecalis adhesin, Ace, mediates attachment to extracellular matrix proteins collagen type IV and laminin as well as collagen type I. Infection and Immunity, 68(9), 5218–5224. https://doi.org/10.1128/IAI.68.9.5218-5224.2000

Nallapareddy, S. R., Weinstock, G. M., & Murray, B. E. (2003). Clinical isolates of Enterococcus faecium exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. Molecular Microbiology, 47(6), 1733–1747. https://doi.org/10.1046/j.1365-2958.2003.03417.x

Navarre, W. W., & Schneewind, O. (1999). Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiology and Molecular Biology Reviews : MMBR, 63(1), 174–229. https://doi.org/1092-2172

Navarro, F., & Muniesa, M. (2017). Phages in the human body. In Frontiers in Microbiology (Vol. 8, Issue APR, p. 566). Frontiers Media SA. https://doi.org/10.3389/fmicb.2017.00566

Negut, I., Grumezescu, V., & Grumezescu, A. M. (2018). Treatment strategies for infected wounds. Molecules, 23(9), 1–23. https://doi.org/10.3390/molecules23092392 Nelson, A., Wright-Hughes, A., Backhouse, M. R., Lipsky, B. A., Nixon, J., Bhogal, M. S., Reynolds, C., & Brown, S. (2018). CODIFI (Concordance in Diabetic Foot Ulcer Infection): A cross-sectional study of wound swab versus tissue sampling in infected diabetic foot ulcers in England. BMJ Open, 8(1), 1–11. https://doi.org/10.1136/bmjopen-2017-019437

Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., Dutilh, B. E., & Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. In Nature Reviews Microbiology (Vol. 16, Issue 12, pp. 760–773). Springer US. https://doi.org/10.1038/s41579-018-0070-8

O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. EMBO Reports, 7(7), 688–693. https://doi.org/10.1038/sj.embor.7400731

O'Neil, J. (2016). Tackling Drug-Resistant Infections Globally: Final Report And Recommendations. Available online at: http://amr-review.org.

Ofir, G., & Sorek, R. (2018). Contemporary Phage Biology: From Classic Models to New Insights. In Cell (Vol. 172, Issue 6, pp. 1260–1270). Elsevier. https://doi.org/10.1016/j.cell.2017.10.045

Oliveira, A., Sillankorva, S., Quinta, R., Henriques, A., Sereno, R., & Azeredo, J. (2009). Isolation and characterization of bacteriophages for avian pathogenic E. coli strains. Journal of Applied Microbiology, 106(6), 1919–1927. https://doi.org/10.1111/j.1365-2672.2009.04145.x

Palmer, K. L., Godfrey, P., Griggs, A., Kos, V. N., Zucker, J., Desjardins, C., Cerqueira, G., Gevers, D., Walker, S., Wortman, J., Feldgarden, M., Haas, B., Birren, B., & Gilmore, M. S. (2012). Comparative genomics of enterococci: Variation in Enterococcus faecalis, clade structure in E. faecium, and defining characteristics of E. gallinarum and E. casseliflavus. MBio, 3(1), 1–11. https://doi.org/10.1128/mBio.00318-11

Paul, K., Merabishvili, M., Hazan, R., Christner, M., Herden, U., Gelman, D., Khalifa, L., Yerushalmy, O., Coppenhagen-Glazer, S., Harbauer, T., Schulz-Jürgensen, S., Rohde, H., Fischer, L., Aslam, S., Rohde, C., Nir-Paz, R., Pirnay, J. P., Singer, D., & Muntau, A. C. (2021). Bacteriophage rescue therapy of a vancomycin-resistant enterococcus faecium infection in a one-year-old child following a third liver transplantation. Viruses, 13(9). https://doi.org/10.3390/v13091785

Périchon, B., & Courvalin, P. (2009). VanA-type vancomycin-resistant Staphylococcus aureus. Antimicrobial Agents and Chemotherapy, 53(11), 4580–4587. https://doi.org/10.1128/AAC.00346-09

Pertics, B. Z., Cox, A., Nyúl, A., Szamek, N., Kovács, T., & Schneider, G. (2021). Isolation and characterization of a novel lytic bacteriophage against the k2 capsule-expressing hypervirulent klebsiella pneumoniae strain 52145, and identification of its functional depolymerase. *Microorganisms*, *9*(3), 1–20. https://doi.org/10.3390/microorganisms9030650

Petrovic Fabijan, A., Lin, R. C. Y., Ho, J., Maddocks, S., Ben Zakour, N. L., & Iredell, J. R. (2020). Safety of bacteriophage therapy in severe Staphylococcus aureus infection. Nature Microbiology, 14(6). https://doi.org/10.1038/s41564-019-0634-z

Pires, D. P., Oliveira, H., Melo, L. D. R., Sillankorva, S., & Azeredo, J. (2016). Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. In Applied Microbiology and Biotechnology (Vol. 100, Issue 5, pp. 2141–2151). https://doi.org/10.1007/s00253-015-7247-0

Piuri, M., & Hatfull, G. F. (2006). A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. *Molecular Microbiology*, *62*(6), 1569–1585. https://doi.org/10.1111/j.1365-2958.2006.05473.x

Plattner, M., Shneider, M. M., Arbatsky, N. P., Shashkov, A. S., Chizhov, A. O., Nazarov, S., Prokhorov, N. S., Taylor, N. M. I., Buth, S. A., Gambino, M., Gencay, Y. E., Brøndsted, L., Kutter, E. M., Knirel, Y. A., & Leiman, P. G. (2019). Structure and Function of the Branched Receptor-Binding Complex of Bacteriophage CBA120. Journal of Molecular Biology, 431(19), 3718–3739. https://doi.org/10.1016/j.jmb.2019.07.022

Plisson, C., White, H. E., Auzat, I., Zafarani, A., São-José, C., Lhuillier, S., Tavares, P., & Orlova, E. V. (2007). Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO Journal, 26(15), 3720–3728. https://doi.org/10.1038/sj.emboj.7601786

Pouget, C., Dunyach-Remy, C., Pantel, A., Boutet-Dubois, A., Schuldiner, S., Sotto, A., Lavigne, J. P., & Loubet, P. (2021). Alternative Approaches for the Management of Diabetic Foot Ulcers. Frontiers in Microbiology, 12(October), 1–13. https://doi.org/10.3389/fmicb.2021.747618

Pouget, C., Dunyach-Remy, C., Pantel, A., Boutet-Dubois, A., Schuldiner, S., Sotto, A., Lavigne, J. P., & Loubet, P. (2021). Alternative Approaches for the Management of Diabetic Foot Ulcers. Frontiers in Microbiology, 12(October), 1–13. https://doi.org/10.3389/fmicb.2021.747618

Prestinaci, F., Pezzotti, P., & Pantosti, A. (2015). Antimicrobial resistance: A global multifaceted phenomenon. Pathogens and Global Health, 109(7), 309–318. https://doi.org/10.1179/2047773215Y.0000000030

Proença, D., Fernandes, S., Leandro, C., Silva, F. A., Santos, S., Lopes, F., Mato, R., Cavaco-Silva, P., Pimentel, M., & São-José, C. (2012). Phage endolysins with broad antimicrobial activity against Enterococcus faecalis clinical strains. *Microbial Drug Resistance*, *18*(3), 322–332. https://doi.org/10.1089/mdr.2012.0024

Qin, X., Galloway-Pĕa, J. R., Sillanpaa, J., Roh, J. H., Nallapareddy, S. R., Chowdhury, S., Bourgogne, A., Choudhury, T., Muzny, D. M., Buhay, C. J., Ding, Y., Dugan-Rocha, S., Liu, W., Kovar, C., Sodergren, E., Highlander, S., Petrosino, J. F., Worley, K. C., Gibbs, R. A., … Murray, B. E. (2012). Complete genome sequence of Enterococcus faecium strain TX16 and comparative genomic analysis of Enterococcus faecium genomes. BMC Microbiology, 12. https://doi.org/10.1186/1471-2180-12-135

Rakhuba, D. V, Kolomiets, E. I., Dey, E. S., & Novik, G. I. (2010). Bacteriophage receptors and its mechanism in host cell. 59(3), 145–155.

Ramos, Y., Sansone, S., & Morales, D. K. (2021). Sugarcoating it: Enterococcal polysaccharides as key modulators of host–pathogen interactions. PLoS Pathogens, 17(9), 1–16. https://doi.org/10.1371/journal.ppat.1009822

Rao, K. N., Bonanno, J. B., Burley, S. K., & Swaminathan, S. (2006). *Crystal Structure of Glycerophosphodiester Phosphodiesterase From Agrobacterium tumefaciens by SAD With a Large Asymmetric Unit*. *518*(April), 514–518. https://doi.org/10.1002/prot

Rao, V. B., & Feiss, M. (2015). Mechanisms of DNA Packaging by Large Double-Stranded DNA Viruses. Annual Review of Virology, 2(1), 351–378. https://doi.org/10.1146/annurev-virology-100114-055212

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., & Finn, R. D. (2018). *The MEROPS database of proteolytic enzymes , their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database*. *46*(November 2017), 624–632. https://doi.org/10.1093/nar/gkx1134

Redgrave, L. S., Sutton, S. B., Webber, M. A., & Piddock, L. J. V. (2014). Fluoroquinolone resistance: Mechanisms, impact on bacteria, and role in evolutionary success. Trends in Microbiology, 22(8), 438–445. https://doi.org/10.1016/j.tim.2014.04.007

Regeimbal, J. M., Jacobs, A. C., Corey, B. W., Henry, M. S., Thompson, M. G., Pavlicek, R. L., Quinones, J., Hannah, R. M., Ghebremedhin, M., Crane, N. J., Zurawski, D. V., Teneza-Mora, N. C., Biswas, B., & Hall, E. R. (2016). Personalized therapeutic cocktail of wild environmental phages rescues mice from acinetobacter baumannii wound infections. Antimicrobial Agents and Chemotherapy, 60(10), 5806–5816. https://doi.org/10.1128/AAC.02877-15

Reid, W. W. (1950). Estimation and Separation of the Pectin-Esterase and Polygalacturonase of Micro-fungi. *Nature*, *166*(4222), 569–569. https://doi.org/10.1038/166569a0

Rhoads, D. D., Wolcott, R. D., Kuskowski, M. A., Wolcott, B. M., Ward, L. S., & Sulakvelidze, A. (2014). Bacteriophage therapy of venous leg ulcers in humans: results of a phase I safety trial. Journal of Wound Care, 18(6), 237–243. https://doi.org/10.12968/jowc.2009.18.6.42801

Richter, C., Chang, J. T., & Fineran, P. C. (2012). Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. In Viruses (Vol. 4, Issue 10, pp. 2291–2311). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/v4102291

Rigottier-Gois, L., Madec, C., Navickas, A., Matos, R. C., Akary-Lepage, E., Mistou, M. Y., & Serror, P. (2015). The surface rhamnopolysaccharide epa of enterococcus faecalis is a key determinant of intestinal colonization. *Journal of Infectious Diseases*, *211*(1), 62–71. https://doi.org/10.1093/infdis/jiu402

Rihtman, B., Meaden, S., Clokie, M. R. J., Koskella, B., & Millard, A. D. (2016). Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. PeerJ, 4, e2055. https://doi.org/10.7717/peerj.2055

Roach, D. R., & Donovan, D. M. (2015). Antimicrobial bacteriophage-derived proteins and therapeutic applications. Bacteriophage, 5(3), e1062590. https://doi.org/10.1080/21597081.2015.1062590

Rodríguez-Rubio, L., Martínez, B., Donovan, D. M., Rodríguez, A., & García, P. (2013). Bacteriophage virion-associated peptidoglycan hydrolases: Potential new enzybiotics. Critical Reviews in Microbiology, 39(4), 427–434. https://doi.org/10.3109/1040841X.2012.723675

Rohde, C., Resch, G., Pirnay, J. P., Blasdel, B. G., Debarbieux, L., Gelman, D., Górski, A., Hazan, R., Huys, I., Kakabadze, E., Łobocka, M., Maestri, A., Almeida, G. M. de F., Makalatia, K., Malik, D. J., Mašlaňová, I., Merabishvili, M., Pantucek, R., Rose, T., … Chanishvili, N. (2018). Expert opinion on three phage therapy related topics: Bacterial phage resistance, phage training and prophages in bacterial production strains. Viruses, 10(4). https://doi.org/10.3390/v10040178

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. Frontiers in Microbiology, 5(APR), 1–17. https://doi.org/10.3389/fmicb.2014.00172

Ross, A., Ward, S., & Hyman, P. (2016). More is better: Selecting for broad host range bacteriophages. Frontiers in Microbiology, 7(SEP), 1–6. https://doi.org/10.3389/fmicb.2016.01352

Rozdzinski, E., Marre, R., Susa, M., Wirth, R., & Muscholl-Silberhorn, A. (2001). Aggregation substance-mediated adherence of Enterococcus faecalis to immobilized extracellular matrix proteins. Microbial Pathogenesis,

30(4), 211–220. https://doi.org/10.1006/mpat.2000.0429
Sadowy, E., & Luczkiewicz, A. (2014). Drug-resistant and hospital-associated Enterococcus faecium from wastewater, riverine estuary and anthropogenically impacted marine catchment basin. BMC Microbiology, 14(1). https://doi.org/10.1186/1471-2180-14-66
Salmond, G. P. C., & Fineran, P. C. (2015). A century of the phage: Past, present and future. Nature Reviews Microbiology, 13(12), 777–786. https://doi.org/10.1038/nrmicro3564
Sanger, F., Air, G. ., & Smith, M. (1977). © 1977 Nature Publishing Group. Nature, 266(April 21), 730–732. https://doi.org/10.1038/266309a0
São-José, C. (2018). Engineering of Phage-Derived Lytic Enzymes: Improving Their Potential as Antimicrobials. Antibiotics, 7(2), 29. https://doi.org/10.3390/antibiotics7020029
Sava, I. G., Heikens, E., & Huebner, J. (2010). Pathogenesis and immunity in enterococcal infections. Clinical Microbiology and Infection, 16(6), 533–540. https://doi.org/10.1111/j.1469-0691.2010.03213.x
Scheurwater, E., Reid, C. W., & Clarke, A. J. (2008). Lytic transglycosylases: Bacterial space-making autolysins. *International Journal of Biochemistry and Cell Biology*, *40*(4), 586–591. https://doi.org/10.1016/j.biocel.2007.03.018
Schleifer, K. H., & Kilpper-Balz, R. (1984). Transfer of Streptococcus faecalis and Streptococcus faecium to the genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus faecium comb. nov. *International Journal of Systematic Bacteriology*, *34*(1), 31–34. https://doi.org/10.1099/00207713-34-1-31
Schuster, C. F., & Bertram, R. (2013). Toxin-antitoxin systems are ubiquitous and versatile modulators of prokaryotic cell fate. FEMS Microbiology Letters, 340(2), 73–85. https://doi.org/10.1111/1574-6968.12074
Schwarzer, D., Stummeyer, K., Gerardy-schahn, R., & Mu, M. (2007). *Characterization of a Novel Intramolecular Chaperone Domain Conserved in Endosialidases and Other Bacteriophage Tail Spike and Fiber Proteins The Journal of biological chemistry 282 (5), 2821–2831.doi:10.1074/jbc.M609543200.*
Shahed-Al-Mahmud, M., Roy, R., Sugiokto, F. G., Islam, M. N., Lin, M.-D., Lin, L.-C., & Lin, N.-T. (2021). Phage φAB6-Borne Depolymerase Combats Acinetobacter baumannii Biofilm Formation and Infection. *Antibiotics*, *10*(3), 279. https://doi.org/10.3390/antibiotics10030279
Sharma, S., Chatterjee, S., Datta, S., Prasad, R., Dubey, D., Prasad, R. K., & Vairale, M. G. (2017). Bacteriophages and its applications: an overview. In Folia Microbiologica (Vol. 62, Issue 1, pp. 17–55). Folia Microbiologica. https://doi.org/10.1007/s12223-016-0471-x
Shi, L., Liu, J. F., An, X. M., & Liang, D. C. (2008). Crystal structure of glycerophosphodiester phosphodiesterase (GDPD) from Thermoanaerobacter tengcongensis, a metal ion-dependent enzyme: Insight into the catalytic mechanism. Proteins: Structure, Function and Genetics, 72(1), 280–288. https://doi.org/10.1002/prot.21921
Shkoporov, A. N., & Hill, C. (2019). Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. Cell Host & Microbe, 25(2), 195–209. https://doi.org/10.1016/J.CHOM.2019.01.017
Silhavy, T. J., Kahne, D., & Walker, S. (2010). The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, *2*(5), a000414–a000414. https://doi.org/10.1101/cshperspect.a000414
Sillankorva, S., Pleteneva, E., Shaburova, O., Santos, S., Carvalho, C., Azeredo, J., & Krylov, V. (2010). Salmonella Enteritidis bacteriophage candidates for phage therapy of poultry. Journal of Applied Microbiology, 108(4), 1175–1186. https://doi.org/10.1111/j.1365-2672.2009.04549.x
Simmonds, P., & Aiewsakun, P. (2018). Virus classification – where do you draw the line? Archives of Virology, 163(8), 2037–2046. https://doi.org/10.1007/s00705-018-3938-z
Singh, K. V., Coque, T. M., Weinstock, G. M., & Murray, B. E. (1998). In vivo testing of an Enterococcus faecalis efaA mutant and use of efaA homologs for species identification. FEMS Immunology and Medical Microbiology, 21(4), 323–331. https://doi.org/10.1016/S0928-8244(98)00087-X
Singh, K. V., Weinstock, G. M., & Murray, B. E. (2002). An Enterococcus faecalis ABC homologue (Lsa) is required for the resistance of this species to clindamycin and quinupristin-dalfopristin. Antimicrobial Agents and Chemotherapy, 46(6), 1845–1850. https://doi.org/10.1128/AAC.46.6.1845-1850.2002
Skrepnek, G. H., Mills, J. L., Lavery, L. A., & Armstrong, D. G. (2017). Health care service and outcomes among an estimated 6.7 million ambulatory care diabetic foot cases in the U.S. Diabetes Care, 40(7), 936–942. https://doi.org/10.2337/dc16-2189
Son, S. T., Han, S.-K., Lee, T. Y., Namgoong, S., & Dhong, E.-S. (2017). The Microbiology of Diabetic Foot Infections in Korea. Journal of Wound Management and Research, 13(1), 8–12. https://doi.org/10.22467/jwmr.2017.00108
Steenbergen, J. N., Alder, J., Thorne, G. M., & Tally, F. P. (2005). Daptomycin: A lipopeptide antibiotic for the

