University of Glasgow

Convergent evolution of PICIs-mediated phage interference

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

By

Nuria Carpena Garcia

BPharm, MRes

Submitted on the 21 of September 2016

Institute of Infection, Immunity and Inflammation
College of Medical, Veterinary and Life Sciences
University of Glasgow

Author's Declaration

Unless explicitly states otherwise, all of the experimental work presented in this thesis was performed by me, either independently or in conjunction with others. No part of this research or any part of this document has or will be presented for the fulfilment of any other degree or qualification.

Acknowledgement

First, I would like to thank my supervisors Professor José R Penadés and Dr Alberto Marina, for the chance to work with them and their good advice and assistance. I would also like to thank the current and past members of the Penadés lab and everybody on Level 2 for their help and support, contributing to my personal and professional growing. I cannot thank my family and friends enough for their support, love, patience and encouragement. A very special thank you to Fabian, I wouldn't be here without you.

Abstract

Staphylococcal pathogenicity islands (SaPIs), the prototype members of the family of phage inducible chromosomal islands (PICIs), are extremely mobile phage satellites, which are transferred between bacterial hosts after their induction by a helper phage. The intimate relationship between SaPIs and their helper phages is one of the most studied examples of virus satellite interactions in prokaryotic cells. SaPIs encode and disseminate virulence and fitness factors, representing a driving force for bacterial adaptation and pathogenesis. Many SaPIs encode a conserved morphogenetic operon, including a core set of genes whose function allows them to parasitize and exploit the phage life cycle. One of the central mechanisms of this molecular piracy is the specific packaging of the SaPI genomes into reduced sized capsid structures derived from phage proteins. *Pac* phages were classically thought to be the only phages involved in the mobilisation of phage-mediated virulence genes, including the transfer of SaPIs within related and non-related bacteria.

This study presents the involvement of *S. aureus cos* phages in the intra- and intergeneric transfer of *cos* SaPIs for the first time. A novel example of molecular parasitism is shown, by which this newly characterised group of *cos* SaPIs uses two distinct and complementary mechanisms to take over the helper phage packaging machinery for their own reproduction. SaPIbov5, the prototype of the *cos* SaPIs, does not encode the characteristic morphogenetic operon found in *pac* SaPIs. However, *cos* SaPIs features both *pac* and *cos* phage cleavage sequences in their genome, ensuring SaPI packaging in small- and full-sized phage particles, depending on the helper phage. Moreover, *cos*-site packaging in *S. aureus* was shown to require the activity of a phage HNH nuclease. The HNH protein functions together with the large terminase subunit, triggering cleavage and melting of the *cos*-site sequence. In addition, a novel piracy strategy, severely interfering with the helper phage reproduction, was identified in *cos* SaPIs and characterised. This mechanism of piracy depends on the *cos* SaPI-encoded *ccm* gene, which encodes a capsid protein involved in the formation of small phage particles, modifying the assembling process via a scaffolding mechanism. This strategy resembles the

ones described for *pac* SaPIs and represents a remarkable example of convergent evolution. A further convergent mechanism of capsid size-reduction was identified and characterised for the *Enterococcus faecalis* EfCIV583 pathogenicity island, another member of the PICI family. In this case, the self-encoded CpmE conducts this molecular piracy through a putative scaffolding function. Similar to *cos* SaPIs, EfCIV583 carries the helper phage cleavage sequence in its genome enabling its mobilisation by the phage terminase complex. The results presented in this thesis show how two examples of non-related members of the PICI family follow the same evolutionary convergent strategy to interfere with their helper phage. These findings could indicate that the described strategies might be widespread among PICIs and implicate a significant impact of PICIs mediated-virulence gene transfer in bacterial evolution and the emergence of pathogenic bacteria.

Table of Contents

List of Tables

List of Figures

List of Accompanying Material

Abbreviations

ϕ	Phage
Δ	Deletion
%	Percent
ºC	Degrees Celsius
µg	Microgram(s)
µl	Microliter(s)
µm	Micrometre(s)
µM	Micromolar
amp	Ampicillin
att	Attachment site
att_C	Attachment site into the bacterial chromosome
att_L	Left attachment site
att_R	Right attachment site
att_S	Attachment site into circular intermediate of the SaPI
blaZ	β-lactamase gene
BHI	Brain-heart infusion medium
bp	Base pair(s)
Bulk	Majority DNA (genomic + DNA phage)
$CaCl_2$	Calcium chloride
CFU	Colony forming units
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
ERC	Excision-Replication-Packaging cycle
Erm	Erythromycin
ermC	Erythromycin resistance gene
g	Gram(s)
GIs	Genomic Islands
H	Hour(s)
HGT	Horizontal Gene Transfer
HPLC	High-performance liquid chromatography
IPTG	Isopropyl-β-D-thiogalactoside
kb	Kilo base pairs
kV	Kilovolt(s)
LB	Luria-Bertani medium
M	Molar
MC	Mitomycin C
$MgCl_2$	Magnesium chloride
MGEs	Mobile Genetic Elements
$MgSO_4$	Magnesium sulphate

min	Minute(s)
ml	Millilitre(s)
mm	Millimetre(s)
mM	Millimolar
MW	Molecular weight
NaCl	Sodium chloride
ng	Nanogram(s)
nM	Nanomolar
nt	Nucleotide
OD (nm)	Optical density at a certain wavelength
ORF	Open reading frame
ori	Origin of replication
P	Promoter
PB	Phage base
PBA	Phage base agar
Pcad	Cadmium inducible promoter
PCR	Polymerase Chain Reaction
PDB	Protein Databank
PEG	Polyethylene glycol
PICIs	Phage Inducible Chromosomal Islands
rpm	Revolutions per minute
SaPIs	*S. aureus* Pathogenicity Islands
SOS	Stress bacterial response
TerS	Terminase small subunit
$TerS_{SP}$	SaPI Terminase small subunit
TerL	Terminase large subunit
Tet	Tetracycline
*tet*A	Tetracycline resistance gene in Gram-negative
*tet*M	Tetracycline resistance gene in Gram-positive
Tris	Tris-(hydroxymethyl) aminomethane
TSA	Tryptic soy agar medium
TSB	Tryptic soy broth
TTC	2,3,5-triphenyltetrazolium chloride
U	Unit
UV	Ultraviolet
V	Volt(s)
wt	Wild type
Xgal	5-Bromo-4-chloro-3-indole-β-D-galactopyranoside

Chapter 1 Introduction

1.1 *Staphylococcus aureus*

Staphylococcus aureus is a Gram-positive, non-motile, coagulase and catalase positive and oxidase negative bacteria, which grows in a spherical shape in irregular clusters. As an emergent opportunistic pathogen, *S. aureus* colonises the skin and mucous membranes of humans and other mammals[1]. Although *S. aureus* is not generally pathogenic, it can cause severe infectious diseases, and the emergence of drug-resistant strains make it a major healthcare burden.

S. aureus is a commensal bacteria of the human nose, with 30% of the population being an asymptomatic carrier[2-4]. Being a carrier increases the probability of staphylococcal infection after cutaneous or mucosal injuries allowing the microorganism to penetrate underlying tissues. From these initial sites of infection, the bacteria can invade the blood, eventually causing organ infections such as pneumonia, endocarditis, arthritis, osteomyelitis and severe symptoms of sepsis[5-8]. Likewise, *S. aureus* can cause toxin-mediated diseases and diseases such as necrotic infections[9,10], food poisoning[11], toxic shock syndrome[12,13] and scalded skin syndrome[14,15]. Moreover, it causes important diseases in animals, including mastitis in ruminants[16], dermatitis in rabbits[17] and skeletal infections in poultry[18]. The number of *S. aureus* infections in hospitals and community settings is increasing, and this bacterium is becoming more virulent and resistant to methicillin, vancomycin and other ß-lactam antibiotics[19]. The ability to colonise is due to the acquisition of virulence factors that provide the bacteria with unique mechanisms for host adaptation, allowing *S. aureus* to infect and ultimately survive in a wide variety of tissues and hosts[20].

Within the last decades, a growing number of *S. aureus* infections had been reported with the appearance of vancomycin-resistant (VRSA)[21], via acquisition of the *vanA* cluster; and methicillin-resistant (MRSA) isolates[22,23], encoding the staphylococcal cassette chromosome *mec* (SCC*mec*)[24]. Recently, two new groups of MRSA have emerged, HA-MRSA (healthcare-associated MRSA) and CA-MRSA (community-associated MRSA)[25-27]. HA-MRSA isolates can

be considered more problematic than MRSA, being resistant to all ß-lactam and fluoroquinolones antibiotics[28]. CA-MRSA isolates normally are less resistant to diverse antibiotic and carry a smaller version of the SCC*mec* cassette. Furthermore, these isolates usually incorporate the phage-encoded Panton–Valentine leukocidin gene (*PV-luk*)[10,29], which is a cytotoxin involved in severe necrotic and haemorrhagic infections. In addition to antibiotic resistances, the infection ability of *S. aureus* is increased by the combination of fitness factors, expressed during the various stages of infection (colonisation, immune evasion, bacterial multiplication and dissemination) that ultimately contribute to the emergence of new superbugs.

The evolution and adaptation of *S. aureus* pathogenicity clones have in part been driven by the acquisition of mobile genetic elements (MGEs) through mechanisms of horizontal gene transfer (HGT). MGEs are discrete DNA elements, encoding genes that promote their independent mobilisation between host genomes[30,31]. MGEs found in *S. aureus* include bacteriophages, pathogenicity islands, plasmids, transposons and SCC*mec* [32,33]. Remarkably, many of the bacterial genes encoding toxins, adhesins, invasins or other virulence and fitness factors are part of MGEs[32,33]. It is estimated that about 20% of the genetic setup of a given bacterial species had been acquired from other bacteria by mechanisms of HGT[34]. Likewise, this interchangeable genetic pool of MGEs and non-mobile genomic islands compose 15-20% of the *S. aureus* genome, playing an important role in its adaptability to new environmental scenarios[35–38].

1.2 Bacteriophages

1.2.1 Background

Bacteriophages are the most abundant and one of the most studied MGEs. Also designated as phages, they are viruses that infect and replicate inside bacteria cells. They constitute the most abundant biological entity, with an estimated number of 10^{31} tailed phages inhabiting the Earth biosphere[39]. Phages are present in all ecosystems inhabited by bacteria, from the aquatic and soil biosphere to animal-linked environments[40–46]. Predicted to be found in

one to two orders of magnitudes higher than their bacterial hosts, phages shape their host cells by self-encoded key proteins and modulate bacterial populations by mediating gene transfer[47]. The influence of the phage-bacteria symbiotic relationship is manifold, having an impact on the carbon cycle of soil and water environments and the viability of global nutrients[43,48], transfer of a genomic pool in-between biomass[49], as well as the acquisition of virulence traits in the evolution of human bacterial pathogens[50,51]. Moreover, phages are vectors of fitness and virulence factors, following different mechanisms such as the direct transfer of MGEs or the ability of lysogenic conversion[50,52]. Hence, phages play a significant role in niche bacterial warfare, providing advantageous traits for bacterial plasticity and evolution in a specific environment.

In addition to the direct supply of the bacterial host with phage-encoded virulence traits, phage-like machinery such as the R-type pyocins and the type VI secretion system (T6SSs) were evolutionary acquired by the bacterium for their benefit. These elements have a phage-like tail structure in common which increases the virulence of the bacteria. In the case of the T6SSs, this tail-like structure transports effector proteins via injection into the outer membrane of another cell[53]. The self-encoded tail structures of R-type pyocins will be released upon cell lysis, binding to other bacteria and causing their death[54].

1.2.2 Lysogenic conversion and transfer of phage-encoded virulence genes

The integration of temperate phage represents a significant degree of the genetic variability between different bacterial species and even between strains of the same bacterial species[50]. Thus, most of the differences between closely related bacterial genomes are frequently the result of prophages insertions. For example, a total of 18 prophages can be found in the *E. coli* Sakai pathogenic strain O157:H7 and up to 10% of the genome of *Streptococcus pyogene* is constituted by phages[50]. Moreover, nucleotide and protein sequence analysis of S. *aureus* strains identified a vast number of bacteriophages, making up to 4-5% of the total genome of the bacterium.

These studies postulate phage acquisition as one of the most important sources of genome variation when comparing different virulent strains[11,55,56].Within the broad spectrum of molecular mechanisms of phages, the transition from the lytic to lysogenic cycle, also call lysogenic conversion, provides the bacterial-host with numerous features to improve its fitness into a new environment[57,58]. Integrated phages encoding toxin-related genes and other finesses traits are one of the most significant contributions to the bacterial evolution towards a pathogenic state[59,60].

Many virulence factors of pathogenic bacteria, such as toxin and antibiotic resistances, are carried by phages[50] (Table 1.1). Bacteriophages encode genes that cause toxin-mediated diseases (toxinoses), including botulism, cholera, diphtheria, scarlet fever and others[61]. In *S. aureus*, examples for these virulence genes are the enterotoxin A (SEA), superantigen present in contaminated food, Panton-Valentine leukocidin (PV-Luk), which is involved in necrotizing pneumonia processes and severe skin diseases, or the staphylokinase toxin (Sak)[59,62–65]. In some cases, the mobilisation of these toxins is linked to the lytic cycle of the phage not only triggering the toxin to be released from the cell but also its expression. The association of toxin release and phage lytic cycle is clinically important as all of these pathogenic strains are treated with antibiotics that activate the SOS response. Phage induction ultimately activates the up-regulation of these toxins and cell lysis will release them from the bacteria host. *E. coli* phages encoding the Shiga toxin (Stx) are one example of toxin production linked to the activation of the lytic cycle of phages[66]. The expression of Stx after the use of quinolones will increase the risk of haemolytic uremic syndrome (HUS)[67]. In *S. aureus* bacteriophages, transcription of toxins such as *seg2*, *sek2*, *sea* and *sak* are also activated following phage induction[59]. Further virulence and fitness factors are also phage-encoded, including superantigens, adhesion factors, DNAase, mitogenic factors, invasion and immune evasion among others[50,68].

Nevertheless, the beneficial effect of an integrated prophage (phage) to its host is not limited to bacterial virulence traits, and some of the mechanism by which non-pathogenic and pathogenic bacteria adapt to their ecological

niches may be the result of phage integrations[69]. Hence, certain prophages are capable of altering the phenotype of the host bacterium by their integration into the coding region of the host chromosome resulting in the disruption of host genes[70-72]. These phages are able to convert and switch the host phenotype by a process known as negative lysogenic conversion. Two examples of this type of conversion are caused by the integration of staphylococci phages into the virulence-related β-haemolysin and lipase genes[73,74]. This mechanism is also found in *Shigella spp*, where phage integrates into the *cad*A gene that encodes the lysine decarboxylase enzyme, which ultimately will increase the virulence[75].

Table 1.1 List of some of the toxins produced by bacteriophages.

Protein	Gene	Phage	Bacterial host
Diphtheria toxin	*tox*	ß-Phage	*C. diphtheriae*
Neurotoxin	*C1*	ϕC1	*C. botulinum*
Cholera toxin	*ctx*AB	ϕCTX	*V. cholerae*
Leukocidin	*pvl*	ϕPVL	*S. aureus*
Shiga-like toxin	*stx1, stx2*	H-19B	*E. coli*
Enterotoxin	*see, sel*	NA	*S. aureus*
Enterotoxin A	*ent*A	ϕ13	*S. aureus*
Enterotoxin A	*sea*	ϕMu50A	*S. aureus*
Exfoliative toxin A	*eta*	ϕETA	*S. aureus*
Enterotoxin P	*sep*	ϕN315	*S. aureus*
Superantigens	*ssa, spe*A1, *spe*C, *spe*I, *spe*H, *spe*M, *spe*L, *spe*K	8232.1	*S. pyogenes*
Type III effector	*sop*E	ϕSopE	*S. enterica*

Likewise, bacteriophages exert protection against infection with other lytic phages[50,76-78] and also provide advantages in the environmental competition between different bacterial species through the induction of other lysogenic strains[79]. Besides, temperate phages can act as an anchor point for genomic reorganisation[50,69]. Moreover, a recent study has shown that phages can

adhere to mucosal surfaces, providing a non-host-derived immunity protection against bacterial colonisation and infection[80].

Bacteriophages are not only significant as MGEs themselves but also as vectors for the HGT of other MGEs among bacteria, playing a critical role in the biology, diversity and evolution of bacteria[50]. There are two mechanisms by which phages can transfer MGEs, with the general transduction, the packaging of foreign bacterial DNA into the phage procapsid, being the most widespread and studied form. It is assumed to be 1% of total phage particles, and it was firstly described in *Salmonella* typhimurium and *E. coli*[81,82]. Although general transduction occurs at a relatively low frequency, about every ~ 10^8 phage infections, the abundance of phage particles and the variety of bacterial hosts make it a common event in nature. Pioneer studies with phages P1 and P22 established that 5% of phage particles carried bacterial DNA, leading up to 10% of stable transductants of those transducing particles[83–85]. Furthermore, it was shown that certain genomic and pathogenic islands, such as the SCC*mec* and the *Vibrio cholerae* pathogenicity island (VPI)[86,87], are transferred by phage general transduction. Other phage satellites, such as *Staphylococcus aureus* pathogenicity islands (SaPIs), the RS1 element of *Vibrio cholerae* or the satellite phage P4[88–90], have developed several mechanisms to take advantage of the phage machinery and to be mobilised using the phage particles. On the other hand, specialised transduction occurs when a prophage aberrantly excises from their integration site of the host bacteria, and flanking regions of the bacterial chromosome are packaged with the phage genome[91,92].

It was reported that phage transduction between biomass occurs at a greater magnitude as previously suggested[49,93]. Recently studies have focused on the analysis of phages transducing bacterial genomes and phageomes (DNA encapsidated in phage particles), such as antibiotic resistance genes. Antibiotic treatment of mice resulted in an enrichment of the phageome with a variety of functional phage-encoded genes related to antibiotic resistance in the mice gut microbiota[94]. Furthermore, longitudinal metagenomic analysis evaluating the dynamics and composition of the human gut virome revealed

that a huge number of sequences belonged to complete or partial phage genomes. The assembling revealed gene sequences for antibiotic resistance, phage lytic and lysogenic function and unexpected CRISPR arrays genes[95]. In another study, 77 out 80 analysed faecal samples of healthy humans were tested positive for the presence of *blaTEM*, *blaCTX-M-1*, *mecA*, *armA*, *qnrA*, and *qnrS* antibiotic resistance genes[96].

In conclusion, bacteriophages are involved in the spread of antibiotic resistance genes, and other virulence and fitness traits, via transduction mechanisms, highlighting the importance of phage transduction in the emergence of novel pathogenic bacteria. Moreover, the fitness of a bacterial population is frequently driven by the actions of temperate phages with its ability to confer the host with a broad number of features to adapt and to survive in various environments.

1.2.3 Phage structure and taxonomy

Over the last decade, the development of whole genome sequence analysis, cryomicroscopy, image reconstruction and X-ray crystal structures have improved understanding of bacteriophages and changed the way they are described and organised in families. Organisations such as The International Committee on Taxonomy of Viruses (ICTV) have listed bacteriophages using various approaches, grouping them by family, genus or host range, in alphabetical or numerical order. With the existing data, capsid morphology, phage genome and viral structure are now used to discriminate between different bacteriophages[97-99]. However, new classification methods were recently developed, using others parameters such as “phage proteome trees”, which use the phage proteome to distinguish phages, or by discriminating phages by their functional-structural domains[100-102].

Most bacteriophages consist of a complex head-tail structure. Conventionally, tailed-phages shared some structural properties such as an icosahedral capsid and a helical tail attached to a hexagonal baseplate structure. All these components are assembled in independent pathways which ultimately join to form a mature viral particle[103-105]. Such complex bacteriophages belong to the

order of the *Caudovirales* and contain a linear double-stranded DNA genome (dsDNA)[106]. This order is further divided into three families *Myoviridae, Siphoviridae and Podoviridae*, differed by their tail structure (Figure 1.1). The *Myoviridae* family, such as the *E. coli* phage P2 or T4, has a long contractile tail. *Siphoviridae* or "λ-like viruses" have long non-contractile tails. λ , HK97 (*E. coli*), SPP1 (*Bacillus subtilis)* and most of the *S. aureus* phages, such as ϕ11[107], belong to this family. The tail of *Podoviridae* is short and non-contractile. Examples for this phage family are the "P22-like viruses" such as phage P22, phi29 or T7.

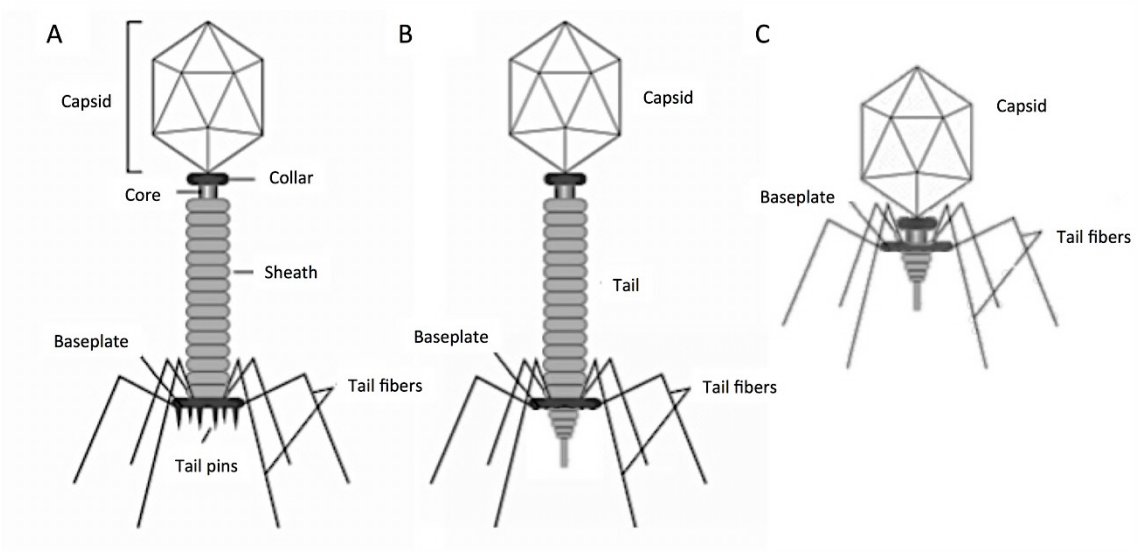


Figure 1.1 Structure of prototypical bacteriophages. (a) *Myoviridae*, T4; (b) *Siphoviridae*, λ; *Podoviridae,* P22. Adapted from Trun *et al.* 2009[108].

1.2.4 The life cycle of a bacteriophage

Bacteriophages can be further classified as lytic or temperate depending on their 'life cycle'. By those means, in the lytic cycle, after induction of the lytic cell program, or after infection of a recipient strain, phages replicate using the host cellular machinery and newly formed phage progeny will be release after the lysis of the bacteria. Characteristically, the initiation of this life cycle depends on the binding of the phage to particular cell surface receptors, followed by the injection of the phage genome into the bacterial cytoplasm and its transcription and replication. Subsequently, the proteins involved in the formation of the tails and the capsids are synthesised, which

assemble to accommodate the phage chromosome, forming new virion particles. Other phage-encoded proteins play a major role triggering the lysis of the host cell, such as the holin protein, which generates pores in the cytoplasmic membrane. Additionally, the phage-encoded endolysin (lysin) will enter to the peptidoglycan layer, triggering peptidoglycan hydrolysis and cell lysis, leading to the discharge of the newly formed infectious phage particles. Once released, the virions can re-infect sensitive bacteria and start the lytic cycle (Figure 1.2). The different phage stages are induced in a controlled and temporal pattern to promote the correct expression of the genes of the lytic cycle, with the genes involved in replication being the first to be expressed, followed by the expression of the packaging and finally the lysis genes.

Temperate phages can adopt either the lytic or the non-lytic (lysogenic) life cycle. In the lysogenic cycle, temperate phages integrate into the host genome as a prophage and its dsDNA replicates concomitantly with the bacterial chromosome during cell division. Hence, phages are transmitted from one generation to another (vertical transfer). The lysogenic state is usually sustained via a phage-encoded repressor (generically designated as CI for the phage λ), inhibiting the transcription of most phage genes, including those necessaries for the lytic cycle. Usually, but not always, the transition from non-lytic to lytic cycle can be triggered by stress to the host cells, such as DNA damage induced by antibiotics, UV irradiation, thermal stress, and others processes that activate the bacterial SOS response[67,109–113]. RecA, a DNA repair protein involved in the SOS response, initiates CI autocleavage, withdrawing the repression of lytic genes and leading to the initiation of the lytic cycle. During induction, the phage is excised from the bacterial chromosome and begins to replicate in the cytoplasm, being subsequently encapsidated and horizontally transferred to other bacterial hosts, as described for the lytic phages (Figure 1.2).

Furthermore, there is a third lifecycle called pseudolysogeny, which occurs when the phage resides inside the bacterial host cell as a plasmid-like element, without inducing the lytic cycle or been integrated[114] (Figure 1.2).

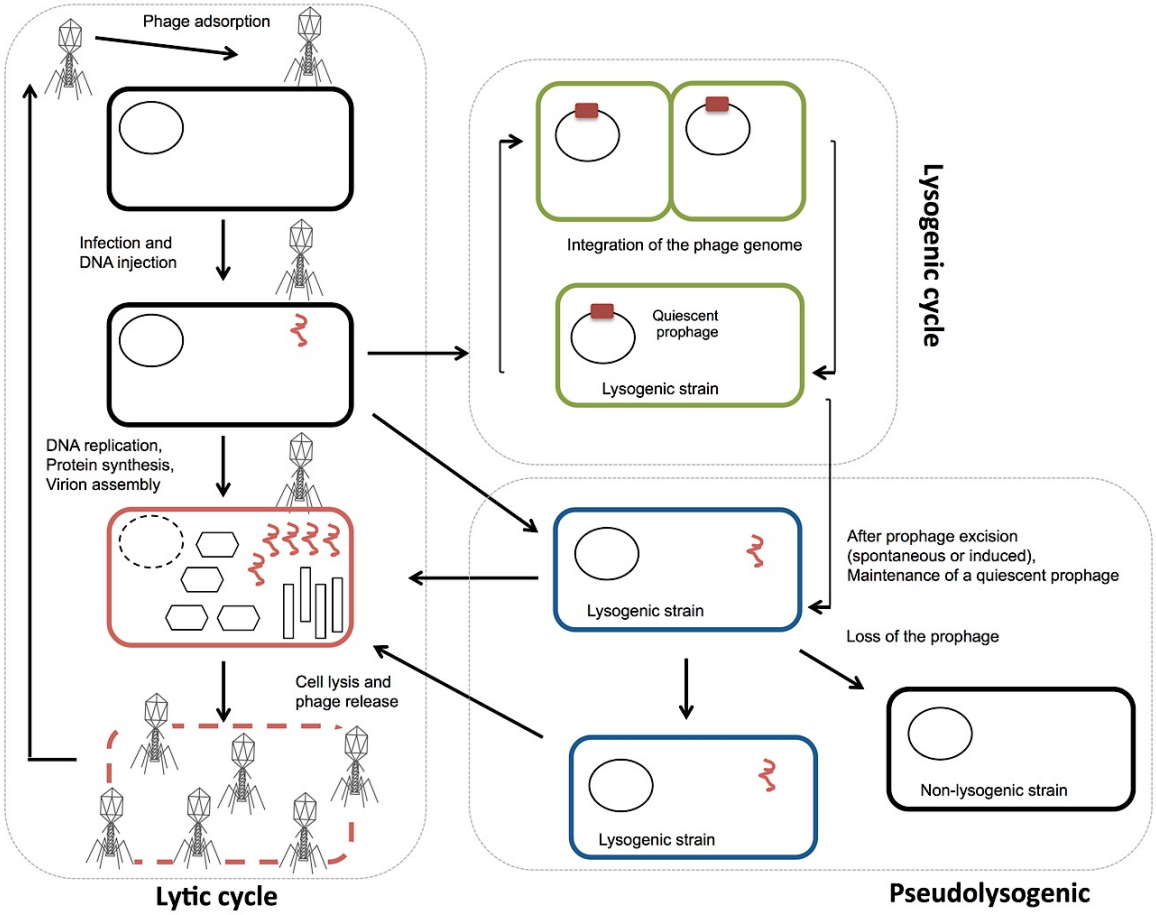


Figure 1.2 Bacteriophage life cycle. After binding of the virion to specific host cell receptors, adsorption and injection of the phage dsDNA occur. Following the injection of the dsDNA, the phage can be integrated into the genome of the bacterium (lysogenic cycle) and remain until the induction of the lytic cycle occurs. Once the phage enters the lytic cycle, it excises from the bacterial chromosome and the virus genome replicates as a concatemer, whilst morphogenetic proteins are synthesised. The capsids are loaded with the viral genome and the mature virus particle formed. Finally, lysis of the bacterial cell will release new virions that will infect other bacterial cells. Adapted from Fortier *et al.* 2013[115].

1.2.5 Genetic and functional organisation of bacteriophages

The genes of bacteriophages are grouped into genomic modules according to their biological function. Hence, when using this modular genetic organisation as a parameter of phage classification, phages belonging to different families are usually clustered together. Thus, using this criterion, phage λ, HK97 (Siphovirus), P22 (Podovirus) and some *S. aureus* bacteriophages are considered members of the same family, called lambdoid bacteriophages. Esther Lederberg first characterised the phage λ around 1951 and since then it was used as a model in phage molecular and genetic research and for the

development of genetic engineering tools[116-120]. Members of this lambdoid group have a very similar mosaic genomic structure, grouping orthologous genes into three functional modules: the early transcription or lysogeny module, the intermediate transcription or replication module and the late transcription or packaging and lysis module. The organisation of phage genes of related function into modular clusters permits their expressions in a controlled and joint pattern, from early gene transcription to cell lysis during the lytic cycle. Thus, the genes included in the lysogeny module will be the first to be expressed after the infection, and will determine the establishment of one of the two distinctive routes (lytic or lysogenic cycle). The module of intermediate gene expression encodes proteins involved in phage replication and DNA metabolism. Finally, genes involved in DNA packaging, capsid and tail morphogenesis processes and phage host-cell lysis are encoded in the late transcriptional module (Figure 1.3).

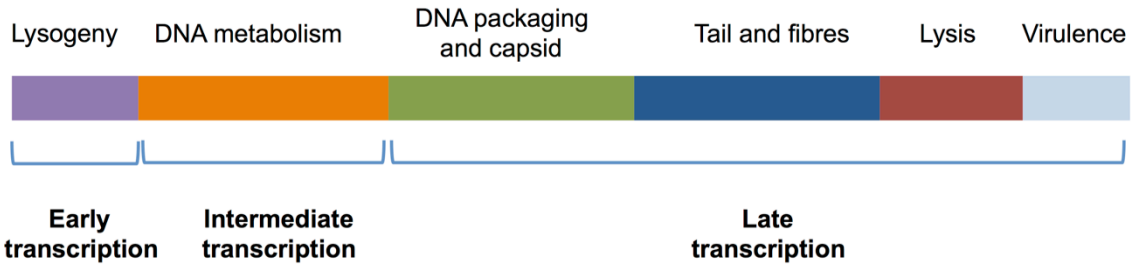


Figure 1.3 Schematic representation of the genetic organisation of the *Siphoviridae* family of staphylococci phages genomes. The coloured boxes represent the different functional modules: purple, lysogeny; orange, DNA metabolism; green, packaging and capsid morphogenesis; blue, morphogenesis of the tail and fibres; pink, cell lysis; light blue, virulence genes.

Furthermore, the organisation in functional clusters allows bacteriophages to interchange entire modules among them[121,122]. The characteristically mosaic structure of phage genomes is generated by non-homologous genomic interchange (illegitimate recombination), occurring at high frequency in between lytic, lysogenic and also defective phages presents in the host chromosome[69,123,124]. This modular transfer is further enabled by the presence of conserved sequences at the end of each group facilitating the exchange between different DNA molecules via phage homologous recombination[125-128].

The organisation of genes in functional clusters and their exchange, often with proteins that shares no identity in sequences but have similar roles, permits a quicker and more plastic response to different ecological factors[129]. Hence, this horizontal exchange enables phages to acquire an advantageous combination of gene(s), or entire functional clusters, from phages infecting the same bacterial cell, allowing for an optimal adaption to a particular niche[50,71].

1.2.6 Control of phage lysogeny

The genes implicated in the establishment and maintenance of the lysogenic state are grouped in the so-called lysogenic module[120]. An on/off mechanism, or genetic switch, has been established for the lambdoid phages, determining the maintenance of lytic or lysogenic cycle after infection of a temperate phage. In most λ-like phages, the lysogenic module consists of two genes encoding transcriptional regulators, CI (the repressor), and Cro (the antirepressor)[130]. The constitutively expressed CI binds to an intergenic promoter region between *c*I and *cro*, repressing the expression of genes involved in the transition to the lytic cycle, including *cro*. Conversely, Cro promotes the lytic cycle mainly by inhibiting the expression of CI synthesis during infection and following induction of a lysogenic phage, activating the expression of lytic genes. There is a site binding competition for the same promoter region between these CI and Cro, and since the repressor can repress transcription of Cro, this competition will determine the establishment of the lytic or lysogenic cycle[131]. In situations of bacterial stress, the bacterial SOS response will trigger the activation of the hosts' protease RecA, which stimulates autolysis of the CI repressor at a particular alanine-glycine bond[132,133]. The withdrawal of CI repression allows Cro to be expressed, which in turn inhibits the expression of CI, enabling the transition to the lytic state. This mechanism is also responsible for host immunity against phage superinfection. Thus, bacteria carrying a lysogenic phage in their chromosome cannot be re-infected by the same phage since transcription of most of the phage genes, except the repressor CI, are repressed.

1.2.7 Phage integration and excision

The integration of the phage genome into the host chromosome is a key event in the lysogenic cycle. After injection, phage dsDNA circularises through complementary 5'-overhangs (*pac* site) or by cohesive ends (*cos* site) sequences. Then, the virion integrates into the host genome through the site-specific recombination of phage attachment site (att_P site) and bacterial attachment site (att_C site), following the model proposed by Campbell[134-136]. The site-specific integration is catalyzed by a phage-encoded integrase (Int), which is encoded in the lysogeny module[137]. As a result, hybrid attachment sites (att_L and att_R) border the integrated phage genome. These junction sequences function as recognition sites for the excision of the phage genome[138]. After induction of the phage, via spontaneous or SOS-induced mechanisms, the Int and the excisionase (Xis), encoded by the ***xis*** gene located next to the *int,* catalyze the excision of the phage DNA from the bacterial genome. Both proteins, Int and Xis, act in a coordinated manner to induce the excision of the prophage[139-141]. Thus, a correct ratio of Int/Xis is vital in the establishment of lytic or lysogenic cycle. When lysogeny is established, the proteins implicated in the lytic cycle are quickly degraded. The proteins responsible for this degradation are host proteases, such as ClpX or ClpP and others[142]. The integration of some phages is further supported by two host-encoded proteins, the integration host factor (IHF) and the Fis protein[143,144].

1.2.8 Phage DNA replication

The replication module is located next to *cro* and the lysogenic module. Besides the genes involved in the phages replication, this cluster also features the single origin of replication (*ori*). In λ phages, bidirectional replication is initiated from this *ori,* generating θ replication (θ-like structures). A switch in the replication pattern to rolling circle mode will generate the concatemeric DNA molecules (σ-like structures), which are the substrate for the packaging process[145,146]. The replication module in the λ phage consists of two genes, *O* and *P,* combined in a binary compound near the *ori* site. The phage protein O, the helicase loader, initiates the replication process by binding to the λ *ori*

site and requires the help of the host-encoded DNA gyrase, which forms negative supercoils. P then binds to the host replicative helicase, DnaB, which directs the interaction of P with the supercoiled unwound region[147]. The DnaB helicase mediates the widening of the single-strand region and the host DnaG primase will allow the replication in both directions via the synthesis of RNA transcript primers. Finally, the host-encoded DNA polymerase III begins the bidirectional replication process[148]. Replication forks of the single strands produce replication intermediates, which resemble the Greek letter theta (θ-like structures). After completion of dsDNA synthesis, the RNA primers will be removed by the PolA and a host DNA ligase will refill the gaps. The resolution of the formed circular DNA dimer is aided by the host topoisomerase IV[149,150]. DNA replication in structures that resemble the Greek letter σ (σ-like structures)[151] starts in the final stages of the early transcription module after an interruption on the replicative fork by a poorly understood mechanism. σ DNA replication is considered to be the transition of the primary viral functions, performed by the lytic module, to the expression of the intermediate and late genes clusters (morphogenesis and lysis modules).

1.2.9 Packaging regulation systems.

Bacteriophages have developed a wide variety of self-encoded factors coordinating the different stages (early, intermediate and late stage) of phage infection. These regulatory mechanisms assure tight control of gene expression at all the stages of the phage life cycle. Furthermore, the late-expressed gene module directs the synthesis of three different function proteins i) structural proteins, comprising the head and tail, ii) proteins involved in the assembly of the virus particle, without being part of the phage structure, and iii) proteins responsible for the lysis of the host cell.

In bacteriophage λ, the positive regulator Q, an anti-terminator transcription protein, is involved in the regulation of late gene expression. Q binds to a region of DNA, termed *Qut* site, which overlaps with the promoter P_R', allowing the expression of the packaging and lysis module. Transcription from P_R' is constitutive, but stops after ~ 200 nucleotides in the absence of Q, at the t_R' terminator gene. When Q is present, transcription initiated at P_R'

extends without stopping at t_R', transcribing the entire morphogenesis and lysis clusters. In phage λ, this transcript has a total length of ~ 26 kb[152].

In the case of *S. aureus* phages and other phages infecting the Gram-positive bacteria, a diverse variety of late transcriptional regulators (Ltr) had been described. Ltr acts as activators of the late transcriptional operons (morphogenetic and lysis genes), being required for the transcription process of these modules[153,154]. To date, there are five different families of these phage-encoded proteins known, designated as RinA, ArpU, LtrA, LtrB and LtrC, functioning as transcriptional activators of the late operons. The Ltr proteins, encoded at the end of the early gene module, regulate the expression of late genes directed by the CI repressor. Activation of the late gene transcription is initiated after binding of the Ltr activators to a tightly regulated promoter region, located upstream of the *ter*S gene. Gene transcription and therefore phage packaging and lysis functions are exclusively controlled by the phage encoded Ltr proteins.