treatment of serious Gram-positive infections. Journal of Antimicrobial Chemotherapy, 55(3), 283–288. https://doi.org/10.1093/jac/dkh546

Stern, A., & Sorek, R. (2011). The phage-host arms race: Shaping the evolution of microbes. BioEssays, 33(1), 43–51. https://doi.org/10.1002/bies.201000071

StockDale, S. R., Mahony, J., Courtin, P., Pijkeren, J. Van, Britton, R. A., Neve, H., Heller, K. J., Aideh, B., Vogensen, F. K., & Sinderen, D. Van. (2013). *The Lactococcal Phages Tuc2009 and TP901-1 Incorporate Two Alternate Forms of Their Tail Fiber into Their Virions for*. *288*(8), 5581–5590. https://doi.org/10.1074/jbc.M112.444901

Studier, F. W., & Rao Movva, N. (1976). SAMase Gene of Bacteriophage T3 Is Responsible for Overcoming Host Restriction. Journal of Virology, 19(1), 136–145. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC354840/pdf/jvirol00223-0146.pdf

Summers, W. C. (2012). The strange history of phage therapy. Bacteriophage, 2(2), 130–133. https://doi.org/10.4161/bact.20757

Sumrall, E. T., Shen, Y., Keller, A. P., Rismondo, J., Pavlou, M., Eugster, M. R., Boulos, S., Disson, O., Thouvenot, P., Kilcher, S., Wollscheid, B., Cabanes, D., Lecuit, M., Gründling, A., & Loessner, M. J. (2019). Phage resistance at the cost of virulence: Listeria monocytogenes serovar 4b requires galactosylated teichoic acids for InlB-mediated invasion. PLOS Pathogens, 15(10), e1008032. https://doi.org/10.1371/journal.ppat.1008032

Sutherland, I. W. (1995). Polysaccharide Lyases. Fems Microbiology Reviews, 16(4), 323–347.

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356–361. https://doi.org/10.1038/nature04160

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. Nature Reviews. Microbiology, 5(10), 801–812. https://doi.org/10.1038/nrmicro1750

Taylor, N. M. I., van Raaij, M. J., & Leiman, P. G. (2018). Contractile injection systems of bacteriophages and related systems. Molecular Microbiology, 108(1), 6–15. https://doi.org/10.1111/mmi.13921

Teng, F., Jacques-Palaz, K. D., Weinstock, G. M., & Murray, B. E. (2002). Evidence that the enterococcal polysaccharide antigen gene (epa) cluster is widespread in Enterococcus faecalis and influences resistance to phagocytic killing of E. faecalis. Infection and Immunity, 70(4), 2010–2015. https://doi.org/10.1128/iai.70.4.2010-2015.2002

Teng, F., Singh, K. V., Bourgogne, A., Zeng, J., & Murray, B. E. (2009). Further characterization of the epa gene cluster and epa polysaccharides of Enterococcus faecalis. Infection and Immunity, 77(9), 3759–3767. https://doi.org/10.1128/IAI.00149-09

Terms, F. (2011). On an invisible microbe antagonistic to dysentery bacilli . Note by M. F. d'Herelle, presented by M. Roux. Comptes Rendus Academie des Sciences 1917; 165:373–5 . *Bacteriophage*, *1*(1), 3–5. https://doi.org/10.4161/bact.1.1.14941

Theilacker, C., Holst, O., Lindner, B., Huebner, J., & Kaczyński, Z. (2012). The structure of the wall teichoic acid isolated from Enterococcus faecalis strain 12030. Carbohydrate Research, 354, 106–109. https://doi.org/10.1016/j.carres.2012.03.031

Thomas, V. C., Thurlow, L. R., Boyle, D., & Hancock, L. E. (2008). Regulation of autolysis-dependent extracellular DNA release by Enterococcus faecalis extracellular proteases influences biofilm development. Journal of Bacteriology, 190(16), 5690–5698. https://doi.org/10.1128/JB.00314-08

Thurlow, L. R., Thomas, V. C., & Hancock, L. E. (2009). Capsular polysaccharide production in Enterococcus faecalis and contribution of CpsF to capsule serospecificity. Journal of Bacteriology, 191(20), 6203–6210. https://doi.org/10.1128/JB.00592-09

Thurlow, L. R., Thomas, V. C., Fleming, S. D., & Hancock, L. E. (2009). Enterococcus faecalis capsular polysaccharide serotypes C and D and their contributions to host innate immune evasion. Infection and Immunity, 77(12), 5551–5557. https://doi.org/10.1128/IAI.00576-09

Tock, M. R., & Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. Current Opinion in Microbiology, 8(4), 466–472. https://doi.org/10.1016/j.mib.2005.06.003

Tomasz, A. (1979). The Mechanism of the Irreversible Antimicrobial Effects of Penicillins: How the Beta-Lactam Antibiotics Kill and Lyse Bacteria. Annual Review of Microbiology, 33(1), 113–137. https://doi.org/10.1146/annurev.mi.33.100179.000553

Top, J., Willems, R., & Bonten, M. (2008). Emergence of CC17 Enterococcus faecium: From commensal to hospital-adapted pathogen. FEMS Immunology and Medical Microbiology, 52(3), 297–308. https://doi.org/10.1111/j.1574-695X.2008.00383.x

Turner, D., Kropinski, A. M., & Adriaenssens, E. M. (2021). A Roadmap for Genome-Based Phage Taxonomy.

Viruses, 13(3), 1–10. https://doi.org/10.3390/v13030506
Twort, F. W. (1915). AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. The Lancet, 186(4814), 1241–1243. https://doi.org/10.1016/S0140-6736(01)20383-3
Uttley, A. C., Collins, C. H., Naidoo, J., & George, R. C. (1988). VANCOMYCIN-RESISTANT ENTEROCOCCI. *The Lancet*, *331*(8575–8576), 57–58. https://doi.org/10.1016/S0140-6736(88)91037-9
Van Belleghem, J. D., Dąbrowska, K., Vaneechoutte, M., Barr, J. J., & Bollyky, P. L. (2019). Interactions between bacteriophage, bacteria, and the mammalian immune system. In Viruses (Vol. 11, Issue 1, p. 10). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/v11010010
Van Belleghem, J. D., Dąbrowska, K., Vaneechoutte, M., Barr, J. J., & Bollyky, P. L. (2019).
Vehreschild, M. J. G. T., Haverkamp, M., Biehl, L. M., Lemmen, S., & Fätkenheuer, G. (2019). Vancomycin-resistant enterococci (VRE): a reason to isolate? Infection, 47(1), 7–11. https://doi.org/10.1007/s15010-018-1202-9
Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M., & Desvaux, M. (2019). Cell wall hydrolases in bacteria: Insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. *Frontiers in Microbiology*, *10*(FEB). https://doi.org/10.3389/fmicb.2019.00331
Veselovsky, V. A., Dyachkova, M. S., Bespiatykh, D. A., Yunes, R. A., Shitikov, E. A., Polyaeva, P. S., Danilenko, V. N., Olekhnovich, E. I., & Klimina, K. M. (2022). The Gene Expression Profile Differs in Growth Phases of the Bifidobacterium Longum Culture. Microorganisms, 10(8). https://doi.org/10.3390/microorganisms10081683
Viazis, S., Akhtar, M., Feirtag, J., Brabban, A. D., & Diez-Gonzalez, F. (2011). Isolation and characterization of lytic bacteriophages against enterohaemorrhagic Escherichia coli. Journal of Applied Microbiology, 110(5), 1323–1331. https://doi.org/10.1111/j.1365-2672.2011.04989.x
Vijayakumar, P. P., & Muriana, P. M. (2015). A microplate growth inhibition assay for screening bacteriocins against listeria monocytogenes to differentiate their mode-of-action. Biomolecules, 5(2), 1178–1194. https://doi.org/10.3390/biom5021178
Vinga, I., Baptista, C., Auzat, I., Petipas, I., Lurz, R., Tavares, P., Santos, M. A., & São-José, C. (2012). Role of bacteriophage SPP1 tail spike protein gp21 on host cell receptor binding and trigger of phage DNA ejection. Molecular Microbiology, 83(2), 289–303. https://doi.org/10.1111/j.1365-2958.2011.07931.x
Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. FEMS Microbiology Reviews, 32(2), 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x
Waldor, M. K., & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. Science, 272(5270), 1910–1913. https://doi.org/10.1126/science.272.5270.1910
Walsh, J. W., Hoffstad, O. J., Sullivan, M. O., & Margolis, D. J. (2016). Association of diabetic foot ulcer and death in a population-based cohort from the United Kingdom. Diabetic Medicine, 33(11), 1493–1498. https://doi.org/10.1111/dme.13054
Wang, I. N., Deaton, J., & Young, R. (2003). Sizing the holin lesion with an endolysin-β-galactosidase fusion. Journal of Bacteriology, 185(3), 779–787. https://doi.org/10.1128/JB.185.3.779-787.2003
Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. FEMS Microbiology Reviews, 28(2), 127–181. https://doi.org/10.1016/j.femsre.2003.08.001 Weiner, L. M., Webb, A. K., Limbago, B., Dudeck, M. A., Patel, J., Kallen, A. J., Edwards, J. R., & Sievert, D. M. (2016). Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. Infection Control and Hospital Epidemiology, 37(11), 1288–1301. https://doi.org/10.1017/ice.2016.174
Werner, G., Klare, I., & Witte, W. (2002). Molecular analysis of streptogramin resistance in enterococci. International Journal of Medical Microbiology, 292(2), 81–94. https://doi.org/10.1078/1438-4221-00194
WHO. (2017). Global priority list of antibiotic-resistant bacteria. https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/
Wicken, A. J., Elliott, S. D., & Baddiley, J. (1963). The Identity of Streptococcal Group D Antigen with Teichoic Acid. Journal of General Microbiology, 31(2), 231–239. https://doi.org/10.1099/00221287-31-2-231
Wicken, A., & Knox, K. (1975). Lipoteichoic acids: a new class of bacterial antigen. Science, 187(4182), 1161–1167. https://doi.org/10.1126/science.46620
Willyard, C. (2017). The drug-resistant bacteria that pose the greatest health threats. *Nature*, *543*(7643), 15. https://doi.org/10.1038/nature.2017.21550

Wohlkönig, A., Huet, J., Looze, Y., & Wintjens, R. (2010). Structural relationships in the lysozyme superfamily: Significant evidence for glycoside hydrolase signature motifs. *PLoS ONE*, *5*(11), 1–10. https://doi.org/10.1371/journal.pone.0015388

Wright, R. C. T., Friman, V., Smith, M. C. M., & Id, A. B. (2018). Cross-resistance is modular in bacteria – phage interactions. 1–22.

Xiang, Y., Morais, M. C., Cohen, D. N., Bowman, V. D., Anderson, D. L., & Rossmann, M. G. (2008). Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage φ29 tail. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(28), 9552–9557. https://doi.org/10.1073/pnas.0803787105

Xu, J., & Xiang, Y. (2017). Membrane Penetration by Bacterial Viruses. Journal of Virology, 91(13). https://doi.org/10.1128/jvi.00162-17

Yasufuku, T., Shigemura, K., Shirakawa, T., Matsumoto, M., Nakano, Y., Tanaka, K., Arakawa, S., Kawabata, M., & Fujisawa, M. (2011). Mechanisms of and risk factors for fluoroquinolone resistance in clinical Enterococcus faecalis isolates from patients with urinary tract infections. Journal of Clinical Microbiology, 49(11), 3912–3916. https://doi.org/10.1128/JCM.05549-11

Yoong, P., Schuch, R., Nelson, D., & Fischetti, V. A. (2004). Identification of a broadly active phage lytic enzyme with lethal activity against antibiotic-resistant Enterococcus faecalis and Enterococcus faecium. Journal of Bacteriology, 186(14), 4808–4812. https://doi.org/10.1128/JB.186.14.4808-4812.2004

Yosef, I., Shitrit, D., Goren, M. G., Burstein, D., Pupko, T., & Qimron, U. (2013). DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. Proceedings of the National Academy of Sciences, 110(35), 14396–14401. https://doi.org/10.1073/pnas.1300108110

Young, R., & Gill, J. J. (2015). Phage therapy redux--What is to be done? Science, 350(6265), 1163–1164. https://doi.org/10.1126/science.aad6791

Zaczek, M., Łusiak-Szelachowska, M., Jończyk-Matysiak, E., Weber-Dabrowska, B., Miedzybrodzki, R., Owczarek, B., Kopciuch, A., Fortuna, W., Rogóz, P., & Górski, A. (2016). Antibody production in response to staphylococcal MS-1 phage cocktail in patients undergoing phage therapy. Frontiers in Microbiology, 7(OCT), 1–14. https://doi.org/10.3389/fmicb.2016.01681

Appendix 1

Table 6.1. List of all annotated CDS of phiSHEF10.

Locus-tag	Start-end	Size (bp)	annotation
phiSHEF10_02	443 - 916	474	Terminase Small Subunit
phiSHEF10_03	1453 - 3177	1725	Terminase Large subunit
phiSHEF10_05	3416 - 4567	1152	Portal protein
phiSHEF10_06	4554 - 5117	564	Prohead protease
phiSHEF10_07	5187 - 6437	1251	Capsid protein
phiSHEF10_09	6807 - 7103	297	Head-Tail Connector Protein
phiSHEF10_10	7075 - 7410	336	Head-tail joining protein
phiSHEF10_13	8252 - 8818	567	Major tail protein
phiSHEF10_15	9581 - 13,951	4371	Tail tape measure protein
phiSHEF10_16	14,034 - 16,115	2082	Tail protein
phiSHEF10_17	16,186 - 18,243	2058	Endopeptidase tail protein
phiSHEF10_19	18,723 - 18,968	246	Hemolysin
phiSHEF10_20	18,983 - 19,219	237	Holin
phiSHEF10_21	19,216 - 20,202	987	Endolysin
phiSHEF10_22	20,284 - 20,511	228	Glutaredoxin-like protein
phiSHEF10_24	21,170 - 23,461	2292	DNA polymerase
phiSHEF10_30	26,098 - 26,874	777	Beta-lactamase superfamily domain protein
phiSHEF10_35	28,539 - 29,057	519	HNH homing endonuclease
phiSHEF10_36	29,011 - 29,751	741	Primase/polymerase protein
phiSHEF10_39	30,197 - 31,492	1296	SNF2 family N-terminal domain protein
phiSHEF10_48	33,797 - 35,377	1581	DNA primase
phiSHEF10_67	41,202 - 41,573	372	HNH endonuclease

Table 6.2. List of all annotated CDS of phiSHEF11.

Locus-tag	Start - end	Size (bp)	annotation
phiSHEF11_01	105 - 476	372	HNH endonuclease
phiSHEF11_16	4984 - 6564	1581	DNA primase
phiSHEF11_25	9161 - 10,456	1296	SNF2 family N-terminal domain protein
phiSHEF11_28	10,888 - 11,631	744	DNA primase/polymerase
phiSHEF11_33	13,196 - 13,717	522	NUMOD4 motif protein
phiSHEF11_34	13,704 - 14,483	780	Beta-lactamase superfamily domain protein
phiSHEF11_40	17,119 - 19,410	2292	DNA polymerase
phiSHEF11_41	19,473 - 20,159	687	DNA methylase
phiSHEF11_42	20,160 - 20,384	225	Glutaredoxin-like protein
phiSHEF11_43	20,477 - 21,574	1098	Endolysin
phiSHEF11_44	21,577 - 21,810	234	Holin
phiSHEF11_45	21,824 - 22,069	246	Hemolysin
phiSHEF11_46	22,252 - 24,480	2229	Endopeptidase tail protein
phiSHEF11_47	24,525 - 26,600	2076	Phage tail protein
phiSHEF11_48	26,683 - 31,053	4371	Phage tail tape measure
phiSHEF11_50	31,816 - 32,382	567	Phage major tail protein
phiSHEF11_53	33,224 - 33,559	336	Phage head-tail joining protein
phiSHEF11_54	33,531 - 33,827	297	head-tail connector protein
phiSHEF11_56	34,185 - 35,438	1254	Major capsid protein
phiSHEF11_57	35,508 - 36,071	564	Prohead protease
phiSHEF11_58	36,058 - 37,209	1152	Phage portal protein
phiSHEF11_61	37,810 - 39,531	1722	Phage terminase, large subunit

phiSHEF11_62	39,904 - 40,377	474	Phage terminase, small subunit

Table 6.3. List of all annotated CDS of phiSHEF13.

Locus-tag	Start-end	Size (bp)	annotation
phiSHEF13_002	921 - 2204	1284	metallophosphatase
phiSHEF13_007	4751 - 5947	1197	Ig domain-containing protein
phiSHEF13_008	5996 - 6274	279	Holin
phiSHEF13_011	7368 - 8615	1284	recombinase A
phiSHEF13_014	10,743 - 13,769	3027	DNA polymerase
phiSHEF13_015	13,852 - 14,163	312	integration host factor
phiSHEF13_020	16,000 - 16,677	678	PDDEXK family nuclease
phiSHEF13_026	18,524 - 19,405	882	trimeric dUTP diphosphatase
phiSHEF13_027	19,415 - 20,470	1056	DNA primase
phiSHEF13_029	21,147 - 23,039	1893	endonuclease
phiSHEF13_031	23,412 - 24,452	1041	metallophosphatase
phiSHEF13_032	24,452 - 25,924	1437	DNA helicase
phiSHEF13_034	27,597 - 29,372	1776	DNA helicase
phiSHEF13_037	30,895 - 34,353	3459	tail protein
phiSHEF13_041	38,091 - 41,273	3183	Ig-like domain containing protein
phiSHEF13_042	41,285 - 42,337	1053	baseplate protein
phiSHEF13_043	42,352 - 43,056	705	baseplate assembly protein
phiSHEF13_048	44,903 - 47,317	2415	DUF859 domain-containing protein
phiSHEF13_050	54,111 - 56,627	2517	tail lysin
phiSHEF13_051	56,677 - 60,330	3654	tail tape measure protein
phiSHEF13_055	62,244 - 63,953	1710	tail sheath protein
phiSHEF13_062	67,804 - 69,225	1422	major capsid protein
phiSHEF13_065	71,246 - 72,934	1689	portal protein
phiSHEF13_067	73,425 - 74,369	945	N-acetylmuramoyl-L-alanine amidase
phiSHEF13_068	74,473 - 75,099	627	LysM domain containing protein
phiSHEF13_069	75,258 - 76,514	1257	N-acetylmuramoyl-L-alanine amidase
phiSHEF13_073	78,888 - 80,723	1836	terminase large subunit
phiSHEF13_080	83,309 - 83,379	71	tRNA-Asn
phiSHEF13_081	83,465 - 83,536	72	tRNA-Gln
phiSHEF13_082	83,646 - 83,734	89	tRNA-Ser
phiSHEF13_083	84,443 - 84,514	72	tRNA-Met
phiSHEF13_084	84,639 - 84,713	75	tRNA-Asp
phiSHEF13_085	84,719 - 84,790	72	tRNA-Glu
phiSHEF13_086	84,997 - 85,070	74	tRNA-Met
phiSHEF13_087	85,281 - 85,352	72	tRNA-Trp
phiSHEF13_088	85,369 - 85,443	75	tRNA-Other
phiSHEF13_089	85,451 - 85,523	73	tRNA-Cys
phiSHEF13_090	85,645 - 85,718	74	tRNA-Phe
phiSHEF13_092	86,877 - 86,948	72	tRNA-Thr
phiSHEF13_093	87,141 - 87,214	74	tRNA-Lys
phiSHEF13_094	87,350 - 87,423	74	tRNA-Lys
phiSHEF13_095	87,526 - 87,598	73	tRNA-Arg
phiSHEF13_096	87,714 - 87,785	72	tRNA-Val
phiSHEF13_097	87,939 - 88,012	74	tRNA-Ile

phiSHEF13_100	88,721 - 88,791	71	tRNA-His
phiSHEF13_101	88,969 - 89,052	84	tRNA-Leu
phiSHEF13_102	89,137 - 89,221	85	tRNA-Leu
phiSHEF13_104	90,397 - 90,483	87	tRNA-Ser
phiSHEF13_105	90,580 - 90,666	87	tRNA-Ser
phiSHEF13_107	91,093 - 91,257	165	LysM domain containing protein
phiSHEF13_108	91,262 - 91,334	73	tRNA-Arg
phiSHEF13_109	91,343 - 91,415	73	tRNA-Ala
phiSHEF13_110	91,736 - 91,808	73	tRNA-Pro
phiSHEF13_111	91,969 - 92,541	573	GIY-YIG nuclease family protein
phiSHEF13_118	97,032 - 99,203	2172	ribonucleotide reductase
phiSHEF13_119	99,274 - 100,089	816	deoxynucleoside kinase
phiSHEF13_120	100,106 - 100,870	765	Nicotinamide mononucleotide transporter
phiSHEF13_121	100,863 - 101,102	240	glutaredoxin-like protein
phiSHEF13_125	102,904 - 104,031	1128	NadR-like protein
phiSHEF13_126	104,033 - 104,776	744	nicotinamide mononucleotide transporter
phiSHEF13_132	107,302 - 108,261	960	thymidylate synthase
phiSHEF13_134	108,547 - 109,161	615	4Fe-4S single cluster domain-containing protein
phiSHEF13_135	109,258 - 110,127	870	prohibitin family protein
phiSHEF13_151	115,255 - 115,785	531	exoribonuclease AS
phiSHEF13_152	115,786 - 116,364	579	metallophosphatase
phiSHEF13_153	116,361 - 116,969	609	metallophosphatase
phiSHEF13_155	117,223 - 117,966	744	metallophosphatase
phiSHEF13_163	121,928 - 122,269	342	peptidase M20 containing protein
phiSHEF13_164	122,359 - 122,637	279	DNA-binding protein
phiSHEF13_169	124,423 - 125,163	741	nickel pincer cofactor biosynthesis protein
phiSHEF13_201	137,194 - 137,526	333	TM2 domain-containing protein
phiSHEF13_205	139,325 - 139,573	249	transcriptional regulator

Table 6.4. List of all annotated CDS of phiSHEF14.

Locus-tag	Start - end	Size (bp)	annotation
PhiSHEF14_01	142 - 429	288	hypothetical protein
phiSHEF14_02	513 -842	330	Single-stranded DNA-binding protein
phiSHEF14_03	917 - 1339	423	hypothetical protein
phiSHEF14_04	1342 - 1800	459	hypothetical protein
phiSHEF14_05	1814 - 3052	1239	DNA encapsidation protein
phiSHEF14_06	3085 - 5424	2340	DNA polymerase
phiSHEF14_07	5499 - 5687	189	hypothetical protein
phiSHEF14_08	5684 - 5836	153	hypothetical protein
phiSHEF14_09	5841 - 6071	231	hypothetical protein
phiSHEF14_10	6068 - 6622	555	HNH homing endonuclease
phiSHEF14_11	6625 - 8634	2010	NlpC/P60 family
phiSHEF14_12	8697 - 9611	915	Endolysin
phiSHEF14_13	9608 - 9874	267	Holin
phiSHEF14_14	9874 - 11,625	1752	Major tail protein
phiSHEF14_15	11,637 - 13,496	1860	hypothetical protein
phiSHEF14_16	13,498 - 14,292	795	hypothetical protein
phiSHEF14_17	14,303 - 15,727	1425	Phosphohydrolase

phiSHEF14_18	15,740 - 16,384	645	lower collar protein
phiSHEF14_19	16,350 - 17,375	1026	Phage Connector protein
phiSHEF14_20	17,395 - 18,567	1173	Major capsid protein
phiSHEF14_21	18,569 - 18,733	165	hypothetical protein
phiSHEF14_22	18,748 - 19,095	348	hypothetical protein

Table 6.5. List of all annotated CDS of phiSHEF16.

Locus-tag	Start-end	Size (bp)	annotation
phiSHEF16_001	5628 - 8147	2520	NLPC/P60 family protein
phiSHEF16_002	8197 - 11,841	3645	tail tape measure protein
phiSHEF16_006	13,759 - 15,468	1710	tail sheath protein
phiSHEF16_013	19,319 - 20,740	1422	major capsid protein
phiSHEF16_016	22,779 - 24,467	1689	portal protein
phiSHEF16_018	24,958 - 25,902	945	NlpC/P60 family protein
phiSHEF16_019	26,005 - 26,667	663	LysM domain containing protein
phiSHEF16_020	26,825 - 28,093	1269	N-acetylmuramoyl-L-alanine amidase
phiSHEF16_024	30,465 - 32,300	1836	terminase large subunit
phiSHEF16_031	34,892 - 34,962	71	tRNA-Asn
phiSHEF16_032	35,054 - 35,125	72	tRNA-Gln
phiSHEF16_034	37,047 - 37,118	72	tRNA-Met
phiSHEF16_035	37,243 - 37,317	75	tRNA-Asp
phiSHEF16_036	37,323 - 37,394	72	tRNA-Glu
phiSHEF16_037	37,602 - 37,674	73	tRNA-Met
phiSHEF16_038	37,886 - 37,957	72	tRNA-Trp
phiSHEF16_039	38,198 - 38,270	73	tRNA-Cys
phiSHEF16_040	38,391 - 38,464	74	tRNA-Phe
phiSHEF16_042	39,621 - 39,692	72	tRNA-Thr
phiSHEF16_043	39,885 - 39,958	74	tRNA-Lys
phiSHEF16_044	40,093 - 40,166	74	tRNA-Lys
phiSHEF16_045	40,269 - 40,341	73	tRNA-Arg
phiSHEF16_046	40,450 - 40,521	72	tRNA-Val
phiSHEF16_047	40,616 - 40,689	74	tRNA-Ile
phiSHEF16_051	41,845 - 41,915	71	tRNA-His
phiSHEF16_052	42,093 - 42,166	74	tRNA-Leu
phiSHEF16_054	43,266 - 43,352	87	tRNA-Ser
phiSHEF16_056	43,939 - 44,011	73	tRNA-Arg
phiSHEF16_057	44,020 - 44,092	73	tRNA-Ala
phiSHEF16_058	44,150 - 44,216	67	tRNA-Arg
phiSHEF16_059	44,537 - 44,609	73	tRNA-Pro
phiSHEF16_060	44,730 - 44,800	71	tRNA-Gly
phiSHEF16_063	47,666 - 47,881	216	helix-turn-helix XRE-family-like protein
phiSHEF16_064	47,982 - 48,287	306	cystathionine beta-lyase family protein
phiSHEF16_065	48,306 - 49,136	831	ParB N-terminal domain-containing protein
phiSHEF16_067	49,568 - 51,739	2172	ribonucleoside-triphosphate reductase
phiSHEF16_068	51,809 - 52,624	816	deoxynucleoside kinase
phiSHEF16_069	52,643 - 53,407	765	nicotinamide mononucleotide transporter
phiSHEF16_070	53,400 - 53,639	240	redoxin family protein
phiSHEF16_071	53,734 - 54,096	363	HAD-like family phosphatase
phiSHEF16_072	54,182 - 55,312	1131	NadR-like protein

phiSHEF16_074	55,622 - 56,452	831	nicotinamide mononucleotide transporter
phiSHEF16_075	56,467 - 56,934	468	NUDIX hydrolase
phiSHEF16_078	58,404 - 59,285	882	thymidylate synthase
phiSHEF16_080	59,571 - 60,185	615	4Fe-4S single cluster domain-containing protein
phiSHEF16_081	60,282 - 61,151	870	prohibitin family protein
phiSHEF16_090	64,967 - 65,497	531	RNase AS
phiSHEF16_091	65,498 - 66,073	576	metallophosphatase
phiSHEF16_092	66,070 - 66,678	609	metallophosphatase
phiSHEF16_094	66,932 - 67,672	741	metallophosphatase
phiSHEF16_104	72,432 - 72,914	483	helix-turn-helix XRE-family protein
phiSHEF16_105	72,917 - 73,231	315	NAD-dependent DNA ligase
phiSHEF16_110	75,778 - 77,022	1245	tRNA-splicing ligase
phiSHEF16_144	89,513 - 89,761	249	transcriptional regulator
phiSHEF16_155	97,862 - 99,589	1728	PcfJ-like protein
phiSHEF16_156	99,589 - 100,437	849	nucleotidyltransferase
phiSHEF16_163	104,290 - 105,573	1284	metallophosphatase
phiSHEF16_169	108,128 - 109,324	1197	putative Ig-like protein
phiSHEF16_170	109,374 - 109,652	279	holin
phiSHEF16_173	110,747 - 111,994	1248	recombinase A
phiSHEF16_174	112,075 - 113,391	1317	large tegument protein
phiSHEF16_176	114,117 - 116,306	2190	DNA polymerase
phiSHEF16_178	117,032 - 117,874	843	DNA polymerase
phiSHEF16_179	117,956 - 118,267	312	integration host factor
phiSHEF16_184	120,104 - 120,781	678	holliday junction resolvase
phiSHEF16_190	122,564 - 123,445	882	dUTP diphosphatase
phiSHEF16_191	123,455 - 124,510	1056	DNA primase
phiSHEF16_193	125,187 - 127,079	1893	endonuclease
phiSHEF16_194	127,198 - 128,238	1041	metallophosphatase
phiSHEF16_195	128,238 - 129,710	1473	DNA helicase
phiSHEF16_197	131,383 - 133,158	1776	DNA helicase
phiSHEF16_202	142,421 - 143,473	1053	baseplate protein

Appendix 2

Virus Research 327 (2023) 199073

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Enterococcal bacteriophage: A survey of the tail associated lysin landscape

Alhassan M. Alrafaie [a,b,*], Graham P. Stafford [a,*]

[a] Integrated BioSciences, School of Clinical Dentistry, University of Sheffield, Sheffield, United Kingdom
[b] Department of Medical Laboratory Sciences, College of Applied Medical Sciences in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

ARTICLE INFO

Keywords:
Bacteriophages
Prophage
Enterococcus
Tail-associated lysin

ABSTRACT

Bacteriophages are viruses that exclusively infect bacteria which require local degradation of cell barriers. This degradation is accomplished by various lysins located mainly within the phage tail structure. In this paper we surveyed and analysed the genomes of 506 isolated bacteriophage and prophage infecting or harboured within the genomes of the medically important *Enterococcus faecalis* and *faecium*. We highlight and characterise the major features of the genomes of phage in the morphological groups podovirus, siphovirus and myovirus, and explore their categorisation according to the new ICTV classifications, with a focus on putative extracellular lysins chiefly within tail modules. Our analysis reveals a range of potential cell-wall targeting enzyme domains that are part of tail, tape measure or other predicted base structures of these phages or prophages. These largely fall into protein domains targeting pentapeptide or glycosidic linkages within peptidoglycan but also potentially the enterococcal polysaccharide antigen (EPA) and wall teichoic acids of these species (i.e., Pectinesterases and Phosphodiesterases). Notably, there is a great variety of domain architectures that reveal the diversity of evolutionary solutions to attack the Enterococcus cell wall. Despite this variety, most phage and prophage possess a putative endopeptidase (70%), reflecting the ubiquity of this cell surface barrier. We also identified a predicted lytic transglycosylase domain belonging to the glycosyl hydrolase (GH) family 23 and present exclusively within tape measure proteins. Our data also reveal distinct features of the genomes of podo-, sipho- and myo-type viruses that most likely relate to their size and complexity. Overall, we lay a foundation for expression of recombinant TAL proteins and engineering of enterococcal and other phage that will be invaluable for researchers in this field.