1.2.10 Packaging initiation, DNA recognition and capsid formation

Structural and lysis proteins are synthesised following the replication of the phage genome. Proteins include those required for head and tail assembling, scaffolding of the viral particle and proteins initiating the lysis of the host cell. Those genes are encoded in the highly conserved morphogenesis and lysis modules, whose mosaic genome structure can be found in lambdoid bacteriophages infecting closely and distantly related bacteria, from *E. coli* to archaebacteria[155,156]. The mechanism of phage capsid assembly is well conserved and can also be found in viruses infecting eukaryote cells such as herpesviruses[157]. Bacteriophages were traditionally used as a model for the process of capsid assembly and the understanding of the function of the ATP phage molecular motor. Among the most studied assembly models for lambdoid bacteriophages are phages such as λ, SPP1, HK97 and P2. Although singularities can be found in some of the assembly processes or the structures of capsid and tail, the general principles of the packaging process apply to most bacteriophages[105,158,159] (Figure 1.4).

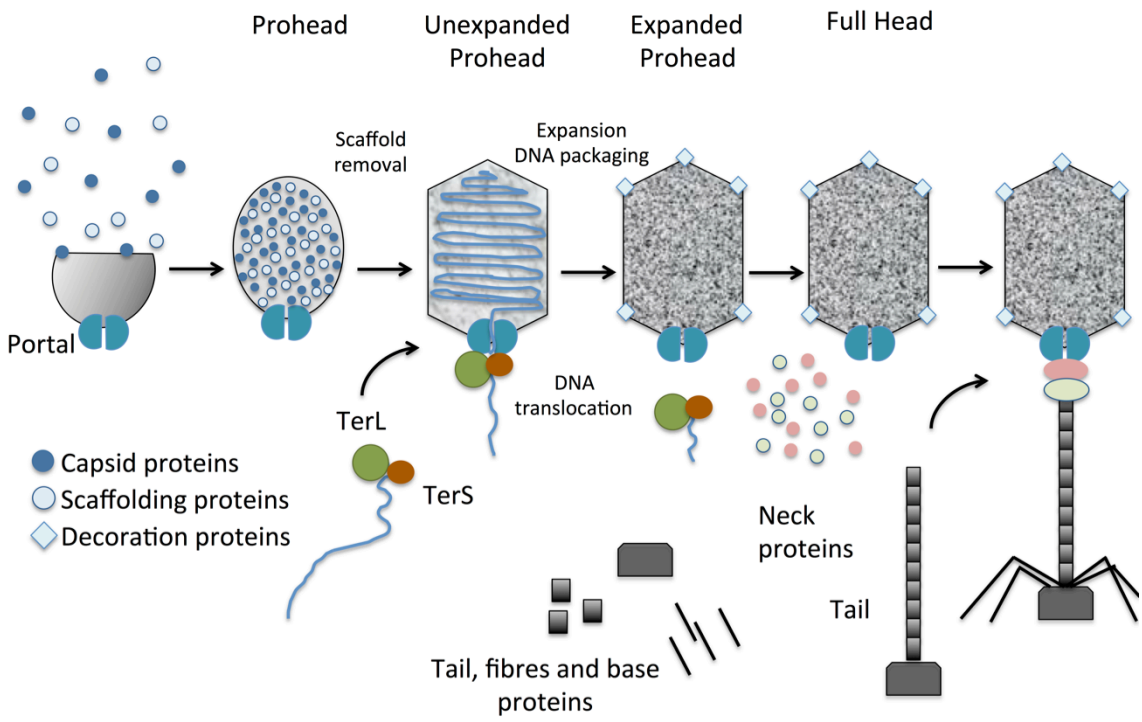


Figure 1.4 Schematic representation of the assembly of the virus particle of dsDNA bacteriophages. 1) Assembly of the procapsid from the portal vertex. 2) Packaging and scaffolding proteins removal. 3) Dissociation of complex terminase and incorporation of the decorative proteins and neck. 4) Assembly of tail proteins. 6) Unification of capsid, tail and tail long fibres and formation of the mature virus particle. Adapted from Rao *et al.* 2015[160].

1.2.10.1 Head assembling

The formation of an empty shell or procapsid, which is subsequently filled with the viral genome, is one of the first steps of the packaging process[161] (Figure 1.4). A portal protein will initiate the shell assembly, being a central component for the DNA translocation into the procapsid. This portal protein is formed by protein monomers that are grouped producing a dodecameric shaped cone with a central channel ring, through which the DNA will pass[162]. The assembly of the capsid shell is initiated at a single 5-fold portal vertex by the co-polymerization of monomers of the major capsid protein and in some cases scaffolding proteins[162-166].

The procapsid is an icosahedron constituted of multiple copies of the major capsid protein also referred as coat protein. Most phages encode just for a single capsid protein, although some phages, such as T4, encode two[159]. All capsid proteins within the family of lambdoid phages fold in a typical HK97

pattern, even though they sometimes share only 10-15% or less sequence identity. This standard folding has some structural features found in most bacteriophage shell proteins. It contains an N-arm in the N-terminal region with α-helical content; an E-loop with two-stranded anti-parallel β-sheets, a P-domain composed of a spine helix and a long β-sheet; and an A-domain with a central β-sheet. In some phages the capsid protein features an additional N-terminal scaffolding, or Δ domain, limiting conformational probabilities and supporting the formation of the procapsid (Figure 1.5). The 102 residues forming the Δ domain are often removed by proteolysis after the assembly of the procapsid. Examples for scaffolding domains can be found in phage T5 or HK97[167,168] (Figure 1.5).

Some tailed phages, such as the P22, encode separate scaffolding proteins or core proteins in order to promote capsid proteins interaction and restrict the conformation of the procapsid structure[169]. The T4 phage encodes six different scaffolding proteins[170,171], whilst most other phages encode just one. Scaffolding proteins also act as an attachment point for the capsid proteins, coordinating the interactions of the monomers and determining the final structure of the procapsid. Scaffolding proteins are typically eliminated from the complex during the process of capsid maturation[172]. Hence, controlling the formation of the capsid structure.

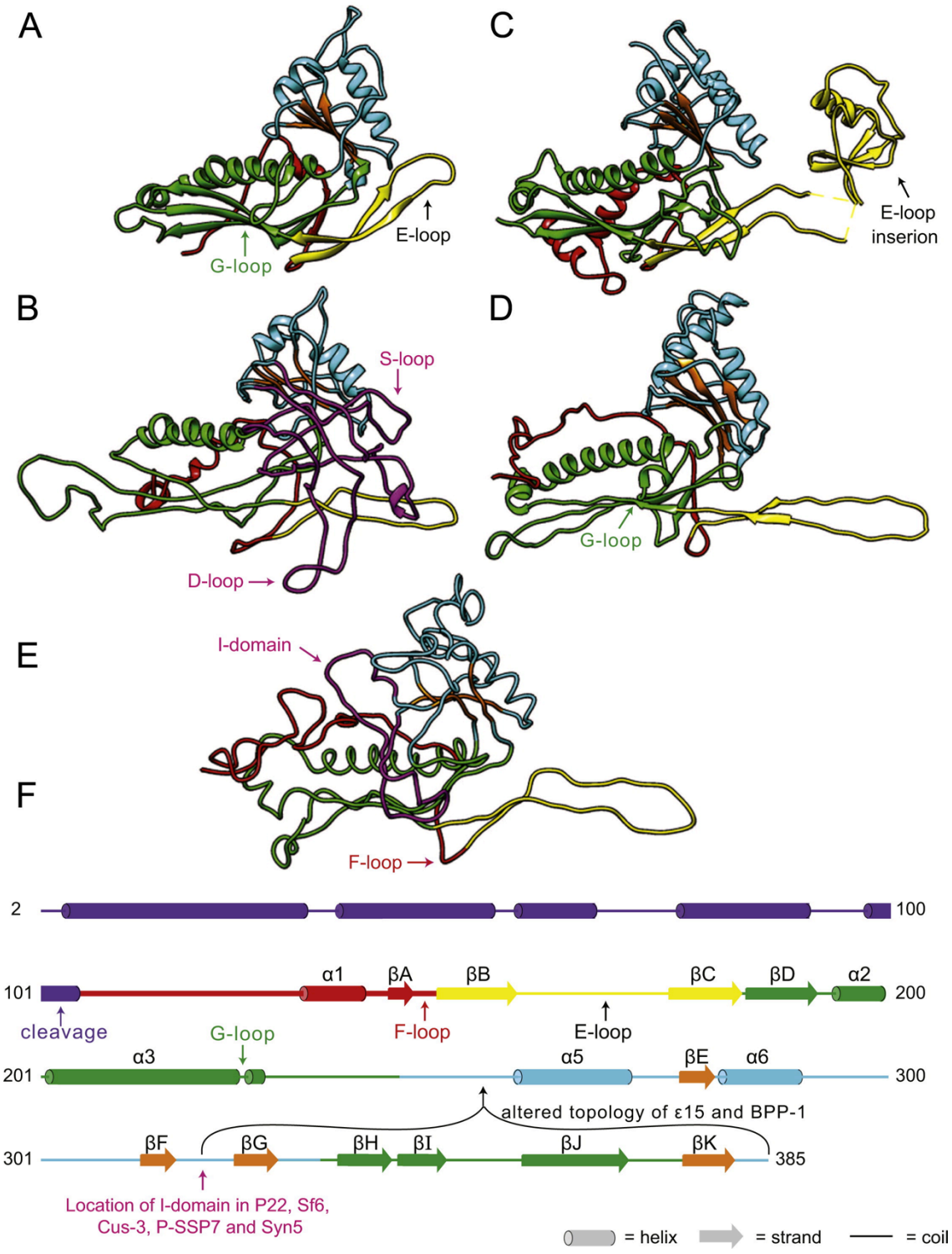


Figure 1.5 HK97 and related coat proteins. HK97 coat protein in the immature state "Prohead II" ((A) PDB entry 3E8K) is compared to P22 ((B) from (Rizzo *et al.*, 2014[173])), T4 ((C) 1YUE), BPP-1 ((D) 3J4U), and P-SSP7 ((E) 2XD8) coat proteins. The common structural elements are coloured correspondingly; N-arm (red), E-loop (yellow), P-domain (green), A-domain (cyan), β−hinge (orange). P22 and P-SSP7 coat proteins have insertions between βF and βG of the β-hinge (magenta in (B) and (E)). (F) Linear diagram of the secondary structure of HK97 coat protein based on PDB 3E8K with predicted secondary structure (PSIPRED) for the Δ-domain. β-strands E, F, G and K make up the β-hinge. The "backbone" is color-coded as in (A)–(E). From Suhanovsky and Teschke 2015[174]. Reproduced with permission of Elsevier.

1.2.10.2 Genome recognition and processing

The phage genome replicates simultaneously with the assembly of the capsid shell. Packaging of the replicated dsDNA requires lineal head-to-tail genome concatemers generated by the rolling-circle mechanism. DNA packaging requires an oligomeric holoenzyme consisting of two subunits, the large subunit of the terminase (TerL) and the small subunit (TerS) (Figure 1.6). TerL provides both the ATPase function, providing the energy for the translocation of the phage genome into the preformed icosahedral shell, and the nuclease function, responsible for the cleavage of individuals copies of the phage genome[175]. The TerL subunit contains several other features that are consistent in all characterised tailed phages. One example is the Walker A/B motif, related to proteins with ATP binding and hydrolysis functions[176]. The nuclease activity of TerL is located in the C-terminal domain and is responsible for dsDNA cleavage during the packaging process[177] (Figure 1.6). This ATPase activity has been shown in phages such as T4, λ, and P22, among others[178–181].

The TerS subunit is responsible for the recognition of the cleavage site and modulates the nuclease and ATPase activities of the large subunit[103,176,182–184]. The DNA binding domain of TerS is located in the N-terminal region (Figure 1.6). It binds to the cognate recognition sequence, directing the cleavage of the phage concatemer performed by the TerL subunit[185]. After the initial cleavage, the terminase complex starts to translocate the phage concatemer under ATP hydrolysis from the portal vertex into the procapsid. Conformational changes in the portal protein and the complete fill of the capsid are the signals for the TerL subunit to produce the second cut of the phage dsDNA, after which the holoterminase complex dissociates from the procapsid[164,186].

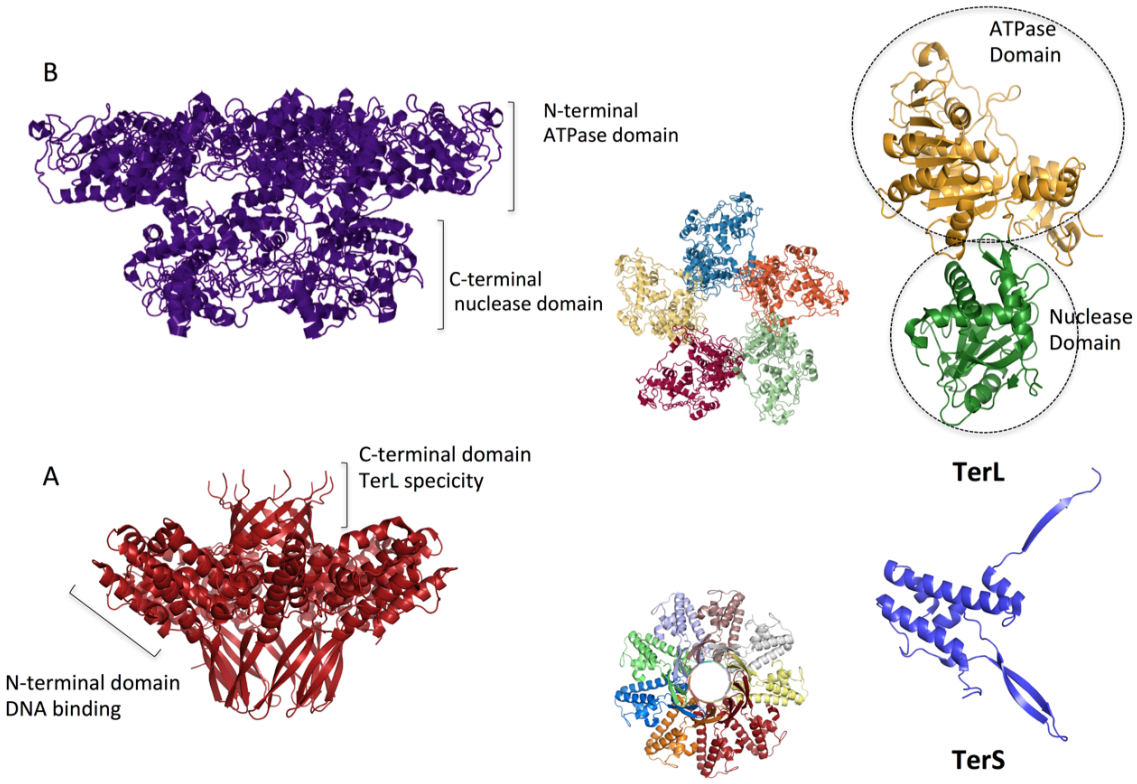


Figure 1.6 The structures of the two proteins involved in the holoterminase complex. (A) Monomer and side and top views of the homo-octamer small terminase subunit (TerS) from phage P22 (PDB 3P9A)[187]. (B) Monomer and side and top views of the pentamer large terminase subunit (TerL) from phage T4 (PDB 3CPE)[186]. The carboxy-terminal nuclease domains and the amino-terminal ATPase and are coloured green and yellow, respectively. Alternate identical subunits are shown in different colours. Parts are not drawn to scale. The ribbon diagrams were created with PyMOL.

Two models have been proposed for the cleavage and packaging of the phage concatemeric dsDNA[159,188-190]. In both models, the nucleoprotein complex TerS/TerL starts the translocation of the genome into a preformed procapsid, after recognition and cleavage of a specific cleavage site. The mechanism of how the packaging is completed differs between the two models. Dependent on the model, phages and cleavage recognition sites are classified as either *pac* or *cos*. For the *pac* phages, such as T4, P22, SPP1 or P1, or the *S. aureus* phages ϕ11 and ϕ80α, TerL will generate a second cut in a non-sequence-specific manner after the capsid is fully loaded. Redundant sequences at the ends of the phage chromosome, 3–10% of the phage chromosome depending on the phage, allow the molecule to circularise upon infection via homologous recombination. In this model, called "headful packaging", more than one unit-length, up to 105%, of the phage genome is packaged (Figure 1.7). Only the

recognition of a single *pac*-site is required to enable *pac* phage packaging. By those means, and since this process only requires the presence of a single *pac*-sites homologue in the bacterial chromosome to start the packaging process, *pac* phages are regarded to be the conductors of general transduction of foreign DNA.

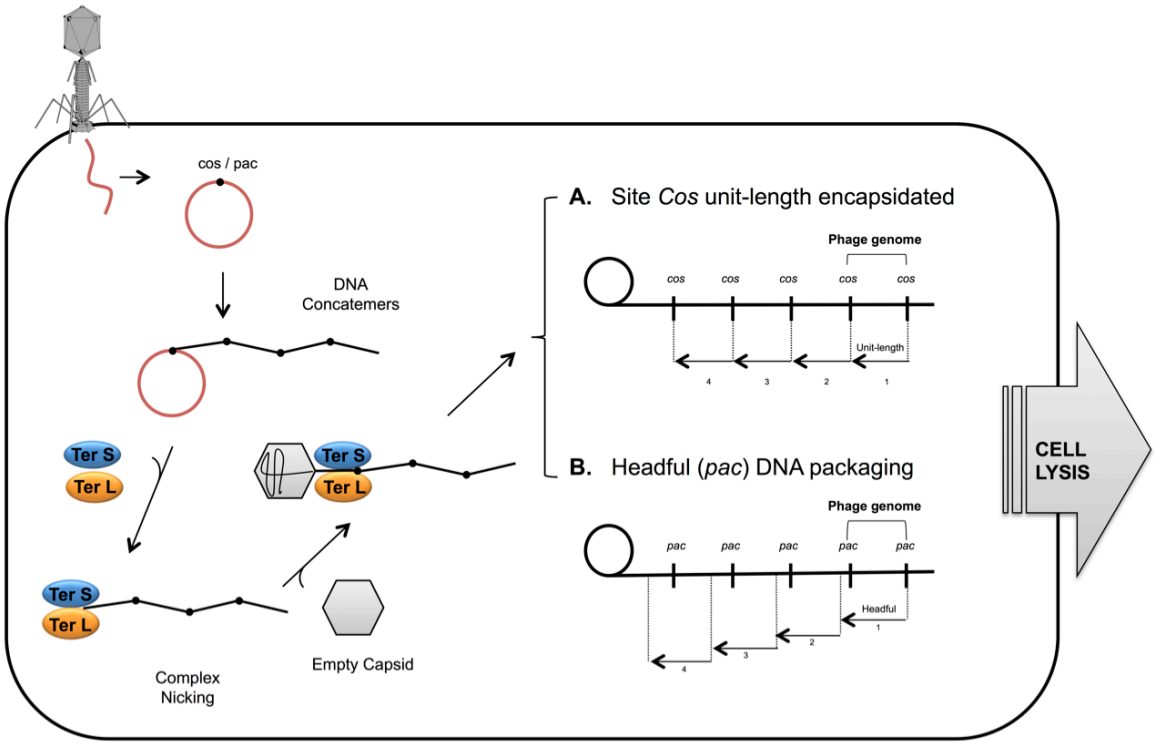


Figure 1.7 Classical models for cleavage and packaging of concatemeric phage dsDNA. Both models are conceptually identical, except in how phage packaging finalises after the procapsid reaches its capacity. In *cos* phages (A) the TerL produces a second specific cut in the same *cos* site, generating unit-length encapsidated molecules, in the *pac* phages (B) the TerL generates a second cutting a sequence-independent manner only when the procapsid is fully loaded.

In *cos* phages, such as λ, P2 or T7, or the *S. aureus* phages ϕ12 and ϕSTL, TerL will produce the second cut of the DNA at a second *cos* sequence, identical to the one the initial cut was performed at[159,188,189]. This mechanism produces dsDNA concatemers with cohesive ends, which allow for the subsequent circularisation in the recipient cell (Figure 1.7). In the case of λ phage, *cos*-site cleavage produces 12 bp base-long cohesive ends, designated as the *cos*N site. *Cos*N provides identical chromosomal ends, which allows the molecule to circularise upon infection generating unit-length encapsidated molecules. Beside from this site, two other sites, *cos*B and *cos*Q, will determine the

initiation and termination of the process. Due to the more restrictive characteristics of this packaging mechanism, *cos* phages are not regarded to participate in generalised transduction, since the likelihood of encountering two identical pseudo *cos*-sites in a unit length distance in other parts of the bacterial genome is low.

1.2.10.3 The HNH endonuclease

HNH domain proteins are wildly prevalent in nature, being found in bacteria, phages, viruses, archaea and eukaryotes. The HNH motif is a nucleic acid binding and cleavage module of around 30 to 40 amino acid residues and a single divalent bounded metal ion, usually a Zn^{+2} ion. The HNH abbreviation corresponds to the histidine and asparagine residues present in the degenerate motif[191]. The HNH motif is found in a diversity of enzymes, vital in a variety of cellular processes, including DNA repair, replication, recombination, bacterial killing[191].

Interestingly, HNH endonucleases are present in a number of *cos* bacteriophages infecting Gram-positive and Gram-negative bacteria[192,193] (Table 1.2). In phage HK97, the holoterminase complex alone cleaves *cos* site sequences only in an inefficient manner and needs the presence of the HNH nuclease to cut the phage genome efficiently[194]. Sequence analysis had also shown that HNH proteins are present not only in many *cos* phage terminase complexes but also in viruses like human cytomegalovirus (HCMV) and herpes simplex virus (HSV). In the HCMV model, the terminase complex binds to the HNH protein forming a heterotrimer, which will be crucial for the correct cleavage and packaging of the virus[195,196]. Furthermore, all genes encoding HNH proteins share a common genetic position in the phage genome, located next to the 5' of the *ter*S genes. Although all analysed phage HNH proteins carry endonuclease activities[192,193], the biological function of these phage-encoded proteins remain unknown and is yet to be identified.

Table 1.2 Phages encoding proteins with HNH endonuclease domains.

Species	Access number
Enterobacteria phage ϕP27	CAC83552.1
Listeria monocytogenes FSL J2-071	ZP_06557004.1
Lactobacillus lactis bIL285	NP_076611.1
Clostridium perfringens B str. ATCC 3626	ZP_02636982.2
Bacillus thuringiensis serovar sotto str. T04001	ZP_04127949.1
Bacillus amyloliquefaciens subsp. plantarum YAU B9601-Y2	YP_005421965.1
Bacillus cereus AH820	YP_002453337.1
Geobacillus kaustophilus HTA426	YP_146379.1
Lactobacillus salivarius ACS-116-V-Col5a	ZP_07206929.1
Paenibacillus larvae subsp. larvae B-3650	ZP_09072400.1
Staphylococcus caprae C87	ZP_07841333.1
Paenibacillus larvae subsp. larvae B-3650	ZP_09072400.1

1.2.11 DNA translocation machinery and packaging termination

The minimal requirements to perform viral packaging include the terminases complex, providing the ATP energy to translocate the viral genome, and the proheads with the portal channel (Figure 1.4). During the replication of the genome and its incorporation into the shell, the capsid undergoes several stages of maturation, including the elimination of the scaffold proteins[197]. In the case of the phages P22 and phi29, these proteins exit the procapsid, through several holes at the centre of the capsid[169,198]. Scaffolding proteins can also be proteolytically cleaved by self-encoded proteases, as it is the case of phage HK97 and T4[199]. In *S. aureus* phages, such as phage ϕ80α, scaffolding proteins are eliminated via bacteria-encoded proteases[200]. Hence, the packaging of DNA causes the exit of the scaffolding proteins triggering capsid maturation, which results in an expansion of the capsid to accommodate the entire viral genome[201] (Figure 1.4). In some cases, the expansion of the capsid is promoted by the formation of covalent bonds between coat subunits[97]. Other tailed phages accomplish stabilisation of their shell through the binding of self-encoded decoration proteins binding to the outer capsid surface. An

example for a decoration protein is gpD of the λ phage, necessary for the packaging of the phage genome[202,203], or the Soc protein in phage T4, required for virion stability in environments of extreme pH and temperature conditions[204]. Furthermore, minor capsid proteins can be incorporated, albeit present in a low copy number and not essential to the formation of the structure of the capsid, they are crucial for the virions infectivity efficiency. The P22 phage, for example, has three minor proteins modulating ejection of the dsDNA during infection[205].

Once the viral DNA is packaged, the terminase dissociates from portal protein and is replaced by gatekeeper proteins, forming the neck of the viral particle[206] (Figure 1.4). The portal protein forms a complex with connector proteins composing a connector channel preventing the premature exit of the DNA before the association of the tail structure completes the viral particle. This tail structure will be crucial, allowing the release of DNA upon binding of the phage fibres to the bacterial host receptors[207].

In the families of *Siphoviridae* and *Myoviridae* the phage tail assembles via a pathway separate from the formation of the capsid[208], whereas in the *Podoviridae* phages, tail proteins are sequentially attached to the newly formed capsid[209,210]. The assembly of the tail in *Myoviridae* and *Siphoviridae* begins with the formation of an initiator complex called motherboard or baseplate, which will constitute the phage absorption tube from its distal end[211]. The cylindrical section of the tail is formed at the baseplate through the assembly and polymerization of the tail protein(s), passing through several conformational changes during this process. The length of tails in phages of the *Myoviridae* and *Siphoviridae* families are determined by a tape-measure protein, which acts as a scaffolding protein during tail polymerization[212]. The tail is generally composed of multiple copies of the major tail protein, although there are exceptions such as for bacteriophage SSP1, which encodes two major tail proteins present in a ratio 3:1[213]. In *Myoviridae* phages, the newly formed tail is subsequently covered by a contractile sheath protein, which contracts during phage infection driving the tail through the host bacterial membrane[214]. Once the tail assembly is finished, completion

proteins interact with tail proteins mediating the binding of tail and baseplate to the mature capsid[215,216] (Figure 1.4). Finally, following the association of capsid and tail complexes, the binding of long tail fibres completes the virion assembly. Phage fibres are responsible for the recognition of host bacterium receptors[217,218] and remain in a retracted position when conditions are unfavourable for the growth of phage, preventing infection[219].

Immediately after the formation of the virion particles, the newly formed phages leave the host to infect new bacterial cells. Bacteriophages have developed multiple strategies to promote the release of virions out of the host cell. Two phage-encoded proteins, the endolysin and the holin protein, can be found in most of the studied phages. The endolysins or lysines are enzymes that degrade the bacterial wall, while the holin is a small protein that accumulates in the cytoplasmic membrane causing small nonspecific holes, making it permeable and allowing the lysin to act on the peptidoglycan in the bacterial wall[220,221].

1.3 Phage Inducible Chromosomal Islands

1.3.1 Overview of the genomic island family

Genomic islands (GIs) are discrete sections of DNA, generally between 10 and 200 kb in length, which are found in bacterial genomes but differ significantly from the rest of it. The nucleotide composition, percentage of GC content, and codon usage can diverge from the rest of the host chromosome, suggesting GIs were acquired via mechanisms of HGT. GIs are usually flanked by 16-20 bp long direct repeats, designated as att_L and att_R sites, which, as seen in phages, act as a recognition site for the specific enzymatic cleavage and excision from the host chromosome[222]. They often encode genes for mobility, such as integrase, plasmid conjugation factor systems or bacteriophage-related genes, suggesting mobility or parasitism of other elements that are themselves mobile. Genes included in their genomes provides a selective advantage to the host bacterium, being the force driving their maintenance in the bacterial chromosomes. Depending on the nature of those genes, the GIs are often described as pathogenic, symbiotic, metabolic,

defence or resistance islands[222]. Under this definition, a vast number of functional elements with different lifestyles are included in this family.

The first described genomic island was the so-called pathogenicity island (PAI) in *E. coli*. J. Hacker and colleagues coined this term, in the late eighties, while investigating the genetic basis of the virulence of *E. coli* uropathogenic strains[223,224]. PAI elements were described as unstable chromosomal regions with varying virulence characteristics that were associated with the phenotype of each strain[225]. The term GIs describes a much broader and diverse group, including PAIs and other elements that do not encode virulence factors, differing highly in size and being widely distributed among bacterial genomes[226–228]. GIs are involved in a wide range of processes such as virulence, symbiosis processes[229], metabolism of sucrose and aromatic compounds[230], mercury resistance and cytochrome *c* biosynthesis[231]. A significant number of GIs were identified in *S. aureus* clinical clones, varying in organisation and gene content, which included staphylococcal cassette chromosome *mec*, pathogenicity islands encoding the toxic shock syndrome toxins (TSST-1) and the vSa (vSAα, vSAß, and vSAγ) genomic islands[35,232–234]. Moreover, it was shown that clinically relevant strains of *S. aureus* carry more than one of these mobile genetic elements[35,89].

One of the most studied groups of GIs is the family of *S. aureus* pathogenicity islands (SaPIs). The SaPIs are phage satellites that maintain a close relation with the temperate phages (helper phage) they parasitize on, in order to be induced and transferred to a new host cell. SaPIs share a common life cycle, similar to the mechanism found for the plasmid P4[88] or the *Sulfolobus* plasmid pSSVx[235], which also exploit their host for their replication. The initial studies on SaPI mobilisation revealed that these genomic islands are excised from the bacterial chromosome and circularise after their induction by some helper phages. Two scenarios can lead to this activation, which could be either the induction of a resident prophage by environmental stress, including certain antibiotics[79,109,110], or the infection by susceptible strains carrying the SaPIs[13,236] (Figure 1.8; Figure 1.9). SaPIs replicate after their excision from the bacterial chromosome and are packaged into virion particles composed of

phage proteins, allowing their intra- and inter-generic transfer to other cells at a high frequency[13,110]. This series of these events is denoted as the excision-replication-packaging (ERP) SaPI cycle (Figure 1.8; Figure 1.9).

The ERP cycle of SaPIs can be easily examined in the laboratory using a strain harbouring a pathogenic island and an inducible helper phage. Activation of the SOS response leads to the induction of the phage and consequently of the SaPI (Figure 1.8). The monomeric form of the islands, packaged in phage smaller particles, can be visualised by Southern blot using a SaPI probe. Correspondingly, the larger band observed, results from the replicating (concatemeric) SaPI, which will be the substrate for the packaging process (Figure 1.8).

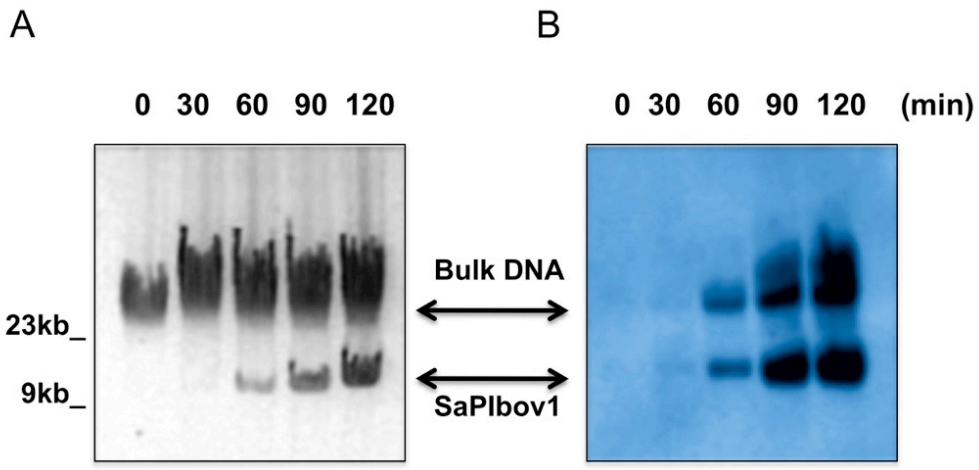


Figure 1.8 SaPIbov1 induction by phage ϕ80α. (A) Separation of DNA samples extracted from agarose gel after induction of the strain with Mitomycin C. (B) Southern blot of these samples using a SaPI specific probe SaPIbov1. Samples were taken at different times.

As shown in Figure 1.10, SaPIs display a prophage-like gene organisation, encoding some phage-homolog proteins such as integrase, excisionase, primase, replication initiator and a small terminase subunit. Hence, SaPI lifestyle is functionally related to certain phages to be transferred to other bacteria (Figure 1.9). SaPI chromosomes are flanked by direct repeats of 16-20 bp, called *att* sites (att_L and att_R). To date, there are five different integration sites (att_C) known in *S. aureus*, which are distinct to those used by the bacteriophages. Most of these SaPIs encode TSST-1 (toxic shock syndrome toxin) together with two or more toxigenic superantigens[61]. Furthermore,

several SaPIs were found to encode additional genes involved in the formation of biofilm (SaPIbov2)[236], antibiotic resistance (SaRIfusB)[237], or host adaptation (SaPIbov4, SaPIbov5, SaPIeq1 and SaPIov2)[238]. The fact that most SaPIs contain these type of genes in their genome highlights the significance of the role of SaPIs in the *S. aureus* virulence.

Although first discovered in *S. aureus*[13], SaPI-like elements are found in other Gram-positive bacteria, being recently described in *Streptococcus pyogenes* (SpyCIM1), *Enterococcus faecalis* V583 (EfCIV583) and *Lactococcus lactis* (LlCIbIL312)[89,239,240] (Figure 1.10). This family of highly mobile genetic elements was referred to collectively as phage inducible chromosomal islands (PICIs). PICIs precisely interfere with and exploit bacteriophages for high-frequency horizontal transfer. This novel family of MGEs has a major role in bacterial evolution and the emergence of new super pathogens.

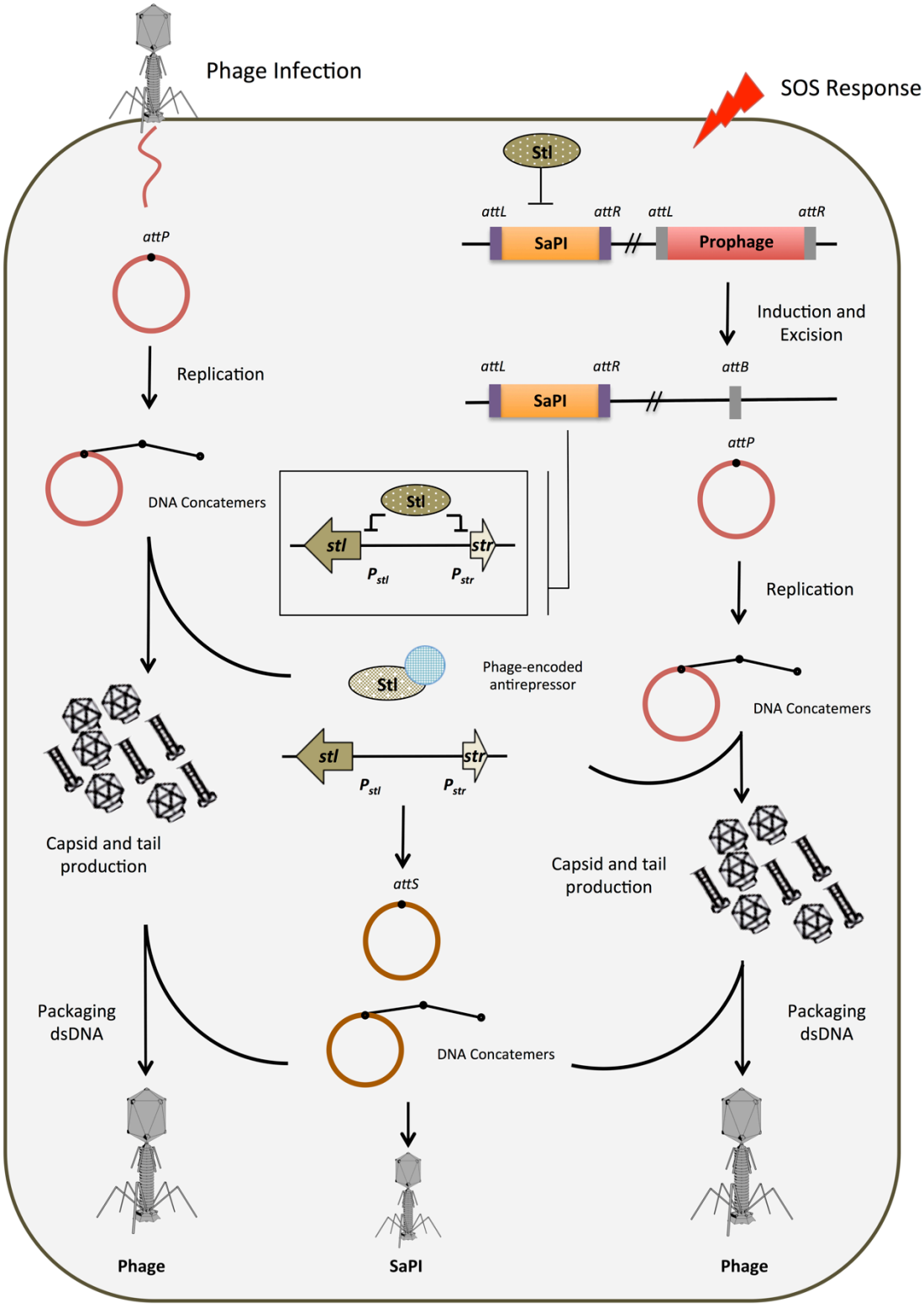


Figure 1.9 The SaPI Excision-Replication-Packaging cycle. Following infection or SOS induction of a helper phage, a phage-encoded antirepressor protein releases the Stl-mediated suppression of the SaPI cycle. SaPI-encoded integrase (Int) and excisionase (Xis) proteins promote SaPI excision by using the Campbell mechanism. After SaPI circularisation, replicates producing concatemers. The concatemeric DNA is substrate to the packaging process catalyzed by the Ter_{SP}-TerL complex. SaPI proteins redirect the phages capsid assembly process to the formation of small capsids in which the SaPI genomes are translocated.

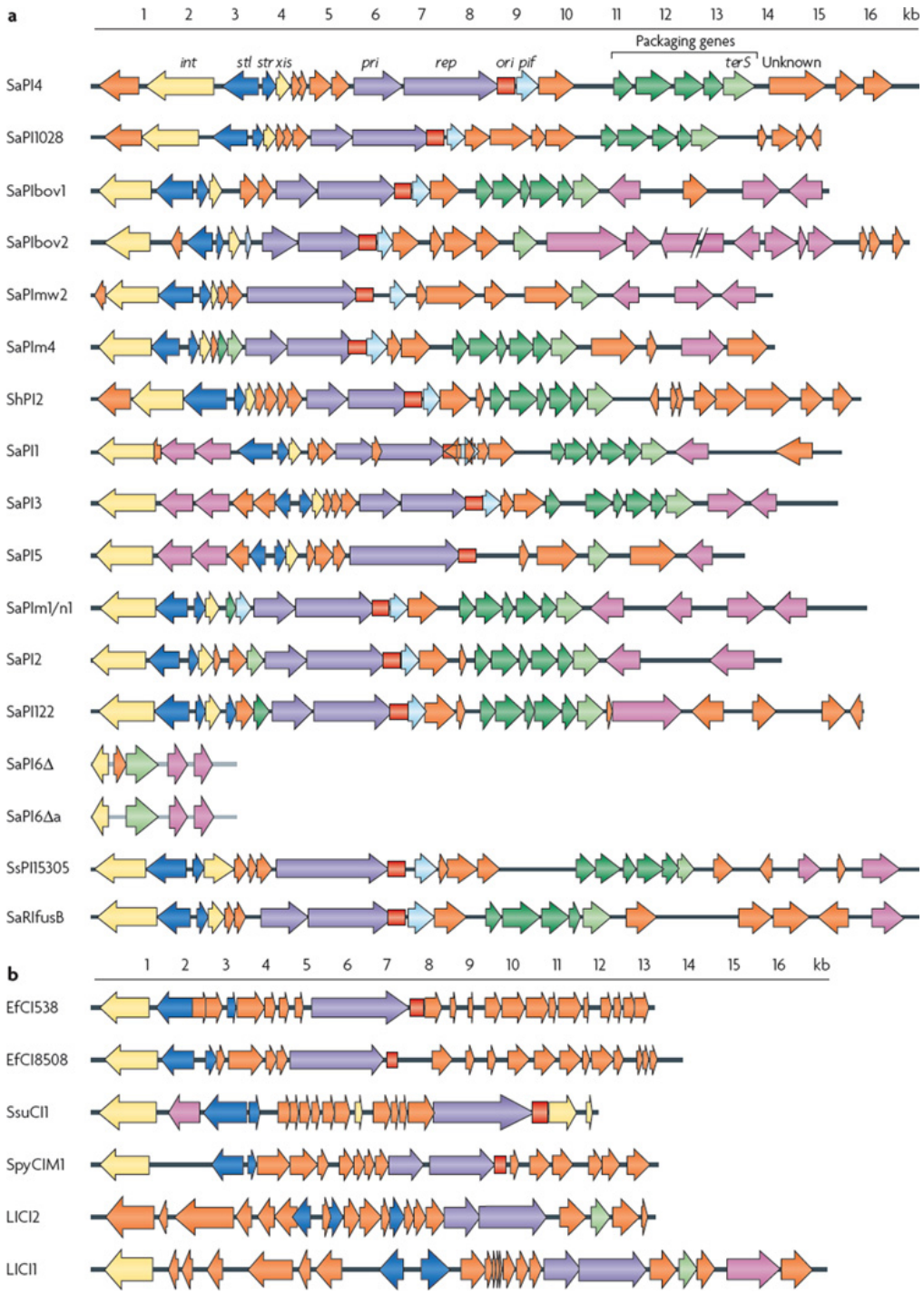


Figure 1.10 Comparison of phage inducible chromosomal island genomes. Genomes are aligned according to the prophage convention, with the integrase gene (*int*) at the left end. Genes are coloured according to their sequence and function: *int* and *xis* (excisionase) are yellow; transcription regulators are dark blue; replication genes (including the primase gene (*pri*) and the replication initiator gene (*rep*)) are purple; the replication origin (*ori*) is red; encapsidation genes are green, with the terminase small subunit gene (*ter*S) in light green; superantigen and other accessory genes are pink; and *ppi* (which functions in phage interference) is light blue. Genes encoding hypothetical proteins are orange. (a) The gene organisation of *S. aureus* pathogenicity islands (SaPIs). (b) Putative phage-related chromosomal islands from genera other than *Staphylococcus.* SaPIm1/n1 indicates SaPIn1 (from *S. aureus* str. n315) and SaPIm1 (from *S. aureus* str. mu50), which are essentially identical. From Novick *et al.* 2010[89]. Reproduced with permission of Nature Publishing Group.

1.3.2 Genome organisation and regulation of the SaPIs

SaPIs are the prototype member of the PICI family, having a highly conserved genetic structure comparable to the one of their host bacteriophages. Hence, genes with related functions are organised in functional modules (transcriptional, integration-excision, replication, morphogenetic and virulence modules)[89,241] (Figure 1.10). Some functions are conserved, but can be carried out by different proteins. SaPIs are not restricted to one particular helper phage and they can exploit the functions of several phages. Once they ERP cycle is activated, the SaPIs can also interfere with different helper phages.

1.3.2.1 SaPIs transcriptional control

The transcriptional module consists of the *stl* and the *str* genes, encoded in a divergent orientation (Figure 1.10). Together *stl* and *str* regulate the transcriptional activation of the SaPI genes. Both encoded proteins feature a turn-helix motif and are poorly conserved between different SaPIs[242]. The *stl* gene encodes a master repressor, exerting its action by binding to the promoter region located between *stl* and *str*. Expression of Stl blocks the excision and replication of the island, allowing the maintenance of the element in a quiescent state, integrated into the bacterial chromosome. Stl is a functional homologue to the phage lambda CI repressor, but unlike the CI, Stl is not degraded through auto-proteolysis upon activation of the SOS response. On the contrary, the Stl repressor is deactivated via protein-protein interactions with phage-encoded proteins produced during the phage cycle, which act as SaPI inducers. These de-repressor proteins bind to Stl, eliminating the Stl-mediated repression and activating transcription of the *stl* and *str* promoters, activating the SaPI cycle[242,243]. In the absence of the repressor, the island excises and replicates autonomously[241] (Figure 1.9). Remarkably, different SaPIs encode different Stl repressors. Therefore, the phage-encoded SaPI inducers differ between SaPIs, depending on the Stl repressor. This implies that the *S. aureus* bacteriophages differ in their ability to induce different SaPIs, and only those encoding the appropriate inducer

proteins, interacting with the repressor of a particular island, will be capable of inducing the SaPI cycle.

The second component of the transcriptional module is the *str* gene. The function of the Str is unknown so far, although unpublished data of our group suggest that Str could mediate rightward transcription of the SaPI genome. Str could act in a fashion similar to Cro in the lambda system, competing with Stl in the binding to the *stl-str* divergent region. Nevertheless, the absence of this protein has no influence on the ERP cycle of SaPIbov1[241] or SaPI1[244]. Further studies will be required to decipher the function of this protein within the SaPI cycle.

1.3.2.2 SaPI integration and excision

The integration-excision module is located next to the transcriptional module, comprising an integrase gene, *int* [236,245-247] and a gene coding for a protein with excisionase activity, ***xis***[248] (Figure 1.10). The SaPI-encoded tyrosine integrase catalyzes the integration of circularised SaPI intermediates into the host chromosome. The integration is catalyzed by the recombination of specific attachment sites in the hosts' chromosomal genome (att_C) and the SaPI (att_S), similar to the phage att_P site. The mechanism creates two hybrid sites, att_L and att_R, flanking the integrated island. Stl keeps SaPIs stable in the bacterial chromosome until they are de-repressed. After SaPI induction, the genes of the module are expressed, and a coordinated expression of integrase (*int*) and excisionase (xis) allows for the correct cleavage of the att_L and att_R, leading to the release of a circularised SaPI intermediate[249] (Figure 1.9).

Interestingly, SaPI integration never occurs twice in the same att_C of a given bacterial chromosome. Not only the presence of the chromosomal attachment site is needed for a correct SaPI integration, but also the flanking regions of this site for a correct excision[248]. However, SaPI integrases have a lower specificity, as deletion of SaPI1 att_C does not impair with its transduction rate *in vitro*, being integrated into non-conserved secondary att_C at the same rate[250]. Moreover, spontaneous excision has been demonstrated in the

absence of an inducing helper phage for SaPIbov1 and SaPIbov2[236]. Meanwhile, other SaPIs resides stables integrated into the host chromosome[247]. Although high levels of spontaneous excision are linked to high rates of transfer by non-inducing helper phages, it also correlates with a higher rate of loss. Differences in the Stl affinity to its binding sites could play a role in the mechanisms of spontaneous excision[248].

1.3.2.3 SaPI replication

The replication module is comprised of a cognate replication origin (*ori*) and the primase (*pri)* and replicase (*rep)* genes (Figure 1.10). The *ori* site is situated 3' of the *rep* gene and has a unique organisation, different to the *ori* present in many phages, consisting of repetitions of divergent iterons flanked with regions rich in A-T nucleotides[251,252]. The product of the *pri* gene has been shown to improve replication but does not seem to be essential[251]. The *rep* gene encodes a protein featuring a helicase function. Rep binds to the AT-rich region of the *ori*, initiating the replication of the circularised SaPI by melting the double strand within this region[251]. Occasionally, these two genes can be found fused. Thus, the structure and function of *pri-rep-ori* is a highly conserved pattern, not just in the SaPIs, but also among PICIs. It is assumed that SaPIs replicate in a rolling circle form, producing a single linear genome concatemer, which is further processed for encapsidation.

1.3.2.4 SaPI packaging and phage exploitation

The SaPI islands do not encode structural proteins for the formation of the virion particle, but instead utilise the structural phages proteins to form capsids for the packaging of their reduced size genomes[253,254]. The proteins associated with the assembly and the packaging of most of the SaPIs are organised in two distinctive gene clusters (Figure 1.10). The first module encodes Ppi (phage packaging interference), which blocks the interaction of the phage-TerS packaging machinery with the phage dsDNA. Ppi does not affect the SaPI-encoded TerS, ensuring advantageous packaging of the island over the phage and ultimately interfering with the life cycle of the inductor phage[255].

Related to the packaging module, most SaPIs encode well conserved morphogenetic and interference genes comprised in the so-called operon I, which is controlled by a LexA-dependent promoter, driving the expression of the genes in this module[256]. Included in this module is the gene encoding the SaPI small terminase subunit ($TerS_{SP}$). This protein forms a complex with the large subunit TerL of the phage terminase, to recognise and cut the specific *pac* sequence encoded in the intergenic region between the two modules of the island[246]. This cluster also encodes two proteins, CpmA and CpmB, which redirect the assembly of the host phages procapsids[254,257], toward the production of smaller capsids, only facilitating the packaging of SaPI genomes[110,247,258] (Figure 1.10). Furthermore, *pti*A, *pti*B, *pti*M genes are located in this morphogenetic module. The expression of these genes interferes with the phage reproduction by blocking the helper phages late gene expression operon[259] (Figure 1.10). As mentioned previously, SaPIs encode a specific *pac* site sequence which is recognised by SaPI-encoded $TerS_{SP}$[260,261], facilitating the specific packaging of the SaPI DNA into the SaPI and phage capsids (Figure 1.9). Interestingly, this LexA-controlled operon I is induced by SOS response, and could be of benefit for SaPI-mediated gene transfer in the presence of a non-inducing phage[261]. However, this mechanism of operon I activation is not required for the normal induction SaPI ERC cycle[262], and after derepression, this operon seems to be transcribed from the *str* promoter[244].

1.3.2.5 SaPI accessory genes

Virulence genes and genes that increase the fitness of bacteria, including nonessential and accessory genes, are also found in SaPIs. They are usually located within two regions in the SaPI genome. One location is left to the *stl* and *int* region and the other at the end of the genome, after the operon I (Figure 1.10). These 'morons' or accessory genes are similar to the ones encoded by the temperate phages and their expression depends on their own promoter regions.

The vast majority of staphylococcal superantigens (SAgs) are encoded by MGEs[62,263]. More specifically, SaPIs are known to encode the TSST-1 protein

and superantigens including enterotoxins such as B, C, K, L and Q[13,36,89] (Table 1.3). Likewise, other virulence factors with different roles, such as the biofilm associated protein Bap[236], the iron transporter FhuD[36], the exfoliative toxin ETA[264], the staphylococcal complement inhibitor (SCIN)[238,265] or the von Willebrand factor-binding protein (vWbp)[238] can be found. The different variations of the SaPI-encoded von Willebrand binding protein allow *S. aureus* to specifically coagulate bovine, ovine, caprine and equine plasma[238]. The fusidic acid antibiotic resistance (FusB) is another SaPI encoded virulence factor, recently found in impetigo clones[237,266]. Another SaPI-like element, *v*Ss15305, found in the *S. saprophyticus* genome encodes antibiotic resistance determinants, such fosfomycin and streptomycin[267].

Table 1.3 Pathogenicity islands of *S. aureus* and its virulence genes.

SaPIs	Size (kb)	Virulence genes	Reference
SaPI1	15,2	*ear, tst, sek, seq*	13
SaPI2	14,7	*tst , eta*	13,264
SaPI3	15,6	*ear, seb, sel, sek*	247
SaPI4	15,1	-	35
SaPIbov1	15,8	*tst, sel, sek*	263
SaPIbov2	27	*bap*	236
SaPIn1 SaPIm1	15	*tst, sel, sec3*	14
SaPI1028	15,6	-	56
SaPI5	14	*ear, sek, seq*	268
SaRIfusB	16,7	*fus*B	237

1.3.3 SaPI-helper phage interactions

SaPIs rely entirely on the presence of a helper phage to be propagated and transferred. They exploit their helper phages in many different manners and developed several interference mechanisms to manipulate the phage to their own benefit. The exploitation of the phage morphogenetic machinery not only enables high transfer rates, but ultimately impairs the phage reproducibility, eventually helping host bacteria by reducing phage superinfections[247,259].

Several complementary strategies have been described by which different SaPIs exploit the phage machinery, not all of them being present in the SaPIs.

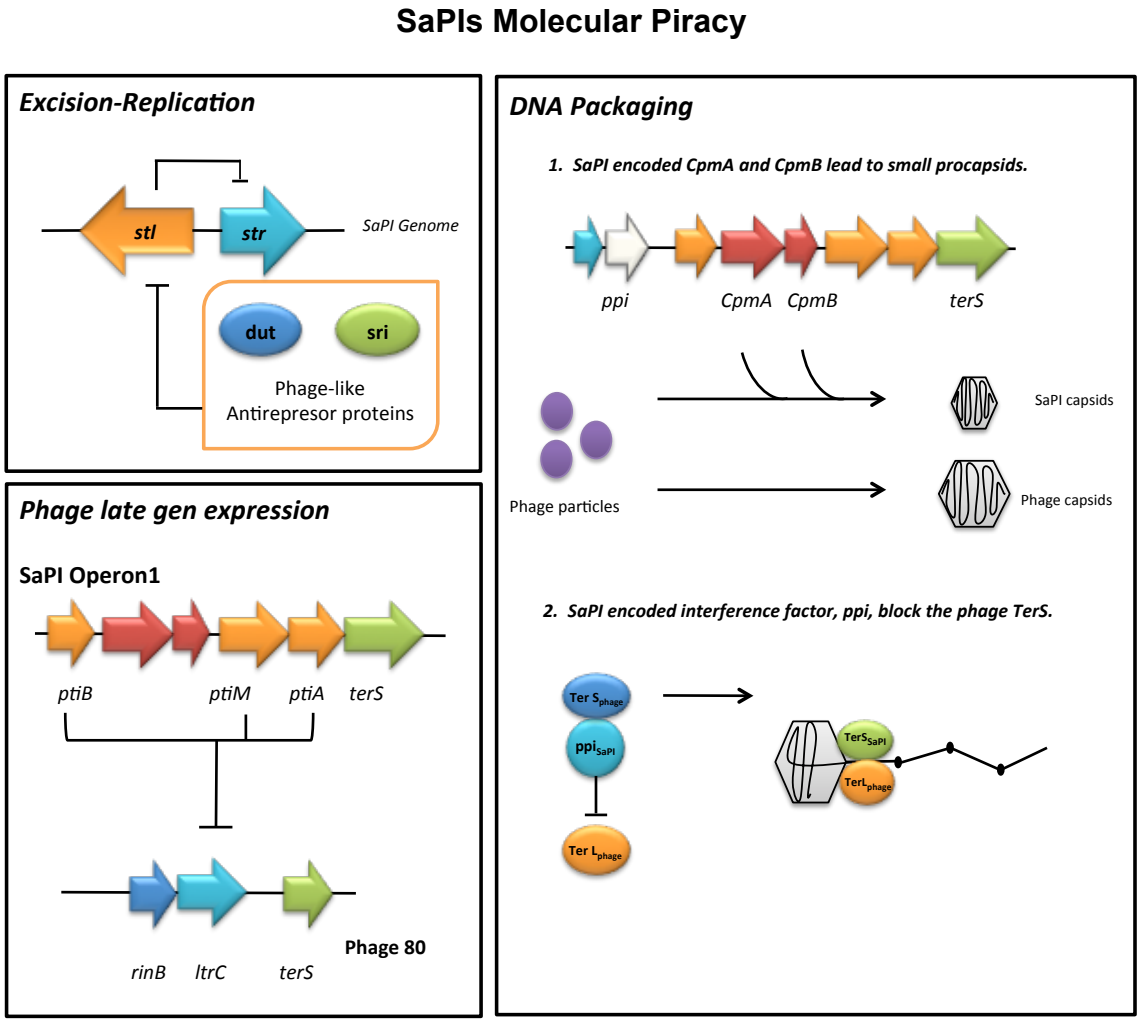


Figure 1.11 Mechanisms of SaPI interference with the phage reproduction. Phages induce the SaPI cycle by expressing proteins that block the activity of the SaPI-encoded Stl repressor. SaPI induction blocks phage reproduction using alternative and complementary mechanisms. These include blockage of the phage LtrC regulator by the PtiABM proteins, generation of SaPI-sized capsids by the expression of the CpmAB proteins, or blockage of the phage TerS by the Ppi protein.

1.3.3.1 SaPI induction and derepression

As previously mentioned, SaPIs reside in a stable state in the host chromosome, maintained by the action of the self-encoded Stl repressor, which binds to a region in between the divergent promoters, blocking the expression of the SaPI genes. This repressor, unlike the one of the temperate phages, is not degraded by auto-proteolysis after induction of the bacterial SOS response. Instead, the island uses a “moonlighting” phage protein as the

inducer of the SaPI cycle. This phage antirepressor protein binds to the Stl binding site of the SaPI, dissociating the Stl-DNA complex, allowing the expression of the SaPI genes involved in its mobilization[242,269] (Figure 1.9, Figure 1.11). As different islands encode different repressor proteins, different phage proteins are required to specifically induce the different islands[242]. Phage ϕ80α can induce the vast majority of SaPIs tested, such as SaPIbov1, SaPIbov2, SaPIbov5, SaPI1 and SaPI2[13,110,242,245,253,270]. Other phages, on the contrary, activate only a limited spectrum of SaPIs. Phage ϕ80 can induce only SaPIbov1 and SaPI2[13,110,264], ϕ11 de-represses SaPIbov1 and SaPIbov5[245,270], ϕNM1 induces SaPI1, SaPIbov1 and SaPIbov5 and ϕNM2 activates SaPI1, SaPIbov1 and SaPIbov5[271]. However, phage ϕ85 does not induce any of the aforementioned islands[245].

Three "moonlighting" phage proteins were described initially for the phage ϕ80α, and each targets different SaPIs[242,270]. These proteins include the phage-encoded trimeric dUTPase (Dut), the inducer for SaPIbov1 and SaPIbov5; the Sri protein, the inducer for SaPI1; and the ORF15, with unknown function, which derepresses SaPIbov2[242,246] (Figure 1.11). The Sri protein was first found in phage 77, where it binds and blocks the host helicase DnaI, inhibiting the bacterial DNA replication[272]. Dut modulates the dUTP levels, but has also been recently proposed that these enzymes can act as signalling molecules, employing dUTP as a second messenger[243,269]. The variety of the proteins involved in the process of SaPIs induction manifests the evolutionary pressure on these elements to overcome the continuous adaptation of phages to avoid infection by SaPIs. The co-evolutionary arms race of these MGEs forces the emergence of novel variants of the antirepressor proteins with lower affinity to the Stl region[242,273]. Additionally, a reduced expression of the phage-encoded antirepressors has been described as an adaption of phages to avoid SaPI interference[273]. In some phages the antirepressor protein was replaced by other proteins with the same biological function but with a completely different sequence, avoiding SaPI activation[273]. In conclusion, SaPIs represent a considerable driving evolutionary force for staphylococcal phages, sparking the generation of a new pool of phage-encoded genes via selective evolutionary pressure.

1.3.3.2 Assembly and capsid size redirection

Modulation of the phage-encoded capsid assembly pathway towards the formation of smaller particles is one of the most observed mechanisms in SaPI-mediated molecular piracy (Figure 1.11). This strategy is not only found in SaPIs and is a widely distributed mechanism amongst the PICI family and other MGEs. This small capsids mechanism not only procures the assembly of particles to commensurate their reduced genome size (~15 kb; Figure 1.10) but also prevents the package of the complete larger helper phage genomes[274-276]. This redirection leads to the production of virion particles with *T*=4 icosahedral symmetry, instead of the *T*=7 phage particles formed of phage capsomeres[253,254,277,278] (Figure 1.12). The presence of several mechanisms redirecting the capsid size suggests that this strategy provides an ultimate and evolutionary advantage that also yields an interference function against the helper phage[255,257].

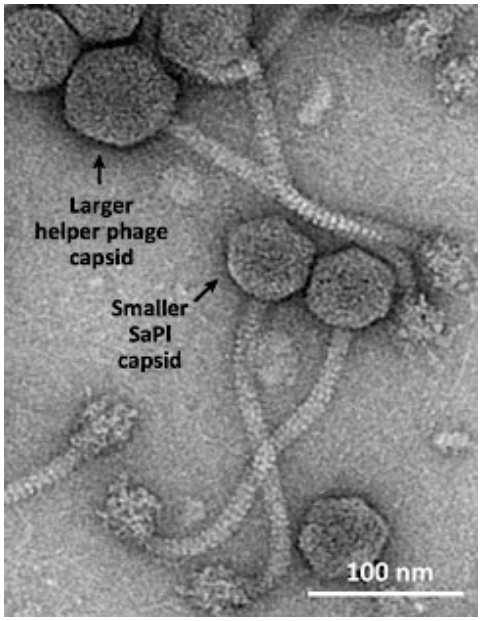


Figure 1.12 Electron micrographs of a mixed helper phage-SaPI lysate. From Novick and Ram 2016[279]. Reproduced with permission of Elsevier.

Some SaPIs control the size-reshaping of the capsid through the expression of the products of the *cpm*A and *cpm*B capsid morphogenetic genes[256,258] (Figure 1.10; Figure 1.11). In the SaPI system, both CpmA and CpmB are necessary for an efficient redirection of the procapsid size and are not associated with the

mature shell, being removed after the formation of the procapsid[257,258]. CpmB is an internal scaffolding protein that forms dimers with the phages shell proteins, competing with the phages own scaffolding proteins, shaping phage capsid geometry into an icosahedral reduced capsid[280]. The role of CpmA in this capsid remodelling is less clear. Due to the low number of copies of this protein participating in the assembly process, it was suggested that CpmA has a transitory function on the capsid assembling[258]. Hence, CpmA could be involved in the regulation of the scaffolding core organisation avoiding the formation of small aberrant capsids. CpmA could also modulate the access of CpmB to the capsid proteins, preventing the binding of phage scaffolding proteins[200]. Nevertheless, CpmB and CpmA alone are not functional and need the presence and assistance of the phage capsid and scaffolding proteins[200,277]. Capsid size redirecting proteins perform the size remodelling only when they are compatible with the helper phage capsid system. It was shown that the efficiency of the formation of the reduce-size virion by the same SaPI was different depending on the phage. This is the case of SaPI2, which is mobilised in small capsids by the phage ϕ80α, but not by ϕ80 due to its incompatibility with the ϕ80 capsid protein, different to the one in phage ϕ80[281]. Moreover, after infection by ϕ80α size modelling of the capsid particle by SaPIbov1 is less efficient than by SaPI1, although their CpmA and CpmB proteins are nearly identical, suggesting that other factors may also play a role in this process[271].

The exploitation of the helper phage P2 by the MGE P4 is another example of molecular piracy depleting proteins unrelated to the previously mentioned CpmAB. The P4 plasmid redirects P2 capsid formation towards the assembly of small virions which only can accommodate P4 genomes[88,275]. The scaffolding protein Sid is responsible for this redirection[282], forming an external scaffolding structure encasing the procapsid with a dodecahedral cage[283]. Trimers of this elongated protein, constituted of bundles of α-helices, are formed around the capsid forcing the shell size remodelling[284]. Sid mutants fail to form the reduced shell[285] and mutants in the loop-helix motif of the phage capsid protein, gpN, responsible for Sid interaction, called Sir (Sid responsiveness), are resistant to the action of the P4 scaffolding protein[284,285].

The formation of P4 capsid is further helped by the phage scaffolding protein, whose role is believed to be related to its N-terminal protease domain, located inside the phage capsid. Furthermore, the P4 plasmid encodes a decoration protein, Psu, incorporated into the mature capsid, which will stabilise the final particle. Psu also acts as a 'moonlighting' protein, suppressing the *rho*-dependent transcription termination[286,287].

Size redirection is an evolutionary convergent mechanism, yet it is not required for an efficient transfer for all the MGEs. In the case of the P2/P4 system, *sid* mutants are perfectly viable and can be packaged in full-size phage capsids[285]. Likewise, SaPI dsDNA can be packaged in both capsids, small and large, depending on the SaPIs. SaPI1, for example, mainly uses small capsids, while SaPIbov1 uses small and phage sized capsids with similar efficiency[256]. Remarkably, SaPIbov1 and SaPI1 *cpm*AB mutants can be transduced at a high frequency into large phage particles, after being packaged[255-257]. Not all SaPIs encode these morphogenetic proteins and several SaPIs lacking some of these genes can be found in nature. This is the case for $SaPI_{pT1028}$, which does not encode the *cpm*AB genes, being packaged in full-size phage particles, allocating three copies of the 15 kb SaPI genome inside each particle. Another example is SaPIbov2 in which the CpmB protein is no present. The genome of SaPIbov2 is also too large (27 kb) to be encapsidated in small capsids and is packaged in phage-sized capsids as dimers[245]. Strikingly, the encoded *ter*S, *cpm*A, *pti*M and *pti*B present in SaPIbov2 have a low sequence identity with the prototype SaPIs, being similar to the ones encoded in the packaging module of some *Staphylococcus epidermidis* pathogenicity islands (SePIs). Moreover, SaPIbov5 (13.5 kb) does not encode protein homologues to the established operon I.

1.3.3.3 DNA packaging and phage interference mechanisms

As previously mentioned, the concatemeric dsDNA of phages is packaged and cleaved via two different mechanisms, both depending on the terminase TerS-TerL complex[159]. After the recognition of a specific region in the phage genome (*pac* or *cos* site) by the phage TerS (TerSϕ), the phage TerL (TerLϕ) facilitates the cleavage and translocation of the dsDNA into the procapsid in

an ATP-dependent manner (Figure 1.7). SaPIs exploit this packaging mechanism of their helper phages through different strategies.

The classical SaPIs use the headful mechanism to package their genomes mediated by the encoded $TerS_{SP}$, which recognises the SaPI *pac* site, directing packaging of the SaPI genome into the SaPI sized small capsids[246,256,260]. SaPI replication generates a concatemer, in the same way as their helper phages. After the cleavage of the *pac* site, redundant terminal DNA allow the molecule to circularise[247] (Figure 1.9; Figure 1.11). $TerS_{SP}$ binds to a specific *pac* sequence in the intergenic region upstream of the SaPI *ter*S_{SP} gene[246,260], redirecting the translocation of the SaPI genome in complex with the phage TerL. The binding of $TerS_{SP}$ to its unique recognition sequence promotes SaPI packaging at a high frequency, whilst mobilisation in its absence is decreased to the levels of general transduction[246,256]. SaPI $TerS_{SP}$ alone does not promote the helper phage interference, as mutants in the $TerS_{SP}$ did not increase the ability of the phage to be packaged[246,256]. Furthermore, TerLϕ and $TerS_{SP}$ have to be compatible in order to promote the specific translocation of the SaPI dsDNA. The *cos* phage ϕ13, for example, is able to induce *pac* SaPI1 but fails to produce SaPI virions, as the DNA packaging machinery of the *cos* phage is not compatible with the *pac* SaPI machinery ($TerS_{SP}$)[247].

Furthermore, SaPIs encode a phage packaging interference (Ppi) protein, giving $TerS_{SP}$ an advantage in the competition for binding to TerLϕ. Ppi binds directly to the phage small terminase subunit, although the exact mechanism of its interference is yet unknown[246,255]. Ppi could directly affect the TerSϕ recognition affinity to its related *pac* site, or could prevent the interaction with the phage TerL. Structural differences between the SaPI and phage TerS proteins prevent Ppi from blocking $TerS_{SP}$. $TerS_{SP}$ has an additional helix domain located in the C-terminal and a supplementary oligomerization domain located in the central part of the protein, being both absent in the phage TerS[253,258]. Two conserved allelic subsets of Ppi have been identified to date, which interfere with the DNA packaging machinery of different helper phages[255]. Class I, represented by the SaPIbov1 encoded Ppi, interferes with phages ϕ80α and ϕ11, whilst class II, represented by SaPI1, blocks ϕ12

packaging. Homologs of Ppi are found in other Gram-positive organisms such as *Eubacterium limosum* KIST612, *Aerococcus viridans* ATCC 11563 and *Clostridium bolteae* ATCC BAA613[255]. Those homologs of Ppi could be encoded by not yet identified PICIs, suggesting that the Ppi interference mechanism could be widespread amongst the PICI family.

Although this mechanism of interference, using the headful packaging, is the most common among the SaPI family, several variants seem to follow a different strategy. In some SaPI elements, the prototypical TerS-morphogenetic operon I is substituted by a ~90 bp fragment encoding a phage *cos* site sequence homolog. The prototypes of these new variants of *cos* SaPIs are SaPIbov4 and SaPIbov5. Since these elements lack the prototypical morphogenetic cluster, they were long believed to be defective. These *cos* SaPIs encode several highly conserved proteins in a cluster named here as 'operon I-like', which has not been characterised and for which no homologs have yet been found. Thus, the mechanisms behind the transfer of *cos* SaPIs is still unknown and will be one of the main focuses of this thesis.

Besides the examples found in PICIs, P4 enables its transfer by featuring the same 55 bp *cos* site as its helper phage P2, which is recognised by the phage terminases, mediating P4 encapsidation. The usage of the bacteriophage machinery allows packaging of P4 genomes at a high frequency[288]. This acquired *cos* site is the major feature by which P4 interferes and reduces the reproducibility of its helper phage P2[289]. The P2-P4 system, unlike most of the bacteriophages, uses covalently closed circular DNA intermediates as the packaging substrate[103,150]. Additional to this mechanism of interference, capsid size determination by P4 produce a decrease of phage reproducibility, depending on the infection condictions[289]. It is believed that a yet unknown function accomplished by Sid or another gene(s) could modulate this interference process. Nevertheless, in *sir* mutants, which no longer produce small capsid, P2 is no longer blocked[290].

1.3.3.4 Phage late gene transcription interference

Novel SaPI-encoded proteins were recently found to play a major role in the interference with the reproduction of their helper phage. These phage transcription inhibitors (Pti) proteins were shown to modulate the expression of the late gene transcription operon of some helper phages (Figure 1.11). This single polycistronic operon encompasses genes related to morphogenetic, assembly, packaging and lysis function, which are required for both phage and SaPI survival[154,259]. Its transcription is regulated by DNA-binding proteins such as RinA, ArpU, LtrA, LtrB and LtrC, which are members of the super family of the late transcription regulators (Ltr)[153,154].

The interference with the helper phages late gene transcription, mediated by the *pti* genes, was described for the gene product of SaPI2 ORF17, named PtiA[259]. PtiA blocks the reproduction of phage ϕ80 by binding to the phage-encoded LtrC protein (analogue of λQ), inhibiting the transcription of the morphogenetic cluster of the phage. Expression of the morphogenetic cluster of phage ϕ80, which depends on LtrC[281], was reduced by the SaPI2 PtiA interference mechanism. Thus, CpmAB or Ppi proteins encoded in the SaPI2 operon did not affect phage ϕ80 reproduction. In a likewise manner, SaPI2 PtiA protein did not interfere with the phage ϕ80α RinA late transcription regulator[153,259]. Similar correlations can be found in other SaPIs which helper phages are affected by the CpmAB or Ppi proteins, such as SaPIbov1, where the encoded PtiA variant does not interfere with phage ϕ80[259]. In conclusion, the different PtiA variants encoded in the SaPIs differently affect the Ltr variants expressed in certain helper phages.

PtiA activity is negatively regulated by another SaPI protein, Pti*M*, which is encoded next to *pti*A in operon I (Figure 1.11). Pti*M* directly binds to PtiA, modulating its repression activity over LtrC[259]. This regulation is paramount, since SaPIs need the packaging machinery of the phage to be encapsidated, and a total blocking of the late phage operon would be detrimental for both SaPIs and phages. Furthermore, another protein encoded in operon I has been

identified to contribute to this interference, named PtiB, which blocks the phage by an unknown mechanism[259].

1.3.4 SaPIs as vehicles for horizontal intra- and intergeneric virulence gene transfer

The close relationship between SaPIs and their helper phages, mimicking their gene organisation and functions, highlights that these elements have evolved in parallel to their helper phages to spread their genome efficiently among the staphylococci. In fact, recent sequencing analysis of a number of pathogenic *S. aureus* strains revealed that these strains are poly-lysogenic and also possess one or more islands in their genome. In addition to the wide distribution of SaPIs within the staphylococcal genus, there is evidence that SaPIs can be horizontally transferred to different genera and distantly related bacterial species. It was shown that *pac* SaPIs, such as SaPIbov2, could be moved from *S. aureus* to other staphylococci, such as *S. epidermidis* and *Staphylococcus xylosus*[110]. Furthermore, SaPI1 and SaPIbov1 can also be mobilised to *Listeria monocytogenes*[245,250]. The transfer of a SaPI does not require the helper phage being able to parasitize the new host bacteria and only requires an appropriate recipient strain to inject the dsDNA, and its circularisation and integration into the new host att_C site (Figure 1.13).

The host range of a phage is usually determined by its ability to lyse certain bacteria, visualised by the formation of plaques. Nevertheless, the capability to inject the dsDNA from the capsid into a cell is sufficient to transfer SaPIs, which might be able to integrate into host genomes that are not accessible for phages. This ability of phages to transfer MGEs into target cells in which they cannot grow themselves, named 'silent transfer', was shown for *pac* phages, transferring *pac* SaPIs to bacterial hosts, where phages themselves were not able to parasitize[245,250].

In addition, SaPIs have other mechanisms to allow mobility by a non-inducer helper phage. These mechanisms are of a significant advantage as not all the phages resident in a bacterial host can induce a SaPI ERC cycle. Hence, conventional generalised transduction of SaPIs, although with a low-frequency

transduction rate, only requires the packaging of the SaPI genome into a phage particle. After injection into a new host cell, SaPIs can integrate into the chromosome at an identifiable att_C site using one of two pathways (Figure 1.13). First, SaPIs are mobilised in an Integrase (Int) and Excisionase (Xis) dependent, but RecA-independent manner. After the mispackaging by generalised transduction of a fragment containing an intact SaPI, the activity of Int and Xis allow the SaPIs to excise, circularise and integrate into the att_C site[245,248]. The other pathway involves homologous recombination of the flanking sequences of the SaPI by RecA. Although the rate of Int-dependent SaPI transfer is higher than generalised transduction, they still have to be induced and packaged by a helper phage to be moved in high frequency.

Additionally, SaPIs contribute to the generalised transduction of chromosomal DNA using the SaPI-encoded $TerS_{SP}$[260,261]. $TerS_{SP}$ binds to pseudo *pac* sequences scattered across the host chromosome and initiates the mispackaging of fragments of chromosomal DNA downstream of the signal sequences[261]. These pseudo-*pac* sites are usually located next to chromosomal genes encoding proteins involved in adaptation or virulence[261]. Furthermore, SaPI packaging module operon I, where $TerS_{SP}$ is located, is SOS-activated by derepression of the LexA promoter. After SOS activation of a host prophage, expression of $terS_{SP}$, as well as the expression of *pti* and *cpm* genes in this operon, will be promoted independent from SaPI induction or transduction. As a consequence, $TerS_{SP}$ will bind and cleave *pac* or pseudo-*pac* sequences scattered across the host DNA and packaged by the phage TerL leading to the formation of SaPI transducing particles[261]. SaPIs and in extension PICIs that encode TerS proteins could contribute to gene transfer and as a consequence modulate bacterial pathogens evolution.