1. Introduction

The antibiotic resistance crisis is a major global health threat, with a predicted potential 10 million deaths annually by 2050 if unchecked (O'Neil, 2016). One particular bacterial genus that has contributed to this crisis is the enterococci, Gram-positive bacteria that normally inhabit the gastro-intestinal tract (GIT) of humans as well as other animals such as reptiles, fish and insects (Bondi et al., 2020). GIT colonization by enterococci is achieved through multiple factors such as genome plasticity, nutritional adaption and antimicrobial production and resistance (Banla et al., 2019). As a pathogen, they are commonly involved in human wound infections (Bowler et al., 2001), bacteraemia, endocarditis, urinary tract infection (UTI) (Ben Braïek and Smaoui, 2019) as well as recalcitrant endodontic dental infections (Love, 2001). Until 1984 enterococci were classified as *Streptococcus* but were reclassified into their own genus, with the two main species associated with human diseases being *Enterococcus faecalis* and *faecium* (Schleifer and Kilpper-Balz, 1984). These organisms the first to become resistant to the glycopeptide antibiotic of last resort Vancomycin in 1986- classed as vancomycin-resistant enterococci (VRE) (Uttley et al., 1988). These have now spread worldwide and have been recognized by the World Health Organisation (WHO) as high priority to be targeted and investigated for new antimicrobials (Willyard, 2017).

One area of interest that has seen a resurgence in the context of fighting multidrug resistant pathogens is in the use of naturally occurring antibacterial viruses known as bacteriophage, often simply called phage (Dion et al., 2020). Phages were discovered in the early 20th century by Frederick Twort in 1915 (Twort, 1915) and Felix d'Herelle in 1917 (Terms, 2011) and are likely to be the most abundant biological entities on our planet with estimates as high as 10^{31} globally at any one time (Suttle, 2005). Phages are thus found almost everywhere in our environment, with abundant phage found in soil, wastewater sewage, ocean sediment and the human body.

Much like viruses that infect humans, multiple infectious lifecycles

* Corresponding authors at: Integrated BioSciences, School of Clinical Dentistry, University of Sheffield, Sheffield, United Kingdom.
E-mail addresses: a.alrafaie@psau.edu.sa (A.M. Alrafaie), g.stafford@sheffield.ac.uk (G.P. Stafford).

https://doi.org/10.1016/j.virusres.2023.199073
Received 12 October 2022; Received in revised form 5 February 2023; Accepted 10 February 2023
Available online 22 February 2023
0168-1702/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

exist with the two major ones being designated lytic and lysogenic lifecycles, with temperate phages being those that can follow both lifecycles. The first step in phage infection is the adsorption of phage to the host surface, followed by genome injection into the bacterial cytoplasm. At this stage, lytic phages take over the infected cell to synthesise virions before a critical load is reached in the cell before virions are released and the bacterial cells lysed from the inside-out- using phage (endo)lysins (Abdelrahman et al., 2021). For temperate phages, the phage genome is either integrated into the bacterial genome (known as a prophage) or remains dormant as a plasmid in the bacterial cytoplasm. The phage genome can be induced to allow the temperate phage to then undergo a lytic cycle. For phage therapy, lytic phages are the first choice since they obligatorily lyse their hosts.

During the phage lifecycle, phage must first adhere to the target bacterial cell surface but must also present their DNA-injection machinery in close enough proximity to the host cell membrane. In some cases, this is a particular challenge as the bacterial strain might be coated with capsular material, in the case of enterococci these include enterococcal polysaccharide antigen (EPA) (Guerardel et al., 2020) as well as peptidoglycan and teichoic acids (Silhavy et al., 2010). As a result phage often contain lysins that are often associated with components of the phage virion structural proteins, such as terminal tail structures or tail tape measure proteins. Hence these are often known as Virion-associated lysins (VAL) or Tail-associated lysins (TAL), we will use the latter term here.

Based on the mode of action, TALs can be classified into three main classes: glycosidases, amidases and endopeptidases (Latka et al., 2017) (Fig. 1). The Glycosidases often target the β−1,4 glycosidic bonds in the sugar moiety of peptidoglycan and are divided into three subtypes. First, N-acetyl- β-D-muramidases that cleave the link between N-acetylmuramic acids (MurNAc) and N-acetylglucosamines (GlcNAc). Secondly, N-acetyl-β-D-glucosidases target the bond between GlcNAc and MurNAc. The third subtype are Lytic transglycosylases which require water molecules for lysing MurNAc-GlcNAc linkages. The second class are amidases (N-acetylmuramoyl-L-alanine amidases) that cleave the bond between MurNAc and the first amino acid (L-alanine) in the peptide stem. Finally, the third class are endopeptidases which cleave the peptide bond either within the interpeptide bridge or stem peptide of peptidoglycan (Elbreki et al., 2014).

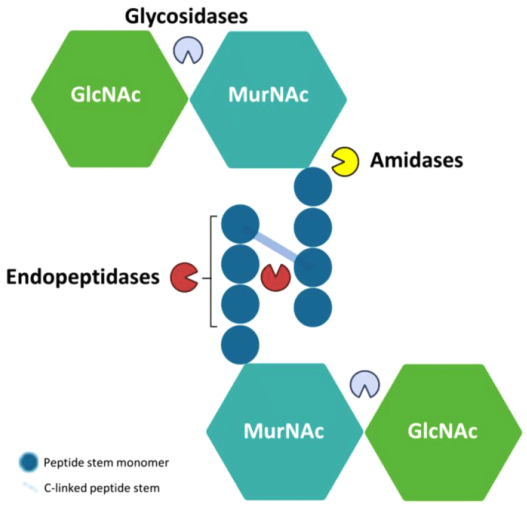


Fig. 1. Schematic representation of the bacterial peptidoglycan structure showing target locations of Glycosidases (grey), Amidases (yellow) and Endopeptidases (red). MurNAc: N-acetylmuramic acids GlcNAc: N-acetylglucosamine. This figure was created with Biorender.com.

In this study we sought to explore the potential TAL landscape of sequenced enterococcal phage, as well as predicted prophage from the genomes of sequenced enterococcal strains of *E. faecalis* and *E. faecium* with a view to potentially using this information to design novel antimicrobials in the future. These studies are made possible since phage genomes are efficiently ordered with functional modules clustered in their genomes (Moura de Sousa et al., 2021). Examples of these modules include packaging, head, tail and lysis modules. Of the over 20,000 sequenced phage genomes, to date 162 infect enterococcal strains (October 2022), but no investigations of the TAL landscape for these have been undertaken to our knowledge. Therefore, we investigated TALs in both enterococcal phage and prophage genomes which resulted in identifying various TALs targeting different bacterial layers. Our work also highlights previously unreported unique features for several enterococcal TALs.

2. Methods

2.1. Phage and prophage genomes

One hundred complete enterococcal phage DNA genomes available on the NCBI GenBank database were obtained as Genbank and Fasta sequences (up to 11/10/2020). The search for these genomes was done on the NCBI virus portal by using "bacteriophage" for virus choice, "Genbank" sequence type, "complete" for genome sequence and "Enterococcus" for host. The GenBank accession numbers of these genomes are included in supplementary file 1. For prophage genomes 203 complete *E. faecalis* and *E. faecium* bacterial genomes available on the NCBI GenBank database were obtained (up to 10–10–2020), accession numbers are included in supplementary files 2 and 3. The online web server PHASTER (Arndt et al., 2016) was used to identify putative intact prophages. All phage and prophage genomes were re-annotated to ensure annotation consistency using RASTtk (new version of Phage Rapid Annotation using Subsystem Technology (RAST) pipeline (Brettin et al., 2015).

2.2. TAL identification and analysis

The tail module was identified between the head and lysis modules in most of the phage and prophage genomes based on RASTtk annotation. All tail proteins were checked for TALs using Pfam and NCBI conserved domains (CDD) databases. Structural analysis was also performed using the PHYRE2 webserver (Kelley et al., 2015). SnapGene (v 5.3.2) and Artemis (Carver et al., 2012) were used for genome visualisation. The identified TAL proteins were aligned using ClustalW (genome.jp) and MultAlin webservers (Corpet, 1988). Phylogenetic trees were constructed using FastTree (genome.jp) and visualised using the ITOL online website (Letunic and Bork, 2021) (https://itol.embl.de/). To check putative peptidase classifications, the MEROPS database was employed (http://www.ebi.ac.uk/merops/) (Rawlings et al., 2018) while the CAZy (Carbohydrate Active Enzymes) database (CAZy; http://www.cazy.org) (Lombard et al., 2014) was used for predicted Glycoenzymes. Genome size analysis was performed using GraphPad Prism version 7,San Diego, California USA, www.graphpad.com.

3. Results and discussion

Using the NCBI virus portal, the genome sequences of 100 phage targeting either *Enterococcus faecalis* or *Enterococcus faecium* were collected for analysis. These comprised 86 phages isolated using *E. faecalis* strains while 10 phages were isolated using *E. faecium* strains. For predicted prophages, a total of 203 *E. faecalis* & *E. faecium* complete genomes were scanned using the PHASTER prediction tool with the default parameters set for classification of "intact" prophage set at

(>90%), "questionable" (scoring 70–90%) and "incomplete" (scoring<70%). In this study we focused on the intact prophages as these have the highest confidence level to maintain a full set of functional modules and allow tail module identification. The PHASTER searches revealed 406 intact prophages in both *E. faecalis* and *E. faecium* bacterial genomes, meaning that in total this study examined 506 phage and prophage genomes.

3.1. Isolated phage genomes for Enterococcus are members of a range of viral classes

The 100 isolated phage genomes showed a variation in size from 16.9 to 156.5 kb (Fig. 2A). Our analysis showed that the phage genomes can be categorised into three main groups based on genome size and phage virion morphology. Phages with small genomes (<30.5 kb) are generally podoviruses (Rountreeviridae), medium sized genomes (31–86.3 kb) siphoviruses (Efquatrovirus, Phifelvirus, Saphexavirus and Andrewesvirinae), and large genomes of over 130 kb myoviruses (Herelleviridae) (Fig. 2A). There were some exceptions, for example, the smallest *Enterococcus* phage EFRM31 (16.9 kb) is a siphovirus (unclassified according the current ICTV classification) with an isometric head and long non-contractile tail (206 nm tail length), whose genome has 35 predicted ORFs (Open Reading Frames) (Fard et al., 2010). Other examples of note within the podoviruses (Autographiviridae) are the *E. faecalis* phages EFA-1 (40.7 kb) and EFA-2 (39.9 kb) which have higher GC contents and number of ORFs (EFA-1 is 50.14%, 52 ORFs and EFA-2 is 48.55%, 49 ORFs) compared to the average value for enterococcal podoviruses (Autographiviridae and Sarlesvirinae) analysed in this study (35.1%, 30 ORFs).

Regarding morphology, all the 100 enterococcal phages are predicted morphologically to be either podoviruses (short tailed), siphoviruses (long non-contractile tail) or myoviruses (contractile tail) based on database entries. While the morphological categorisation of podo- myo- and sipho-virus has been widely used for many years, the recent increase in genomic information has identified a number of differences and allowed continual improvement of phage taxonomy (Turner et al., 2021). However, we will in some places use the commonly used morphological terms to simplify discussions. All of the 18 small genomed predicted podoviruses are classified as Copernicusvirus or Minhovirus within the Rountreeviridae or belong to Autographiviridae according to the new ICTV classifications and have a genome size of 17.9 to 40.7 kb (Fig. 2A) (Turner et al., 2021). The number of ORFs encoded in these genomes ranged from 22 to 52 with an average of 30.

Siphoviruses make up 64% of the isolated phage with genomes ranging from 16.9 to 86.3 kb. Based on the genome size and TAL analysis, siphoviruses can be classified into two groups: group-1 (21–43 kb, Efquatroviruses or Phifelviruses) and group-2 (55–86 kb, Saphexavirus or Andrewesvirinae) (Fig. 2A). The average number of predicted ORFs in group-1 is 62 while this is 104 for Group-2. Lastly, we analysed 18 myovirus-type genomes (Herelleviridae, Schiekvirus or Kochikohdavirus) where the genomes varied from 130.9 to 156.5 kb (average 146.5 kb). The new classifications are further supported by these data since the genome sizes alone can indicate likely species membership according to our data.

Unsurprisingly, a positive correlation was seen between phage genome size and the number of ORFs with the small podoviruses having the lowest number of ORFs and the largest genomes (myoviruses) the highest ORFs number (Fig. S1A). The number of tRNAs also shows a positive correlation with the genome size, with podovirus genomes having no tRNAs while larger genomes of siphoviruses and myoviruses contain several putative tRNA genes (Fig. S1B). In contrast, there is no clear correlation between the genome size and GC content (Fig. S1C).

Of the 100 enterococcal phage genomes we identified several temperate phages, based on the presence of integrase and repressor genes that are necessary for phage integration and maintenance during the lysogenic cycle into the bacterial genome. In our data, 16% of viruses are likely to be temperate, as they contain integrase and/or repressor genes (Fig. 2A). Of these temperate phages, only one is reported to be pseudotemperate, and was identified in the genome of *E. faecalis* 62 and contains a toxin-antitoxin system (Brede et al., 2011).

3.2. Enterococcal prophages

For prophages, only predicted intact prophage genomes were chosen and analysed. In our study, a total of 406 putative intact prophages (93 from *E. faecalis* & 313 from *E. faecium* genomes) were identified, with the most in one genome being five from 203 genomes that were scanned. These showed large variation in the predicted genome size (6.9–91.1 kb) (Fig. 2B), with the smallest prophage containing 10 ORFs and the largest 121 ORFs with an average GC content of 35.9%. It is worth mentioning here that not all identified intact prophage genomes possessed all necessary genes to complete phage lifecycle indicating a limitation of the PHASTER webserver. Our analysis also showed a positive correlation between the number of ORFs and prophage genomes size (Fig. S2A). The analysis of the number of tRNA genes showed no correlation (Fig. S2B).

3.3. Five types of predicted tail associated lysins exist in enterococcal phage genomes

Phage genomes are generally organised in modules where related functional genes are grouped together such as packaging, head, tail and lysis functions. For example, the tail module of siphovirus type phage is considered to generally comprise of three main genes in the following order: "Tape measure protein"- TMP, "Distal tail protein"- Dit and "Tail associated Lysin"- Tal (Goulet et al., 2020) (Fig. 3B). In this study, the term "TAL" means any lysin in the tail module while "Tal" is referring to the third gene in the siphovirus tail unit.