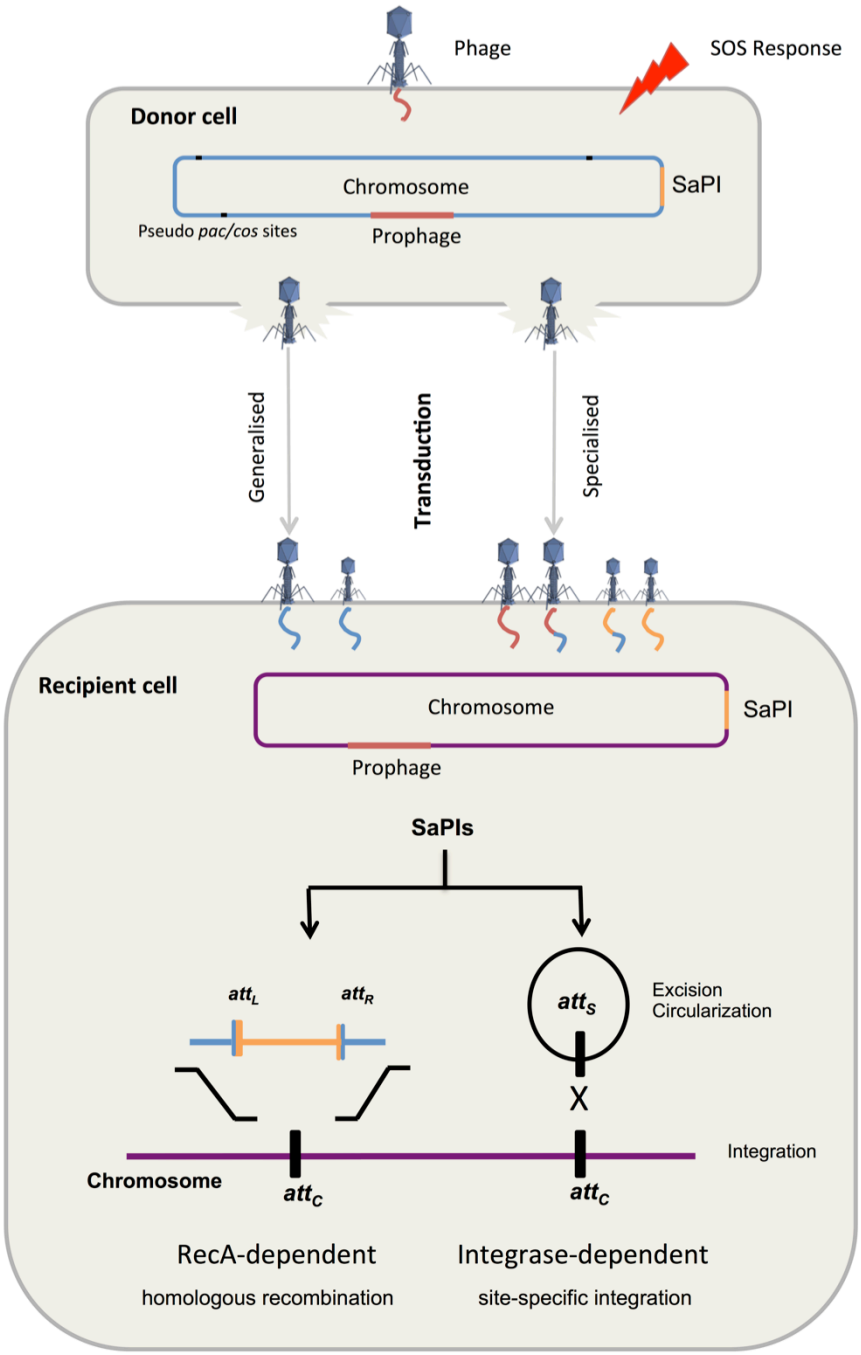


Figure 1.13 SaPIs general transfer pathways. After infection of a phage or SOS activation of a resident helper phage (red), dsDNA excise from the host genome (blue). Both phage and SaPI will replicate and finally will package its DNA (specialised transduction) or host DNA (generalised transduction). Cell lysis is induced, and the infectious particles will infect new recipient cells in which the new DNA recombines into the recipient host cell chromosome (purple).

1.3.5 Biology of non-SaPI PICI elements

Elements similar to SaPIs, designated as phage-inducible chromosomal islands (PICIs), were first identified in staphylococci other than *S. aureus*, and

recently described in different Gram-positive genera such as *Streptococcus pyogenes* (SpyCIM1), *Enterococcus faecalis* V583 (EfCIV583) and *Lactococcus lactis* (LlCIbIL312)[89,239,240]. This new cohesive PICI family has a similar gene cluster structure as the earlier described SaPIs. Thus, these elements have several features in common with the SaPIs; starting with an *int* gene, a CI-Cro–like divergent transcription regulatory system, the *pri-rep* and in some cases the presence of a *ter*S homolog. Nevertheless, the vast majority of the studies investigating PICIs are mainly based just on sequence analyses.

EfCIV583 was first identified as a PICI by Novick, Gail and Penadés[89], after it was initially erroneously regarded to be a prophage and named ϕV7 by Duerkop *et al.*[291]. In this study the interaction between "ϕV7" and an *E. faecalis* V583 ϕV1 prophage was shown. The authors suggested that the two phages were defective and that a composite phage, with both elements, was constituted (ϕV1/ϕV7). Although this conclusion was not true, the study showed that the interaction between these two elements was similar to the one described for SaPIs and their inductor helper phages. EfCIV583 was further investigated by Matos *et al.*[239]. The authors identified the element as a PICI and demonstrated that the phage ϕp1 was the helper phage, while refuting the existence of the proposed ϕV1/Vϕ7 intermediate. They further showed that EfCIV583 was also packaged in smaller capsids than ϕp1. However, they propose that the element was SOS-induced, stating that the helper phage was only required for packaging, but not for the induction of the PICI cycle[291]. These data contradict a recent study which shows that the EfCIV583 Stl repressor is not induced the SOS response and that phage ϕp1 is required for both packaging and induction of the EfCIV583 island[240]. This study conclusively confirms EfCIV583 as a novel PICI element.

The $SpyCI_{M1}$ element is another PICI which was first described in *S. pyogenes* M1 genome strain SF370[292]. The integration of the $SpyCI_{M1}$ element at its att_C site separates the hosts DNA mismatch repair genes *mut*S and *mut*L, providing a molecular switch mechanism for the host. Activation of *mut*L and the downstream genes is only achieved when the element is induced and excises from the host chromosome[292,293]. Excision of the $SpyCI_{M1}$ element during

exponential host cell division guarantees DNA replication fidelity. On the other hand, integration at the att_C site promotes the generation of mutated strains with more ability to adapt to different environmental conditions.

$SpyCI_{M1}$-like elements are also present in the *Streptococcus* genera[294]. Despite having the typical PICI structural characteristics, $SpyCI_{M1}$-like elements encode a repressor protein with a LexA- S24-like (cd06529) domain, which is inducible by the SOS response[292,293]. None of these elements encodes an identifiable terminase subunit[294] and the mechanism by which their dsDNA is package remains to be elucidated.

Recently, numerous elements were proposed to be part of the PICI family, such as the PICI-like element (PLE) in the Gram-negative bacterium *Vibrio cholerae*[295]. Although this island encodes an integrase gene, no homologues of genes of the regulatory, replication or packaging modules could be identified. Nevertheless, *V. cholerae* PLE has an uncharacterized mechanism by which it interferes with the phage reproduction. More studies have to be done to clarify if this PICI-like element can be included into the PICI family.

1.4 Aims

Extensive studies were accomplished to analyse and characterise the role of proteins involved in the packaging process of specific bacteriophages and PICIs. However, most proteins encoded in the *cos* phages and *cos* SaPIs morphogenesis modules have not been characterised yet. It was considered as one of the objectives of this thesis to comprehend which roles these selected proteins play in the packaging process of the dsDNA genomes of phages and SaPIs. The central purpose of this thesis was therefore to determine the essential molecular requirements to produce phage and PICIs virion particles.

Consequently, the overall aims of this thesis were:

- To unravel the biological function of phage-encoded HNH proteins in the life cycle of *cos* phages infecting Gram-positive and Gram-negative bacteria.

- To describe the mechanism by which SaPIbov5, which does not encode a $TerS_{SP}$ or CpmAB homolog, parasitizes its helper phages. Further, to verify whether other members of the recently identified PICI family could use this novel mechanism.

- To determine whether this new family of SaPIs, of which SaPIbov5 is the prototype, could be subject to an efficient horizontal gene transfer and whether SaPIbov5 can be mobilised by *cos* phages to other Gram-positive bacteria, such as non-*aureus* staphylococci and *Listeria monocytogenes*.

- To characterise the mechanism of SaPIbov5 packaging, and to investigate the function of the operon I-like genes, which have not been characterised to date and no homologs have yet been found. It was hypothesised that this selected set of genes could be responsible for the processes of SaPI-mediated phage interference and SaPI packaging.

- Finally, to characterise the mechanisms by which the PICI element EfCIV583 is packaged. EfCIV583 exploits its helper ϕp1 in order to be encapsidated in small phage particles. The goals of this study were to characterise the genes involved in the capsid size redirection and to identify the proteins involved in the EfCIV583-mediated mechanism of interference with ϕp1.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Unless indicated otherwise, all reagents used in this thesis were acquired through major suppliers including Fisher Scientific or Sigma-Aldrich.

2.1.2 Enzymes

Restriction endonucleases and buffers used for DNA manipulations were purchased from New England Biolabs (NEB). The DNA polymerase used for PCR reactions in this study was the high fidelity Pfu DNA Polymerase from Promega.

2.2 Bacteria strains and vectors

2.2.1 Bacterial strains

All bacterial strains used in this project are listed in Table 2.1. Cloning constructs were transformed and tested in *E. coli* DH5α. The obtained plasmids were introduced in different *S. aureus* or *E. faecalis* strains by transformation[296]. Protein expression was carried out in BL-21 (DE3) cells.

2.2.2 Storage and bacterial strains growth conditions

All bacteria strains were long term stored as glycerol stocks held at -80ºC. Single colonies were grown in the appropriate selective media supplemented with the correct antibiotic and stored in 1ml aliquots with 10% glycerol.

S. aureus strains were routinely grown at 32°C, 37°C or 43°C in tryptic soy agar (TSA) or tryptic soy broth (TSB) liquid medium (Oxoid) on an orbital shaker at 120 rpm. Media were supplemented with the appropriate antibiotics such as erythromycin (10 μg ml^{-1} or 2.5 μg ml^{-1}), tetracycline (3 μg ml^{-1}), chloramphenicol (20 μg ml^{-1}) or the X-gal product (20 mg ml^{-1}, Roche), according to the transfected plasmid.

E. faecalis strains were grown routinely at 32°C, 37°C or 43°C on brain heart infusion (BHI, Oxoid) agar plates or BHI liquid medium on an orbital shaker at 120 rpm. Media were supplemented with the appropriate antibiotics such as tetracycline (3 μg ml^{-1}), chloramphenicol (20 μg ml^{-1}) or the X-gal product (20 mg ml^{-1}, Roche), according to the transformed plasmid.

E. coli strains DH5α were grown routinely at 28°C, 37°C or 43°C on Luria-Bertani (LB, Sigma-Aldrich) agar plates or LB liquid medium on an orbital shaker at 120 rpm. Media were supplemented with antibiotic such as ampicillin (100 μg ml^{-1}), chloramphenicol (20 μg ml^{-1}) or kanamycin (100 μg ml^{-1}) to maintain the correspondent constructs in *E. coli*.

2.2.3 Plasmids vectors

All plasmids vectors used in this project are listed in Table 2.2. Plasmid construction was accomplished using standard techniques including polymerase chain reaction (PCR), PCR purification, restriction enzyme digestion, and gel extraction, as described under Section 2.3. In general, inserts were amplified through PCR and cloned into pre-digested plasmid vectors.

All expression vectors used for complementation in *S. aureus* were generated by cloning the target gene into the pCN51 plasmid, which has a cadmium-inducible promoter (P_{Cad})[297]. *E. coli* ϕP27 mutants were complemented by cloning each gene in the pBAD-18 plasmid, with an arabinose-inducible promoter (P_{BAD})[298]. The *bla*Z transcriptional fusions in *S. aureus* were generated using plasmid pCN42[297]. *In vitro* nuclease activity assays were performed using a pPROEX-HTa plasmid expressing the target proteins in *E. coli*. The pCU1[299] derivatives containing the SaPI or phage *cos* sites were used to analyse the functionality of these regions in the *cos* phage-mediated transfer.

All generated plasmids were verified, using primers listed in Appendix 1, by Sanger sequencing performed by Eurofins Genomics. Sequence data was confirmed by examination of the chromatogram using SnapGene, GSL Biotech LLC sequence-viewing program.

Table 2.1 Bacterial strains used in this study.

***Cos*-site packaging and phage-encoded HNH endonucleases**		
Strains	**Description**	**Reference**
RN4220	Restriction-defective derivative of RN450	[300]
RN451	RN450 lysogenic for ϕ11	[301]
RN10359	RN450 lysogenic for ϕ80α	[251]
JP10435	RN4220 lysogenic for ϕ12	This work
JP1794	RN451 SaPIbov1 *tst*::*tet*M	[254]
JP7085	RN451 SaPIbov5 *vwb*::*tet*M	This work
JP3603	RN10359 SaPIbov1 *tst*::*tet*M	[254]
JP7084	RN10359 SaPIbov5 *vwb*::*tet*M	This work
JP11041	JP10435 SaPIbov1 *tst*::*tet*M	This work
JP11010	JP10435 SaPIbov5 *vwb*::*tet*M	This work
JP11215	RN4220 SaPIbov5 *vwb*::*tet*M Δ*cos* site	This work
JP11229	JP10435 SaPIbov5 *vwb*::*tet*M Δ*cos* site	This work
JP11228	RN451 SaPIbov5 *vwb*::*tet*M Δ*cos* site	This work
JP3377	RN451 Δ*ter*S	[254]
JP3378	JP3377 SaPIbov1 *tst*::*tet*M	[254]
JP10764	JP3377 SaPIbov5 *vwb*::*tet*M	This work
JP11401	JP3377 SaPIbov5 *vwb*::*tet*M pJP1570	This work
JP10971	JP10435 Δ*hnh*	This work
JP11011	JP10971 SaPIbov5 *vwb*::*tet*M	This work
JP11406	JP11011 pJP1514	This work
LUG1170	SH1000 ϕSLT	[9]
JP11195	LUG1170 SaPIbov1 *tst*::*tet*M	This work
JP11194	LUG1170 SaPIbov5 *vwb*::*tet*M	This work
JP11230	LUG1170 SaPIbov5 *vwb*::*tet*M Δ*cos* site	This work
JP11402	LUG1170 ϕSLT Δ*hnh*	This work
JP11403	LUG1170 ϕSLT Δ*hnh* SaPIbov5 *vwb*::*tet*M	This work
JP11404	JP11402 pJP1077	This work
JP11405	JP11403 pJP1077	This work
JP10817	RN4220 pJP1523	This work
JP10818	RN4220 pJP1524	This work
JP6847	RN4220 pJP836	Nuria Quiles-Puchalt
JP6848	RN4220 pJP837	Nuria Quiles-Puchalt
JP5011	RN4220 lysogenic for ϕSLT *pvl*::*tet*M	[153]
JP8240	JP5011 ϕSLT *pvl*::*tet*M Δ*hnh*	Nuria Quiles-Puchalt
JP9104	JP5011 ϕSLT *pvl*::*tet*M Δ*ter*S	Nuria Quiles-Puchalt
JP9105	JP5011 ϕSLT *pvl*::*tet*M Δ*ter*L	Nuria Quiles-Puchalt

Strains	Description	Reference
JP9106	JP5011 ϕSLT *pvl*::*tet*M ΔORF40	Nuria Quiles-Puchalt
JP9107	JP5011 ϕSLT *pvl*::*tet*M ΔORF41	Nuria Quiles-Puchalt
JP9108	JP5011 ϕSLT *pvl*::*tet*M ΔORF42	Nuria Quiles-Puchalt
JP9951	JP5011 ϕSLT *pvl*::*tet*M ΔORF47	This work
JP9666	JP5011 ϕSLT *pvl*::*tet*M HNH H57A	This work
JP9667	JP5011 ϕSLT *pvl*::*tet*M HNH H58A	This work
JP10934	JP5011 ϕSLT *pvl*::*tet*M TerL E208A	This work
JP10976	JP5011 ϕSLT *pvl*::*tet*M TerL D363A	Nuria Quiles-Puchalt
JP11367	JP5011 pCN51	This work
JP9059	JP8240 pJP1077	Nuria Quiles-Puchalt
JP9119	JP9104 pJP1245	Nuria Quiles-Puchalt
JP9120	JP9105 pJP1246	Nuria Quiles-Puchalt
JP9121	JP9106 pJP1247	Nuria Quiles-Puchalt
JP9122	JP9107 pJP1248	Nuria Quiles-Puchalt
JP9123	JP9108 pJP1249	Nuria Quiles-Puchalt
JP10024	JP9951 pJP1520	This work
JP11370	JP8240 pJP1537	This work
JP11371	JP8240 pJP1538	This work
JP9959	JP9666 pJP1077	This work
JP9960	JP9667 pJP1077	This work
JP10903	JP9105 pJP1521	Nuria Quiles-Puchalt
JP10904	JP9105 pJP1522	Nuria Quiles-Puchalt
JP11307	JP10934 pJP1246	Nuria Quiles-Puchalt
JP11308	JP10976 pJP1246	Nuria Quiles-Puchalt
JP11074	JP10971 pJP1514	This work
JP11204	RN451 pCU1	This work
JP11198	RN451 pJP1525	This work
JP11199	RN451 pJP1526	This work
JP11196	RN451 pJP1527	This work
JP11197	RN451 pJP1528	This work
JP11200	RN451 pJP1529	This work
JP11201	RN451 pJP1530	This work
JP11202	RN451 pJP1531	This work
JP11203	RN451 pJP1532	This work
JP10974	JP10435 pCU1	This work
JP10907	JP10435 pJP1525	This work
JP10879	JP10435 pJP1526	This work
JP10968	JP10435 pJP1527	This work
JP10878	JP10435 pJP1528	This work
JP10875	JP10435 pJP1529	This work

Strains	Description	Reference
JP10876	JP10435 pJP1530	This work
JP10908	JP10435 pJP1531	This work
JP10880	JP10435 pJP1532	This work
JP9818	STEC strain 2771/97 ϕP27	[66]
JP10045	STEC strain 2771/97 ϕP27 *stx*::*tet*A	This work
JP10363	*E. coli* strain MG1655	Lab strain
JP10819	JP10363 ϕP27 *stx*::*tet*A	This work
JP10960	JP10819 ϕP27 ΔORF34	This work
JP10961	JP10819 ϕP27 ΔORF35	This work
JP10962	JP10819 ϕP27 ΔORF36	This work
JP10963	JP10819 ϕP27 ΔORF38	This work
JP10964	JP10819 ϕP27 ΔORF39	This work
JP10965	JP10819 ϕP27 ΔORF40	This work
JP10967	JP10819 ϕP27 ΔORF47	This work
JP11327	JP10819 pBAD18	This work
JP11328	JP10960 pJP1539	This work
JP11329	JP10961 pJP1540	This work
JP11330	JP10962 pJP1541	This work
JP11331	JP10963 pJP1542	This work
JP11332	JP10964 pJP1543	This work
JP11333	JP10965 pJP1544	This work
JP11334	JP10967 pJP1545	This work
JP10025	BL21 (DE3) pJP1533	This work
JP10566	BL21 (DE3) pJP1534	Nuria Quiles-Puchalt
JP10567	BL21 (DE3) pJP1535	Nuria Quiles-Puchalt
JP10317	BL21 (DE3) pJP1253	This work

Intra- and inter-generic SaPI transfer by *cos* phages

Strains	Description	Reference
RN4220	Restriction-defective derivative of RN450	[300]
JP10435	RN4220 lysogenic for ϕ12	This work
JP11010	JP10435 SaPIbov5 *tet*M	This work
JP11229	JP10435 SaPIbov5 *tet*M Δ*cos* site	This work
JP7422	*L. monocytogenes* SK1351	[250]
JP7432	*L. monocytogenes* EGDe	[250]
JP1220	*S. xylosus* C2a	[245]
JP4226	*S. aureus*	[238]
JP11545	JP829 (SaPIbov5 *tst*::*tet*M)	This work
JP11553	JP830 (SaPIbov5 *tst*::*tet*M)	This work

Intra- and inter-generic SaPI transfer by *cos* phages		
Strains	**Description**	**Reference**
JP11543	JP7422 (SaPIbov5 *tst*::*tet*M)	This work
JP11544	JP7432 (SaPIbov5 *tst*::*tet*M)	This work
JP11542	JP1220 (SaPIbov5 *tst*::*tet*M)	This work
JP11554	JP4226 (SaPIbov5 *tst*::*tet*M)	This work
JP10812	*E. coli* DH5α (pJP1511)	This work

SaPIbov5-mediated phage interference via small capsid production		
Strains	**Description**	**Reference**
RN4220	Restriction-defective derivative of RN450	[300]
RN451	RN450 lysogenic for ϕ11	[301]
RN10359	RN450 lysogenic for ϕ80α	[241]
JP10435	RN4220 lysogenic for ϕ12	This work
JP12520	RN4220 lysogenic for ϕ12 evolved 1	This work
JP12521	RN4220 lysogenic for ϕ12 evolved 2	This work
JP12522	RN4220 lysogenic for ϕ12 evolved 3	This work
JP12523	RN4220 lysogenic for ϕ12 evolved 4	This work
JP5011	RN4220 lysogenic for ϕSLT *pvl*::*tet*M	[153]
JP9108	JP5011 ΔORF42 ϕSLT *pvl*::*tet*M	This work
LUG1170	SH1000 ϕSLT	[9]
JP11194	LUG1170 SaPIbov5 *vwb*::*tet*M original	This work
JP6894	RN4220 SaPIbov5 *vwb*::*tet*M original	This work
JP11010	JP10435 SaPIbov5 *vwb*::*tetM* original	This work
JP12942	JP12523 SaPIbov5 *vwb*::*tet*M original	This work
JP12218	RN4220 SaPIbov5 *vwb*::*tet*M evolved	This work
JP12684	JP10435 SaPIbov5 *vwb*::*tet*M evolved	This work
JP12943	JP12523 SaPIbov5 *vwb*::*tet*M evolved	This work
JP11425	LUG1170 SaPIbov5 *vwb*::*tet*M evolved	This work
JP11634	RN4220 SaPIbov5::*erm*C	This work
JP13161	RN4220 SaPIbov5::*erm*C ΔORF8	This work
JP12648	RN4220 SaPIbov5::*erm*C ΔORF9	This work
JP13048	RN4220 SaPIbov5::*erm*C ΔORF10	This work
JP13087	RN4220 SaPIbov5::*erm*C ΔORF11	This work
JP12618	RN4220 SaPIbov5::*erm*C ΔORF12	This work
JP13192	RN4220 SaPIbov5::*erm*C ΔORF8-ORF9-ORF10-ORF11-ORF12	This work
JP12419	JP10435 SaPIbov5 *vwb*::*erm*C	This work
JP13171	JP10435 SaPIbov5 *vwb*::*erm*C ΔORF8	This work
JP12688	JP10435 SaPIbov5 *vwb*::*erm*C ΔORF9	This work
JP13162	JP10435 SaPIbov5 *vwb*::*erm*C ΔORF10	This work

Strains	Description	Reference
JP13110	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF11	This work
JP12636	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF12	This work
JP13092	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF8-ORF9	This work
JP13304	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF9-ORF10	This work
JP13205	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF8-ORF9-ORF10	This work
JP13112	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF10-ORF12	This work
JP13168	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF11-ORF12	This work
JP13298	JP10435 SaPIbov5 *vwb*::*ermC* ΔORF8-ORF9-ORF10-ORF11-ORF12	This work
JP13443	JP13298 pJP1769	This work
JP12941	JP12523 SaPIbov5 *vwb*::*ermC*	This work
JP13172	JP12523 SaPIbov5 *vwb*::*ermC* ΔORF8	This work
JP12796	JP12523 SaPIbov5 *vwb*::*ermC* ΔORF9	This work
JP13163	JP12523 SaPIbov5 *vwb*::*ermC* ΔORF10	This work
JP13111	JP12523 SaPIbov5 *vwb*::*ermC* ΔORF11	This work
JP12795	JP12523 SaPIbov5 *vwb*::*ermC* ΔORF12	This work
JP13299	JP12523 SaPIbov5 *vwb*::*ermC* ΔORF8-ORF9-ORF10-ORF11-ORF12	This work
JP12685	LUG1170 SaPIbov5 *vwb*::*ermC*	This work
JP11905	JP5011 SaPIbov5 *vwb*::*ermC*	This work
JP11906	JP5011 ΔORF42 ϕSLT SaPIbov5 *vwb*::*ermC*	This work
JP12694	RN4220 SaPIbov5::*ermC* Small	This work
JP12686	LUG1170 SaPIbov5::*ermC* Small	This work
JP12736	JP10435 SaPIbov5::*ermC* Small	This work
JP12867	JP12523 SaPIbov5::*ermC* Small	This work
JP12380	RN4220 pCN51	This work
JP12945	RN4220 pJP1749	This work
JP12800	RN4220 pJP1748	This work
JP12968	RN4220 pJP1747	This work
JP11423	RN4220 pJP1730	This work
JP12647	RN4220 pJP1745	This work
JP12405	RN4220 pJP1744	This work
JP12936	RN4220 pJP1750	This work
JP11367	JP5011 pCN51	This work
JP11559	JP5011 pJP1730	This work
JP12293	RN10359 pCN51	This work
JP11556	RN10359 pJP1730	This work
JP12343	RN451 pCN51	This work
JP11557	RN451 pJP1730	This work
JP12294	JP10435 pCN51	This work
JP13006	JP10435 pJP1749	This work

Strains	Description	Reference
JP13007	JP10435 pJP1748	This work
JP13008	JP10435 pJP1747	This work
JP11558	JP10435 pJP1730	This work
JP13009	JP10435 pJP1745	This work
JP13010	JP10435 pJP1744	This work
JP13396	JP10435 pJP1750	This work
JP13012	JP12523 pCN51	This work
JP13013	JP12523 pJP1749	This work
JP13014	JP12523 pJP1748	This work
JP13015	JP12523 pJP1747	This work
JP13016	JP12523 pJP1730	This work
JP13017	JP12523 pJP1745	This work
JP13018	JP12523 pJP1744	This work
JP13397	JP12523 pJP1750	This work

Characterisation of the putative morphogenetic cluster in EfCIV583

Strains	Description	Reference
JP10983	VE18590 (pp-)	[239]
JP11027	JP10983 EfCIV583::*tet*M (pp7+)	This work
JP13121	JP10983 EfCIV583::*tet*M ΔEF2947	This work
JP13122	JP10983 EfCIV583::*tet*M ΔHPI	This work
JP13123	JP10983 EfCIV583::*tet*M ΔEF2946	This work
JP14450	JP10983 EfCIV583::*tet*M ΔEF2945	This work
JP14449	JP10983 EfCIV583::*tet*M ΔEF2944	This work
JP13124	JP10983 EfCIV583::*tet*M ΔEF2943	This work
JP13125	JP10983 EfCIV583::*tet*M ΔEF2942	This work
JP14459	JP10983 EfCIV583::*tet*M ΔEF2941	This work
JP13126	JP10983 EfCIV583::*tet*M ΔEF2940	This work
JP13127	JP10983 EfCIV583::*tet*M ΔEF2939	This work
JP13128	JP10983 EfCIV583::*tet*M ΔEF2938	This work
JP13129	JP10983 EfCIV583::*tet*M ΔEF2937	This work
JP14460	JP10983 EfCIV583::*tet*M ΔEF2936	This work
JP13130	JP10983::*tet*M ΔI574	This work
JP10984	VE18562 lysogenic for ϕp1	[239]
JP11028	JP10984 EfCIV583::*tet*M	This work
JP12587	JP10984 EfCIV583::*tet*M ΔEF2947	This work
JP12590	JP10984 EfCIV583::*tet*M ΔHPI	This work
JP12588	JP10984 EfCIV583::*tet*M ΔEF2946	This work
JP12589	JP10984 EfCIV583::*tet*M ΔEF2945	This work
JP13798	JP10984 EfCIV583::*tet*M ΔEF2944	This work

Strains	Description	Reference
JP12591	JP10984 EfCIV583::*tet*M ΔEF2943	This work
JP12592	JP10984 EfCIV583::*tet*M ΔEF2942	This work
JP13308	JP10984 EfCIV583::*tet*M ΔEF2941	This work
JP12630	JP10984 EfCIV583::*tet*M ΔEF2940	This work
JP12690	JP10984 EfCIV583::*tet*M ΔEF2939	This work
JP12735	JP10984 EfCIV583::*tet*M ΔEF2938	This work
JP12631	JP10984 EfCIV583::*tet*M ΔEF2937	This work
JP13307	JP10984 EfCIV583::*tet*M ΔEF2936	This work
JP12693	JP10984 EfCIV583::*tet*M ΔI574	This work
JP11867	JP10984 ΔEF3032	This work
JP11737	JP10984 ΔEF3033	This work
JP11878	JP11867 EfCIV583::*tet*M	This work
JP12475	JP11737 EfCIV583::*tet*M	This work
JP12426	JP10983 pJP1775	This work
JP12427	JP10983 pJP1774	This work
JP12411	JP11867 pJP1775	This work
JP12412	JP11737 pJP1774	This work
JP11457	JP10984 pJP1708	This work
JP12428	JP10984 pJP1775	This work
JP12429	JP10984 pJP1774	This work
JP12416	JP12428 EfCIV583::*tet*M	This work
JP12490	JP12429 EfCIV583::*tet*M	This work

Table 2.2 Plasmids used in this study.

Cos-site packaging and phage-encoded HNH endonucleases		
Plasmid	**Description**	**Reference**
pMAD	Vector for efficient allelic replacement	[302]
pJP1510	pMAD derivative, deletion of ϕ12 *hnh*	This work
pJP1076	pMAD derivative, deletion of ϕSLT ORF 37	This work
pJP1240	pMAD derivative, deletion of ϕSLT ORF 38	This work
pJP1241	pMAD derivative, deletion of ϕSLT ORF 39	This work
pJP1242	pMAD derivative, deletion of ϕSLT ORF 40	This work
pJP1243	pMAD derivative, deletion of ϕSLT ORF 41	This work
pJP1244	pMAD derivative, deletion of ϕSLT ORF 42	This work
pJP1509	pMAD derivative, deletion of ϕSLT ORF 47	This work
pJP1254	pMAD derivative, mutation H57A in ϕSLT ORF 37	This work
pJP1255	pMAD derivative, mutation H58A in ϕSLT ORF37	This work
pJP1512	pMAD derivative, mutation E208A in ϕSLT ORF 39	This work
pJP1513	pMAD derivative, mutation D363A in ϕSLT ORF39	This work

Plasmid	Description	Reference
pJP730	pMAD derivative, insertion of the *tet*M cassette in the *vwb* gene of SaPIbov5	[238]
pJP1557	pMAD derivate, deletion *cos* site SaPIbov5	This work
pCN51	Expression vector	[297]
pJP1514	pCN51-*hnh* ϕ12	This work
pJP1077	pCN51-*hnh* ϕSLT	This work
pJP1245	pCN51-*ter*S ϕSLT	This work
pJP1246	pCN51-*ter*L ϕSLT	This work
pJP1247	pCN51-p40 ϕSLT	This work
pJP1248	pCN51-p41 ϕSLT	This work
pJP1249	pCN51-p42 ϕSLT	This work
pJP1520	pCN51-p47 ϕSLT	This work
pJP1570	pCN51-*ter*S ϕ11	This work
pCN42	Used in transcriptional fusions to the staphylococcal β-lactamase *bla*Z. Contains the *Pcad* promoter	[297]
pJP1523	Transcriptional analysis of ϕ12 *hnh* in the presence of RinA, pCN42 derivate	This work
pJP1524	Transcriptional analysis of ϕ12 *hnh* in the absence of RinA, pCN42 derivate	This work
pJP836	Transcriptional analysis of ϕSLT *hnh* in the presence of RinA, pCN42 derivate	This work
pJP837	Transcriptional analysis of ϕSLT *hnh* in the absence of RinA, pCN42 derivate	This work
pKD46	Plasmid with Red system of lambda phage.	[303]
pCP20	Plasmid for cassette replacement using FRTs in *E. coli*.	[303]
pBAD18	Expression vector	[298]
pJP1539	pBAD18-*hnh* ϕ27	This work
pJP1540	pBAD18-*ter*S ϕ27	This work
pJP1541	pBAD18- *ter*L ϕ27	This work
pJP1542	pBAD18-p38 ϕ27	This work
pJP1543	pBAD18-p39 ϕ27	This work
pJP1544	pBAD18-p40 ϕ27	This work
pJP1545	pBAD18-p47 ϕ27	This work
pCU1	Cm^r. Cloning vector	[299]
pJP1525	pCU1 *cos* site ϕSLT	This work
pJP1526	pCU1 Δ*cos* site ϕSLT	This work
pJP1527	pCU1 *cos* site ϕ12	This work
pJP1528	pCU1 Δ*cos* site ϕ12	This work
pJP1529	pCU1 *cos* site ϕSaPIbov5	This work
pJP1530	pCU1 Δ*cos* site ϕSaPIbov5	This work
pJP1531	pCU1 *cos* site ϕP27	This work
pJP1532	pCU1 Δ*cos* site ϕP27	This work

Plasmid	Description	Reference
pJP1533	Expression in *E. coli* of His-HNH ϕSLT, pPROEX HTa derivative	This work
pJP1534	Expression in *E. coli* of His-HNH H57A mutant ϕSLT, pPROEX HTa derivative	This work
pJP1535	Expression in *E. coli* of His-HNH H58A mutant ϕSLT, pPROEX HTa derivative	This work
pJP1253	Expression in *E. coli* of His-TerL ϕSLT, pPROEX HTa derivative	This work

Intra- and inter-generic SaPI transfer by *cos* phages		
Plasmid	**Description**	**Reference**
pJP1511	pMAD derivative, deletion of ϕS12 ORF 29	This work

SaPIbov5-mediated phage interference via small capsid production		
Plasmid	**Description**	**Reference**
pBT_2-βgal	Vector for efficient allelic replacement	[304]
pJP1712	pBT_2-βgal derivative, insertion of the *erm*C cassette in SaPIbov5^{adjusted}	This work
pJP1761	pBT_2-βgal derivative, insertion of the *erm*C cassette in SaPIbov5^{small}	This work
pJP1757	pBT_2-βgal derivative, ochre mutation in SaPIbov5 ORF8	This work
pJP1758	pBT_2-βgal derivative, ochre mutation in SaPIbov5 ORF9	This work
pJP1759	pBT_2-βgal derivative, ochre mutation in SaPIbov5 ORF10	This work
pJP1717	pBT_2-βgal derivative, ochre mutation in SaPIbov5 ORF11	This work
pJP1760	pBT_2-βgal derivative, ochre mutation in SaPIbov5 ORF12	This work
pJP1749	pCN51-*orf8* SaPIbov5	This work
pJP1748	pCN51-*orf9* SaPIbov5	This work
pJP1747	pCN51-*orf10*SaPIbov5	This work
pJP1730	pCN51- *ccm* SaPIbov5	This work
pJP1745	pCN51-*orf12* SaPIbov5	This work
pJP1744	pCN51-orf11-*orf12* SaPIbov5	This work
pJP1750	pCN51-*ppi*-orf9-orf10-*ccm*-*orf12* SaPIbov5	This work
pCN51-*cat194*	Derivative of pCN51 with chloramphenicol cassette. Expression vector	This work
pJP1769	pCN51-*cat194*-*orf11* SaPIbov5	This work

Characterisation of the putative morphogenetic cluster in EfCIV583		
Plasmid	**Description**	**Reference**
pBT_2-βgal	Vector for efficient allelic replacement	[304]
pJP1716	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2947	This work
pJP1731	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M HPI	This work

Plasmid	Description	Reference
pJP1718	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2946	This work
pJP1733	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2945	This work
pJP1734	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2944	This work
pJP1735	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2943	This work
pJP1736	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2942	This work
pJP1737	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2941	This work
pJP1738	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2940	This work
pJP1739	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2939	This work
pJP1740	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2938	This work
pJP1741	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2937	This work
pJP1742	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M EF2936	This work
pJP1743	pBT_2-βgal derivative, deletion of EfCIV583::*tet*M I574	This work
pJP1710	pBT_2-βgal derivative, deletion of ϕp1 EF3032	This work
pJP1711	pBT_2-βgal derivative, deletion of ϕp1 EF3033	This work
pJP1708	Derivative of pCU1 (P_{Cad}) with chloramphenicol cassette. Expression vector	This work
pJP1775	pCU1 (P_{Cad}) ϕp1 EF3032	This work
pJP1774	pCU1 (P_{Cad}) ϕp1 EF3033	This work

2.3 General molecular biology techniques

Routine DNA manipulations were performed using standard procedures[305,306].

2.3.1 Chromosomal DNA extraction

Chromosomal DNA from *S. aureus, E. faecalis* and *E. coli* strains was extracted using the 'GenElute Bacterial Genomic DNA' kit (Sigma-Aldrich), following the manufacturer's instructions, except for the lysis of *S. aureus* and *E. faecalis* bacterial strains, which prior to the extraction were incubated with lysostaphin (12.5 μg ml^{-1}, Sigma-Aldrich) or lysozyme (12.5 μg ml^{-1}, Sigma-Aldrich) respectively, at 37°C for 1 hour (h).

2.3.2 Plasmid DNA extraction

Plasmid DNA was extracted from strains of *E. coli* using the QIAprep®Spin Miniprep Kit (Qiagen) according to the manufacturer's instructions. Plasmid DNA of *S. aureus* and *E. faecalis* was extracted using the same protocol,

except the cells were lysed with lysostaphin (*S. aureus*, 12.5 µg ml^{-1}) or lysozyme (*E. faecalis*, 10 µg ml^{-1}), at 37°C for 1 h.

2.3.1 Polymerase Chain Reaction

PCR reactions were performed using an Eppendorf® Mastercycler Personal machine. Primers listed in Appendix 1 were obtained from Thermo Fisher. The lyophilized primers were reconstituted with HPLC-grade water to a final concentration of 200 µM. Primers working stock aliquots were prepared by 1:100 dilutions with HPLC-grade water to 10 µM.

PCR products to be used for cloning were obtained using Pfu DNA Polymerase (Promega) in the following manner: 1x Pfu DNA Polymerase 10X Buffer with $MgSO_4$, 200 µM dNTPs (Invitrogen), 5-50 ng vector DNA input template, 0.2 µM of both the forward and reverse primers, and 1.25 Units (U) of Pfu DNA polymerase, and brought to a final reaction volume of 50 µl. All PCR reaction master mixes were scaled up as needed with 10% excess to account for pipetting error. The following PCR parameters were followed for inserts <10 kb: initial denaturation for 1 minutes (min) at 95°C, followed by 35 cycles of denaturation for 1 min at 95°C, primer annealing for 30 seconds, and primer extension for 2 min per kb of target DNA at 72°C. A final 10 min primer extension step was performed at 72°C. All reactions were kept at 4°C following the final extension before being stored at -20°C. Annealing temperatures for each primer were optimised based on the melting temperature (Tm) of each primer and were the average of Tm-5°C.

Where PCR products were used to detect generated mutants, single colonies were selected as DNA template as follows: 1 isogenic colony was picked and smeared in an even layer, transferred into a sterile PCR tube and thoroughly vortexed with 40 µl of HPLC grade water.

2.3.2 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate genomic and plasmid DNA. Different percentages of agarose (0.6%, 0.7% and 1%; Invitrogen) were used to

validate PCR reactions and to verify cloning steps by analytic restriction digests. Gels were prepared with 1X Tris-acetate-EDTA (TAE) buffer and SybrSafe® (0.1 µg/ml; Thermo Fisher).

The appropriate volume of 5X loading dye (50% v/v glycerol, 50% v/v 1X TAE, bromophenol blue added to the desired colour) was added to DNA samples prior loading and gels were run at 120 volts (V) for approximately 40 min, or until ladder bands were easily visualised. The 1 kb^+ size marker (Invitrogen) was used to determine the size of sample bands. DNA was visualised using an uviPRO gold Uvitec transilluminator (Uvitec).

Gel images were edited using Adobe Photoshop in a standardised manner. All gels were converted to grey scale without any additional manipulation.

2.3.3 Restriction digestion

Restriction endonucleases and restriction enzyme buffers used for DNA manipulations were purchased from NEB and were used according to manufacturer's instructions. Restriction digests were performed in a volume of 40 µl, or multiples thereof, depending on the following application of the DNA product. Samples for analytical and preparative restriction digest were prepared in the following manner: 1-3 µg of plasmid DNA was digested with the appropriate restriction buffer and 10 units of the desired restriction enzyme(s). The mixture was incubated for 1-2 h at 37°C. When digested DNA was used for cloning, the mixture was purified after incubation.

2.3.4 Gel extraction and PCR purification

DNA fragments intended for cloning were purified before ligation. To obtain DNA fragments generated by preparative plasmid digest, the corresponded bands were cut out from the agarose gel and gel extraction was performed using the QIAquick® PCR Extraction Kit (Qiagen) according to the manufacturer's instructions. PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen) directly from the PCR reaction mix. All DNA

concentrations were determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2.3.5 Ligation reactions

Ligation reactions contented 50-150 ng of digested plasmid backbone, with the insert added at a 1:3 molar ratio (plasmid: insert), 10x DNA ligase buffer and 5 U of T4 DNA ligase (NEB) in a total volume of 20 µl. The reactions were incubated at 22 °C for 1 h, and the enzyme then inactivated for 10 min at 65°C.

2.3.6 Precipitation of Genomic DNA

For DNA precipitation, a ratio of 1/10 of sodium acetate 3M, pH=5.2 (NaAc: DNA) was added to the subject DNA. Subsequently, two volumes of cold 100% ethanol were added to the mix, inverted several times and incubated at -20°C for 2 h or -80°C for 30 min. Precipitated DNA was pelleted by centrifugation at 14,000 rpm, 20 min at 4°C, the supernatant was discarded, and pellets were washed with 1 ml of cold 70% ethanol to remove the excess of salts. Samples were then centrifuged at 14,000 rpm, 5 min at 4°C, supernatant discarded and dried for 10 min in an Eppendorf Concentrator 5301 for 15 min at 45°C, being re-suspended in 20 µl of HPLC-grade water.

2.3.7 Overexpression of pProOEX-HTa vectors

Overexpression of the different proteins was performed as described by *Chien et al. 1999*[307] using pProOEX-HTa vectors. To perform *in vivo* endonuclease assays, constructed vectors with the cloned proteins were induced using 1µl of 0.4 M isopropyl-β-D-thiogalactoside (IPTG, Invitrogen) and incubated at 37°C at 120 rpm for approximately 3-4 h. Plasmid DNA was extracted and digested with *Bam*HI, which cuts the plasmid once and samples were run on an agarose gel to test for degradation of plasmid DNA.

2.3.8 Southern blot

For Southern blot hybridization, chromosomal DNA was purified, digested if necessary, and separated by agarose gel electrophoresis. DNA was transferred from the gels to nylon membranes (0.45 mm hybond-N pore diameter, Amersham Life Science) using standard methods[305,308]. The labelled probes were obtained using primers shown in Appendix 1 and hybridized according to the protocol provided by the DNA Labelling Kit and PCR DIG-chemiluminescence detection (Roche).

2.3.8.1 DNA extraction for replication studies

In order to identify the three consecutive and definable stages (excision-induction, replication and packaging and transfer; ERP) of the virus life cycle, samples of induced (as described under Section 2.6.1) bacteria culture were taken at different time points and subsequently centrifuged to pellet the cells. The precipitates were re-suspended in TES-sucrose (10 mM Tris, 1 mM EDTA, 100 mM NaCl, 0.5 M sucrose at pH 8) and lysed with lysostaphin (5 µg ml^{-1}) in the presence of pancreatic RNase (12.5 µg ml^{-1}) for 30 min. at 37°C. The obtained lysate was incubated with proteinase K (12.5 µg ml^{-1}) and SDS (1% v/v) for 30 min. at 55°C. Subsequently, loading buffer was added and samples were vortexed for at least 20 min. The samples were frozen in an ethanol dry ice bath and subsequently thawed in a water bath at 65°C. This procedure was repeated three times. Finally, the samples were loaded on a 0.7% agarose gel and allowed to run overnight at 20 V to improve separation of genomic DNA of phage and SaPI induction. Lastly, DNA was transferred from the gel to a nylon membrane to perform Southern blot protocol (Section 2.3.8).

2.4 Bacterial transformation

2.4.1 Preparation and transformation of chemically competent *E. coli* cells

In order to propagate newly generated constructs, plasmids were transformed into competent *E. coli* DH5α cells. To prepare chemically competent cells, *E. coli* cells were grown overnight in LB liquid medium at 37°C and 120 rpm. 100 ml of fresh LB were inoculated with 1 ml of the overnight culture. The culture was incubated at 37°C and 120 rpm for approximately 1.5 h until OD_{600} = 0.350 was reached. The flask was immediately transferred to ice for 10 min to stop bacterial growth. The culture was split between two 50 ml conical tubes, kept on ice at all times, and both were pelleted at 4°C at 4000 rpm for 10 min. The supernatant was decanted and cell pellets re-suspended in 15 ml ice-cold, sterile 0.1 mM $CaCl_2$ solution before being re-pelleted at 4000 rpm, 4°C for 15 min. The pellets were each re-suspended in 2 ml of 0.1 mM $CaCl_2$ solution and 200 µl frozen at -80°C.

For transformation, aliquots of competent cells were thawed on ice before adding 10 µl of the ligation mix. The reaction mixture was incubated on ice for 15 min before being heat-shocked in a 42°C water-bath for 90 seconds and then immediately placed on ice. After 2 min, 800 µl of LB liquid medium was added and incubated for 1 h at 37°C on an orbital shaker set at 120 rpm. Subsequently, 100 µl of the transformation mixture were spread onto LB plates supplemented with 100 µg/ml ampicillin or kanamycin. The remaining volume of the transformation was centrifuged at 4000 rpm for 1 min. The supernatant was decanted and the pellet resuspended in the residual media by vortexing. This concentrate was also plated on the appropriate agar plates. All plates were incubated at 37°C for 24 h.

2.4.2 Preparation and transformation of electrocompetent *S. aureus* and *E. faecalis* cells

To prepare electrocompetent cells, preculture of *S. aureus* were grown in TSB and preculture of *E. faecalis* in BHI, at 37°C and 120 rpm. 100 ml of fresh TSB or BHI, respectively, were inoculated with 1 ml of the interest overnight culture strain. The cultures were shaken at 37°C at 120 rpm for approximately 1.5 h until OD_{540} = 0.2 was reached. The flask was immediately transferred to ice for 15 min to stop bacterial growth. The culture was split between two 50 ml conical tubes, kept on ice at all times, and both were pelleted at 4°C at 4000 rpm for 10 min. The supernatant was decanted and cell pellets washed twice by re-suspending them in 15 ml ice-cold, sterile 0.5 mM sucrose solution and centrifugation at 4000 rpm, 4°C for 15 min. Each cell pellet was then re-suspended in 1 ml ice-cold 0.5 mM sucrose solution and incubated at 4°C for 20 min. The cells were pelleted by centrifugation at 4°C for 10 min at 4000 rpm. The supernatant was decanted and the pellets re-suspended in 400 µl of 0.5 mM sucrose solution.

Plasmids obtained from *E. coli* were introduced by transformation into staphylococci and enterococci as previously described by Cucarella *et al.*, 2001[296]. *S. aureus* electrocompetent cells, RN4220 derivatives, and *E. faecalis* cells, VE18590 (curated of MGEs; pp-) or VE 18316 (lysogenic for ϕp1 and EfCIV583) derivatives, were transformed with the correspondent plasmids by electroporation.

For electroporation, aliquots of competent cells were thawed on ice before adding the plasmid DNA. 50 µl aliquots of electrocompetent cells and 6 µl (0.3- 0.6 µg) plasmid DNA were added to a 0.1 cm Gene Pulser® cuvette. A MicroPulserTM electroporation apparatus on STA setting (2.5 kV, 25 µF, 100 Ω) was used for electroporation. Immediately after the pulse, 1 ml of TSB or BHI broth, respectively, was added directly to the bacteria in the cuvette and then transferred to a 1.5 ml microcentrifuge tube. Cells were incubated at 32°C, 120 rpm for 1 h, before being plated on TSA or BHI plates supplemented with appropriate antibiotics and/or X-gal, as needed. Generally, 100 µl and

the cell pellet concentrate (pelleted at 4000 rpm for 1 min, the supernatant decanted, pellet re-suspended in residual supernatant) was spread and incubated for 24 h at 32°C.

2.5 Biochemical methods

2.5.1 Enzyme assay for the quantification of the β-lactamase activity in transcriptional fusion plasmids

Overnight cultures of strains carrying pCN42 derivative plasmids were diluted 1:50 and grown to a density of OD_{540} = 0.3-0.4. One 100 µl aliquot (T_0) was collected and then frozen for subsequent analysis. Following, depending on the experimental setup, Mitomycin C (2 µg ml^{-1}; MC, Sigma-Aldrich) was added and incubated under slow agitation at 32°C on an orbital shaker at 80 rpm. Aliquots of 50 µl were taken at 90 and 180 min after induction, and sodium azide (5 mM) added. Samples were transferred to a flat bottom polystyrene 96 plate to measure optical density on a THERMOmax apparatus (Molecular Devices) at a wavelength of OD_{650} and OD_{490}. Finally, 50 µl Nitrocefin solution were added to each well (132 mg/ml in a 0.1M sodium phosphate solution, pH 5.8) and the plate back in the THERMOmax plate reader performing measurements in 20-second intervals for 20 min. The following formulas were used to calculate the units of ß-lactamase activity:

Unit ß-lactamase activity = $[(slope)(V_d)/(\varepsilon_m)\ (l)(s)]$

An increase in absorbance of 0.001 OD_{490} was defined as 1 unit of ß-lactamase activity.

2.6 Methodology used in the induction studies of PICIs and phages

2.6.1 Phage propagation: Induction

Lysogenic strains were induced using MC to generate bacteriophage lysates. That technique takes advantage of the SOS response pathway that happens in direct response to DNA damage and triggers the lytic cycle of the phages. To induce the *S. aureus* cells, a 1/50 dilution was made in TSB from a stationary culture and grown until it reached a density of OD_{540} = 0.2. Then MC (2 μg ml^{-1}) was added causing cellular damage and activating the SOS response. Subsequently, the induced cultures were grown on an orbital shaker set up to 80 rpm at 32°C. The lysis of the cells was usually achieved 4 h after induction for a 5 ml of lysate. In Sothern blot studies, 1 ml samples of the induced cultures were taken at different time points before lysis occurred for DNA extraction (see Section 2.3.8.1). To remove unlysed bacteria, lysates were filtered using 0.2 μm filters and stored at 4°C until use.

Alternatively, lysates were also obtained by phage infection of susceptible bacteria. Phages enter the bacteria and start the lytic cycle, amplifying the phage population leading to bacteria lysis. To perform phage infection a 1/50 dilution of the stationary culture was grown to a density of OD_{650} = 0.15. Subsequently, the culture was centrifuged, re-suspended in TSB-phage buffer (PB; 1 mM NaCl, 0.05 M Tris pH 7.8, 1 mM $MgSO_4$, 4 mM $CaCl_2$) and infected with a phage lysate at a multiplicity of phages/bacteria of 3:1 or 1:1. Lysis normally occurred at 4 h post infection for a 5 ml of lysate.

The same induction and infection protocols were used for *E. faecalis* bacteriophages, using BHI liquid and agar medium, instead of TSB. For activation of the SOS response, *E. faecalis* strains were grown until it reached a density of OD_{540} = 0.15.

2.6.2 Bacteriophage Titering Assay

The strain RN4220 was used as recipient to perform infecting studies in *S. aureus*. RN4220 is restriction-deficient and highly sensitive to infections due to a mutation in the *sau1 hsdR* gene[300]. VE18590 (pp-) strain, curate of any MGE, was used for the virulence analysis in *E. faecalis*[239].

To quantify the number of phage particles contained in an interest lysate, a stationary culture of the acceptor strain was diluted 1:50 in TSB or BHI and grown to a density of OD_{540} = 0.3-0.4. 50 µl of this culture were infected with 100 µl of serial dilutions of phage lysate in PB at room temperature for 10 min. The mixture of phage and bacteria were plated on phage base agar plates (PBA; 25 g of Nutrient Broth No. 2, Oxoid; 7g agar) supplemented with $CaCl_2$ to a final concentration of 10 mM. PBA plates were incubated at 37°C for 24 h before plaques were counted. Plaque forming units were calculated using the following equation:

$$\frac{\text{PFU}}{\text{ml}} = \frac{\text{number of plaques}}{(\text{d x v})}$$

Where, d = dilution of plate counted

v = volume of lysate dilution plated

To stain the phage plaques and enhance plaque visibility, a 0.1% (w/v) solution of 2,3,5-triphenyltetrazolium chloride (TTC) was applied to culture plates. Only living bacteria will reduce TTC dye to red formazan and plaques will remain unstained in a surrounding red background. 5 mL of 0,1% TTC was added to the plate with the phage plaques already formed and were left for 30 min or until stained.

2.6.3 Transduction Titering Assay

All PICI, ϕSLT and ϕP27 derivatives used in this study contained a *tet*M, *tet*A or *erm*C antibiotic cassette used as markers, allowing to select phage and PICI particles by plating on TSA or BHI plates supplemented with 10 µg/ml

tetracycline or erythromycin. Transductants were quantified following a modify phage-titre protocol, as described above.

Serial dilutions of the interest lysates were performed in PB. 100 µl of these dilutions were added to 1 ml of acceptor bacteria at OD_{540} = 1.4 and $CaCl_2$ added to a final concentration 4.4 mM. The lysate mixture was incubated at 37°C for 20 min to allow the island to infect the bacteria. The mixture was then plated on TSA or BHI plates with antibiotics as required. Media contained sodium citrate (17 mM) to prevent the phage infection of bacteria by the acceptor lysate. For the *S. aureus* strains, the RN4220 strain bacterium was used as acceptor. In the case of *E. faecalis* strains, VE18590 (pp-) was used as the acceptor for enterococci PICI infection.

2.6.4 Precipitation of phage and SaPI particles

To obtain virion dsDNA from the SaPI or phage capsids, and after a large-scale infection (see Section 2.6.1), the bacteria lysates were centrifuged (9,000 rpm for 15 min) to remove unlysed cell debris. The supernatant was treated with DNase and RNase at a concentration of 1 µg/ml, to remove non-encapsidated DNA and ribosomal RNA. After incubation at room temperature for 30 min, NaCl was added to a final concentration of 1 M and incubated on ice for 1 h and then centrifuged at 11,000 rpm for 10 min at 4°C. Subsequently, the supernatant was precipitated with PEG 8000, adding 50 gr of PEG-8000 for each 500 ml of lysate and kept overnight at 4°C. Following, the purified lysate was centrifuged using a Beckman 72 Ti rotor (110,000 g for 4 h at 4°C) and the pellet resuspended in PB. Pellets obtained from the initial lysate of 500 ml were re-suspended in a final volume of 2 ml of PB. Genomic DNA was purified from phage and SaPIs capsids by phenol-chloroform extraction and subsequent ethanol precipitation (See Section 2.3.6).

2.6.4.1 DNA purification from phage and SaPI particles

To extract dsDNA from phage and SaPI particles, 100 µl of lysis buffer (9.5 µl SDS 20%, 4.5 µl Proteinase K (20 mg/ml), 90 µl of H_2O) were mixed with 100 µl of precipitated lysate and incubated on ice for 60 min. Subsequently, an equal

volume of Phenol:Chloroform:Isoamyl Alcohol ratio was added and mixed. Pellets were centrifuged at 14,000 rpm, 5 min at 4°C. The upper (aqueous) phase was then removed and further purified by ethanol precipitation (see Section 2.3.6).

2.6.5 Electron Microscopy

To obtain samples to be used for electron microscopy, lysogenic strains containing phage ϕSLT and its derivatives, were induced with MC (2.5 μg ml^{-1}). 15 ml of these lysates were treated with 5 μl of RNase (10 mg/ml) for 20 min at room temperature and filtered using 0.2 μm filters. Subsequently, 10 ml of filtered lysate were centrifuged at 81.374 x g for 2 h at 15°C (Beckman 50 Ti rotor) and the pellet was re-suspended overnight in 200 μl PB at 4°C. 5 μl of each sample were negatively stained with 1% uranyl acetate, and observed in a CM12 Phillis using 53,000x and 110,000x magnification.

ϕ12 phage and SaPIbov5 transducing particles from strains JP10435 and JP12419, respectively, were induced with MC (2.5 μg ml^{-1}) at OD_{600} = 0.5, and grown for an additional 3 hrs. Cell pellets were extracted with lysostaphin before collecting the supernatants, which were further purified by PEG precipitation and cesium chloride density-gradient centrifugation, as previously described[278]. The purified phage and transducing particles were negatively stained with 1% uranyl acetate and observed in an FEI Tecnai F20 electron microscope operated at 200 kV with a typical magnification of 65,500x. Images were captured on a Gatan Ultrascan 4000 CCD camera. Electron microscopy images were performed at Department of Microbiology, University of Alabama at Birmingham by Professor Terje Dokland.

2.7 Allelic exchange

Introduction of selected mutations into the studied prophages genomes was carried by allelic exchange by using different strategies.

2.7.1 Deletion mutagenesis using pMAD

Allelic exchange of selected genes of interest in the *S. aureus* chromosome was performed by homologous recombination using the pMAD vector as described by *Arnaud et al., 2004*[302]. The pMAD system facilitates recombination dependent on homology arms in in a two-step manner, an initial integration of the entire vector and a subsequent deletion of the vector backbone.

pMAD contains a thermosensitive origin of replication for Gram-positive bacteria, an erythromycin resistance cassette and also encodes the ß-galactosidase enzyme, allowing the identification of clones carrying the plasmid using blue-white screening by X-gal. To generate specific targeting vectors, the flanking regions of the desired gene to be deleted were cloned into pMAD to serve as homology arms. Two flanking regions were amplified and fused by double PCR reaction (Figure 2.1). Two sets of primers were used to amplify the 5' homology arm (*primer 1 and 2*) and 3' homology arm (*primer 3 and 4*) individually in a first PCR (PCR 1). The 5' end of *primer 3* contains a 20 bp overlap, homologue to the 3' end of the 5' homology arm, which allows both PCR products to be fused in a subsequent second PCR (PCR 2). To perform the fusion PCR reaction, 1 µl of each product, PCR 1 and PCR 2, and the external *primer 1 and 4* were used to amplify the final fused sequence. Restriction sites were added to allow subsequent cloning of the fusion PCR product to the 5' ends of *primer 1 and 4*. The amplified fragments were digested with the appropriate enzyme pair, cloned into pMAD vector and sequenced.

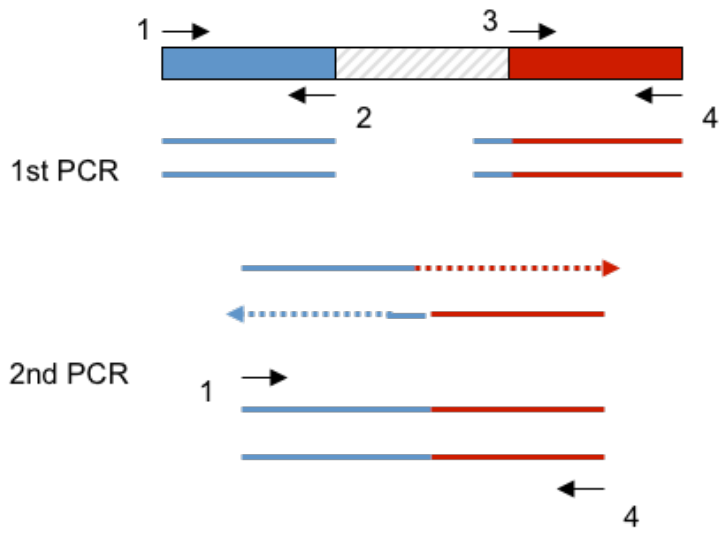


Figure 2.1 Schematic representation of the PCR performed to obtain double deletion mutants. The hatched box indicates the chromosomal sequence to be deleted. The red and blue boxes represent the flanking 5' and 3' homology arms. Arrows indicate the position where the primers hybridise.

S. aureus strains were transformed with the corresponding pMAD plasmid using the electroporation technique and selected by plating on TSA plates supplemented with erythromycin (2.5 µg/ml; Erm) and X-gal (200 µg/ml), incubated overnight at 32°C, a permissive growth temperature. Single blue candidate colonies were inoculated in 5 ml of TSB supplemented with Erm (2.5 µg/ml) and shaken overnight on a 37°C orbital shaker at 120 rpm. The cultures were 10-fold diluted and plated on TSA plates supplemented with Erm (5 µg/ml) and X-gal (200 µg/ml), incubated overnight at 44°C. Selection for antibiotic resistance at this temperature promotes the co-integration of the plasmid, which triggers homologous recombination of the flanking regions located in the plasmid with the ones present in *S. aureus* chromosome target gene (Figure 2.2).

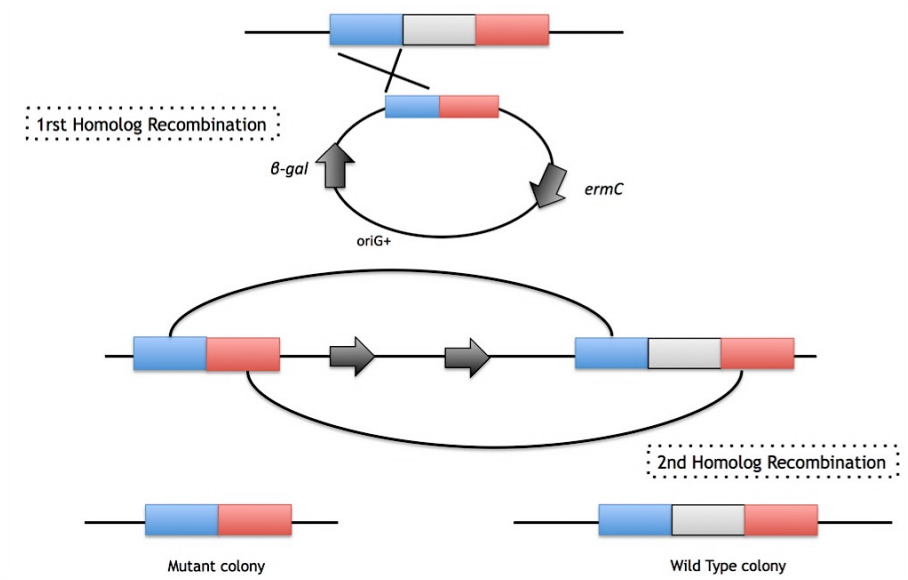


Figure 2.2 Scheme representative of pMAD mediated mutagenesis. On 1st homologous recombination (plasmid integration) colonies are blue and grow in the presence of Erm. On 2nd homologous recombination (deletion plasmid) bacteria is sensitive to Erm and are white. In this 2nd recombination, two genotypes can be produced: deletion mutant or restoration of the wild genotype.

Isolated blue and Erm resistant colonies were then used to inoculate 5 ml of TSB without antibiotic and shaken overnight at 120 rpm at 32°C, a permissive temperature that supports plasmid resolution. The cultures were 10-fold serially diluted, 100 µl was spread on TSA plates supplemented with X-gal (200 µg/ml), and incubated for 48 h at 44°C. Using this two-step strategy, homologous sequences carried on the plasmid, without the target gene, will recombine with the ones present in the chromosomal DNA.

Individual white colonies were patched into TSA plates supplemented or not with Erm (5 µg/ml) and X-gal (200 µg/ml), incubated overnight at 37°C. To screen for white colony sensitive to Erm, which implies loss of the plasmid. Candidate colonies were tested using colony PCR (see Section 2.3.1) to confirm that the insert had been specifically recombined. To verify candidates for the allelic exchange, PCR products were purified using QIAquick PCR Purification Kit (Qiagen) and sequenced through Eurofins MWG Operon.

2.7.2 Deletion mutagenesis using pBT$_2$-βgal

An alternative strategy, using the pBT2-ßgal vector[309], was used for the allelic exchange of selected genes in *S. aureus* and *E. faecalis* chromosomes by homologous recombination. Mutagenesis by pBT$_2$-ßgal follows the same mechanism and protocol as described for pMAD, but features different selection markers. The pBT$_2$-ßgal plasmid contains a thermosensitive origin of replication for Gram-positive bacteria, and confers of chloramphenicol resistance to Gram-positive strains. Additionally, ß-lactamase-encoding gene allows for identification of bacteria carrying the plasmid by blue-white screening using X-gal selection.

The procedure for PCR fusion fragments and homologous recombination is performed as described in Section 2.7.1 for pMAD[302].

2.7.3 Introduction of a marker cassette in SaPIbov5

To simplify SaPI transduction studies, a 1.2 kb fragment of SaPIbov5 chromosomal genome was replaced by an Erm marker gene (*erm*C) using a derivative of the pBT$_2$ plasmid (Section 2.7.2). To generate the targeting vector, the *erm*C gene was inserted into the multi-cloning site of pBT$_2$ and subsequently flanked with homology arms. For this purpose, the homology regions flaking the chromosomal sequence to be deleted were amplified and clone at both sides of *erm*C gene.

The procedure for PCR fusion fragments and homologous recombination is performed as described in Section 2.7.1 for pMAD[302] (Figure 2.3).

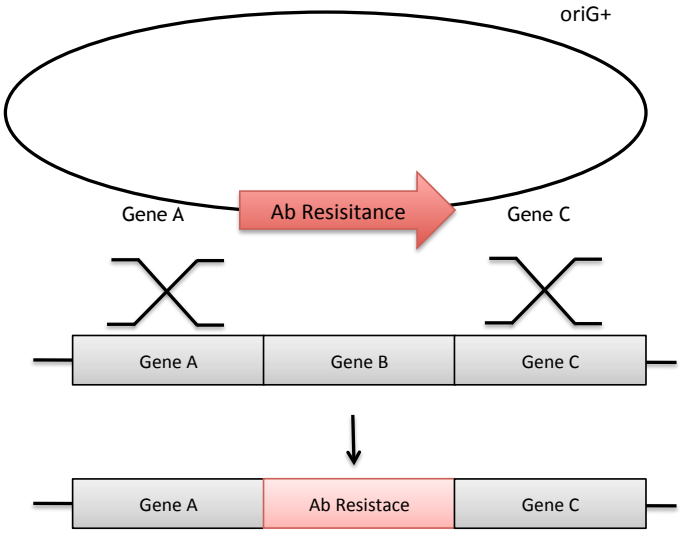


Figure 2.3 Insertion of a marker cassette in SaPIbov5. On 1st homologous recombination (plasmid integration), colonies grow in the presence of Erm and chloramphenicol. On 2nd homologous recombination (deletion plasmid), bacteria are sensitive to chloramphenicol and are Erm resistant. This method generates mutants with antibiotic resistance gene markers.

2.7.4 Gene knockout using Lambda red-mediated recombineering

Single-gene knockouts were generated in *E. coli* by *in vivo* recombineering using a phage λ red based system as described by *Datsenko & Wanner 2000*[303]. Recombineering was used to generate ϕP27 knockouts directly in the bacterial chromosome by placing a selection cassette within the phage genome. The selection marker can subsequently removed by FLP mediated deletion. The red system is based on the activity of the λ recombinases alpha, beta and gamma, which trigger the recombination of short fragments of homology, allowing sequence specific integration of DNA fragments in the desired chromosomal gene. The recombinase system is encoded on the plasmid pKD46. Simple plasmid transfection can be used to introduce the system into the strain in which recombineering is to be performed. The phage λ Red recombinases on pKD46 are controlled by the arabinose inducible *ara*BAD promoter, allowing to induce recombinase expression by the addition of L-arabinose. Additionally, pKD46 encodes an ampicillin resistance cassette and contains a thermosensitive origin of replication, allowing for selection of bacteria carrying the plasmid and of the plasmid after recombineering has been carried out.

Targeting constructs were obtained by PCR. Homology arms were added to a FRT flanked chloramphenicol resistance cassette template using bi-partite primers. Those primers contained (from 5' to 3') 50 bp homologue to the flanking region of target gene followed by 20 bases homologue to the drug-resistant cassette (Figure 2.4). FRT sites allow the elimination of the cassettes once inserted into the bacterial chromosome by using a FLP helper plasmid. Targeting PCR products were electroporated into *E. coli* strains, previously transformed with the plasmid pKD46. To generate electrocompetent cells, bacteria were grown at 30°C in 100 ml of LB media supplemented with ampicillin (10 μg/ml) and 10 mM L-arabinose to a density of OD_{600} = 0.6. Cells were washed three times by centrifugation (4000 rpm, 4ºC, 10 min) and subsequent resuspension in cold 10% glycerol (v/v). The pellet was eventually re-suspended in 1 ml of cold 10% glycerol solution.

50 μl aliquots of electrocompetent cells and 6 μl (10-100 μg) PCR product were added to a 0.1 cm Gene Pulser® cuvette. A MicroPulserTM electroporation apparatus on STA setting (2.5 kV, 25 μF, 200 Ω) was used for electroporation. Immediately after the pulse, 1 ml of LB was added directly to the cuvette and transferred to a 1.5 ml microcentrifuge tube. Cells were incubated at 37°C, 120 rpm overnight, before being plated on LB plates supplemented with 100 μg/ml of chloramphenicol. Generally, 100 μl and the cell pellet concentrate (pelleted at 4000 rpm for 1 min, the supernatant decanted, pellet re-suspended in residual supernatant) was spread and incubated for 24 h at 30°C. Individual possible mutant colonies were verified by PCR amplifications to check the correct insertion of the antibiotic in the target gene. The pKD46 plasmid was removed by patching candidate colonies in LB plates at 43ºC.

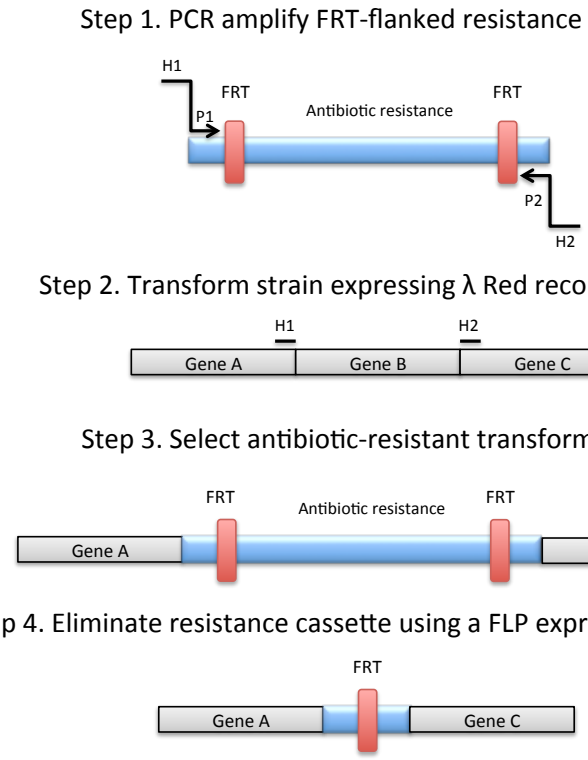


Figure 2.4 Gene replacement knockouts and deletions mutants. (Step 1) A chloramphenicol cassette is amplified by PCR using primers with homology extensions (H1 and H2) and priming sites (P1 and P2). Primer design consisted in 50 bp homology to the phage dsDNA followed by 20 bp with homology to the drug-resistance template gene (FRT). (Step 2) The chloramphenicol cassette is transformed into target competent cells where the Red system had been induced and recombination occurs. (Step 3) The recombinant strain will have inserted a drug-cassette into the target gene. (Step 4) After selection, and using plasmid pCP20 expressing the FLP recombinase, it is possible to eliminate the resistance cassette having at the end a chromosomal deletion of the targeted gene. Adapted from *Datsenko & Wanner 2000*[303].

The FRT sites flanking the chloramphenicol gene allow the elimination of the cassettes once inserted into the bacterial chromosome, using a FLP helper plasmid pCP20 (Figure 2.4). This plasmid encodes the FLP recombinase under the control of a thermosensitive replication origin (43ºC) and an ampicillin resistance cassette. Chloramphenicol mutants were transformed with pCP20, and ampicillin-resistant colonies were selected at 28°C. The obtained transformants were plated on non-selective 43ºC LB solid medium to encourage plasmid loss. To verify candidate colonies, PCR products were purified using QIAquick PCR Purification Kit and sequenced through Eurofins MWG Operon.

2.8 Data Analysis

Raw data were organised in Microsoft® Excel® 2013 and phage and PICI titres were calculated in PFU/ml or transductants/ml, respectively. Titres were then imported into GraphPad Prism® (La Jolla) to be analysed and graphed. Phage and PICI titre assays were performed in biological and technical triplicate. Significant results were averaged and the standard deviation calculated.

The NCBI BLAST server programme www.ncbi.nlm.nih.gov[310] was used to perform homology comparisons of the obtained sequences against the GenBank database. Phylogenetic analysis were performed using the programme MEGA7[311]. Praline server programme[312] was used in the multiple sequence alignment and to detect remote homologies between proteins a hidden Markov model method was used[313].

2.8.1 Statistical analysis

For comparisons of phage and PICI titre the data were first assessed for normal distribution using the D'Agostino-Pearson normality test on sample sizes of seven or more using GraphPad Prism®. Low *p*-values were found, indicating that the data were not normally distributed and hence non-parametric statistical analysis was carried out. The mean and standard deviation (SD) were calculated and the statistical significance determined by Kruskal-Wallis test with Dunn's Multiple Comparison Test to determine significant grouping compared to the controls. In all experiments; $p \leq 0.05$ was considered significant.

2.8.2 Structural modelling of $HNH_{\phi STL}$ and $TerL_{\phi STL}$ nucleases

Models of the three-dimensional structure of $HNH_{\phi STL}$ and $TerL_{\phi STL}$ nucleases were generated using the I-TASSER server[314]. Higher C-score were chosen (-0.55 and -1.21 for $HNH_{\phi STL}$ and $TerL_{\phi STL}$, respectively) and selected for further structural analysis using CCP4 suite[315] and Coot software[316]. Three-dimensional

structure models were analysed at Instituto de Biomedicina de Valencia (IBV-CSIC) in conjunction with Dr Alberto Marina.

2.8.3 Structural modelling of SaPIbov5 operon I-like proteins

Models of the three-dimensional structure of $\phi12_{gp33}$ with SaPIbov5 Ccm and SaPIbov5 ORF12 with SaPI2 PtiM were generated using RaptorX (default mode)[317], Phyre2 (intensive mode)[318] and the I-TASSER servers[314]. Higher C-score were chosen (-0.88 for Ccm, -3.07 for ptiM and -2.84 for ORF12) and selected for further structural analysis using PyMOL software[319]. RaptorX and Phyre2 servers generated models with low confidence for the N-terminal portions and high confidence for the C-terminal portions of $\phi12_{gp33}$ and SaPIbov5 Ccm (Appendix 5). C-terminal portions of gp33 and Ccm were structurally aligned with Mustang[320] and this alignment was rendered with ESPript 3.0[319]. Three-dimensional structure models were analysed at Instituto de Biomedicina de Valencia (IBV-CSIC) in conjunction with Dr Alberto Marina.

2.8.4 Structural modelling of EfCIV583 and ϕp1 proteins

Secondary structural models of the $TerS_{\phi pI}$ (EF0332) $TerL_{\phi pI}$ (EF0333), ϕp1 capsid protein (EF3039), ϕp1 scaffolding protein (EF3038) and EfCIV583 CpmE (EF2940) were generated using the I-TASSER server[314]. Higher C-score were chosen (-1.79 for EF0332; -0.77 for EF033; -1.97 for EF0339; -3.02 for EF0338; -2.10 for EF2940) and selected for further structural analysis using PyMOL software[319].

Chapter 3 *Cos*-site packaging and phage-encoded HNH endonucleases

Results from this section are included in:

Quiles-Puchalt N*, **Carpena N***, Alonso JC, Novick RP, Marina A, Penadés JR. (2014) Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. *Proc Natl Acad Sci U S A*. 111(16):6016-21.

*Denotes joint first authorship.

Disclaimer on work performed:

Structural modelling data in this chapter was performed in conjunction with and under the guidance of Dr Alberto Marina.

Dr Nuria Quiles Puchalt kindly provided some strains in this chapter, and full details are listed in Table 2.1 of Chapter 2.

The group of José R Penadés kindly provided some strains mentioned in this chapter, and full details are listed in Table 2.1 of Chapter 2.

3.1 Background

Bacteriophage-mediated transfer of MGEs, including the mobilisation of SaPIs, is one of the best-studied examples of HGT, and is highly dependent on the efficient encapsidation of a particular genome into a preformed procapsid. Phages are classified by the mechanism by which their concatemeric dsDNA is cleaved and packaged, defined by either *cos*- or *pac*- sites in the dsDNA concatemeric form[159,188,189]. In both, *cos*- or *pac*-phages, the packaging mechanism is driven by the activity of the nucleoprotein terminase complex, TerS/TerL. The TerL subunit is classically associated with the nuclease functions for the cleavage of specific *pac*- or *cos*- sites, generating phage genome monomers, and the ATPase function for their translocation into the capsid structure. The small subunit TerS is believed to mediate the specific binding of the complex to *cos*- or *pac*- sequences[159,160].

As previously highlighted, transduction of MGEs is assumed to be driven only by *pac* phages[245,250,261]. The only requisite to start the packaging process for *pac* phages is the presence of one single *pac* site homolog on the bacterial chromosome. For *cos* phages, the requirement of a second matching *cos* sequence in the host genome to complete packaging drives this process in a more selective manner and limits *cos* phage contribution to HGT.

Furthermore, a novel HNH endonuclease protein has been described to be encoded in *cos* bacteriophages infecting Gram-positive and Gram-negative bacteria[192,193]. Interestingly, HNH endonucleases share a common genetic position inside the phage genome, been side by side with genes encoding the terminases and other morphogenetic proteins. Moreover, the HNH motif is present in numerous enzymes playing a diversity of functions in different cellular processes[191]. Although all the analysed phage HNH proteins carried endonuclease activity[192,193], the biological role of these phage-encoded proteins remains unknown.