The TMP is usually one of the longest genes in phage genomes, and plays a role in controlling tail length, with the length of its translated protein approximately indicating the length of the phage tail (1 amino acid= 0.15 nm) (Mahony et al., 2016). The TMP also helps facilitate genome ejection toward the bacterial cytoplasm, although mechanistic details are unclear (Mahony et al., 2016). This is evidenced by identifying domains with potential cell wall degrading function in TMPs as well as DNA-binding domains (Piuri and Hatfull, 2006; Stockdale et al., 2013). TMPs are thought to be located in the lumen of the tail and interact with termination and initiation proteins as well as the polymeric Major Tail Protein (MTP)(Cornelissen et al., 2016; Kizziah et al., 2020).

The Dit is part of the baseplate and connects the tail with the tail tip as well as providing in some cases the site of attachment for a RBP "receptor-binding protein"- which may be housed on a fibrous protein (Kizziah et al., 2020). The RBPs are responsible for the specific recognition of bacterial receptors that may include outer membrane proteins, bacterial capsule, teichoic acids, pili and flagella (Bertozzi Silva et al., 2016; Letarov and Kulikov, 2017).

In this study we assessed all genes in the putative tail modules (not only the putative Tal) for the presence of lysin- like domains, so as not to exclude any that might be associated directly with TMPs, RBP or tail fibres since many lysins used by phages in the first steps of phage infection are associated with the tail and baseplate structure (Latka et al., 2017). After obtaining and reannotating the enterococcal phage and prophage genomes, all the predicted tail genes were scanned for the presence of predicted lysin domains by using Pfam, the NCBI Domain database and the Phyre2 webserver. As a result, multiple types of lysins were identified in both phage and prophage genomes (Table 1) (Fig. 3A).

Our analysis showed that presence of a predicted endopeptidase is the most common lysin associated with tail proteins (70.4%), while lytic transglycosylase domains were present in 18.0% of the total identified lytic proteins. These two types of lysins are preferentially carried by phage infecting Gram-positive bacteria (São-José, 2018). Other proteins were also observed to carry other potential lysins, namely peptidases of the NLPC/P60 family (6.2%), GDPD (4.0%) and lastly Pectinesterases

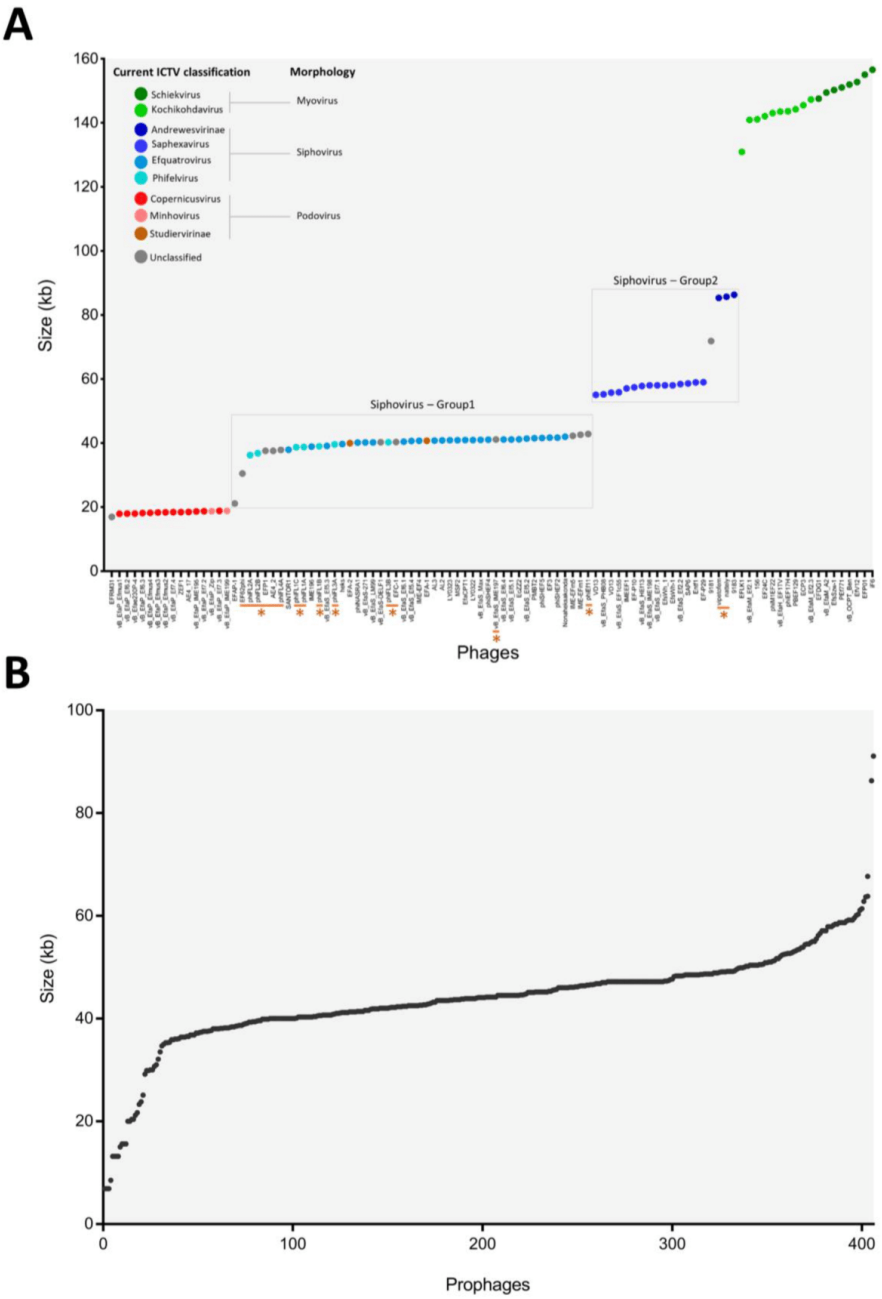


Fig. 2. (A) 100 enterococcal phage genomes were plotted against genome size. The genomes are labelled in accordance with the new ICTV classification as follow: Schiekvirus (dark green), Kochikohdavirus (light green), Andrewesvirinae (dark blue), Saphexavirus (blue), Efquatrovirus (Azure), Phifelvirus (sky), Copernicusvirus (red), Minhovirus (orange), Studiervirinae (brown),. The grey colour indicates unclassified genomes regarding the current ICTV classification and further details are included in the supplementary file1. Phage morphologies are also included according to the ICTV classification. Temperate phages are underlined and labelled with asterisks. (B) 406 intact prophage genomes were plotted against genome size. The genomes are in ascending order in both figures.

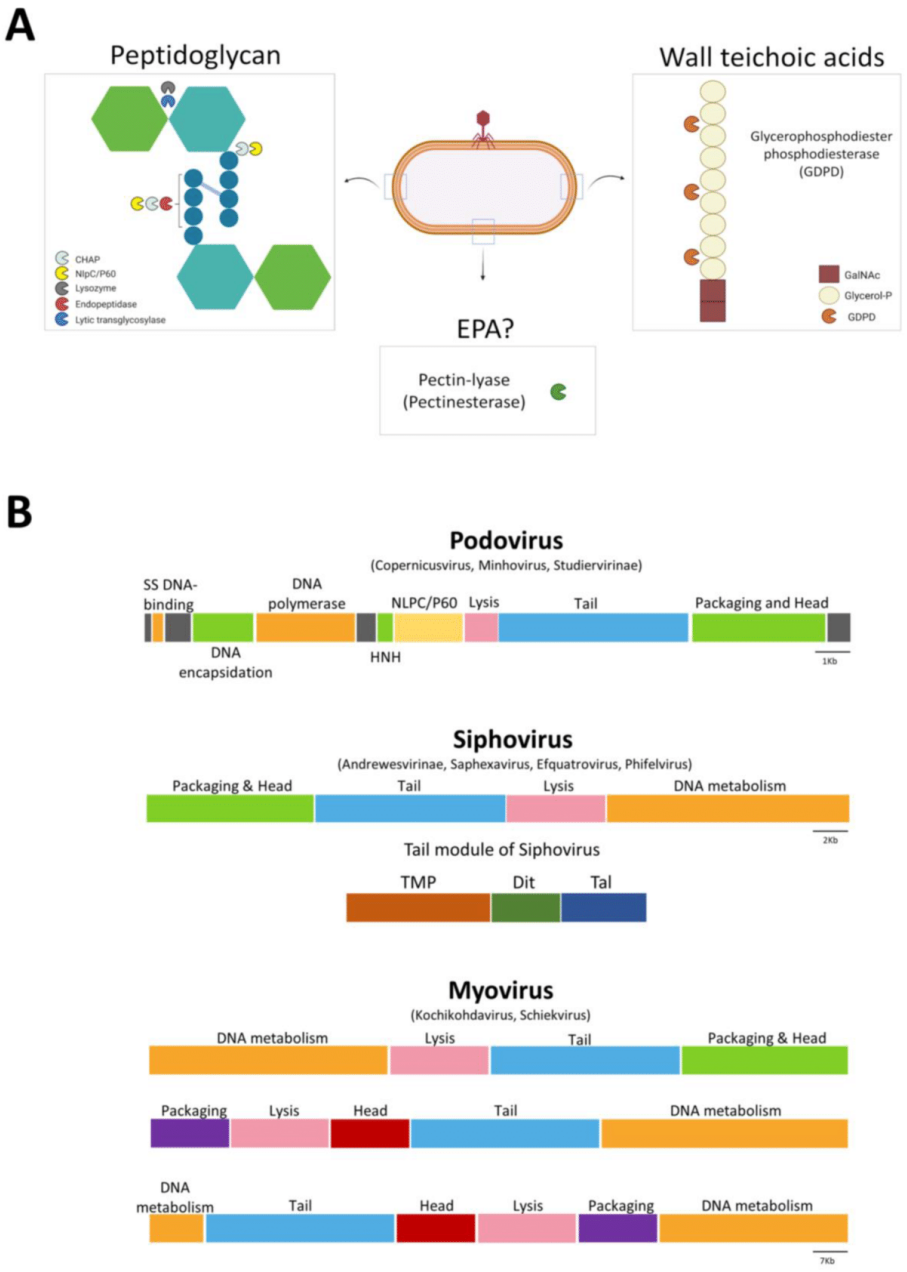


Fig. 3. (A) Schematic representation of the identified lysins and their targets in this study. This part was created with Biorender.com. (B) The order of the Functional modules of the enterococcal Phage genomes. Modules and specific genes are coloured as follow: Green= DNA packaging and head, Red= Head alone, Purple= Packaging alone, Blue= Tail, Pink= lysis, Orange= DNA metabolism. Of note, podovirus general genome organisation contains some labelled gene like yellow= NLPC/P60 gene. HNH stands for homing endonuclease. The general scheme of the siphovirus tail module is also drawn Brown=TMP, Dark green= Dit and Dark blue=Tal. The new ICTV classification is also indicated regarding each phage morphology.

Table 1
Summary of predicted lytic domains associated mainly with the tail module of our study set.

	Domain	Activity	# Sequences (% total 544)
Lysins	Endopeptidase	Endopeptidase	383 (70.4%)
	Lytic transglycosylase	Lytic transglycosylase	98 (18.0%)
	NLPC/P60	Endopeptidase or Amidase	34 (6.2%)
	Glycerophosphodiester phosphodiesterase (GDPD)	Phosphodiesterase	22 (4.0%)
	Pectinesterase	Pectinesterase	7 (1.3%)

(1.3%). Each one of these lysins are further discussed in the following sections.

3.3.1. *Enterococcal endopeptidases (TAEP) display a range of domain architectures*

The **t**ail proteins **a**ssociated with **e**ndo**p**eptidase activity (TAEP) were identified in both phage and prophage genomes. We identified 383 TAEP proteins via homology with predicted phage endopeptidase domains (Pfam: PF06605). These TAEP proteins were then assessed for domain architectures (DA). This revealed 5 different groups (DA-EP), all containing a phage endopeptidase domain (Pfam: PF06605) located at the N-terminal end of the predicted protein (Fig. 4). These domains are all found in the Tal position (i.e. TMP-Dit-Tal), although it is not clear if they have an endopeptidase activity themselves or are involved in forming active complexes or act in a structural manner. Catalytically, endopeptidases target peptide bonds within peptidoglycan- either in the peptide stem or cross-bridge. Of our identified TAEP proteins 60.5% are within the DA1 architecture group and only contain an endopeptidase domain (Fig. 4). This has also been noticed previously as most Tal proteins harboured a single lysin (Latka et al., 2017). To further analyse the endopeptidase domains, the MEROPS database was used to check the peptidase family of these sequences. To do this, three representative sequences from each DA were screened against the MEROPS_scan dataset, resulting in highlighting two types (M23B and C104) with high E-value ($<10^{-10}$).

The other DAs showed various lysin domains in addition to the endopeptidase domain. In DA2,3 and 4 a predicted lysozyme domain (Pfam: PF18013) was identified which is a structural homologue of a cell wall degrading enzyme in the bacteriophage phi29 tail (established using Phyre2 analysis) (Xiang et al., 2008). Besides the phage tail lysozyme, DA2 contains a CHAP domain (cysteine, histidine-dependant amidohydrolases/peptidase), while DA3 harbours a peptidase M23 domain (thought to target the peptide bonds in the peptidoglycan layer (Vermassen et al., 2019). DA4 also contains predicted amidase domains likely attacking the amide bond between MurNAc and the first amino acid L-alanine leading to separation of the glycan and peptide units. Finally, DA5 contains a domain with homology to endosialidase chaperones. Of these domains, all have been associated with cell wall degradation or in the case of DA5- stabilisation of other catalytic domains. For example, CHAP domains (Pfam:PF05257) have been shown to act as endopeptidases (e.g. LysK CHAP) (Becker et al., 2009) or amidases Proença et al., 2012). The Peptidase M23 domains (Pfam: PF01551) in DA3 are located at the C-terminal region as well as the predicted amidase domains (Pfam: PF05382) in DA4. The chaperone of endosialidase in DA5 has shown to facilitate the folding and assembly of endosialidases and other phage proteins as well, as it is eventually cleaved off to ensure the stability of the native protein (Schwarzer et al., 2007).

3.3.2. *Enterococcal new lipoprotein C/Protein of 60-kDa (NLPC/P60) are grouped based on phage morphology and genomic classification*

It is known that many phage proteins contain domains belonging to the NLPC/P60 family (New Lipoprotein C/Protein of 60-kDa). The NLPC/P60 family is a large group of papain-like cysteine proteases present in bacteria, like *Escherichia coli* (NLPC) and *Listeria monocytogenes* p60 (Anantharaman and Aravind, 2003). The members of the NLPC/P60 family can have (endo)peptidase as well as other activities such as amidase, transglutaminases and acetyltransferase and often contain a conserved catalytic N-terminal cysteine and C-terminal Histidine residue (Anantharaman and Aravind, 2003). In bacteria the NLPC/P60 peptidases are likely to be involved in the bacterial cell cycle and morphogenesis by hydrolysing the peptidoglycan layer while in phage they likely aid in local peptidoglycan degradation and hence promoting genome injection (Fukushima et al., 2018; Griffin et al., 2022).

Our analysis identified 34 tail proteins that appear to be within the NLPC/P60 family (accession no.cl21534). Based on a phylogenetic tree made from amino acid alignments, these 34 sequences are classified into two main groups (Fig. 5B). Group1 includes sequences from podovirus-type genera while group 2 are from predicted myovirus subtypes.

Of these, Group 1 can be divided into three subgroups: 1A (Genus: Copernicusvirus) consists of 13 sequences which represent DA1 (Fig. 5A). The 1B subgroup includes only one sequence (EF62phi) which is clearly diverged from the 1A sequences and contains an additional

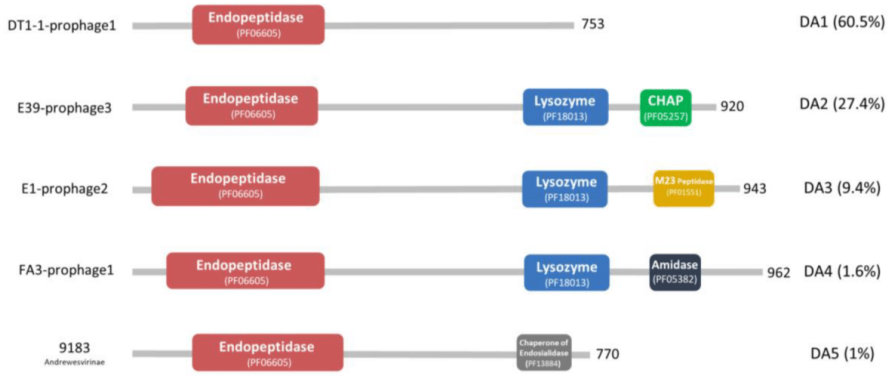


Fig. 4. Domain architectures of TAEP proteins based on Pfam. Five DAs are shown with coloured domains. Red= endopeptidase, Blue= lysozyme, Green= CHAP, Yellow= M23 peptidase, Dark blue= amidase, Grey= chaperone of endosialidase. The left side contains the phage or prophage's name and ICTV classification while the right side contains the DA type and its abundance in percentage. The length of the protein is also indicated on the right side.