SaPIs are a family of ~15 kb MGEs, carrying genes encoding virulence factors and antibiotic resistance. SaPIs survive through parasitism of the life cycle of a temperate (helper) phage. Although SaPIs are not SOS-induced, their helper

phages are, and after activation of the SOS response, de-repression of Stl is achieved by the intervention of a phage-encoded protein[109,110]. Furthermore, most of the studied SaPIs use a *pac* mechanism to encapsidate their DNA. These SaPIs encode a homolog of the small terminase subunit ($TerS_{SP}$) which complexes with the phage-encoded TerL, redirecting SaPI packaging and completing encapsidation of their reduced genome into small particles[256]. SaPI parasitism is aided through other encoded proteins such as Ppi, CpmA or CpmB[255-257], that ultimately redirects the process to its advantage. However, not all SaPIs that have been described in the literature encode a recognisable TerS-morphogenetic module, and for some time these SaPIs were thought to be defective. In recent times, some studies have suggested that these 'defective' SaPIs could be mobilised[238]. In those studies, SaPIbov5 was proposed to be the prototype of these novel identified SaPI islands[238] (Figure 3.1).

The aims of these studies were first to determine the relationship between this novel family of pathogenicity islands, without a particular morphogenetic module, and the different *cos* and *pac* helper phages and to demonstrate that they are indeed mobile. Accordingly, a goal for these studies was to decipher the uncharacterised mechanism by which they are packaged and transferred using these helper phages. Subsequently, the role of the phage-encoded HNH protein in the phage cycle was determined. Ultimately, and to generalise our results, ϕP27, an *E. coli* phage encoding an HNH protein was characterised, and the endonuclease function was compared with that present in Gram-positive phages.

3.2 Results

3.2.1 Phage-specific SaPI packaging and transfer

As previously highlighted, SaPIs have to be induced by a helper phage to be packaged and moved to a new host cell. Although most SaPIs encode a SaPI-specific TerS protein, $TerS_{SP}$, to allow their packaging into small and large capsids[256], some SaPIs lack a recognisable $TerS_{SP}$, suggesting that they are

transferred using an unknown packaging mechanism. To investigate this hypothesis, the phage-specific mobilisation of SaPIbov5 was examined. SaPIbov5 is the prototype of the SaPIs missing a morphogenetic module[238]. The transfer frequency of SaPIbov5 was compared with SaPIbov1, carrying the standard morphogenetic module[241] (Figure 3.1). SaPIbov1 was chosen as a control, not only because it is the prototype of the *pac* SaPIs using a headful mechanism, but also because it encodes the same Stl repressor as SaPIbov5, and therefore it is predicted that the same helper phages induce both islands[242,269].

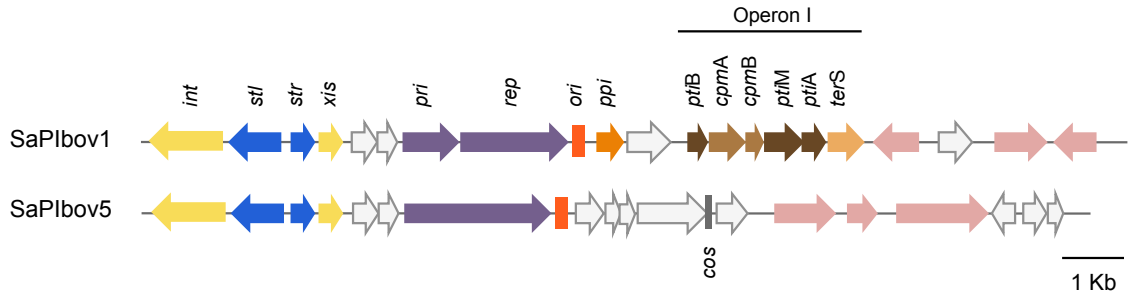


Figure 3.1 Comparison of SaPI genomes. Genomes are aligned according to the prophage convention with the integrase gene at the left end. Gene colour code: *int* and *xis*, yellow; transcription regulators, blue; replication genes, purple; replication origin, red; genes affecting expression (*pti*) or assembly (*cpm*) of helper phage virion components, dark brown and medium brown, respectively; the terminase small subunit gene (*ter*S), light brown; *ppi* (phage interference), orange; superantigen and other accessory genes, pink. Genes encoding hypothetical proteins, white. The *cos* site is shown in dark grey.

Derivatives of SaPIbov5 and SaPIbov1 islands, carrying a *tet*M marker were employed to analyse these distinctive SaPIs for the three definable stages of the SaPI ERP cycle (excision-induction, replication and packaging and transfer). The derivative islands were then introduced into lysogenic strains containing either *pac* (ϕ11 and ϕ80α) or *cos* (ϕ12 and ϕSLT) phages.

The generated strains were then MC induced, and as shown in Figure 3.2, all tested phages induced both families of SaPIs. SaPIbov1 induction by phages ϕ11 and ϕ80α, which are known to be inducers of *pac* SaPIs, generated the characteristic SaPI-specific small band. This band represents the SaPI monomers packaged into the small capsid produced after expression of the SaPI-encoded packaging module[241]. The absence of the SaPI-specific band in the other sample was initially explained by the lack of the SaPI-specific

packaging module in SaPIbov5. However, this initial assumption was wrong, and this question will be addressed in Chapter 5 when the mechanisms by which SaPIbov5 is packaged both in large and small capsids by *cos* phages will be identified.

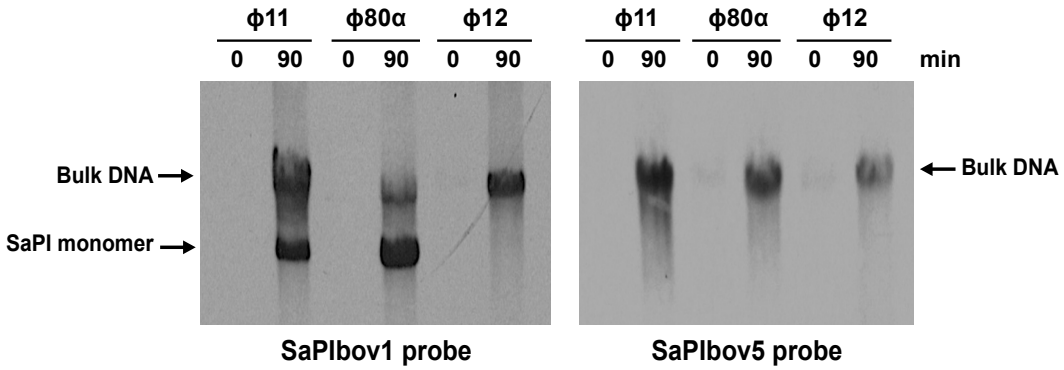


Figure 3.2 Replication and encapsidation analysis of SaPIbov5. Southern blots were performed after SOS induction of the different lysogenic strains carrying either SaPIbov1 (left) or SaPIbov5 (right). Samples were taken at 0 or 90 min after MC induction, separated on agarose and blotted with SaPIbov1- or SaPIbov5-specific probes. The upper band represents 'bulk' DNA, including chromosomal, phage, and replicating SaPI DNA; the lower band represents SaPI linear monomers released from phage heads.

Interestingly, although using different mechanisms, phages ϕ11 and ϕ80α efficiently transferred both islands. Whilst the transfer of SaPIbov1 was dependent on the SaPI-encoded $TerS_{SP}$, SaPIbov5 used the phage-encoded TerS for its packaging and transfer (Table 3.1). As previously published, and confirmed in this experiment, deletion of ϕ11 *ter*S eliminated phage but not SaPIbov1 transfer[254] (Table 3.1). Thus, ϕ11 *ter*S mutation abolished SaPIbov5 but not SaPIbov1 transfer, which confirms that SaPIbov5 requires the phage terminase to be packaged (Table 3.1). This result suggests that SaPIbov5 might contain a *pac* site, likely to be similar to the one present in ϕ11, which could explain the ability of SaIbov5 to be mobilised by a helper phage despite the lack of any identifiable *ter*S gene. Surprisingly, SaPIbov5 but not SaPIbov1 was packaged and transferred by phages ϕ12 and ϕSLT, which have not thus far been proposed to be helper phages (Table 3.1). This transfer can only be explained by the ability of SaPIbov5 to hijack ϕ12 and ϕSLT encoded proteins, including TerS, indicating once more that SaPIbov5 exploits the helper phage reproduction cycle for its own reproduction. As a consequence of the SaPI

parasitism, the phage titre of the *cos* phages is reduced after the induction and packaging of SaPIbov5 (Table 3.1).

Since both ϕSLT and ϕ12 are *cos* phages[321], and in order to clarify how SaPIbov5 is packaged by these phages, sequence analyses were performed to search for a putative *cos* site in the SaPIbov5 genome. Interestingly, several SaPIs lacking the classical morphogenetic module carried a *cos* sequence matching the putative *cos* site found in several staphylococcal phages, including phage ϕ12 (Figure 3.3).

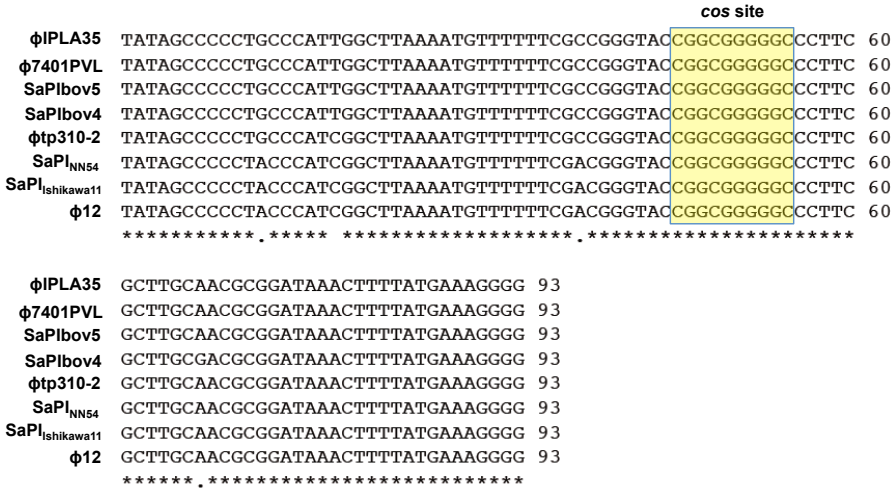


Figure 3.3 The *cos* site alignment for different phages and SaPIs. The predicted SaPI and phage *cos* sites and their flanking sequences were aligned using ClustalW2. The *cos* sites are shaded in yellow.

To determine the functionality of the putative *cos* sites present both in the helper phages and in the SaPIs, the region containing the *cos* site and the flanking sequences was cloned into plasmid pCU1[299]. Derivative plasmids in which the *cos* sites were deleted were also generated as controls. As expected, ϕ11 transferred all these plasmids at equal frequency by generalised transduction. Related to the *cos* phage ϕ12, only plasmids carrying the phage or the SaPI *cos* sites, but not the flanking sequences, were transferred (Table 3.2). Moreover, a mutant of SaPIbov5 *cos* site was generated and it was found that this deletion abolished transfer of the island by ϕ12, but not by ϕ11 (Table 3.1). Together these data confirm that SaPIbov5 mobilisation is phage TerS dependent, which is supported by the functionality of the SaPIbov5 *cos* sequence.

Table 3.1 Effects of phage mutations on phage and SaPI titres[a].

Phage	SaPI	pCN51[b]	Phage titre[c]	SaPI titre[d]
ϕ11	-	-	6.6×10^9	-
ϕ11 Δ*ter*S	-	-	<10	-
ϕ11	SaPIbov1	-	2×10^7	1×10^7
ϕ11	SaPIbov5	-	2×10^9	1.3×10^7
ϕ11	SaPIbov5 Δ*cos*	-	1.6×10^9	1.1×10^7
ϕ11 Δ*ter*S	SaPIbov1	-	<10	1.9×10^7
ϕ11 Δ*ter*S	SaPIbov5	-	<10	<10
ϕ11 Δ*ter*S	SaPIbov5	pCN51-*ter*S$_{\phi 11}$	1.8×10^5	2.3×10^5
ϕ80α	-	-	8×10^{10}	-
ϕ80α	SaPIbov1	-	3.4×10^9	1×10^7
ϕ80α	SaPIbov5	-	1.5×10^{11}	1.2×10^7
ϕ12	-	-	8×10^9	-
ϕ12 Δ*hnh*	-	-	<10	-
ϕ12 Δ*hnh*	-	pCN51-*hnh*$_{\phi 12}$	2.2×10^4	-
ϕ12	SaPIbov1	-	1.1×10^9	<10
ϕ12	SaPIbov5	-	1.1×10^6	1.1×10^4
ϕ12	SaPIbov5 Δ*cos*	-	1.4×10^7	<10
ϕ12 Δ*hnh*	SaPIbov5	-	<10	<10
ϕ12 Δ*hnh*	SaPIbov5	pCN51-*hnh*$_{\phi 12}$	1.4×10^4	8.1×10^4
ϕSLT	-	-	6.2×10^7	-
ϕSLT Δ*hnh*	-	-	<10	-
ϕSLT Δ*hnh*	-	pCN51-*hnh*$_{\phi SLT}$	3.1×10^4	-
ϕSLT	SaPIbov1		2.7×10^6	<10
ϕSLT	SaPIbov5		8.2×10^5	8.7×10^3
ϕSLT	SaPIbov5 Δ*cos*		3.3×10^6	<10
ϕSLT Δ*hnh*	SaPIbov5		<10	<10
ϕSLT Δ*hnh*	SaPIbov5	pCN51-*hnh*$_{\phi SLT}$	5.7×10^3	1.3×10^3

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]Complemented both the donor and the recipient strains.
[c]Phage forming units PFU/ml induced culture, using RN4220 as recipient strain.
[d]No. of SaPI transductants/ml induced culture, using RN4220 as recipient strain.

Table 3.2. Analysis of the putative phage and SaPI *cos* sites[a].

Donor strain	Phage	Cloned site *cos*[b]	Plasmid titre[c]
JP10974	ϕ12	Empty vector	< 10
JP10968	ϕ12	ϕ12	2×10^3
JP10878	ϕ12	ϕ12 Δ*cos*	< 10
JP10907	ϕ12	ϕSLT	6×10^2
JP10879	ϕ12	ϕSLT Δ*cos*	< 10
JP10875	ϕ12	SaPIbov5	1×10^3
JP10876	ϕ12	SaPIbov5 Δ*cos*	< 10
JP10908	ϕ12	ϕP27[d]	< 10
JP10880	ϕ12	ϕP27 Δ*cos*	< 10
JP11204	ϕ11	Empty vector	6.2×10^3
JP11196	ϕ11	ϕ12	5×10^3
JP11197	ϕ11	ϕ12 Δ*cos*	9.6×10^3
JP11198	ϕ11	ϕSLT	2.4×10^3
JP11199	ϕ11	ϕSLT Δ*cos*	9.6×10^3
JP11200	ϕ11	SaPIbov5	3.1×10^3
JP11201	ϕ11	SaPIbov5 Δ*cos*	7×10^3
JP11202	ϕ11	ϕP27[d]	2.1×10^3
JP11203	ϕ11	ϕP27 Δ*cos*	1.1×10^3

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]In addition to the *cos* site, the cloned DNA contains ~50 bp of the region flanking the *cos* site. The fragments cloned in the Δ*cos* plasmids are mutant in the *cos* site (maintaining the flanking sequences).
[c]No. of plasmid transductants/ml induced culture, using RN4220 as recipient strain.
[d]ϕP27 is an *E. coli* phage carrying an entirely different *cos* site.

3.2.2 The phage-encoded HNH endonuclease is required for the packaging and transfer of both phages and SaPIs

The genome of *cos* phages is arranged in modules clustering genes of related function. Analysis of the packaging module of helper phage ϕ12 was performed to decipher the mechanisms by which SaPIbov5 was transferred. In these experiments, phage ϕSLT was also included, which has the same packaging module as ϕ12 and encodes the clinically relevant Panton-Valentine

leukocidin (PVL) toxin, lately implicated in serious staphylococcal infections[64]. Furthermore, phages ϕSLT and ϕ12 encode small and large terminases, located next to the *rin*A gene (Figure 3.4). RinA controls the expression of the late morphogenetic operon genes and is located at the end of the early gene cluster in the ϕSLT and ϕ12 phage genomes[153,154].

Interestingly, located between the *rin*A and the *ter*S genes of both ϕSLT and ϕ12 phages, is a gene encoding an uncharacterised protein (Figure 3.4). Following sequence analysis, it was found that both proteins, ϕ12p28 and ϕSLTp37, share a high sequence identity (98%), and are related to the PFAM family of HNH endonuclease proteins. As both genes are located within the late morphogenetic operon, it was hypothesised that they might be involved in the packaging process. If so, RinA should control their expression, as it regulates genes involved in this process. To determine whether RinA regulates the expression of ϕ12p28 and ϕSLTp37, transcriptional analyses were performed using plasmid pCN42. Thus, a set of reporter gene fusions with and without *rin*A was generated and tested for reporter gene expression (see scheme in Figure 3.5A). As shown in Figure 3.5B, the expression of the *hnh* gene was RinA dependent. These data confirmed that the *hnh* gene from ϕ12 and ϕSLT are transcriptionally regulated by RinA, and are likely to belong to the same morphogenetic cluster.

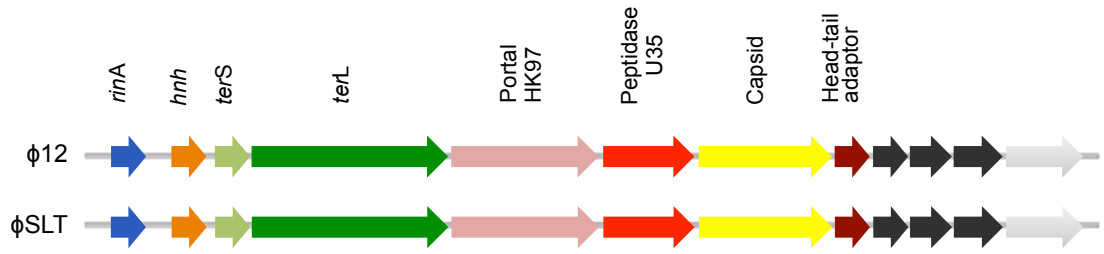


Figure 3.4 Alignments of the morphogenetic module of phages encoding HNH proteins. Genes are coloured according to their sequence and function: *rin*A*:* blue; *hnh: orange*; *ter*S: light green; *ter*L*:* green; portal: pink; peptidase: red; capsid: yellow; head-tail adaptors: brown; hypothetical proteins: black.

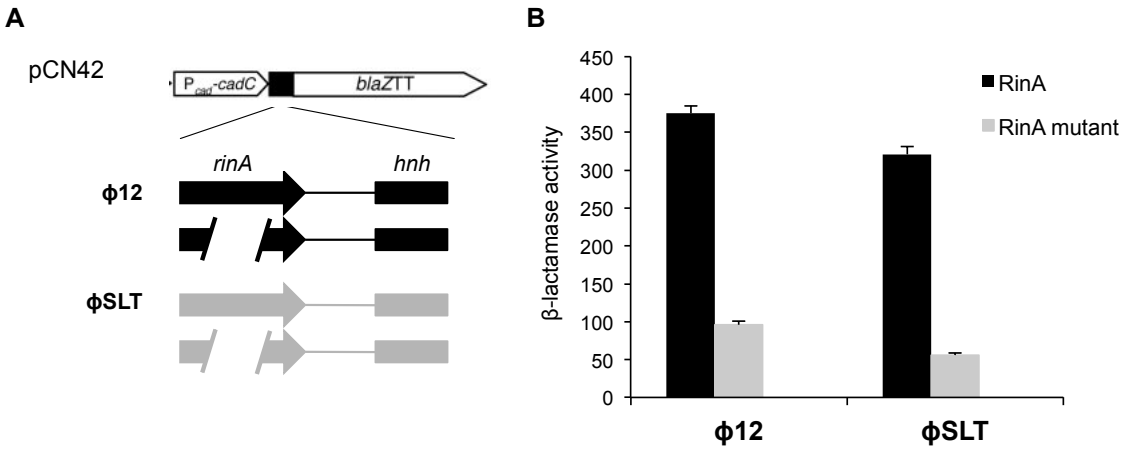


Figure 3.5 RinA proteins control *hnh* expression. (A) Schematic representation of the generated *bla*Z transcriptional fusions. (B) Derivatives of strain RN4220 containing each of the indicated plasmids were assayed at mid-exponential phase for β-lactamase activity under standard conditions. Samples were normalised for total cell mass. Error bars shown SEM.

It was hypothesised that the HNH proteins of ϕSLT and ϕ12 phages could play a role in phage DNA packaging, not only because of their position in the genome, next to the terminase genes, but also because they were regulated by the master regulator of the late morphogenetic cluster. To test for a potential contribution of the *hnh* genes in the packaging process, knockout mutants of the *hnh* genes from both ϕ12 and ϕSLT prophages were generated and then tested for their ability to form phage particles following induction. As shown in Table 3.3, none of the derivate mutants was able to produce infectious phage particles. Subsequently, the ϕ12*hnh* and ϕSLT*hnh* mutants were analysed for its ability to mobilise SaPIbov5 and found that this mutation completely abolished SaPI transfer (Table 3.1). Complementation of the *hnh* mutants restored both phage and SaPI titres (Table 3.1, Table 3.3). In conclusion, ϕ12p28 and ϕSLTp37 are indispensable not only for ϕ12 and ϕSLT reproduction but also for SaPIbov5 transfer.

As previously described, HNH family proteins are known to display endonuclease activity, which is usually nonspecific[191,192]. However, several studies have shown that proteins with the HNH motif display nuclease activity upon recognition of a specific binding site sequence[193,322]. In view of this, it was investigated if the aforementioned proteins displayed nuclease activity and if this activity was essential for phage packaging.

To test this hypothesis, 3-dimensional structural models were generated to determine whether the ϕSLTp37 protein displayed the classical features of HNH proteins. For that purpose, the iterative threading assembly refinement I-TASSER server[314] was used (Figure 3.6). Despite low sequence identity, ϕSLTp37 displayed a high degree of structural similarity to the HNH proteins GS-15 of *Geobacter metallireducen*s (yellow; PDB 4H9D)[323] and *Pac*I of *Pseudomonas alcaligenes* (blue; PDB 3LDY)[324].

Structural alignments with GS-15 and *Pac*I HNH nucleases with the ϕSLTp37 protein uncover the conserved catalytic HNH ßßα-metal motif correctly placed for catalysis between residues 38 and 90 (Figure 3.7). Within this motif, a His residue was found at position 58, corresponding to the His which activates a water molecule that performs the nucleophilic attack on the sugar-phosphate backbone. Likewise, the metal-binding residues positioned at His83 and His57 could chelate the catalytic divalent cation (Mg^{2+} or Mn^{2+}) in ϕSLTp37. Residue Asn74 was positioned for the suitable orientation of the His residue for the catalysis role. The requirement for a tetrahedral coordination of a Zn^{2+} cation is another characteristic feature of the HNH structure. This was observed in the model for ϕSLTp37, where the Zn^{2+} is coordinated by Cys41, Cys44, Cys79 and Cys82.

To demonstrate that this protein does indeed have nuclease activity, the pPROEX-HTa plasmid was used for the expression of the ϕSLTp37 protein in *E. coli*. The purified protein assessed for its ability to digest DNA *in vitro* (Figure 3.8). For that purpose, ϕSLTp37 protein and ϕSLTp37 mutants, in which the two putative catalytic His residues, H57 and H58, were replaced by Ala, were analysed. As a control, the putative $TerL_{\phi SLT}$ was also tested for the nuclease activity, classically associated with the phage large terminase subunit. As reported for other phage-encoded HNH proteins[192], the expressed ϕSLTp37 protein cleaves the DNA non-specifically, and therefore, this protein does not have a recognition domain, suggesting that other proteins will control the specific cleavage of the phage *cos* site. The tested constructs mutant in the catalytic His H57 and H58 lacked nuclease activity, confirming ϕSLTp37 as the $HNH_{\phi SLT}$ protein as an endonuclease. Overexpression of the putative $TerL_{\phi SLT}$

did not result in DNA degradation; this could suggest a possible coordination of the HNH, TerS and the TerL being required for a precise nuclease function. Altogether, this model supports the hypothesis that $HNH_{\phi SLT}$ and $HNH_{\phi 12}$ are functional HNH nucleases.

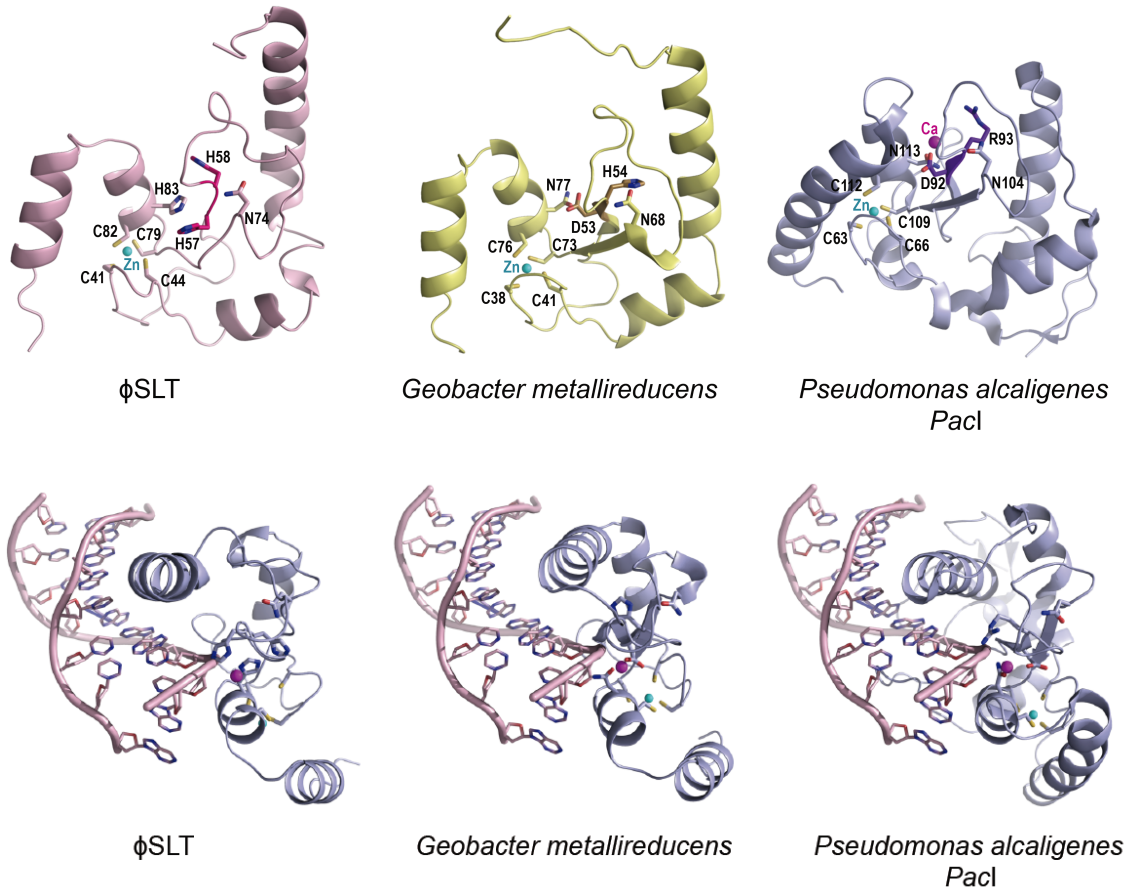


Figure 3.6 ϕSLTp37 shows the characteristic HNH nuclease sequence and fold. ϕSLTp37 structural model (pink) generated with I-TASSER is compared with the experimentally determined structures of *Geobacter metallireducens* GS-15 (yellow; PDB 4H9D) and *Pseudomonas alcaligenes Pac*I (blue; PDB 3LDY) HNH nucleases. In the upper panel, the ribbon representations of the three structures are shown in the same orientation with catalytic and structurally relevant residues represented as sticks and coloured by atoms (nitrogen, oxygen and sulphur in blue, red and yellow, respectively) with carbon in the same colour as the parent structure. The residues mutated to Ala in ϕSLT HNH (H57 and H58) and the correspondent residues in GS-15 and *Pac*I are highlighted in darker blue. The catalytic Zn^{2+} ion is displayed as a cyan sphere. In the lower panel, the DNA-HNH complex is modelled for ϕSLTp37 and GS-15 by superimposing these protein structures with *Pac*I HNH in the *Pac*I-DNA experimentally determined structure (PDB 3LDY).

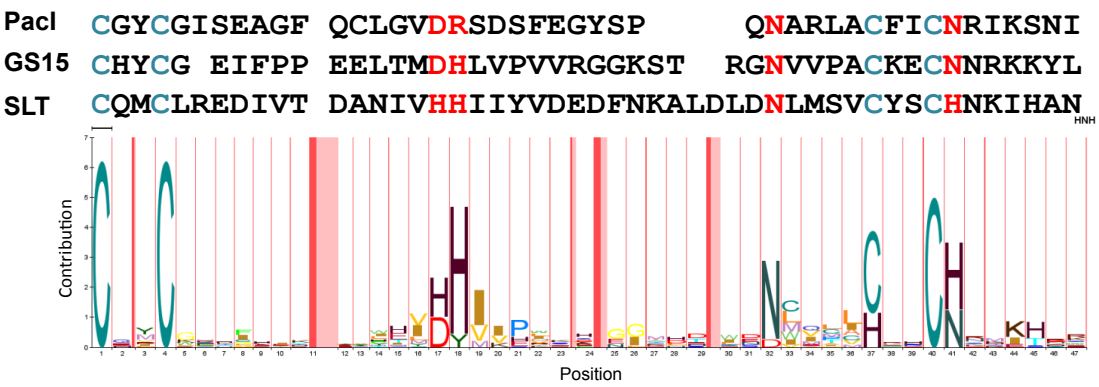


Figure 3.7 Sequence conservation in HNH catalytic domains. HMM logo representation of PFAM family HNH (PF01844) (http://pfam.sanger.ac.uk/family/PF01844.18) generated by the alignment of 7400 sequences. A stack designates each position in the sequence. The overall height of the stack indicates the sequence conservation at that position while the length of the symbols within the stack designates the relative frequency of each amino at that position. The sequence corresponding to the HNH domains of ϕSLTp37 (SLT), GS-15 (GS15) and *Pac*I are aligned with the HMM logo and the catalytic and Zn^{2+} chelating residues displayed as sticks in Figure 3.6 are highlighted in red and cyan, respectively.

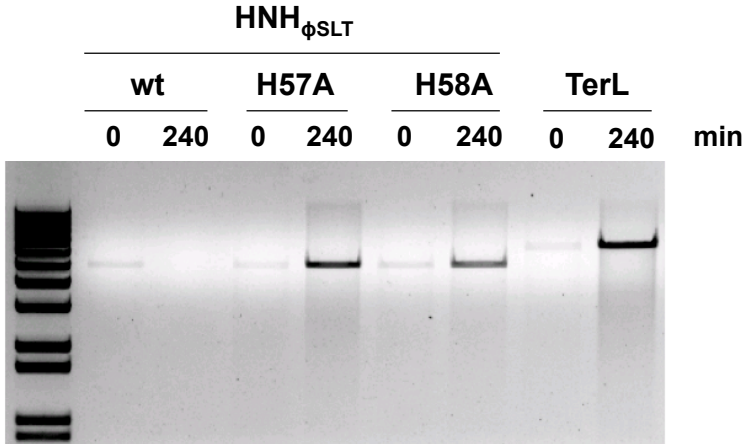


Figure 3.8 Plasmid degradation by the overexpressed HNH proteins. Plasmid DNA from *E. coli* strains expressing the different proteins was extracted and digested with *Bam*HI, which cuts the plasmid once. Note the absence of plasmid in the sample obtained 240 min after IPTG induction of the wt HNH protein, which is indicative of unspecific nuclease activity. By contrast, neither the HNH mutants nor the TerL proteins showed nuclease activity.

Table 3.3 Effect of ϕSLT mutations on phage titre[a].

ϕSLT	**Complemented[b]**	**Phage transductants[c]**
Wild type		5.0×10^7
Δ*hnh*		< 10
Δ*hnh*	pCN51-$hnh_{\phi SLT}$	1.9×10^3
HNH H57A		< 10
HNH H57A	pCN51-$hnh_{\phi SLT}$	2.2×10^2
HNH H58A		< 10
HNH H58A	pCN51-$hnh_{\phi SLT}$	1.3×10^2
Δ*ter*S		< 10
Δ*ter*S	pCN51-$terS_{\phi SLT}$	1.6×10^6
Δ*ter*L		< 10
Δ*ter*L	pCN51-$terL_{\phi SLT}$	4.6×10^5
TerL E208A		< 10
TerL E208A	pCN51-$terL_{\phi SLT}$	1.0×10^4
TerL D363A		<10
TerL D363A	pCN51-$terL_{\phi SLT}$	1.1×10^4
ΔPortal		< 10
ΔPortal	pCN51-$p40_{\phi SLT}$	1.2×10^5
Δ Prohead protease		< 10
Δ Prohead protease	pCN51-$p41_{\phi SLT}$	2.4×10^4
Δ Major capsid protein		< 10
Δ Major capsid protein	pCN51-$p42_{\phi SLT}$	1.4×10^5
Δ Tail protein		< 10
Δ Tail protein	pCN51-$p47_{\phi SLT}$	2.8×10^5
ϕ12		**Phage transductants[c]**
Wild type		8×10^9
Δ*hnh*		< 10
Δ*ter*S		< 10

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]Complemented in donor strain.
[c]No. of phage transductants/ml induced culture, using RN4220 as recipient strain.

3.2.3 TerL features the classical ATPase and nuclease domains

$TerL_{\phi SLT}$ and $TerL_{\phi 12}$ belong to a family of TerL proteins (Terminase_1; PF03354), which has not been characterised yet. The results mentioned above suggested that HNH proteins could provide the nuclease activity for the packaging process, classically associated to the TerL proteins. Since $TerL_{\phi SLT}$ nuclease activity was not detected after overexpression of the $TerL_{\phi SLT}$ protein with the pPROEX-HTa plasmid (Figure 3.8), further analyses were performed to assess the role of the TerL family in the phage cycle.

Structural modelling was performed to characterise the $TerL_{\phi SLT}$ protein and showed that the overall fold of $TerL_{\phi SLT}$ closely resembles that of phage T4 gp17 (PDB 3CPE)[186] and less closely that of phage Sf6 gp2 (PDB 4IEE)[325] large terminases (Figure 3.9). Similar to these structures, the structural model of the protein folds in two well-differentiated domains, an N-terminal region (residues 1-330), which encompasses the ATPase (residues 38-271) and linker (residues 1-37 and 272-330) domains, and a C-terminal region (residues 331–563) corresponding to the nuclease domain (Figure 3.9). The N-terminal ATPase domain displays the canonical ASCE fold observed in other terminases and ATPases, including the Walker A and B motifs that form the active centre[176]. $TerL_{\phi STL}$ Glu208 in the Walker B motif would be the general base that polarises a water molecule for ATP nucleophilic attack, and Lys117 and Thr118 in Walker motif A would participate in the catalysis by coordinating ATP phosphates and an Mg^{2+} cation, respectively (Figure 3.10). In the same way, the C-terminal portion shows the highly conserved features of the canonical phage nuclease domains including a triad of acidic residues (Asp363, Asp447 and Asp528 in $TerL_{\phi STL}$) (Figure 3.10). This acidic triad corresponds to the catalytically essential residues and their mutation results in loss of nuclease activity in T4 gp17, SPP1 gp2 and Sf6 gp2 phage terminases[186,325–327].

To assess whether the $TerL_{\phi SLT}$ protein has the two classical motor and nuclease domains, $TerL_{\phi SLT}$ derivatives were generated in which the key residues in the putative domains, E208 (motor) and D363 (nuclease activity), were mutated to Ala (Figure 3.9, Figure 3.10). A $TerL_{\phi SLT}$ mutant deficient in

the full protein was also generated. In all cases, these mutations affected the phage transfer dramatically (Table 3.3). Overall these results confirm that $TerL_{\phi SLT}$ has the two classical domains present in all characterised TerL proteins and suggest that packaging of this family of phages requires the nuclease activity of both TerL and HNH proteins.

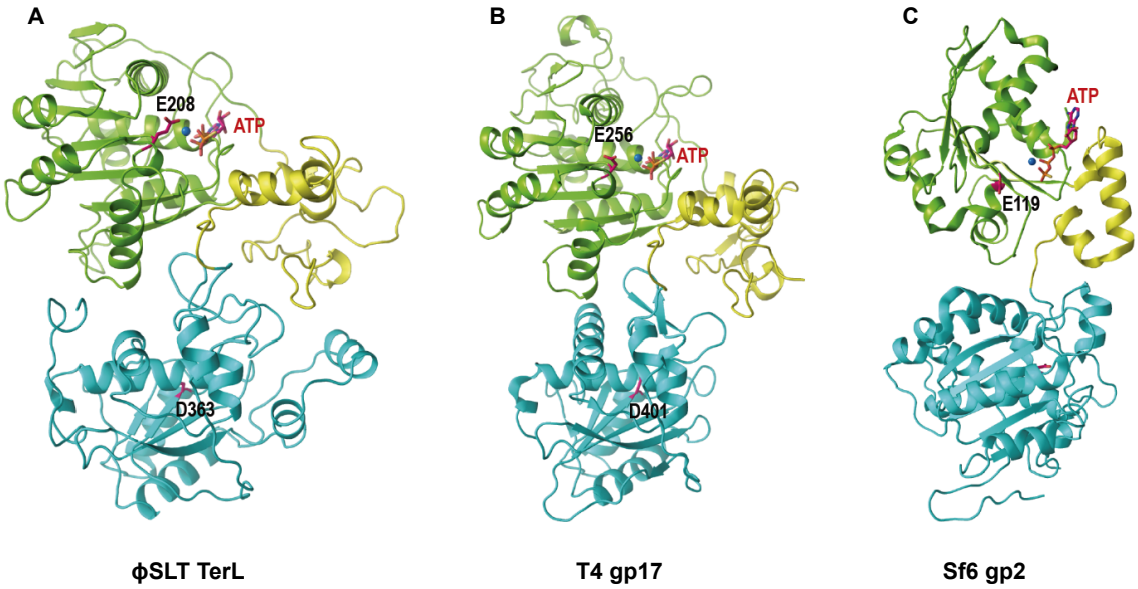


Figure 3.9 Overall modelled three-dimensional structure of $TerL_{\phi SLT}$. (A) The three-dimensional structure of $TerL_{\phi SLT}$ modelled with I-TASSER is shown in ribbon representation with the motor N-terminal domain coloured yellow (subdomain I) and green (subdomain II) and the C-terminal nuclease domain cyan. The $TerL_{\phi SLT}$ model has an overall fold that closely resembles large terminases from phage T4 gp17 (PDB 3CPE) (B) and less closely to phage Sf6 gp2 (PDB 4IEE) (C), for whose structures are coloured as for $TerL_{\phi SLT}$. Catalytic residues mutated in $TerL_{\phi SLT}$ and the corresponding residues in the large terminases of T4 gp17 and Sf6 gp2 are shown as sticks, labelled and coloured magenta. The ATP molecule present in the Sf6 gp2 structure was placed in the $TerL_{\phi SLT}$ and T4 gp17 structures by superimposition of the corresponding motor domains and is shown as sticks magenta.