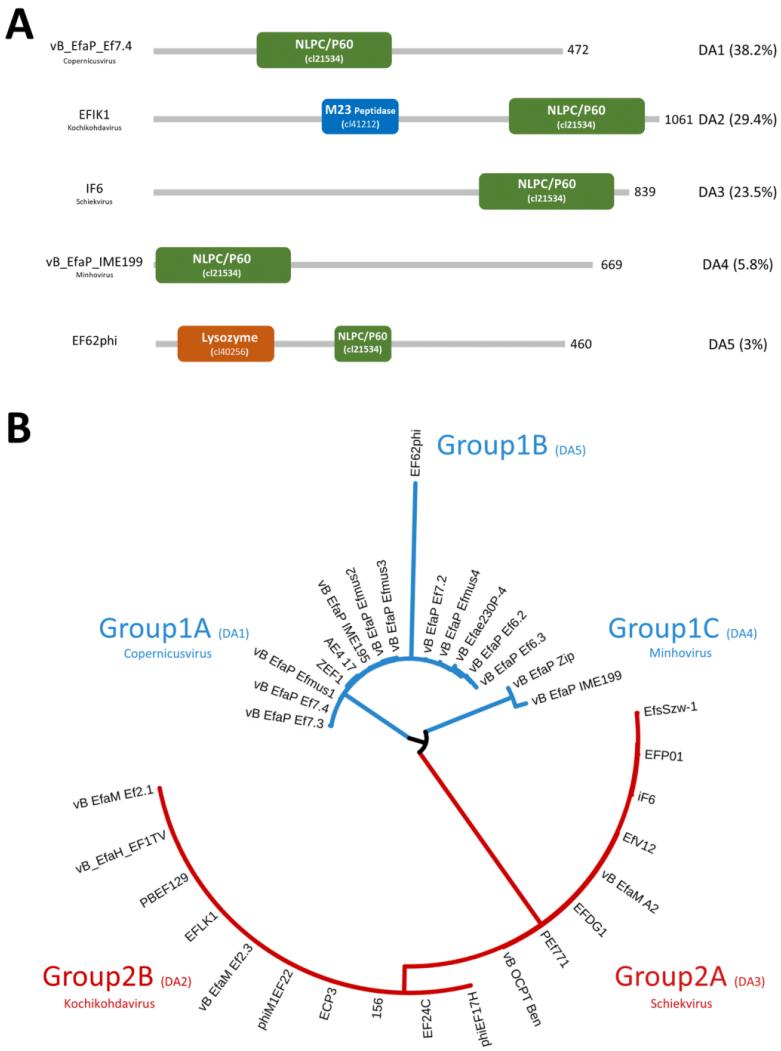


Fig. 5. (A) Domain architectures of NLPC/P60 containing proteins. Five DA are shown with coloured domains. Green= NLPC/P60, Blue= M23 peptidase, Orange= lysozyme. The domain type and abundance (%) are indicated on the right side while phage or prophage name and the ICTV classification is on the left side. The length of the protein is also indicated. (B) a phylogenetic tree of all identified NLPC/P60 containing proteins show two main groups based on phage genomic classification and morphology: sequences from podoviruses labelled Blue while myoviruses labelled red. The tree was constructed using FastTree and visualised using the ITOL online website.

lysozyme domain as shown in DA2. The 1C subgroup (Genus: Minhovirus) consists of sequences from podoviruses that were isolated using *E. faecium* strains in contrast to subgroups 1A and 1B (host strains are *E. faecalis*) as well as the NLPC/P60 domain at the N-terminal region as represented in DA4 (Fig. 5A).This may indicate differences in substrate specificity given the differing crosslinks between these spp. (Lys-Ala-Ala-for *E. faecalis* and Lys-Asx for *E. faecium*) and is the subject of current work in our lab (Arbeloa et al., 2004)

For group 2, these NLPC/P60-containing proteins were all found in myoviruses (Herelleviridae). This group can also be further subdivided based on the phylogenetic tree and domain architecture into subgroups: 2A and 2B (Fig. 5B). The 2A subgroup (Genus: Schiekvirus) contains only the NLPC/P60 domain, while the 2B subgroup (Genus: Kochikohdavirus) contain an additional M23 peptidase domain besides the NLPC/P60 domain.

Overall, we have revealed the presence of a range of lysins in tail modules of enterococcal phage, many of which that contain multiple lysin domains can be attributed to the need to degrade the various cell wall components and might work together to facilitate entry into enterococci. Finally, our data seem to indicate that the putative lysins group according to viral genus classification, suggesting shared function, but whether this relates to host-range is yet to be established.

3.3.3. Tailed enterococcal phage contain tail tape measure proteins with lytic transglycosylase domains (TMP-LT)

Our next step of analysis analysed putative proteins containing Lytic transglycosylases (LT) domains, enzymes that degrade the peptidoglycan layer by cleaving the β−1,4-glycosidic bond between N-Acetylmuramic acid (MurNAc) and N-Acetyl-D-glucosamine (GlcNAc) (Holtje et al., 1975). Of the 98 LT detected, 97 are contained within putative Tail tape-measure proteins (TMPs) from the genomes of viruses with contractile or non-contractile tails. These contained a broad variety

of predicted domain architectures but all had the LT domain at the C-terminal end (Fig. 6A). The TMP proteins are usually the longest proteins in the phage genomes (Piuri and Hatfull, 2006) and the predicted length here varied from 1180 to 2254aa (Fig. 6A). Moreover, our analysis found that the location of the TMP-LT proteins in predicted enterococcal siphovirus genomes is always the same (i.e. TMP-Dit-Tal) (Goulet et al., 2020). Other studies from phage infecting other species have also identified LT within the TMPs (Piuri and Hatfull, 2006; Stockdale et al., 2013). In our study, we do not see any other putative lysins than LT in TMP proteins.

Of note, the N-terminal region of the analysed TMPs often includes domains putatively involved in DNA binding or cleavage such as SCP-1, SMC,endonuclease and SbcC (Fig. 6A). This coincides with the putative proposed function of TMP as facilitating DNA delivery and injection into bacterial cells (Mahony et al., 2016).

Since lytic transglycosylases are carbohydrate targeting enzymes, the CAZy database was used to reveal that all the identified TMP-LT proteins belong to the specific glycosyl hydrolase (GH) family 23. The GH23 family includes lysozyme type G (EC 3.2.1.17), peptidoglycan lytic transglycosylase (EC 4.2.2.n1) and chitinase (EC 3.2.1.14). Amino acid sequence alignment and consensus analysis of our TMP-LTs revealed the presence of the GH23 conserved Glutamic acid (E) active site proton donor (Fig. 6B, red arrow). Previous studies assigned LT enzymes into 8 families based on sequence motifs (Dik et al., 2017). Our identified TMP-LT sequences shared motifs with family 1A: motif I includes the catalytic residue E-S, motif II contains the G-L-M-Q residues, motif III consists of A/G-Y-N residues and motif IV is a conserved Y residue flanked by a hydrophobic residue (Fig. S3A) (Dik et al., 2017). Indeed others have reviewed LTs and noted that the GXXQ of motif II is conserved amongst GH23 enzymes (Blackburn and Clarke, 2001; Dik et al., 2017; Scheurwater et al., 2008; Wohlkönig et al., 2010). Our data also revealed novel conserved motifs in the identified TMP-LT sequences that are not present in the family 1A i.e. T_{46}, F_{47}, $G_{54}I_{59}$, L_{67}, A_{68} (Fig. 5B). Of note, the enterococcal phage LTs examined here contain extra conserved residues not present in the LTs in the literature- from either other Gram positives, Gram negative bacteria or phage from Gram negatives. Hence, we propose a new family that we label 1P (for phage) (Fig. S3A).

Outside of the TMP-LTs discussed here, one outlier was found, this time in a predicted Studiervirinae genome EFA-2 (39.9 kb). EFA-2 has a large genome and an unusual genome organisation compared with the Sarlesvirinae podovirus genomes (Fig. 3B). However, the longest gene in this genome showed a predicted GH23-LT domain which is unusually located at the N-terminal end as opposed to the TMP-LTs which have the LT at the C-terminal end. CAZy database analysis showed that this LT has closest homology with Gram negative infecting phage lysins such as the *E. coli* T7 phage gp16 lytic transglycosylase protein (97% aa similarity) and is likely a member of family 1E as described by Dik et al. (2017) (Fig. S3B).

3.3.4. Prophage pectinesterases potentially targeting EPA

After analysing the enterococcal phage and prophage genomes, seven predicted tail proteins from prophage, but none from lytic phage, were found to harbour a Tail-associated pectinesterase domain. Pectinesterases are enzymes that target pectin via demethylation of galacturonosyl residues (Reid, 1950). Pectin is a main component of plant cell walls and is made up of three types: a homopolymer of galacturonic acid, and two forms of a rhamnogalacturonan (RG-I and RG-II) made up repeating Gal-Rha disaccharides (Mohnen, 2008). Importantly, in enterococci, the cell wall contains a specialised polysaccharide called enterococcal polysaccharide antigen (EPA) that is made up of repeating rhamnose units, interspersed with other sugars and decorated with various modifications (Dale et al., 2017; Guerardel et al., 2020; Rigottier-Gois et al., 2015; Teng et al., 2009). Therefore, we hypothesise that the EPA structure in enterococci could be the target of these phage pectinesterases.

The pectinesterase domains in the seven sequences were located approximately in the central region of the proteins and no other predicted domains were identified (Fig. 7A). The pectinesterase genes were all located right after the common Tal position in siphovirus-type genomes (Efquatroviruses, Phifelviruses, Saphexavirus or Andrewesvirinae)(i.e. TMP-Dit-Tal) and of note were also present in concert with TAEP and TMP-LTs in 4 prophage genomes (Fig. 7B). To further confirm our annotation, structural homology using Phyre2 was performed which identified structural homologues in pectinesterase 1 or rhamnogalacturonan lyase families indicating that these putative genes may well be novel phage pectinesterases or EPA targeting enzymes. Similar pectinesterase/pectin lyase domains were also found in other phages targeting *Klebsiella pneumoniae* (Li et al., 2021; Pertics et al., 2021) and *Acinetobacter baumannii* (Shahed-Al-Mahmud et al., 2021) which all have showed a depolymerase activity upon expression and purification.

3.3.5. Glycerophosphodiester phosphodiesterases (GDPD) often found in tail modules associated with TAEPs

The final type of predicted lysin observed are glycerophosphodiester phosphodiesterases (GDPD), confirmed using Pfam, NCBI domain database and Phyre2. These GDPD enzymes can degrade the phosphodiester bonds holding wall teichoic acids to sn-glycerol 3-phosphate (Gro3P) and their corresponding alcohol (Cornelissen et al., 2016). Our analysis revealed 22 gene predictions carrying the GDPD domain in both phage and prophage genomes. The GDPD proteins display three domain architectures (DA-PD)(Fig. 8A). The first and most common DA-PD (59.1%) contains only a GDPD domain (PF03009.17) with most sequences having a protein size of around 240aa. The second DA harbours a GDPD domain and a predicted membrane domain (PF10110.9) (Fig. 8A) homologues of which have found in *Streptococcus* bacterial genomes (Chuang et al., 2015). The third DA contains the GDPD domain at the C-terminus with a predicted baseplate upper protein (BppU) located at the N-terminal end indicating that this is likely a multifunctional baseplate-lyase protein in phage 9183. Some of these proteins were found within the tail module (e.g., phage 9183 (Andrewesvirinae), vB_EfaS_IME197, SRCM103470-prophage2 (Fig. 8B) while others were spotted throughout the genomes (e.g., BA17124, E39 and E745 prophages). It is also of note here that all the GDPD seen in the tail modules were in concert with a TAEP protein, suggesting potential synergy (Fig. 8B).

The GDPD activity of other phage encoded enzymes been investigated and showed that five conserved residues are required: 2 catalytic Histidines that act as a general acid and general base in catalysing the hydrolysis of the 3′−5′ phosphodiester bond (Rao et al., 2006; Shi et al., 2008) and 3 divalent metal-ion-binding residues (2 Glutamic acid residues and an Aspartic acid residue) (Cornelissen et al., 2016; Shi et al., 2008). In our analysis, the alignment of the enterococcal GDPD domains showed the presence of these highly conserved residues (Fig. 8C).

3.4. Common patterns in arrangement of lysins within genomes follow viral types

During our analysis we noted several patterns in the arrangement of potential lysins within phage genomes and specifically their tail modules. For the siphoviruses, the tail proteins usually follow the TMP-Dit-Tal tail order (Goulet et al., 2020), and this is replicated in the enterococcal phage analysed here. Notably we observed that the type of lysin identified correlates with the phage genome size. Specifically, the smaller genome group (1, 21–43 kb, Efquatroviruses or Phifelviruses) contains TAEP proteins as well as Tape-measure LTs (TMP-LT) (Fig. 9B) while the larger siphoviruses (Saphexavirus or Andrewesvirinae) (55–86 kb) have only a single predicted protein with a lytic domain (TAEP) (Fig. 9C). Despite not identifying the Dit protein bioinformatically, in many cases we observed small Hypothetical proteins (HP) that we assume is the Dit protein in these phages. For the myoviruses (Schiekvirus or Kochikohdavirus), a TMP-LT protein and another

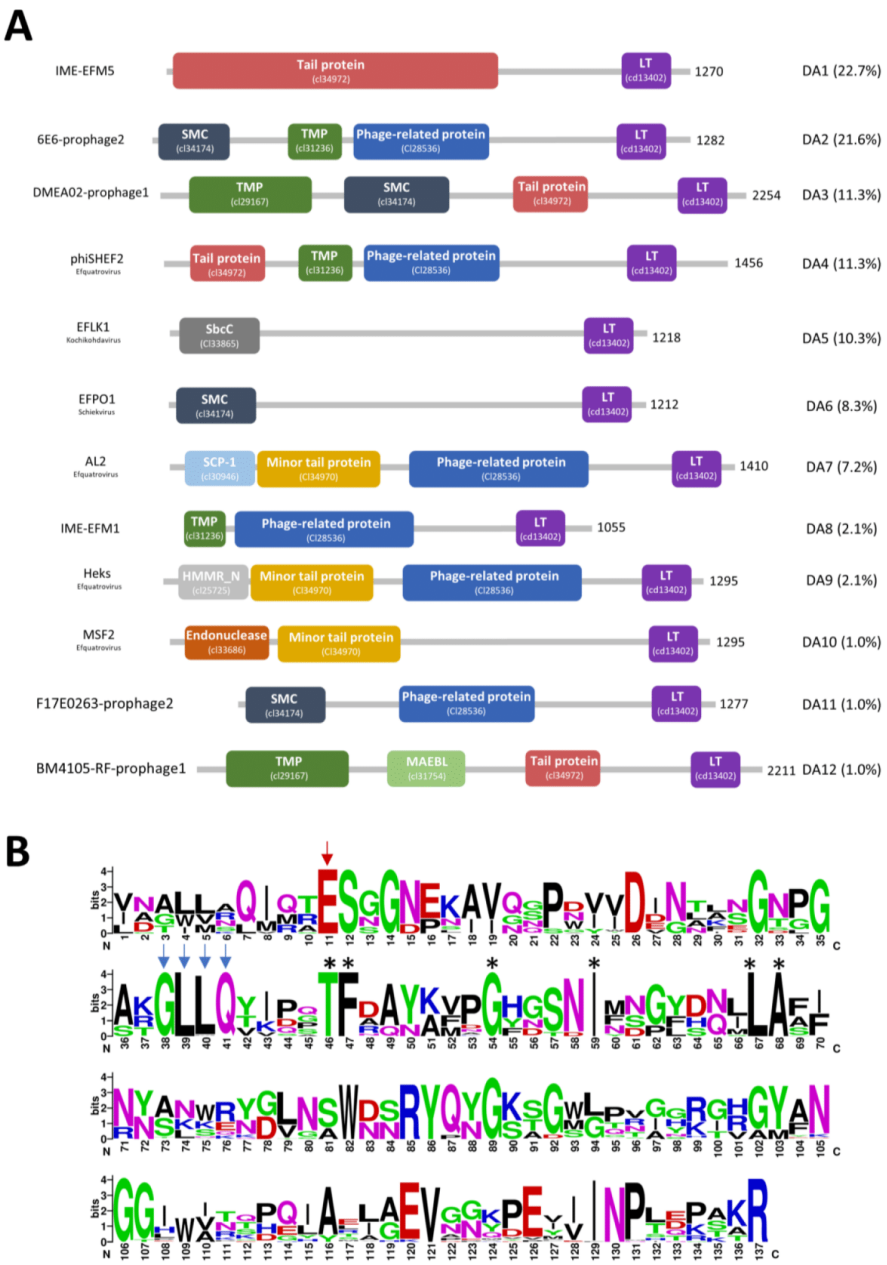


Fig. 6. (A) Domain architecture of TMP-LT proteins. Based on CDD, 12 DA are shown with coloured domains. Purple= lytic transglycosylase-like domain (LT), Green= tape measure protein domain (TMP), Dark blue= structural maintenance of chromosomes (SMC), Red= Tail protein, Sky blue= Synaptonemal complex protein (SCP-1), Orange= Minor tail protein, Dark Grey=SbcC, Blue= Phage-related protein, Light green= Merozoite Apical Erythrocyte Binding-ligand (MAEBL), Brown= Endonuclease, light grey= Hyaluronan mediated motility receptor N-terminal (HMMR_N). The domain type and abundance (%) are indicated on the right side while phage or prophage name and the ICTV classification is on the left side. The asterisks indicate conserved residues specific for analysed sequences. The length of the protein is also mentioned. (B) Weblogo of the TMP-LT domains showing highly conserved domains including the catalytic residue Glutamic acid (red arrow) and GH23 specific motif (Blue arrows). Sequence logos were created using Weblogo (http://weblogo.berkeley.edu/logo.cgi).

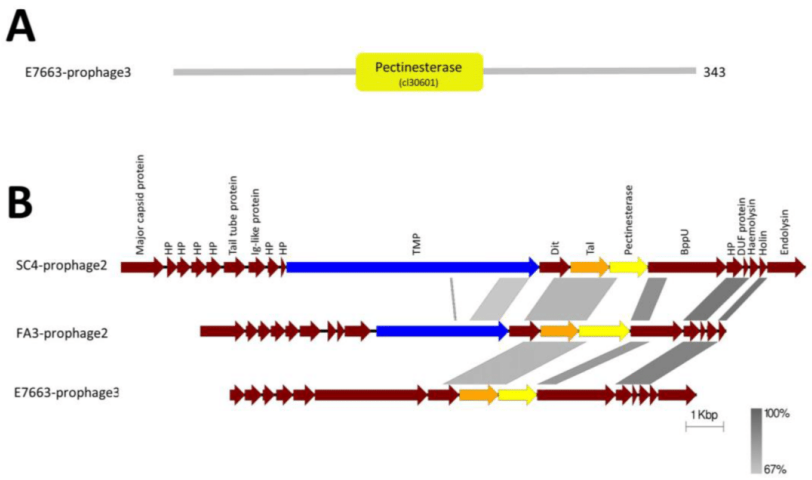


Fig. 7. (A) Domain architecture of a Pectinesterase containing protein from E7663-prophage3 genome based on NCBI domain database which showed Pectinesterase domain. (B) MSA of tail modules showing Pectinesterase protein (yellow), TMP-LT (Blue), TAEP (Orange). HP= Hypothetical proteins. The level of identity is indicated by the grey region between genomes.

adjacent tail protein containing an NLPC/P60 domain were identified in all analysed genomes (Fig. 9D).

Finally, enterococcal podoviruses (Copernicusvirus or Minhovirus), contain a morphology where the head seems to be connected to a baseplate with what one assumes is an infection system (and no tail measure protein is present). Hence, it is unsurprising that they do not display the TMP, Dit, Tal paradigm. Our data indicate that adjacent to the head and endolysin/holin pair most podoviruses genomes contain a predicted NLPC/P60 family protein (Fig. 9A). The location of this protein is highly conserved amongst the analysed podovirus genomes and is likely part of a potential tailspike protein (unpublished data, personal communication, GP Stafford).

As discussed earlier, we have observed a correlation between phage genome size and phage morphology (i.e. small genomes are usually podoviruses while larger genomes are myoviruses). However, the exceptions to this correlation such as the siphovirus phage EFRM31 (16.9 kb) showed no TAL while the siphovirus phage EFAP-1 (21.1 kb) contains both the TAEP and TALT proteins. For podoviruses, phage EFA-2 (39.9 kb) contains an LT containing protein (largest protein in the genome), which could also act similarly to a TMP.

Enterococcal prophages genomes seem to conform to the pattern of siphovirus (Efquatroviruses, Phifelviruses, Saphexavirus or Andrewesvirinae) type tail modules (TMP-Dit-Tal) (Fig. S4). For TAL, the majority (86.7%) of the prophage genomes have the endopeptidase TAEP in the Tal position, i.e. TMP-Dit-Tal(TAEP). The other genomic organisation observed is TMP with LT activity alone (6.9%). TMP and Tal with LT and TAEP activities, respectively, are observed in 6.3% of prophage genomes containing TAL (Fig. S4). Lastly, 59 of the prophage genomes did not contain predicted lysins associated with tail proteins and the functional modules in some of these genomes were not conventionally organised albeit they are predicted to be intact prophages by PHASTER.

Furthermore, we found that genome organisation within the enterococcal prophages coincides with that within isolated phages in terms of module order (i.e. Packaging, Head, Tail, Lysis, DNA Metabolism) (Fig. 3B). Additionally, the genome size of most analysed prophage was between 30 and 60 kb (Fig. 2B) and the majority possess the typical tail module arrangement seen in siphoviruses (Efquatroviruses, Phifelviruses or Saphexavirus) (i.e. TMP-Dit-Tal). As expected, we observed several lysogenic genes such as integrase, repressor and antirepressor in these predicted prophages. Collectively, we propose that these prophages are likely to be Efquatroviruses, Phifelviruses or Saphexavirus.

4. Conclusion

In this study, we have surveyed the lysin landscape of enterococcal bacteriophage genomes. The most commonly identified TAL domains were those targeting peptidoglycan, namely endopeptidases (TAEP), NLPC/P60 (endo)peptidases as well TMP located GH23 lytic transglycosylases- present within tail-tape measure proteins in all predicted tailed viruses surveyed. Lastly other domains potentially target EPA (pectinesterases) and teichoic acids (GDPD). Overall, one predicts that these all target different parts of the cell wall of these enterococci and that the differences in domain and sequence indicate differences in strain specificity that are not yet elucidated. Additionally, the finding that many phages contain multiple potential lysin domains suggests a layer of co-operation between these domains *in vivo* that we have yet to elucidate.

Finally, our data reveal the extent and variety of enterococcal lytic domains as candidates for recombinant production as potential novel antimicrobials, either in isolation or in combination with each other or as potentiators of antibiotics. Finally, we have also laid a platform for potential engineering of enterococcal phage akin to the recent refactoring study on T7 (Liang et al., 2022). Our group are also currently working on expressing examples of a range of these genes recombinantly with a view to production of novel antimicrobials.

Author statement

Manuscript: VIRUS-D-22–00,735

Both authors (AA and GS) carried out study conception and design, analysis and interpretation of data. They also drafted the manuscript and read and approved the final manuscript. Authors: Alhassan M Alrafaie (AA), Graham P Stafford (GS)

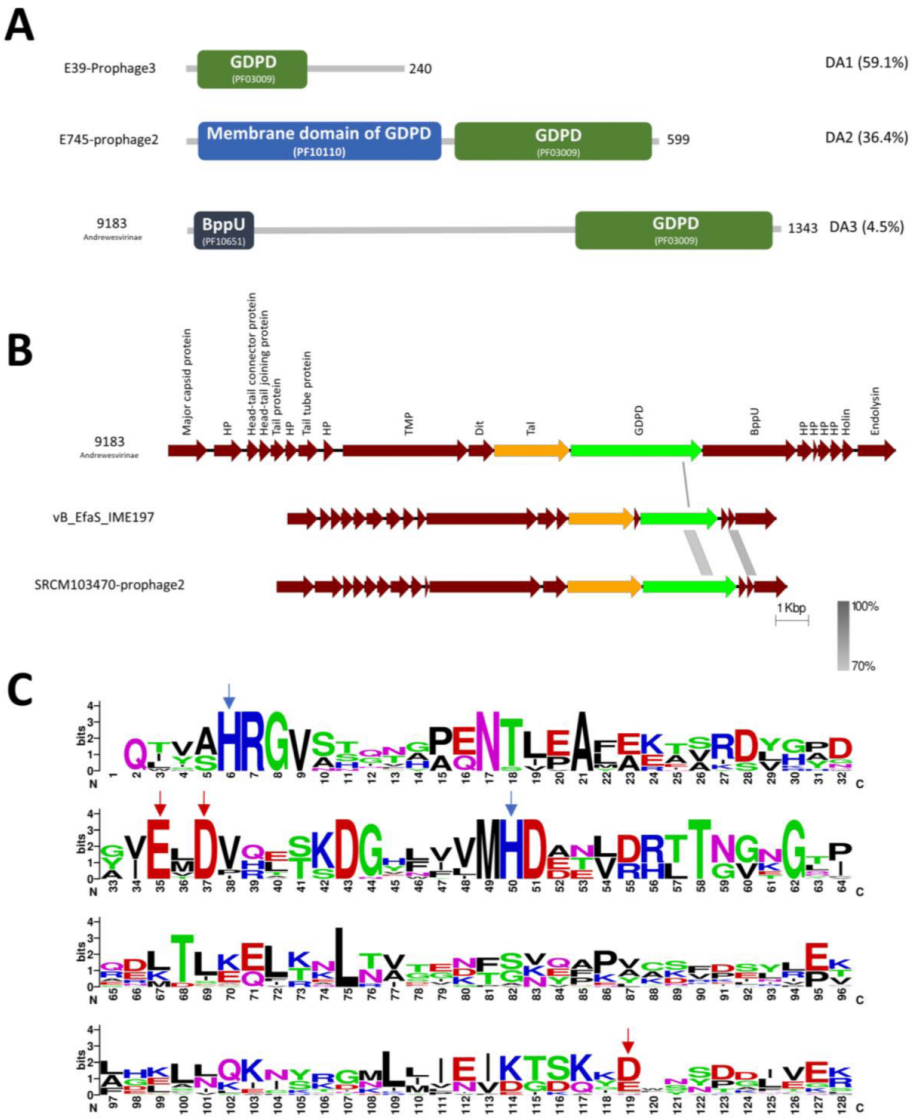


Fig. 8. (A) Domains architectures of the GDPD containing proteins based on Pfam. Three DA are shown with coloured domains. Green= GDPD, Blue= membrane and dark blue=baseplate upper protein. The domain type and abundance (%) are indicated on the right side while phage or prophage name and the ICTV classification is on the left side. The length of the protein is also indicated. (B) MSA of tail modules showing GDPD protein (Green), TAEP (Orange), HP= Hypothetical proteins. The level of identity is indicated by the grey region between genomes. C) weblogo of the aligned GDPD domains which shows catalytic residues (Blue arrows) and metal binding residues (Red arrows).

Funding

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

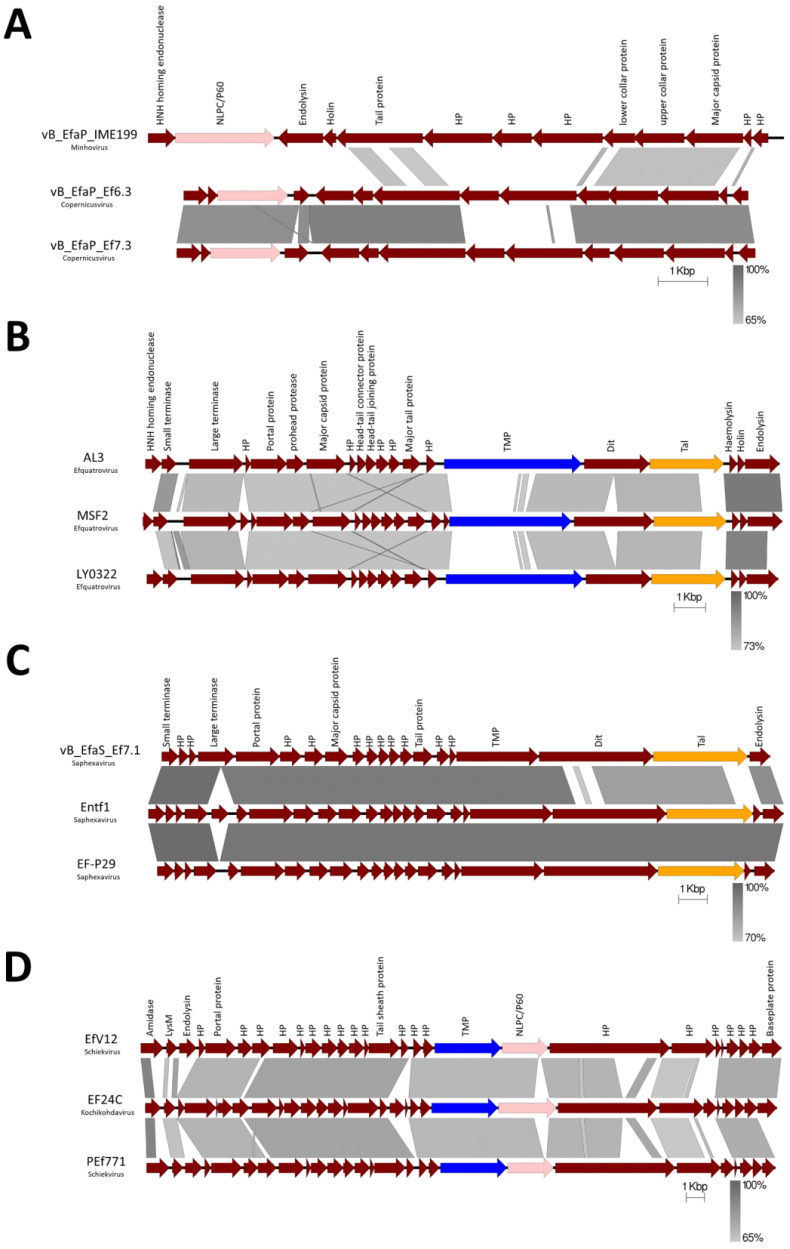


Fig. 9. The general organisation of Tail modules in enterococcal phage genomes. (A) examples of podovirus genomes harbour NLPC/P60 containing protein (pink colour). (B) Group1 of siphovirus genomes contain both TAEP (orange) and TMP-LT (blue) proteins while Group2 (C) contains only TAEP proteins (orange). (D) myovirus genomes harbour TMP-LT and NLPC/P60 proteins. The phage names and the ICTV classification are indicated on the left side. The level of identity is indicated by the grey region between genomes.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2023.199073.

References

Abdelrahman, F., Easwaran, M., Daramola, O.I., Ragab, S., Lynch, S., Oduselu, T.J., Khan, F.M., Ayobami, A., Adnan, F., Torrents, E., Sanmukh, S., El-Shibiny, A., 2021. Phage-encoded endolysins. Antibiotics 10 (2), 1–31. https://doi.org/10.3390/antibiotics10020124.

Anantharaman, V., Aravind, L., 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. Genome Biol. 4 (2), 1–12. https://doi.org/10.1186/gb-2003-4-2-r11.