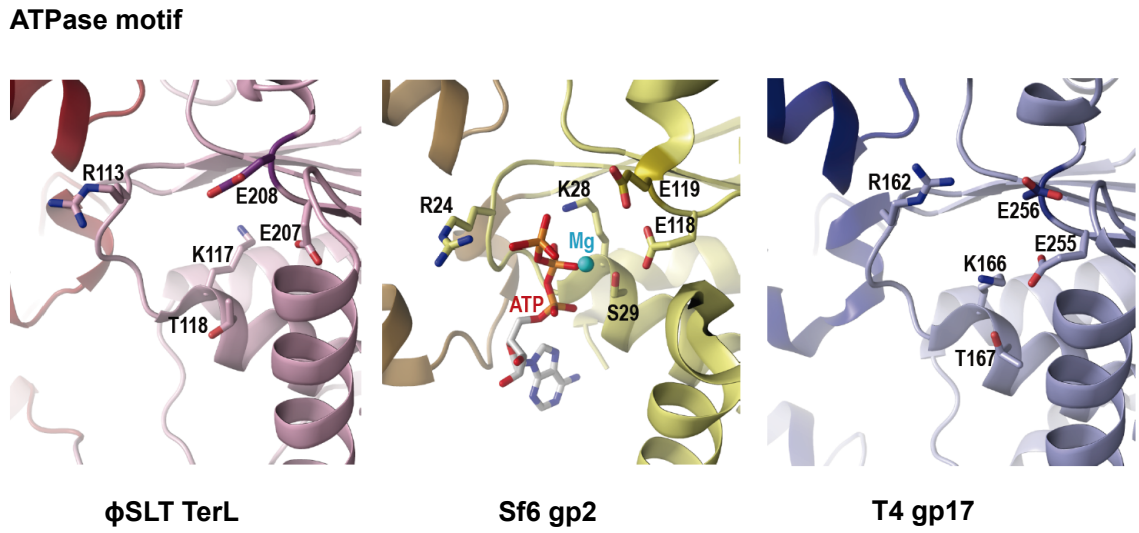


Figure 3.10 ATPase and nuclease active sites of TerL$_{\phi SLT}$. Close-up view of the ATPase (upper) and nuclease (lower) active sites of TerL$_{\phi SLT}$, Sf6 gp2 and T4 gp17 with key catalytically residues shown as sticks. The ATP molecule found in the ATPase site of Sf6 gp2 is shown as sticks with the carbon atoms in white. Cations found in the Sf6 gp2 structure or modelled in the nuclease site of TerL$_{\phi SLT}$ and T4 gp17 structures are shown as cyan spheres. In the TerL$_{\phi SLT}$ active site, the catalytic residues mutated in this work, Glu208 and Asp363 are highlighted in darker hues. Glu208 in the ATPase active site would correspond to the general base that polarises a water molecule for ATP nucleophilic attack. Asp363 is one of the three acidic residues forming the catalytic triad.

3.2.4 Role of *ter*S, *ter*L or *hnh* in *cos*-site cleavage

The aforementioned results support the hypothesis that the HNH and TerL proteins play a key role in the phage packaging process. Since the phage holoterminase complex (TerS-TerL) is involved in the cleavage of the

concatemeric phage genome, the ability to generate the *cos*-site cleavage reaction of *ter*S, *ter*L and *hnh* ϕSLT mutant derivatives was examined. In addition, ϕSLT derivatives in which the proposed HNH catalytic His57 and His58 residues were mutated to alanine were also generated. Both Histidines are responsible for the catalytic activity, and their mutation abolishes the catalytic role of the nuclease domain[324] (Figure 3.6, Figure 3.8). Further, the TerL mutations generated based on the comparative TerL structural analysis[327], in which the residues E208 (motor) and D363 (nuclease activity) were mutated to Ala, were likewise tested (Figure 3.10).

A detoxified derivative of phage ϕSLT, containing a tetracycline resistance marker (*tet*M) inserted into the Panton-Valentine leukocidin (PVL) locus, was used to simplify the transfer studies. With this tetracycline resistance marker phage, and scrutinising for transductants instead of plaques, any single packaged molecule would be detected in the RN4220 recipient strain. Thus, to analyse the *cos*-site cleavage reaction of the respective mutants, intracellular DNA obtained 90 min after SOS induction of the different lysogenic strains was digested with *Xho*I/*Shp*I, which cut the phage genome flanking the *cos* site (see scheme in Figure 3.11A). Southern blotting was performed using *rin*A and *ter*S-specific probes which flank the *cos* site and the induced cultures were tested for phage and SaPI titres. To exclude possible *cos*-site duplexes and due to the high GC content of the sequence, the samples were treated under conditions ensuring *cos*-site melting.

After phage induction, phage concatemers are generated, and *cos*-site cleavage is performed. When cleavage occurs at the *cos* site, it is predicted that the *rin*A probe hybridises with a 2.15 kb fragment and the *ter*S probe with a 3.47 kb fragment, as shown in Figure 3.11A. Whereas, in the absence of *cos*-site cleavage, a 5.62 kb fragment is expected. The appearance of the small fragment, produced by cleavage at the *cos* site, was observed in the WT phage, but not for the derivatives with mutations in the *ter*L, *ter*S or *hnh* genes (Figure 3.11B). Whilst cleavage was partial in the wt sample, a faint band corresponding to processed *cos* sites was observed in the deletion mutants, implying that some cleavage occurs in the absence of any of the

three proteins. However, this processing was not enough to generate detectable packaging in any of the phage mutants (Table 3.3). Complementation of the different mutants restored the phage titre partially, confirming that the observed phenotypes were caused by the mutations. When complementing the single point mutants, the titre was slightly lower compared with the levels of the complemented deletion mutants. This observation suggests that proteins with single point mutations probably interact with other proteins yielding non-functional complexes, interfering with the functionality of the phage machinery (Table 3.3).

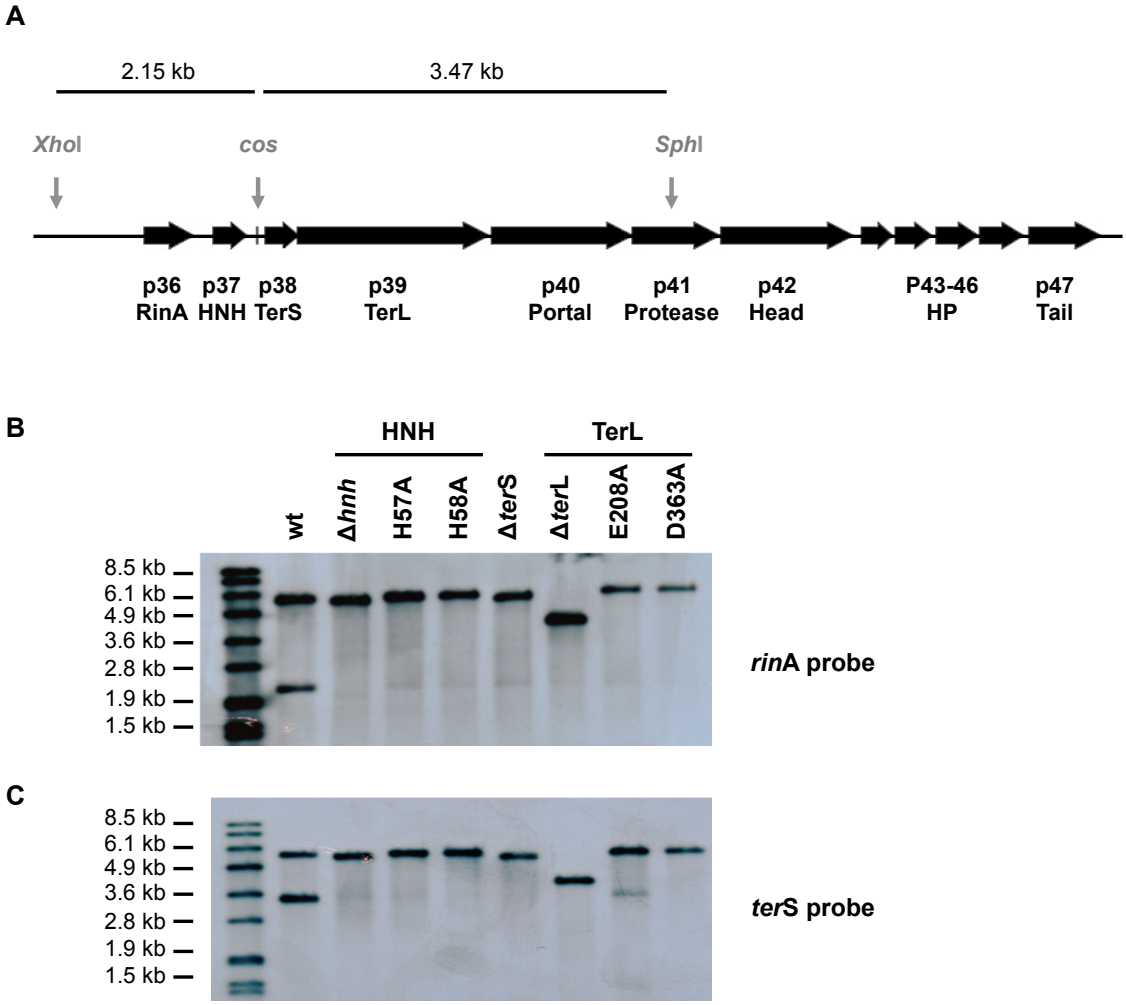


Figure 3.11 Role of the different *ter*L, *ter*S or *hnh* gene mutants in phage ϕSLT packaging. (A) ϕSLT map. The relevant genes and the *cos* site are shown. (B-C) Lysogenic strains carrying the different phage mutations were exposed to MC, then incubated in broth at 32°C. Samples were taken after 90 min and DNA was isolated from them and then digested with *Xho*I/*Sph*I. DNA was separated by agarose gel electrophoresis and transferred to a nylon membrane. Southern blots of these samples were hybridised overnight with a *rin*A (B; left to the *cos* site) or *ter*S (C; right to the *cos* site) phage-specific probes.

The central aim of this study was to clarify whether these novel HNH nuclease proteins were involved in the *cos* cleavage process. Cleavage analysis confirmed that this new family of *cos* phages and their parasite SaPIs requires the nuclease activities both from the TerL and HNH proteins. Nevertheless, the mutant in TerS also indicates that this protein is likewise essential for *cos*-site processing and its function might be determining the cleavage site. It was proposed that a cooperative nuclease complex with HNH and TerL would be requisite for *cos*-site cleavage and ultimately packaging of the dsDNA, and that TerS protein would be involved in the recognition of and binding to the *cos* sites. However, that non-specific nuclease activity could not be detected for the overexpressed $TerL_{\phi SLT}$ protein in *E. coli* (Figure 3.8). A putative model for the HNH-TerS-TerL complex is further discussed.

3.2.5 A supramolecular complex controls HNH and TerL dependent nuclease activities

It is established that the replication of the phage genome and the head assembly pathways are synchronised. The presence of a prohead is not only necessary for the packaging of the dsDNA inside the capsid, but also plays additional roles in the packaging process. In phage λ, a preformed capsid will regulate the cleavage process by coordinating *cos* cleavage and nuclease activity in order to prevent premature *cos*-site cleavage[328]. If the same mechanism is found in the HNH *cos* phages, some proteins involved in capsid formation will control HNH and TerL nuclease activities. To test this hypothesis, the putative genes involved in capsid formation were deleted in the lysogenic strain carrying phage ϕSLT (Figure 3.12A). These included the genes coding for the portal (p40), protease (p41) and major capsid (p42) proteins, in addition to the previously characterised genes encoding the HNH (p37), TerS (p38) and TerL (p39) proteins. The gene encoding the major tail protein (p47) was also deleted serving as a control for this experiment.

To analyse the *cos*-site cleavage reaction of the respective mutants, intracellular DNA obtained 90 min after SOS induction of the different lysogenic strains was digested with *Eco*RI, which cut the phage genome flanking the *cos* site (see scheme in Figure 3.12A). Southern blotting was

performed using a *rin*A specific probe which flanks the *cos* site and the induced cultures were tested for phage and SaPI titres. When cleavage occurs at the *cos* site, it is predicted that the *rin*A probe hybridises with a 2.75 kb fragment. Whereas, in the absence of *cos*-site cleavage, a 8.78 kb fragment is expected. As shown in Figure 3.12B, none of the mutants involved in capsid formation or DNA packaging was able to allow *cos*-site cleavage, while the tail mutant did, indicating that the HNH, TerS and TerL proteins are active as nucleases only once the capsid structure is formed. The complementation of these mutants using pCN51 derivative plasmids restored the abolished phage titres (Table 3.3).

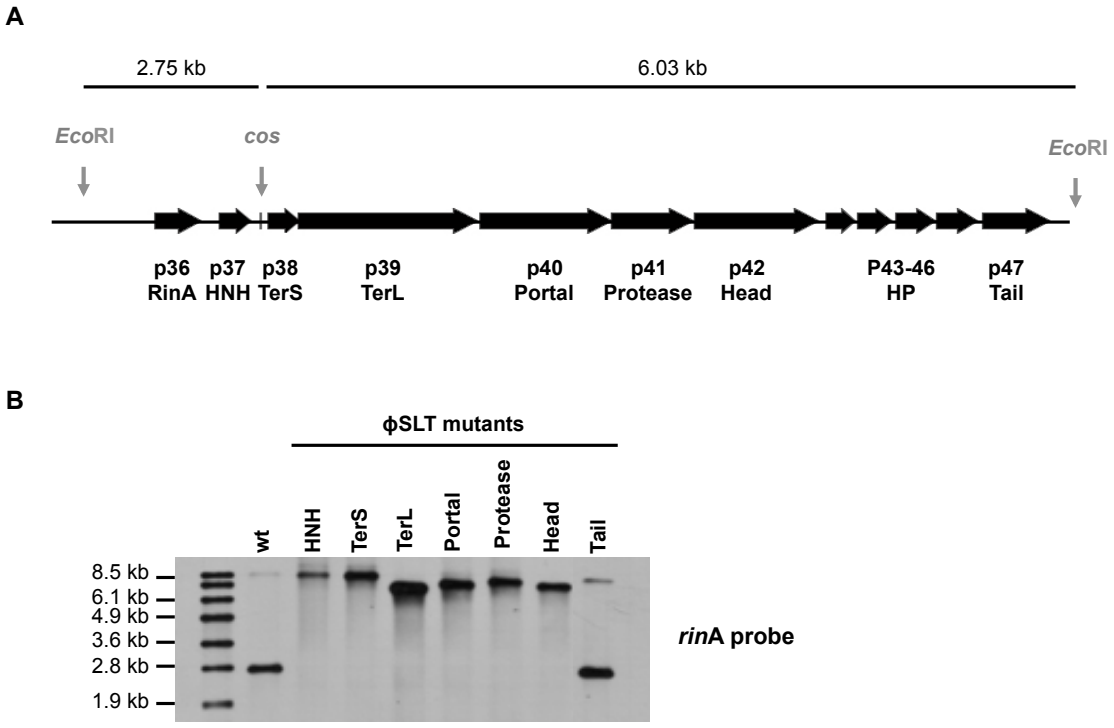


Figure 3.12 Role of the different ϕSLT mutants in phage packaging. (A) ϕSLT map. The relevant genes and the *cos* site are shown. (B) Lysogenic strains carrying the different phage mutations were exposed to MC, then incubated in broth at 32°C. Samples were taken after 90 min and DNA was isolated from them and then digested with *Eco*RI. DNA was separated by agarose gel electrophoresis and transferred to a nylon membrane. Southern blots of these samples were hybridised overnight with a *rin*A phage-specific probe.

To further confirm the predicted roles for each of the genes, electron microscopy studies of the different phage mutant lysates were performed. As expected, while the HNH, TerS, TerL and portal mutants produced capsid and tails, which were not assembled because of the absence of DNA packaging,

mutants in the protease and capsid proteins were only able to produce tails (Figure 3.13). Unexpectedly, some of the tails observed in the protease and major capsid protein mutants were extremely long suggesting that the capsid structure would also control tail length formation.

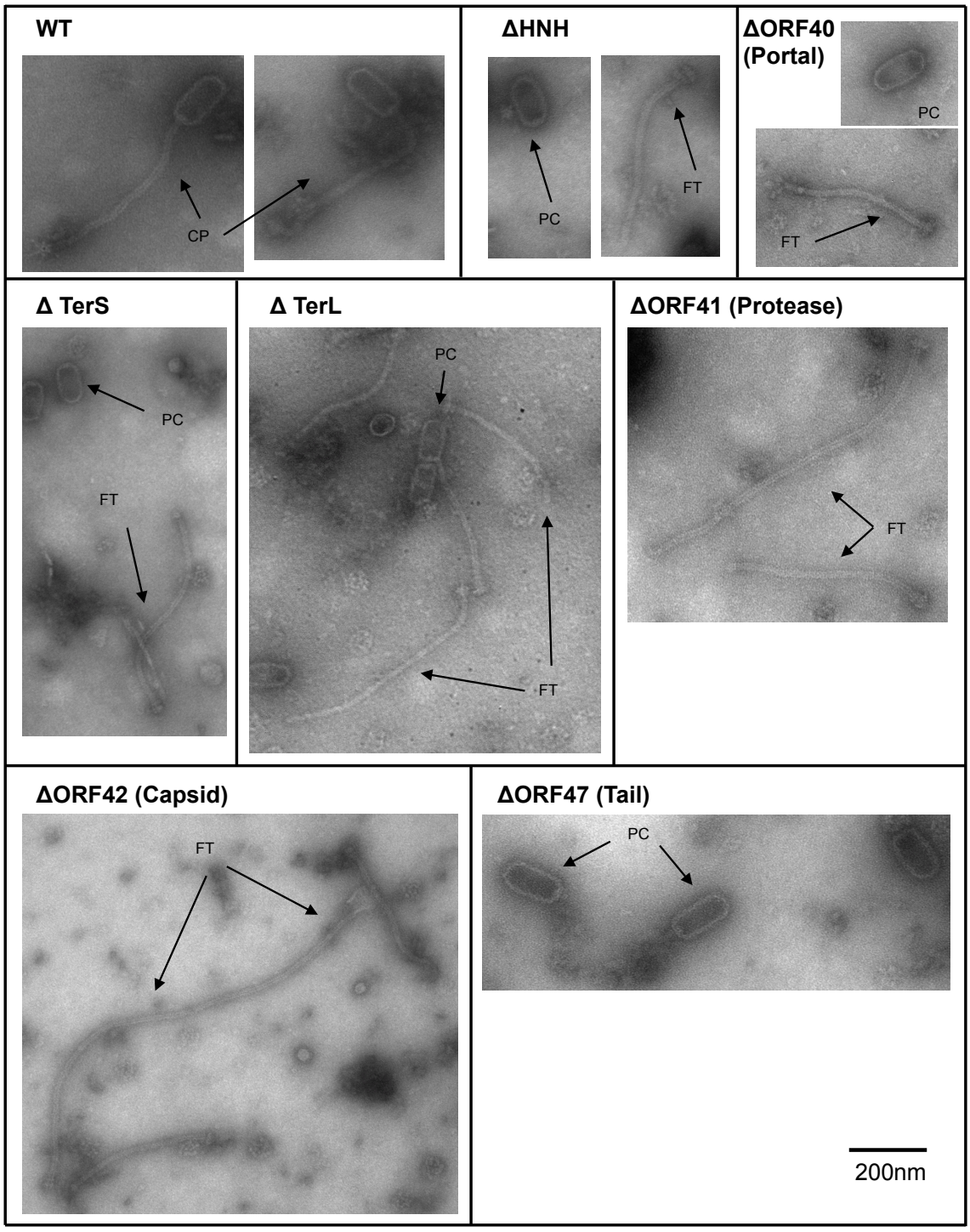


Figure 3.13 Electron micrographs of ϕSLT mutant lysates. Electron microscopic analysis of sedimented particles of wt phage ϕSLT and ϕSLT mutants. CP, complete phage ϕSLT particles; FT, free phage tails; PC, phage capsids. Scale bars are 200 nm.

3.2.6 The novel packaging system involving HNH proteins is widespread in nature

Previous publications have highlighted the significance of HNH proteins in nature, being present in all kinds of *cos* phages infecting Gram-positive or Gram-negative bacteria[193]. The genomic location of the gene encoding the HNH protein is always next to the genes encoding the TerS, TerL and portal proteins[192]. The mechanisms described here established that a major complex was needed for *cos* cleavage, involving not only the HNH, TerS and TerL but also proteins involved in capsid formation and maturation (major capsid, portal and protease). The results of this study suggest that the aforementioned proteins define the minimal module required for this novel packaging model and that this cluster will be present in all phages encoding HNH proteins.

To further support these data, the proteins involved in this module were examined to establish if they belong to the same family as those previously characterised for the *pac* and *cos* phages. It was anticipated that those proteins involved in the packaging process present in *cos* phages encoding HNH proteins will belong to the same family and will be entirely distinct from those present in the previously characterised *cos* or *pac* phages. This was the case, and as shown in Appendix 2 and Table 3.4, the different HNH-encoding phages contained the same structurally related family domains and therefore were completely different from those present in the classical λ (*cos*), and SPP1 and ϕ11 (*pac*) phages.

In addition, a putative HNH protein was identified in the *E. coli* phage ϕP27. The coliphage ϕP27 is a clinically relevant phage that encodes the Shiga toxin (Stx)[66]. A detoxified derivate of phage ϕP27 was used to further characterise the role of this HNH-phage protein and determine whether HNH-*cos*-packaging is widespread in nature. The *stx* gene of this strain was replaced with a tetracycline resistance marker (*tet*A). Furthermore, derivative mutants of the genes encoding the HNH, TerS, TerL, portal, protease and major capsid proteins were generated in the ϕP27 lysogenic strain (Figure 3.14).

Table 3.4 Pfam domains found in the proteins defining different phage packaging mechanisms[a].

Protein	HNH phages	ϕ11	SPP1	λ
HNH	PF01844	-	-	-
TerS	Terminase_4 PF05119	Terminase_2 PF03592	Terminase_2 PF03592	Phage_Nu1 PF07471
TerL	TermInase_1 PF03354	Terminase_3 PF04466	Terminase_3 PF04466	Terminase_GpA PF05876
Portal	Phage-portal PF04860	Phage_prot_Gp6 PF05133	Phage_prot_Gp6 PF05133	Phage_portal_2 PF05136
Peptidase	*Peptidase_U35* PF04586	-	-	Peptidase_S49 PF01343
Capsid	Phage_capsid PF05065	Phage_capsid PF05065	GP13 No Pfam domain	Phage_cap_E PF03864

[a]Pfam is a comprehensive collection of protein domains and families, represented as multiple sequence alignments and as profile hidden Markov models. http://pfam.sanger.ac.uk/

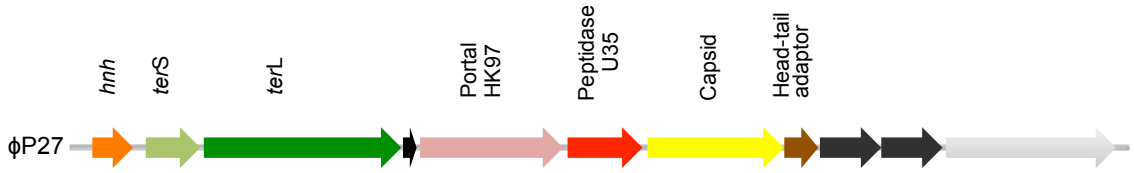


Figure 3.14 Alignments of selected genes from phage ϕP27. Genes are coloured according to their sequence and function: *hnh: orange*; *ter*S: light green; *ter*L*:* green; portal: pink; peptidase: red; capsid: yellow; head-tail adaptors: brown; hypothetical proteins: black.

Deletion of the individual genes did not affect phage replication but eliminated phage packaging and phage transfer (Figure 3.15; Table 3.5). Complementation of the different mutants using pBAD-18 derivative plasmids with an arabinose-inducible promoter (araBAD) partially restored the phage titres in all cases (Table 3.5).

Moreover, the *cos* site cleavage reaction of the respective mutants was assessed. Intracellular DNA obtained 90 min after SOS induction of the different lysogenic strains was digested with *Eco*RV, which cut the phage genome flanking the *cos* site (see scheme in Figure 3.15A). Southern blotting was performed using a *ter*S-*ter*L-specific probe which flanks the *cos* site and

the induced cultures were tested for phage titre. When cleavage occurs at the *cos* site, it is predicted that the *ter*S-*ter*L probe hybridises with a 2.4 kb fragment, as shown in Figure 3.15A. Whereas, in the absence of *cos*-site cleavage, a 4.7 kb fragment is expected. As previously established for the staphylococcal phages, *cos* cleavage *in vivo* was detected only after induction of the wt and tail mutant phages, but not in any other of the analysed mutants (Figure 3.15B). These results confirmed that a supramolecular complex involving phage packaging and capsid morphogenesis controls HNH, TerS and TerL functions.

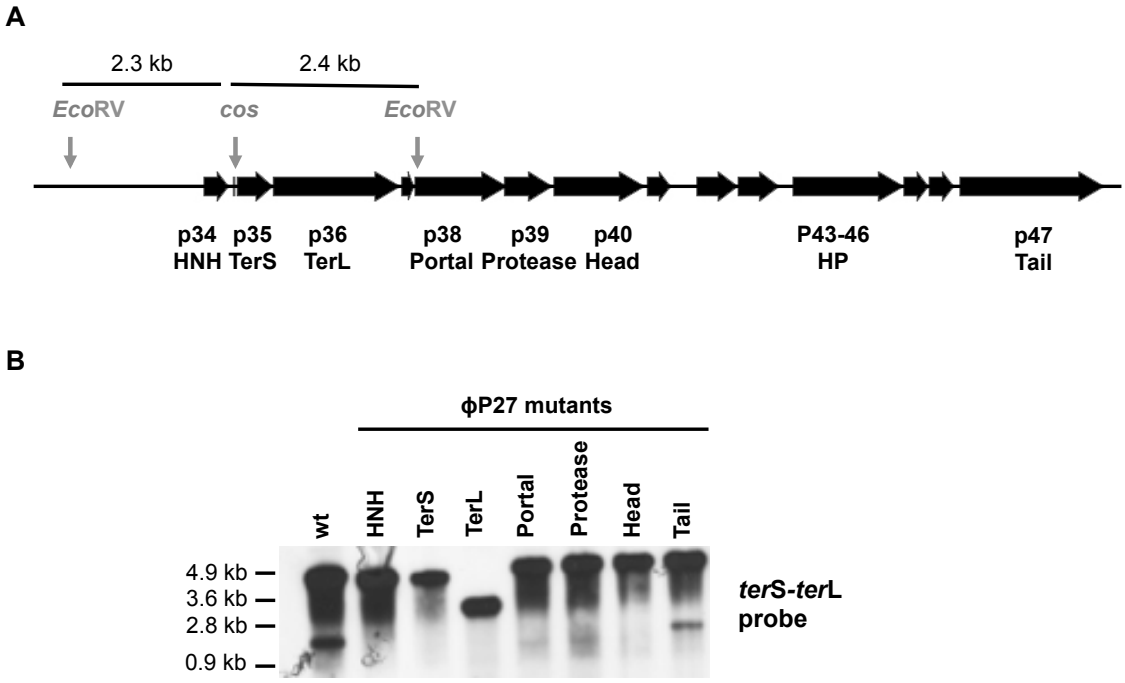


Figure 3.15 Role of the different ϕP27 mutants in phage packaging. (A) ϕP27 map. The relevant genes and the *cos* site are shown. (B) Lysogenic strains carrying the distinctive phage mutations were exposed to MC, then incubated in broth at 32°C. Samples were taken after 90 min and used to isolate DNA was isolated from them an then digested with *Eco*RV. DNA was separated by agarose gel electrophoresis and transferred to a nylon membrane. Southern blots of these samples were hybridised overnight with a *ter*S-*ter*L phage-specific probe.

Table 3.5 Effect of ϕP27 mutations on phage titre[a].

ϕP27	Mutant in	Complemented[b]	Transductant titre[c]
Wild type			1.8×10^8
ΔORF34	*hnh*		< 10
ΔORF34	*hnh*	pBAD18-$hnh_{\phi P27}$	1.6×10^7
ΔORF35	*ter*S		< 10
ΔORF35	*ter*S	pBAD18-$terS_{\phi P27}$	1.5×10^5
ΔORF36	*ter*L		< 10
ΔORF36	*ter*L	pBAD18-$terL_{\phi P27}$	4.7×10^6
ΔORF38	Portal		< 10
ΔORF38	Portal	pBAD18-$p38_{\phi P27}$	2.8×10^3
ΔORF39	Prohead protease		< 10
ΔORF39	Prohead protease	pBAD18-$p39_{\phi P27}$	2.2×10^5
ΔORF40	Major capsid		< 10
ΔORF40	Major capsid	pBAD18-$p40_{\phi P27}$	5.5×10^6
ΔORF47	Tail protein		< 10
ΔORF47	Tail protein	pBAD18-$p47_{\phi P27}$	1.2×10^6

[a]The means of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]Complemented in donor strain.
[c]No. of phage transductants/ml induced culture, using MG1655 as recipient strain.

Finally, GenBank database sequence analyses were performed in order to localise phages with an organisation similar to those analysed here. As expected, this novel packaging module is widespread in nature, being present in phages infecting Gram-positive or Gram-negative bacteria (Figure 3.16). Furthermore, this module was found to be present in several members of the recently discovered PICI family of MGEs. As shown in Figure 3.16, these elements encode some of the genes included in the studied packaging module, but not the tail proteins. This suggests that this module not only contains all the proteins required for the packaging of the PICI elements, as demonstrated for the staphylococcal phages but also hijacks the phage proteins required for tail production. In summary, the presence of HNH proteins in different MGEs indicates that the packaging module described here is widespread in nature.

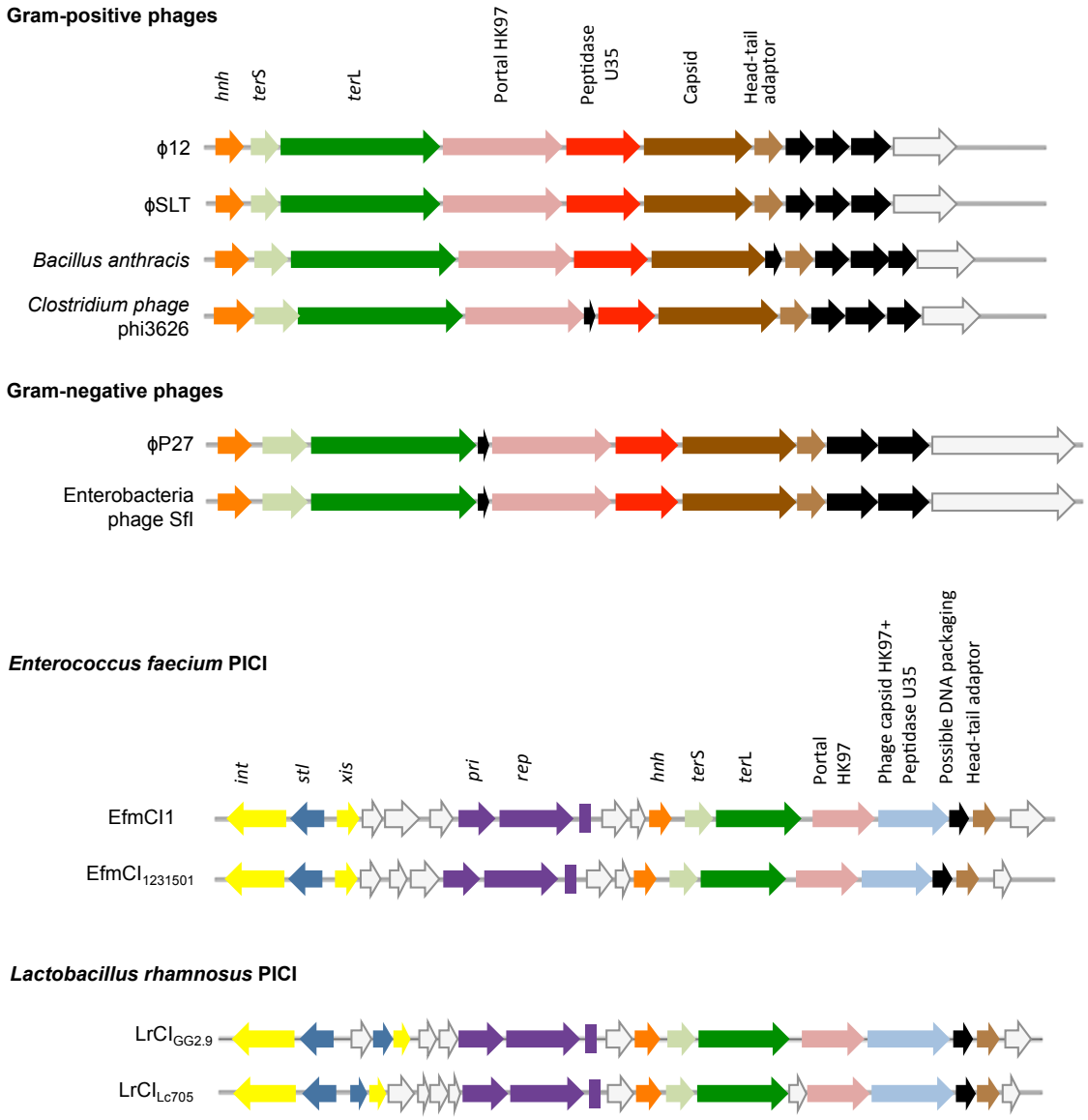


Figure 3.16 Comparison of a subset of phage and PICI genomes encoding HNH proteins. Genes are coloured according to their sequence and function: *int* and *xis*, yellow; transcription regulators, blue; replication genes and replication origin, purple; *hnh*: orange; *ter*S: light green; *ter*L: green; portal: pink; peptidase: red; capsid: brown; head-tail adaptors: light brown; capsid+peptidase: light blue; hypothetical proteins: black.

3.3 Discussion

The *S. aureus* phage-inducible pathogenicity islands are one of the best-studied models of parasitism. SaPIs exploit and manipulate different processes of their helper phages to enable their survival. Proteins involved in these piracy mechanisms are crucial to SaPI reproduction, packaging in phage-like particles and mobilisation to other bacterial hosts. Most of the characterised SaPIs and their helper phages use the *pac* 'headful mechanism' to package

their dsDNA. Due to its nature, this mechanism requires the presence of only a single *pac* sequence to start the loading of the capsid. No specific termination signal is required to end the process and the second cut is performed when the capsid is fully loaded. Due to these features, *pac* phages and SaPIs are seen as the most likely candidates for horizontal gene transfer of MGEs.

As previously highlighted, SaPIs perform their piracy by encoding different proteins; one example is the self-encoded $TerS_{SP}$, a homolog of the phage TerS. This protein will form a complex with the phage-encoded TerL to enable packaging of the SaPI dsDNA in infectious particles composed of phage proteins[254,256]. Additionally, there are other proteins involved in the SaPI phage-interference mechanism, such as those encoded in the morphogenesis module (*cpm*AB). These CpmAB proteins modulate the capsid assembling process into small particles, where only the SaPI genome will be packaged[256]. However, although this packaging mechanism is well established in the vast majority of *pac* SaPIs, several islands do not encode a TerS or CpmAB homolog[238]. SaPIbov5 is one of the SaPIs without a standard morphogenetic module. Therefore, it was often considered to be defective. In this chapter, the mechanism by which this non-classical SaPI is packaged was investigated, including its ability to be induced, packaged and transferred at a high frequency by *pac* and *cos* helper phages. Hence, this element is no longer defective and by association neither will be other SaPIs related to SaPIbov5.

This novel SaPI subfamily, for which SaPIbov5 is the prototype, uses two mechanisms to be mobilised, both of them different from the ones described for the classical *pac* SaPIs. Depending on the phage inductor, *pac* or *cos*, this SaPI can be packaged in full-sized phage particles by using the phage encoded TerS. However, contrary to the results obtained in this chapter concerning the SaPIbov5 packaging process, this element is in fact packaged both in large and small capsids by *cos* phages. The initial observation of SaPIbov5 being packaged exclusively in large capsids was wrong, and this question will be addressed in Chapter 5 when the mechanism by which SaPIbov5 is packaged will be identified. Nevertheless, the data obtained in this study establish that SaPIbov5 is the first element described being mobilised by both *pac* and *cos*

phages using two unrelated packaging sequences, which increases the transfer rate in different bacterial scenarios.

This parasitism is reminiscent of a well-known strategy used by P4 plasmid, seizing upon coliphage P2 to be encapsidated[88]. P4 uses the *cos* packaging mechanism of its helper phage by carrying the ϕP2 *cos* site sequence and enabling packaging by the P2 phage terminase complex. The P4/P2 molecular piracy resembles the one described here, as the analysed *cos* SaPI relies on the phage not only for structural viability but also to redirect the packaging process to its benefit.

Furthermore, a phylogenetic tree was generated using MEGA7 to assess the evolutionary relationships between *pac* SaPIs and cos SaPIs other SaPIs. This study revealed that *cos* and *pac* SaPIs form two distinctive clusters (Figure 3.17). Interestingly, both branches diverge, possibly from a common ancestor, by acquiring either the operon I (*pac* SaPIs) or the operon I-like (*cos* SaPIs) morphogenetic clusters (see Chapter 5), which at the end define their dsDNA packaging mechanisms. Both, *cos* and *pac* SaPIs, have produced a myriad of SaPIs variants that differ in toxin-encoded carriage, integrase type and replicase modules, hypothetically generated by several recombinational inter-exchanges events.