Arbeloa, A., Hugonnet, J.E., Sentilhes, A.C., Josseaume, N., Dubost, L., Monsempes, C., Blanot, D., Brouard, J.P., Arthur, M., 2004. Synthesis of mosaic peptidoglycan cross-bridges by hybrid peptidoglycan assembly pathways in gram-positive bacteria. J. Biol. Chem. 279 (40), 41546–41556. https://doi.org/10.1074/jbc.M407149200.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucle. Acids Res. 44 (W1), W16–W21. https://doi.org/10.1093/nar/gkw387.

Banla, L.I., Salzman, N.H., Kristich, C.J., 2019. Colonization of the mammalian intestinal tract by enterococci. Curr. Opin. Microbiol. 47, 26–31. https://doi.org/10.1016/j.mib.2018.10.005.

Becker, S.C., Dong, S., Baker, J.R., Foster-frey, J., Pritchard, D.G., & Donovan, D.M. (2009). LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells. 294, 52–60. 10.1111/j.1574-6968.2009.01541.x.

Ben Braïek, O., Smaoui, S., 2019. Enterococci: Between Emerging Pathogens and Potential Probiotics. BioMed Research International 2019, 1–13. https://doi.org/10.1155/2019/5938210.

Bertozzi Silva, J., Storms, Z., Sauvageau, D., 2016. Host receptors for bacteriophage adsorption. In FEMS Microbiol. Lett. 363 (4), 1–11. https://doi.org/10.1093/femsle/fnw002.

Blackburn, N.T., Clarke, A.J., 2001. Identification of four families of peptidoglycan lytic transglycosylases. J. Mol. Evol. 52 (1), 78–84. https://doi.org/10.1007/s002390010136.

Bondi, M., Laukova, A., De Niederhausern, S., Messi, P., Papadopoulou, C., Economou, V., 2020. Controversial Aspects Displayed by Enterococci: probiotics or Pathogens? *BioMed Res. Int., 2020*. https://doi.org/10.1155/2020/9816185.

Bowler, P.G., Duerden, B.I., Armstrong, D.G., 2001. Wound microbiology and associated approaches to wound management. Clin. Microbiol. Rev. 14 (2), 244–269. https://doi.org/10.1128/CMR.14.2.244-269.2001.

Brede, D.A., Snipen, L.G., Ussery, D.W., Nederbragt, A.J., Nes, I.F., 2011. Complete genome sequence of the commensal *Enterococcus* faecalis 62, isolated from a healthy norwegian infant. J. Bacteriol. 193 (9), 2377–2378. https://doi.org/10.1128/JB.00183-11.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., Xia, F., 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5 https://doi.org/10.1038/srep08365.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28 (4), 464–469. https://doi.org/10.1093/bioinformatics/btr703.

Chuang, Y., Peng, Z., Tseng, S., Lin, Y., Sytwu, H., & Hsieh, Y. (2015). Impact of the glpQ2 gene on virulence in a streptococcus pneumoniae serotype 19A sequence type 320 strain. 83(2), 682–692. 10.1128/IAI.02357-14.

Cornelissen, A., Sadovskaya, I., Vinogradov, E., Blangy, S., Spinelli, S., Casey, E., Mahony, J., Noben, J.P., Dal Bello, F., Cambillau, C., Van Sinderen, D., 2016. The baseplate of Lactobacillus delbrueckii bacteriophage Ld17 harbors a glycerophosphodiesterase. J. Biol. Chem. 291 (32), 16816–16827. https://doi.org/10.1074/jbc.M116.728279.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucle. Acids Res. 16 (22), 10881–10890. https://doi.org/10.1093/nar/16.22.10881.

Dale, J.L., Nilson, J.L., Barnes, A.M.T., Dunny, G.M., 2017. Restructuring of Enterococcus faecalis biofilm architecture in response to antibiotic-induced stress. NPJ Biofilms Microbiom. 3 (1), 1–9. https://doi.org/10.1038/s41522-017-0023-4.

Dik, D.A., Marous, D.R., Fisher, J.F., Mobashery, S., 2017. Lytic transglycosylases: concinnity in concision of the bacterial cell wall. Crit. Rev. Biochem. Mol. Biol. 52 (5), 503–542. https://doi.org/10.1080/10409238.2017.1337705.

Dion, M.B., Oechslin, F., Moineau, S., 2020. Phage diversity, genomics and phylogeny. Nat. Rev. Microbiol. 18 (3), 125–138. https://doi.org/10.1038/s41579-019-0311-5.

Elbreki, M., Ross, R.P., Hill, C., O'Mahony, J., McAuliffe, O., Coffey, A., 2014. Bacteriophages and Their Derivatives as Biotherapeutic Agents in Disease Prevention and Treatment. *J. Viruses, 2014* 1–20. https://doi.org/10.1155/2014/382539.

Fard, R.M.N., Barton, M.D., Arthur, J.L., Heuzenroeder, M.W., 2010. Whole-genome sequencing and gene mapping of a newly isolated lytic enterococcal bacteriophage EFRM31. Arch. Virol. 155 (11), 1887–1891. https://doi.org/10.1007/s00705-010-0800-3.

Fukushima, T., Uchida, N., Ide, M., Kodama, T., Sekiguchi, J., 2018. DL-endopeptidases function as both cell wall hydrolases and poly-γ-glutamic acid hydrolases. Microbiology (United Kingdom) 164 (3), 277–286. https://doi.org/10.1099/mic.0.000609.

Goulet, A., Spinelli, S., Mahony, J., Cambillau, C., 2020. Conserved and diverse traits of adhesion devices from siphoviridae recognizing proteinaceous or saccharidic receptors. Viruses 12 (5), 1–21. https://doi.org/10.3390/v12050512.

Griffin, M.E., Klupt, S., Espinosa, J., Hang, H.C., 2022. Peptidoglycan NlpC/P60 peptidases in bacterial physiology and host interactions. Cell Chem. Biol. 2022. https://doi.org/10.1016/j.chembiol.2022.11.001.

Guerardel, Y., Sadovskaya, I., Maes, E., Furlan, S., Chapot-Chartier, M.P., Mesnage, S., Rigottier-Gois, L., Serror, P., 2020. Complete structure of the enterococcal polysaccharide antigen (EPA) of vancomycin-resistant enterococcus faecalis v583 reveals that EPA decorations are teichoic acids covalently linked to a rhamnopolysaccharide backbone. MBio 11 (2). https://doi.org/10.1128/mBio.00277-20.

Holtje, J.V., Mirelman, D., Sharon, N., Schwarz, U., 1975. Novel type of murein transglycosylase in escherichia coli. J. Bacteriol. 124 (3), 1067–1076. https://doi.org/10.1128/jb.124.3.1067-1076.1975.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10 (6), 845–858. https://doi.org/10.1038/nprot.2015.053.

Kizziah, J.L., Manning, K.A., Dearborn, A.D., Dokland, T, 2020. Structure of the host cell recognition and penetration machinery of a Staphylococcus aureus bacteriophage. PLoS Pathog. (2), 16. https://doi.org/10.1371/journal.ppat.1008314.

Latka, A., Maciejewska, B., Majkowska-Skrobek, G., Briers, Y., Drulis-Kawa, Z., 2017. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. Appl. Microbiol. Biotechnol. 101 (8), 3103–3119. https://doi.org/10.1007/s00253-017-8224-6, 10.1111/1751-7915.13593.

Letarov, A.V., Kulikov, E.E., 2017. Adsorption of Bacteriophages on Bacterial Cells. *Biochemistry*. Biokhimiia 82 (13), 1632–1658. https://doi.org/10.1134/S0006297917130053.

Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucle. Acids Res. 49 (W1), W293–W296. https://doi.org/10.1093/nar/gkab301.

Li, J., Sheng, Y., Ma, R., Xu, M., Liu, F., Qin, R., Zhu, M., Zhu, X., He, P., 2021. Identification of a depolymerase specific for k64-serotype klebsiella pneumoniae: potential applications in capsular typing and treatment. Antibiotics 10 (2), 1–18. https://doi.org/10.3390/antibiotics10020144.

Liang, J., Zhang, H., Tan, Y.L., Zhao, H., Ang, E.L., 2022. Directed evolution of replication-competent double-stranded DNA bacteriophage toward new host specificity. ACS Synth. Biol. 11 (2), 634–643. https://doi.org/10.1021/acssynbio.1c00319.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., & Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. 42(November 2013), 490–495. 10.1093/nar/gkt1178.

Love, R.M., 2001. Enterococcus Faecalis - a mechanism for its role in endodontic failure. Int. Endod. J. 34 (5), 399–405. https://doi.org/10.1046/j.1365-2591.2001.00437.x.

Mahony, J., Alqarni, M., Stockdale, S., Spinelli, S., 2016. Functional and Structural Dissection of the Tape Measure Protein of Lactococcal Phage TP901-1, August. Nature Publishing Group, pp. 1–10. https://doi.org/10.1038/srep36667.

Mohnen, D., 2008. Pectin structure and biosynthesis. Curr. Opin. Plant Biol. 11 (3), 266–277. https://doi.org/10.1016/j.pbi.2008.03.006.

Moura de Sousa, J.A., Pfeifer, E., Touchon, M., Rocha, E.P.C., 2021. Causes and consequences of bacteriophage diversification via genetic exchanges across lifestyles and bacterial taxa. Mol. Biol. Evol. 38 (6), 2497–2512. https://doi.org/10.1093/molbev/msab044.

Pertics, B.Z., Cox, A., Nyúl, A., Szamek, N., Kovács, T., Schneider, G., 2021. Isolation and characterization of a novel lytic bacteriophage against the k2 capsule-expressing hypervirulent klebsiella pneumoniae strain 52145, and identification of its functional depolymerase. Microorganisms 9 (3), 1–20. https://doi.org/10.3390/microorganisms9030650, 10.1007/s00253-015-7247-0.

Piuri, M., Hatfull, G.F., 2006. A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol. Microbiol. 62 (6), 1569–1585. https://doi.org/10.1111/j.1365-2958.2006.05473.x.

Proença, D., Fernandes, S., Leandro, C., Silva, F.A., Santos, S., Lopes, F., Mato, R., Cavaco-Silva, P., Pimentel, M., São-José, C., 2012. Phage endolysins with broad antimicrobial activity against Enterococcus faecalis clinical strains. Microbial Drug Resist. 18 (3), 322–332. https://doi.org/10.1089/mdr.2012.0024.

Rao, K.N., Bonanno, J.B., Burley, S.K., Swaminathan, S., 2006. Crystal Structure of Glycerophosphodiester Phosphodiesterase from Agrobacterium Tumefaciens By SAD With a Large Asymmetric Unit, 518, pp. 514–518. https://doi.org/10.1002/prot.

O'Neil, J. (2016). Tackling Drug-Resistant Infections Globally: Final Report And Recommendations. Available online at: http://amr-review.org.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., & Finn, R.D. (2018). *The MEROPS database of proteolytic enzymes,* their substrates inhibitors in 2017 and a comparison with peptidases in the PANTHER database. 46(November 2017), 624–632. 10.1093/nar/gkx1134.

Reid, W.W., 1950. Estimation and separation of the pectin-esterase and polygalacturonase of micro-fungi. Nature 166 (4222), 569. https://doi.org/10.1038/166569a0. –569.

Rigottier-Gois, L., Madec, C., Navickas, A., Matos, R.C., Akary-Lepage, E., Mistou, M.Y., Serror, P., 2015. The surface rhamnopolysaccharide epa of enterococcus faecalis is a key determinant of intestinal colonization. J. Infect. Dis. 211 (1), 62–71. https://doi.org/10.1093/infdis/jiu402.

São-José, C., 2018. Engineering of phage-derived lytic enzymes: improving their potential as antimicrobials. Antibiotics 7 (2), 29. https://doi.org/10.3390/antibiotics7020029.

Scheurwater, E., Reid, C.W., Clarke, A.J., 2008. Lytic transglycosylases: bacterial space-making autolysins. Int. J. Biochem. Cell Biol. 40 (4), 586–591. https://doi.org/10.1016/j.biocel.2007.03.018.

Schleifer, K.H., Kilpper-Balz, R., 1984. Transfer of Streptococcus faecalis and Streptococcus faecium to the genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus faecium comb. nov. Int. J. Syst. Bacteriol. 34 (1), 31–34. https://doi.org/10.1099/00207713-34-1-31.

Schwarzer, D., Stummeyer, K., Gerardy-Schahn, R., Mühlenhoff, M., 2007. Characterization of a novel intramolecular chaperone domain conserved in endosialidases and other bacteriophage tail spike and fiber proteins. The Journal of biological chemistry 282 (5), 2821–2831. https://doi.org/10.1074/jbc.M609543200.

Shahed-Al-Mahmud, M., Roy, R., Sugiokto, F.G., Islam, M.N., Lin, M.D., Lin, L.C., Lin, N. T., 2021. Phage φAB6-borne depolymerase combats acinetobacter baumannii biofilm formation and infection. Antibiotics 10 (3), 279. https://doi.org/10.3390/antibiotics10030279.

Shi, L., Liu, J.F., An, X.M., Liang, D.C., 2008. Crystal structure of glycerophosphodiester phosphodiesterase (GDPD) from Thermoanaerobacter tengcongensis, a metal ion-dependent enzyme: insight into the catalytic mechanism. Proteins Struct. Funct. Genet. 72 (1), 280–288. https://doi.org/10.1002/prot.21921.

Silhavy, T.J., Kahne, D., Walker, S., 2010. The Bacterial Cell Envelope, 2. Cold Spring Harbor Perspectives in Biology, a000414. https://doi.org/10.1101/cshperspect.a000414. –a000414.

Stockdale, S.R., Mahony, J., Courtin, P., Van Pijkeren, J., Britton, R.A., Neve, H., Heller, K.J., Aideh, B., Vogensen, F.K., & Van Sinderen, D.. (2013). The lactococcal phages Tuc2009 and TP901-1 incorporate two alternate forms of their tail fiber into their virions for. 288(8), 5581–5590. 10.1074/jbc.M112.444901.

Suttle, C.A., 2005. Viruses in the sea. Nature 437 (7057), 356–361. https://doi.org/10.1038/nature04160.

Teng, F., Singh, K.V., Bourgogne, A., Zeng, J., Murray, B.E., 2009. Further characterization of the epa gene cluster and epa polysaccharides of Enterococcus faecalis. Infect. Immun. 77 (9), 3759–3767. https://doi.org/10.1128/IAI.00149-09.

Terms, F., 2011. On an invisible microbe antagonistic to dysentery bacilli. Note by M. F. d'Herelle, presented by M. Roux. Comptes Rendus Academie des Sciences 1917; 165: 373–5. Bacteriophage 1 (1), 3–5. https://doi.org/10.4161/bact.1.1.14941.

Turner, D., Kropinski, A.M., Adriaenssens, E.M., 2021. A roadmap for genome-based phage taxonomy. Viruses 13 (3), 1–10. https://doi.org/10.3390/v13030506.

Twort, F.W., 1915. An investigation on the nature of ultra-microscopic viruses. Lancet 186 (4814), 1241–1243. https://doi.org/10.1016/S0140-6736(01)20383-3.

Uttley, A.C., Collins, C.H., Naidoo, J., George, R.C., 1988. Vancomycin-resistant enterococci. Lancet 331 (8575–8576), 57–58. https://doi.org/10.1016/S0140-6736(88)91037-9.

Vermassen, A., Leroy, S., Talon, R., Provot, C., Popowska, M., Desvaux, M., 2019. Cell wall hydrolases in bacteria: insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan. Front. Microbiol. (FEB), 10. https://doi.org/10.3389/fmicb.2019.00331, 10.1128/JB.185.14.4022-4030.2003.

Willyard, C., 2017. The drug-resistant bacteria that pose the greatest health threats. Nature 543 (7643), 15. https://doi.org/10.1038/nature.2017.21550.

Wohlkönig, A., Huet, J., Looze, Y., Wintjens, R., 2010. Structural relationships in the lysozyme superfamily: significant evidence for glycoside hydrolase signature motifs. PLoS ONE 5 (11), 1–10. https://doi.org/10.1371/journal.pone.0015388.

Xiang, Y., Morais, M.C., Cohen, D.N., Bowman, V.D., Anderson, D.L., Rossmann, M.G., 2008. Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage φ29 tail. In: Proceedings of the National Academy of Sciences of the United States of America, 105, pp. 9552–9557. https://doi.org/10.1073/pnas.0803787105.