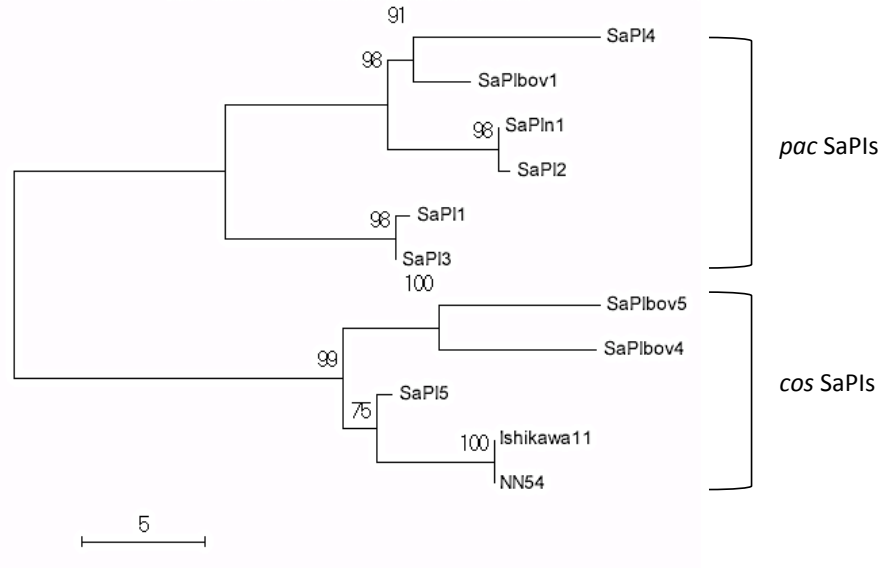


Figure 3.17 Bootstrapped Neighbour-Joining tree of different SaPIs. The tree was generated using MEGA7 version with MUSCLE sequence alignment.

Another interesting result was found upon analysing the packaging process of SaPIbov5 by its *cos* helper phages ϕ12 and ϕSLT. After examination of the packaging module of these *cos* phages, a protein with the HNH endonuclease domain was found. Remarkably, only *cos* phages encode this HNH nuclease, being present in phages infecting Gram-negative as well as Gram-positive bacteria[192,193]. HNH domains are a small nucleic acid binding and cleavage modules widespread in nature. HNH proteins were found to share a common position in the phage genomes, being next to the genes encoding the terminases and other morphogenetic proteins. To date, all the analysed phage HNH proteins carried endonuclease activity. Nevertheless, the biological function of these phage-encoded proteins was unknown.

HNH nucleases encoded by ϕ12 and ϕSLT were found to display non-specific nuclease activities and were essential for both phage and SaPIbov5 packaging. It was also established that HNH, TerS and TerL work as a complex in the *cos*-site cleavage reaction. Since a previous study demonstrated that some HNH family members have single-stranded nicking activity[193], the *cos*-site cleavage in our model could involve both proteins, HNH and TerL, each one of these proteins nicking one different DNA strand; since melting of the 100% GC content of the *cos* site duplex could not be produced in a spontaneous manner, HNH and TerL complex could be modulating this melting process in the *cos*-site cleavage reaction. However, different phage-encoded genes could also be playing a role in the *cos*-site melting process. In the case of phage λ, the role of the self-encoded gpFI protein is to facilitate DNA-binding to the prohead. However, gpFI has also been shown to be involved in increasing the turnover of the *cos* cleavage reaction by the terminase complex[328]. Moreover, and supporting our hypothesis, a recent study has shown that the dsDNA cleavage activity of the *cos* phage HK97 is HNH-dependent and relies on the binding of HNH_{HK97} to $TerL_{HK97}$ for correct DNA processing[194]. Besides, in the SPP1 phage a TerS-TerL complex modulates the nuclease role and viral maturation, protecting the concatemer to be degraded before packaging[329]. Since our data established that both ϕSLT DNA ends are unaffected by host nucleases, it can be hypothesised that the HNH–TerS–TerL complex could also be involved in this process.

The presence of this HNH protein in some *cos* phages, but not all of them, suggests that the role of TerL could be different in both scenarios. While in *cos* phages encoding only TerL, this protein is involved in the cutting of both strands of the dsDNA, it was suggested here that in those phages encoding HNH and TerL, both proteins could be essential to support the generation of the double-stranded cleavage. This hypothesis is sustained by the observation that 90% of occurrences associated with phage-HNH proteins involve the family of phage-TerL Terminase_1 (PF03354.10)[194]. This variation in the classical packaging mechanism, using HNH endonucleases, emphasises the idea of how MGEs evolve to adapt to different niches, leading to the appearance of diverse packaging strategies.

HGT of MGEs, including pathogenicity islands encoding virulence factors, has been typically assumed to be driven by *pac* phages. In this chapter, a new packaging mechanism is proposed in which not only *pac* but also *cos* phages play a major role in spreading MGEs and key virulence factors. Further studies have to be done with a view to deciphering the *cos* SaPI packaging and transfer mechanisms by the *cos* helper phages.

3.4 Conclusions and future work

SaPIs were the first members of the PICI family to be studied and are known to carry superantigen virulence genes within their genome that eventually increase the host genome plasticity and adaptability. In this chapter, a novel mechanism is described by which *cos* SaPIs can exploit the packaging machinery of both *pac* and *cos* phages. This new SaPI subfamily is characterised by the absence of the classical SaPI operon I, which has been substituted by a *cos* sequence. By using this mechanism, they not only parasitize their helper phages exploiting their packaging machinery, but also reduce the amount of functional phage particles that are released. Remarkably, the members of this subfamily also carry an uncharacterised phage *pac*-site in their genomes. By having both *pac* and *cos* sequences, these SaPIs assure their transfer by any class of staphylococcal phages. The next chapters will analyse the mechanism by which this new subfamily of SaPIs interferes with phage reproduction to be mobilised.

Chapter 4 Intra- and inter-generic SaPI transfer by *cos* phages

Results from this section are included in:

Chen J, **Carpena N**, Quiles-Puchalt N, Ram G, Novick RP, Penadés JR. (2014) Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by *cos* phages. *ISME Journal*. doi: 10.1038/ismej.2014.187.

Disclaimer on work performed:

Dr John Chen kindly provided some strains in this chapter, and full details are listed in Table 2.1 of Chapter 2.

The group of José R Penadés kindly provided some strains mentioned in this chapter, and full details are listed in Table 2.1 of Chapter 2.

4.1 Background

Bacteriophages and SaPIs belong to the family of MGEs. Phages are one of the most important factors driving bacterial evolution, shaping and redirecting the routes of HGT. Through interactions between them and the bacteria they parasitize, they contribute to host diversity, plasticity and virulence adaptation, which ultimately controls and drives the evolution of prokaryotic genomes[50,79].

As previously described, *pac*-type phages can transfer any virulence gene from one bacterium to another by HGT. *Pac* phages are encapsidated by the headful mechanism, initiated after the recognition and cleavage of *pac* sites in the phage genome by the terminase complex. After the procapsid is fully loaded, a second cut of the dsDNA is performed in a sequence-independent manner[188]. For *cos* phages, completion of the packaging process by the terminase complex requires a identical second *cos* site at the precise distance of a *cos* phage genome[188]. Since it is unlike that randomly two *cos* sites appear in the host genome at the exact distance, it is assumed that *cos* phages do not contribute to the transfer of MGEs.

Phages moreover act as vectors for the horizontal transfer of MGEs, such as SaPIs. SaPIs, members of the PICI family, encode virulence factors and antibiotic resistant cassettes, among other genes[89]. They parasitize their helper phages, which allows them to be transferred between host cells[242,247,256,269]. Several studies have indicated that SaPIs can be transferred both intra- and inter-generically at high frequency. *Pac* SaPIs were shown to be mobilised from *S. aureus* to other staphylococci (*S. epidermidis* and *S. xylosus*) and also to *Listeria monocytogenes*[245,250,330,331]. As described in Chapter 3, the novel subfamily of *cos* SaPIs, lacking the prototypical morphogenesis module, was able to be efficiently transferred to *S. aureus* strains by using the machinery of *pac* and *cos* phages for the encapsidation of their own DNA[270].

The aim of this study was to determine if this new subfamily of *cos* SaPIs, of which SaPIbov5 is the prototype, can be transferred intra- or inter-generically by *cos* phages. Since *cos* phages have not been previously involved in the transfer of virulence genes, this study could change the view of *cos* phages as contributors of gene transfer. As will be discussed at the end of the chapter, this could have a substantial impact on phage therapy, in which *cos* phages are generally used.

4.2 Results

4.2.1 Intra- and inter-generic SaPIbov5 transfer

It is well established that *pac* phages can transfer SaPIs to non-aureus staphylococci and the Gram-positive pathogen *Listeria monocytogenes*[245,250]. Likewise, it was hypothesised that *cos* phages would also be able to transfer *cos* SaPIs both intra- and inter-generically. SaPIbov5 was shown to be able to be transduced by *cos* phages to *S. aureus* strains (Chapter 3). In order to investigate if this island can also be transferred to species other than *S. aureus*, transduction studies were performed with a SaPIbov5 carrying a tetracycline resistance (*tet*M) marker. Antibiotic selection would detect any single packaged molecule in the different recipients' strains.

Lysogens of *cos* ϕ12 and ϕSLT phages, carrying SaPIbov5 with the tetracycline marker, were SOS induced with MC. SaPIs and phage particles are released from the bacteria after induction. Potential recipient strains were exposed to the lysate and SaPI infected host bacteria detected by tetracycline selection. In order to confirm whether *cos* phages were capable of mobilising *cos* SaPIs to other species, diverse bacteria were selected and used as recipients strains for SaPIbov5 infection and transfer studies (Table 4.1).

Table 4.1 List of bacteria selected for transduction studies.

Trans-specific coagulase-negative staphylococci	Trans-generic transduction
S. aureus JP4226	*L. monocytogenes* SK1351
S. epidermidis JP829	*L. monocytogenes* EGDe
S. epidermidis JP830	
S. xylosus C2a	

The conditions for the manipulation and analysis of phages lysates and procedures for transduction studies in coagulase-negative staphylococci species and *L. monocytogenes* strains were established by *Maiques et al. 2007* and *Chen & Novick 2009*[241,250], and performed as previously described in Chapter 2. As can be observed in Table 4.2, SaPIbov5 was transferred to *S. xylosus, S. epidermidis*, and to *L. monocytogenes* strains, at frequencies that are similar to those seen in *S. aureus* strains. These results clearly show that *cos* SaPIs were able to be transferred to other species and genera.

Furthermore, to test if the complete SaPIbov5 island was mobilised, and not just the *tet*M marker, different PCR analyses were performed. In these analyses, SaPIbov5 integration in its cognate att_C site and also the presence of different SaPIbov5 regions, including the *stl* and *vwb* genes, and the *cos* site were checked (Figure 4.1A). As indicated, a specific PCR confirmed integration of SaPIbov5 at the cognate att_C site in the new hosts' chromosomes. Amplification of the other selected regions was also confirmed in the different bacterial strains (Figure 4.1B).

To further support this data, and to demonstrate that this mechanism of transduction was *cos* site dependent, a *cos* deficient derivative of SaPIbov5 was generated and tested for its replication and transfer. As shown in Figure 4.2, this mutant had no effect on SaPI replication but completely abolished the ability of SaPIbov5 to be transferred to other strains (Table 4.2).

Table 4.2 Intra- and intergeneric SaPIbov5 transfer[a].

Donor strain			
Phage	**SaPI**	**Recipient strain**	**SaPI titre[b]**
ϕ12	SaPIbov5	*S. aureus* JP4226	8.3×10^4
		S. epidermidis JP829	2.4×10^4
		S. epidermidis JP830	4.7×10^4
		L. monocytogenes SK1351	6.6×10^3
		L. monocytogenes EGDe	2.1×10^4
		S. xylosus C2a	7.1×10^4
ϕ12	SaPIbov5 Δ*cos*	*S. aureus*	<10
		S. epidermidis JP829	<10
		S. epidermidis JP830	<10
		L. monocytogenes SK1351	<10
		L. monocytogenes EGDe	<10
		S. xylosus C2a	<10
ϕ12 Δ*ter*S	SaPIbov5	*S. aureus* JP4226	<10
		S. epidermidis JP829	<10
		S. epidermidis JP830	<10
		L. monocytogenes SK1351	<10
		L. monocytogenes EGDe	<10
		S. xylosus C2a	<10
ϕSLT	SaPIbov5	*S. aureus* JP4226	4.1×10^3
		S. epidermidis JP829	1.1×10^3
		S. epidermidis JP830	2.1×10^3
		L. monocytogenes SK1351	3.6×10^2
		L. monocytogenes EGDe	3.1×10^3
		S. xylosus C2a	4.0×10^3
ϕSLT	SaPIbov5 Δ*cos*	*S. aureus*	<10
		S. epidermidis JP829	<10
		S. epidermidis JP830	<10
		L. monocytogenes SK1351	<10
		L. monocytogenes EGDe	<10
		S. xylosus C2a	<10

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases. [b]No. of transductants/ml induced culture.

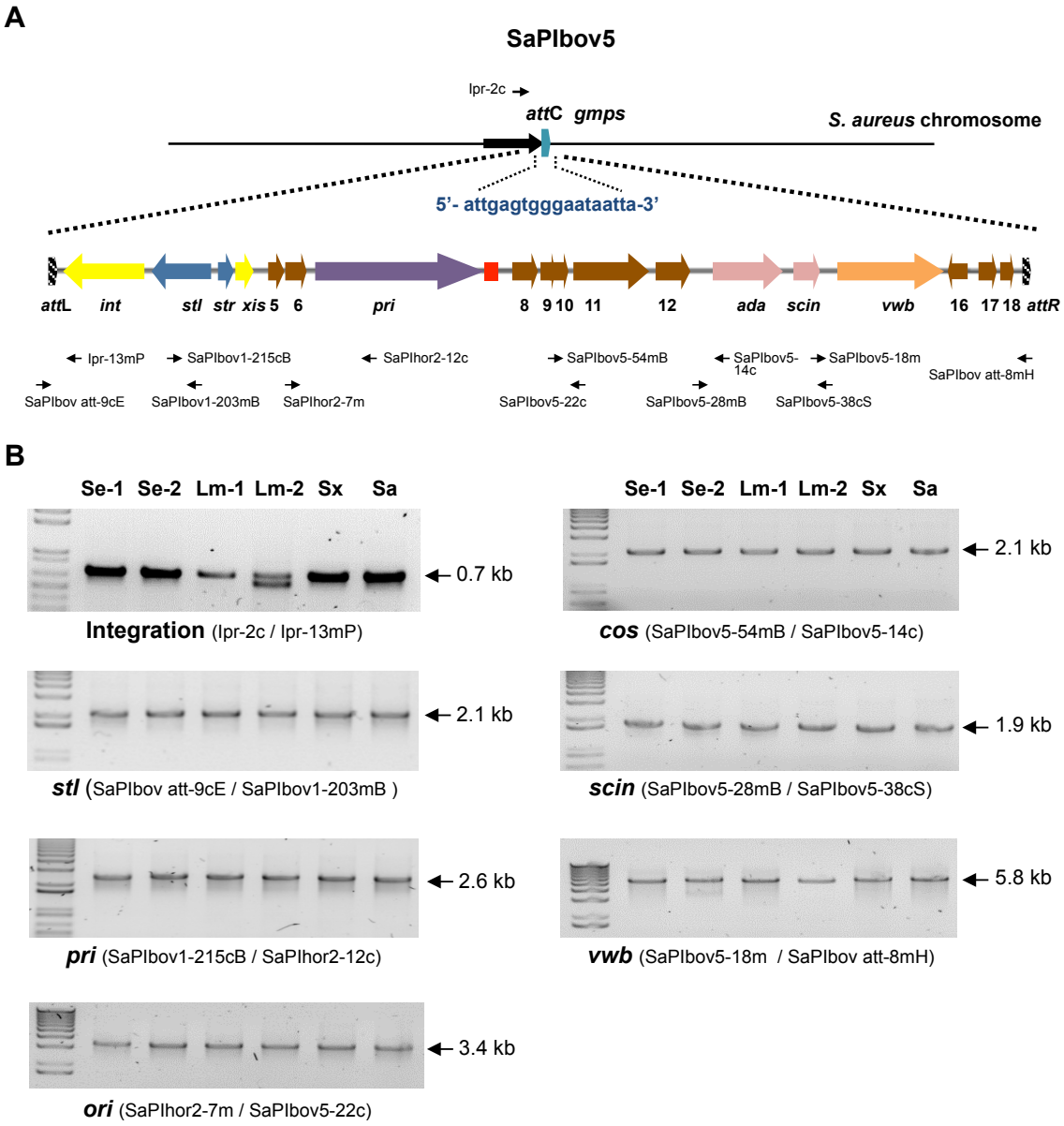


Figure 4.1 (A) Map of SaPIbov5. Arrows represent the location and orientation of ORFs greater than 50 amino acids in length. Rectangles represent the position of the *ori* (in purple) or *cos* (in red) sites. Locations of different primers described in the text are shown. **(B) Amplicons generated for detection of SaPIbov5 in the distinctive recipient strains.** SaPIbov5 was detected in six different bacterial strains: *S. epidermidis* JP829 (Se-1), *S. epidermidis* JP830 (Se-2), *L. monocytogenes* SK1351 (Lm-1), *L. monocytogenes* EGDe (Lm-2), *S. xylosus* C2a (Sx) and *S. aureus* JP4226 (Sa).

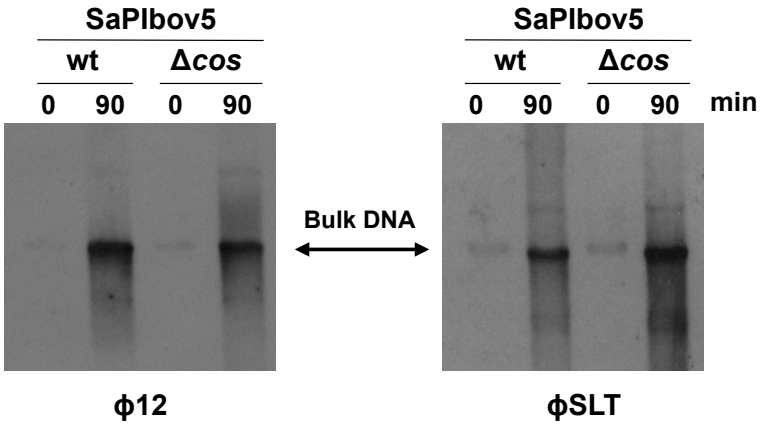


Figure 4.2 Replication analysis of SaPIbov5 Δ*cos*. Southern blot of ϕ12 and ϕSLT lysates, from strains carrying SaPIbov5 or SaPIbov5 Δ*cos*, as indicated. Samples were isolated 0 or 90 min after induction with MC, separated on agarose and blotted with a SaPIbov5-specific probe. The upper band is 'bulk' DNA, and represents replicating SaPIbov5.

There are other mechanisms by which MGEs can be mobilised. To exclude the possibility that another mechanism of gene transfer could have lead to the transfer of SaPIbov5, a ϕ12 mutant, lacking the small terminase gene (*ter*S), was generated as a control. As described in Chapter 3, the TerS protein of ϕ12 is essential for both phage and SaPIbov5 packaging but not for phage-mediated lysis[270]. Capsid packaging, and therefore HGT of SaPIbov5 would be completely blocked in the *ter*S mutant. As seen in Table 4.2, the transfer of the island was *cos* sequence and TerS dependent.

Overall these results indicate that intra- and inter-generic transfer of the SaPIbov5 island was *cos* phage-mediated. Likewise, as a consequence of this transfer, SaPI related proteins such as the integrase or proteins with toxin function in *S. aureus* strains could also be moved to non-*aureus* strains, revealing a new pool of virulence genes to other pathogen bacteria.

4.2.2 Phage silent transfer

Classically, it was believed that phage transfer only occurred within a limited host range of bacteria and that only *pac* phages are involved in the transfer of MGEs. Data from this thesis would indicate that these ideas are flawed. Another common assumption is that MGEs are just transferred within the host range of the phage they parasitize. A general approach to determine the

viability and the host range of phages, hence their capability to transfer MGEs, is the observation of cell lysis and plaque formation, which reflects their ability to reproduce in these recipient strains. It was hypothesised that this paradigm might not be accurate for all phage infection scenarios and *cos* phages could transduce MGEs into strains outside their host range, despite being unable to trigger lysis in these strains.

To verify the ability of phage ϕ12 and ϕSLT to parasitize and form plaques on *S. xylosus, S. epidermidis,* and *L. monocytogenes* strains, standard tests for phage susceptibility were performed. Interestingly, neither phage ϕ12 nor phage ϕSLT were able to lysate non-*aureus* strains (Figure 4.3), although both phages were able to mobilise SaPIbov5 to these strains. As a control, the phages were tested in their established *S. aureus* host, where they triggered the lysis of their host and plaque formation. The fact that SaPIbov5 could be transferred by both phages into strains they could not parasitize themselves led to the conclusion that determining phage host range using these common test will not necessarily identify the boundaries by which SaPIs and other MGEs are transferred by these phages[250]. Phages mobilise DNA to a wider range of species, and infection, but not plaque formation, will determine host range, which can be only evaluated by gene transfer. This mechanism, hereafters refer as 'silent transfer', eventually, could play a major role in evolution and transfer of virulence factors and resistances genes between different host-bacteria.

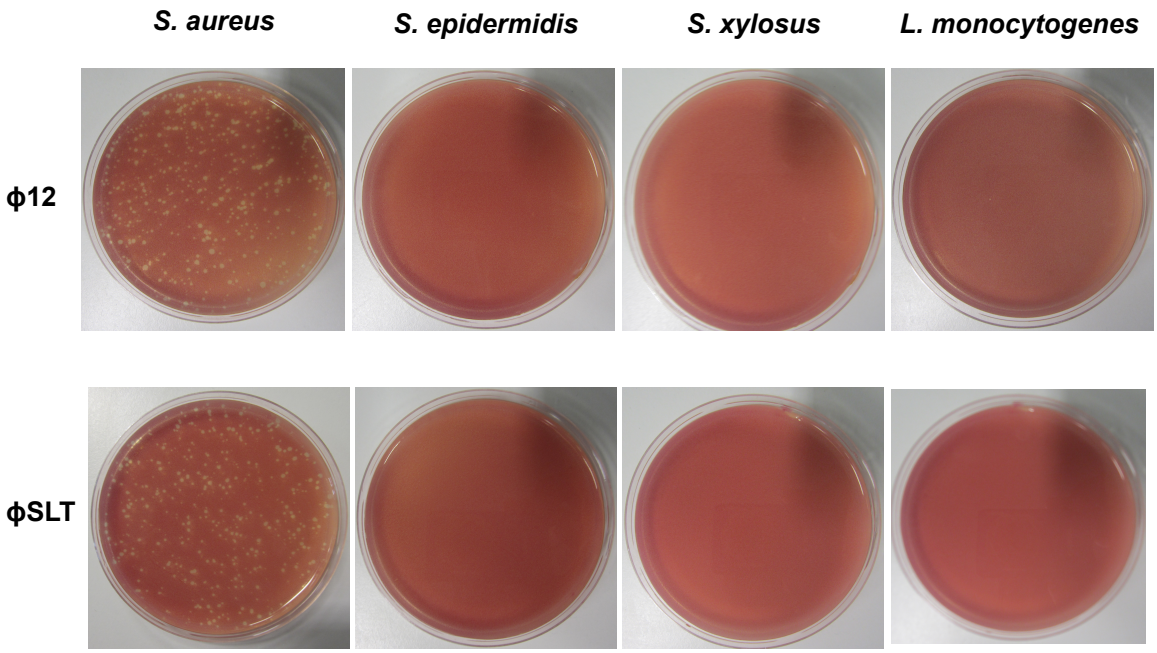


Figure 4.3 Efficiency of plating of *cos*-phages ϕ12 and ϕSLT against various bacterial species. Approximately 10^8 bacteria were infected with 400 PFU of phage ϕ12 (upper panel) or ϕSLT (lower panel), plated on phage bottom agar, and incubated 24 h at 32 °C. Plates were stained with 0.1% TTC in TSB and photographed.

4.3 Discussion

Bacterial pathogens are becoming increasingly resilient against the different strategies used to control them. One reason for this is the fact that pathogenic strains easily interchange virulence factors to enhance competitive advantage of the strains in different environmental scenarios. This genetic exchange and the spreading of antibiotic resistances are creating super potent pathogens that are harder to eradicate. It is becoming clear that the mechanisms behind gene transfer of toxins, antibiotic cassettes and others virulence factors are rather broad and diverse, and unknown strategies may be directing bacterial evolution.

The results described in Chapter 2 showed that *cos* phages, contrary to common belief, are involved in HGT of pathogenicity islands. In this chapter, further evidence was provided that such phages facilitate both intra-generic and inter-generic gene transfer. It was demonstrated that *cos* phages can transduce SaPI particles at high levels to strains they are unable to parasitize themselves. Therefore, the previous assumption that plaque formation can

determine the host range of a phage is no longer accurate, and silent transfer of pathogenicity factors by *cos* phages has to be recognised as a mechanism pathogens could benefit from, as it has been acknowledged for the *pac* phages[245,250].

For instance, species such as *L. monocytogenes,* which have yet not been described to encode superantigen genes, could benefit from the silent transfer of virulence genes by *cos* SaPIs. *Cos* SaPIs carry genes such as the von Willebrand factor-binding protein (vWbp), staphylococcal complement inhibitor (SCIN), enterotoxin L (SEL) or enterotoxin P (SEP) among other superantigens[13,238,263,332]. In lights of the results obtained in this chapter, and in a scenario of co-evolution, *cos* SaPIs could horizontally move these staphylococcal-encoded toxins to *L. monocytogenes*. This gene transfer contribution of *S. aureus* isolates carrying *cos* SaPIs to other unrelated bacteria will be limited by the role that those toxins could have the new host cell.

One possible feature allowing this silent transfer might be the teichoic acids found in the bacterial cell wall. Teichoic acids act as surface ligands mediating the initial attachment of bacteriophages to the host cell[331]. Various genera of microorganisms were found to form teichoic acids of similar structure, possibly allowing bacteriophages to attach and inject their genome into a wide range of bacteria, independent of the range of bacteria they can actually replicate within[331]. Several scenarios could determine the inability of a certain phage to parasitize a given host bacteria. These include adsorption resistance, which reduces interaction between phage and host bacteria due to, as an example, of a different composition of wall structure; abortive infections, where bacteria and phage die; DNA restriction mechanisms, including CRISPR and restriction modifications; to failure to lysate the bacterial cell wall[333,334].

Phage therapy has been used as an alternative therapy to conventional drugs in the treatment of several bacterial infections. It is believed to be an innocuous strategy that is effective against multi-resistant drugs isolates[335–339]. Lytic phages from the *Myoviridae, Podoviridae* family or *Siphoviridae* with low

levels of transduction of foreign DNA, such as *cos* phages, have been proposed as phage therapy for *S. aureus*. The specificity of these phages would limit their impact on the human microbiota whilst not contributing to the spreading of virulence factors[340-342]. The findings reported here suggest that the actual risk of using lytic phages may be more significant than originally considered. In particular, data from this thesis would suggest that old paradigms such as the limitation of phage-host ranges, or the phage ability to promote HGT by *cos* phages are not correct. The better understanding of phage gene content and function, and phage interaction with both host cell and other MGEs are crucial for addressing safer and efficient phage therapy applications.

As previously highlighted, several studies have already characterised the mechanisms of *pac* phage-mediated transfer of SaPIs, not only among *S. aureus* but also between other bacterial species[245,250,343]. The results of this study propose that *cos* phages are also able to mobilise *cos* SaPIs and refute the idea that *cos* phages are harmless, suggesting that they could play a role in the horizontal transfer of other MGEs. Accordingly, *cos* phages are predicted to contribute to the generation of novel bacterial lineage by spreading MGEs carrying virulence genes, resistance cassettes and other important key features.

4.4 Conclusions and future work

The *pac*-type phages have previously been thought to be the sole participant in the transference of virulence genes, including the transfer of MGEs, not only among closely related bacteria but also to other genera. The results of this chapter revealed a so far not described contribution of *cos* phages to bacterial gene transfer in between different species. SaPIbov5, the prototype member of the *cos* SaPI family with a non-$TerS_{SP}$ morphogenetic module, is moved by *cos* phages to non-*aureus* staphylococcal species and furthermore to different genera. This is the first study where *cos* phages have been shown to be involved in intra- and inter-generic transfer of SaPIs elements. Further studies will address the mechanisms involved in this *cos* SaPI transfer by the helper *cos* phages.

Chapter 5 SaPIbov5-mediated phage interference via small capsid production

Results from this section are included in:

Carpena N, Manning KA, Dokland T, Marina A, Penadés JR. (2016) A remarkable example of convergent evolution involving pathogenicity islands in helper *cos* phages interference. *Phil. Trans. R. Soc. B.* doi: 10.1098/rstb.2015.0505.

Disclaimer on work performed:

Structural modelling data in this chapter was performed in conjunction with and under the guidance of Dr Alberto Marina.

Keith A. Manning and Professor Terje Dokland kindly provided the electron microscopy images of ϕ12 and SaPIbov5 particles for this chapter.

The group of José R Penadés kindly provided some strains mentioned in this chapter, and full details are shown in Table 2.1 of Chapter 2.

5.1 Background

SaPIs are closely related to some helper phages whose life they cycle parasitize[344], driving the evolution of both phages and their bacterial host[273]. SaPIs are highly evolved parasites that have developed advanced mechanisms to manipulate the phage biology to package their dsDNA. To date, most of the described mechanisms by which SaPIs interfere with their helper phages are connected to morphogenesis functions. These strategies allow SaPIs to be highly transferred, both intra- and inter-generically[245,250,330,331].

As described previously for phages, SaPIs can also be packaged using two different strategies, *cos* and *pac*, with *pac* SaPIs being better characterised. In these SaPIs, a self-encoded small terminase ($TerS_{SP}$) facilitates the recognition of the unique SaPI *pac* site, promoting the packaging of the SaPI genomes into phage-like capsids[260,261]. This process is further helped by the SaPI-encoded Ppi protein, which blocks the function of phage TerS protein[246,255], impairing phage packaging. Other SaPI proteins are further involved in the reshaping process of the capsid, redirecting the packaging and assembly mechanism to exploit the phage machinery[110,247]. These genes are encoded in a conserved gene cluster, designated operon I in *pac* SaPIs, the expression of which is controlled by the bacterial repressor LexA[256]. Operon I includes the *cpm*A and *cpm*B genes, involved in the capsid size redirection mechanism; *pti*A, *pti*M and *pti*B genes, the products of which modulate interference with the helper phages by interacting with the late gene transcriptional regulator LtrC; and the aforementioned *ter*S_{SP}, which requires to be expressed for the specific and preferred packaging of the SaPIs[246,256,258,259].

As previously described in Chapter 3, SaPIbov5 was characterised as the prototype of the *cos* SaPI elements[18]. In these *cos* SaPIs, the classical *pac* SaPI morphogenetic operon I has been replaced by a ~90 bp fragment containing the phage *cos* site sequence (Figure 5.1). This *cos* site sequence mediates efficient packaging of SaPIbov5 after induction by a *cos* helper phage, leading to a favourable SaPI transfer, both intra- and inter-generically[270,330]. Flanking the *cos* site, these SaPIs also encode several highly conserved proteins in a

cluster that was described here as 'operon I-like'. This operon encodes proteins which are as yet uncharacterised. Thus, the function of these genes in the SaPI cycle remains unclear. It was hypothesised that these unknown genes could be involved in the mechanisms leading to phage interference and SaPI packaging[270]. The central aim of this study was to characterise the mechanisms behind the packaging of SaPIbov5 by *cos* helper phages and to determine the roles that these uncharacterised SaPI-encoded proteins have in the packaging and transfer of these elements.

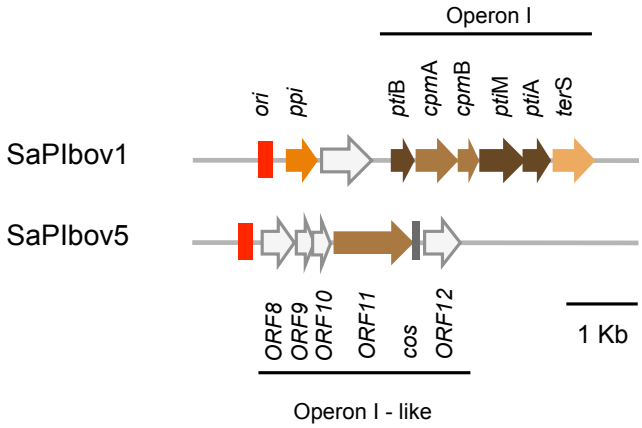


Figure 5.1 Comparison of the different SaPI morphogenetic operons. Gene colour code: replication origin, red; genes affecting expression (*pti*) or assembly (*cpm*) of helper phage virion components, dark brown and medium brown, respectively; the terminase small subunit gene (*ter*S), light brown; *ppi* (phage interference), orange. Genes encoding hypothetical proteins, white. The *cos* site is shown in grey.

5.2 Results

5.2.1 SaPIbov5 is packaged in small capsids

Sequence analysis of SaPIbov5 revealed a putative *cos* site identical to the one defined in several *cos* phages[20]. Thus, when analysing the transfer of the SaPIbov5 island using distinctive *cos* phages, ϕSLT or ϕ12, it was observed that the ability of SaPIbov5 to be transduced changed depending on which phage was used, whereas the SaPIbov5 replication rate was similar after induction by these phages (Figure 5.2; Table 5.1). Analysis of the *cos* site sequence of the island revealed that it was identical to that present in ϕ12, but slightly different to that present in ϕSLT (Figure 5.3). It was supposed that the differences in the *cos* sequences were responsible for the disparities in the

SaPIbov5 transfer efficiency. To test this hypothesis, evolutionary experiments were performed in which SaPIbov5 was exposed several times to ϕSLT. The proposed hypothesis was that the SaPIbov5 *cos* site would evolve to match that used by ϕSLT and that this genome rearrangement would ultimately increase the transfer frequency of the island.

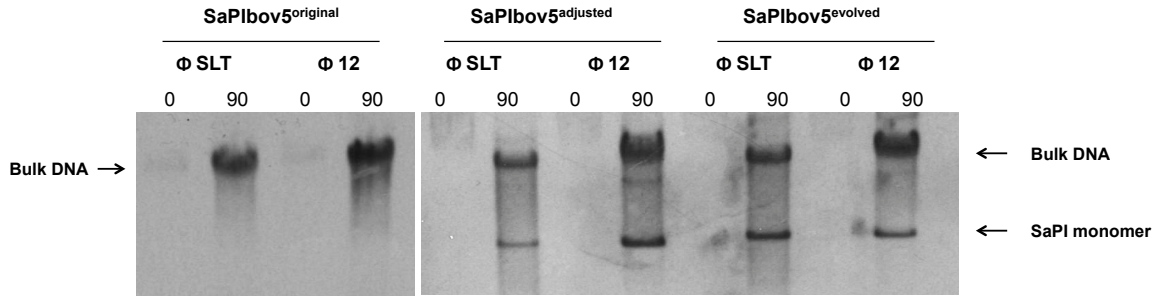


Figure 5.2 Replication analysis of the different SaPIbov5 derivative islands. Southern blot of ϕ12 and ϕSLT lysates, from strains carrying SaPIbov5^{original}, SaPIbov5^{adjusted} and SaPIbov5^{evolved} as indicated (see text for details). Samples were isolated 0 or 90 min after induction with MC, separated on agarose gels and blotted with a SaPIbov5-specific probe. The upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids.

```
phiSLT          tatgccccccctgcccatcggcttaaaatgtttttttcgccgggtaccggagaggccaaaac
phi12           tatagcccccctacccatcggcttaaaatgtttttttcgacgggtaccggcgggggcccttc
SaPIbov5        tatagcccccctgcccattggcttaaaatgtttttttcgccgggtaccggcgggggcccttc
                ***. ******.*****.****************** ********** *.** **   *

phiSLT          gctagcaacgcggataaatttttcatgaaaggggg
phi12           gcttgcaacgcggataaactttt-atgaaaggggg
SaPIbov5        gcttgcaacgcggataaactttt-atgaaaggggg
                *** **************.**** ***********
```

Figure 5.3 The *cos* site alignment of ϕ12 and ϕSLT and SaPIbov5. The SaPIbov5 and phage *cos* sites and their flanking sequences are aligned using MUSCLE. The *cos* sites are shaded in yellow.

For this purpose, a ϕSLT lysogen strain carrying SaPIbov5 with a tetracycline marker was used for island evolution. After SOS induction by MC, the island was transferred to a ϕSLT lysogen. Subsequently, all transferred colonies with tetracycline resistance, and therefore SaPIbov5-positive, were collected and induced again. This procedure was repeated four more times. After the fifth passage, individual colonies were randomly isolated, induced and the SaPI titre was assessed and compared with the titre obtained with the original

SaPIbov5. Selected SaPIbov5 colonies were found to have transfer titres up to 10^2-fold higher compared to the wt island (Table 5.1). This result suggested that the evolved SaPIbov5 islands were being packaged more efficiently by phage ϕSLT. To further investigate whether modification of the *cos* sequence or other genetic rearrangements had increased the transfer titres, the SaPIbov5 island from three different colonies carrying the evolved islands were sequenced. No mutations in the *cos* site were detected. However, in all cases, the SaPIbov5 genome was reduced in size due to the loss of some of the SaPIs virulence genes (Figure 5.4). All three clones were identical in sequence.

The size-reduction outcome can be clarified by understanding the *cos* packaging mechanism, which SaPIbov5 deploys for its mobilisation. As previously highlighted, *cos* phages, and by extension *cos* SaPIs, require the occurrence of an identical *cos* site sequence at a defined unit length distance to initiate and finish the packaging process. This requirement limits the size of the genome that *cos* phages and SaPIs can package. The SaPI strain used for these studies was a derivative of the SaPIbov5 wild type with a *tet*M marker integrated into its genome. Hence, this tag insertion added an extra 2.9 kb to the wild type genome size, which ultimately reduced the transfer efficiency of the island. As can be seen in Figure 5.4, the evolved SaPIbov5 strain had reduced its genome in order to match with the original SaPIbov5 size, and this had increased its ability to be highly transferred by *cos* phages.

The genome reduction of the evolved SaPIbov5 suggested that *cos* SaPIs could be exploiting the same strategy deployed by *pac* SaPIs to be packaged into small capsids and, therefore, be capable of redirecting the *cos* phage capsid machinery to its benefit. Furthermore, nucleotide sequence analysis revealed that other *cos* SaPIs have the same organisation and a similar genome size (~14 kb, Figure 5.5). This size is similar to those described in previous studies for the classical *pac* SaPIs, normally around ~15 kb, which are packaged into small capsids[110,247].

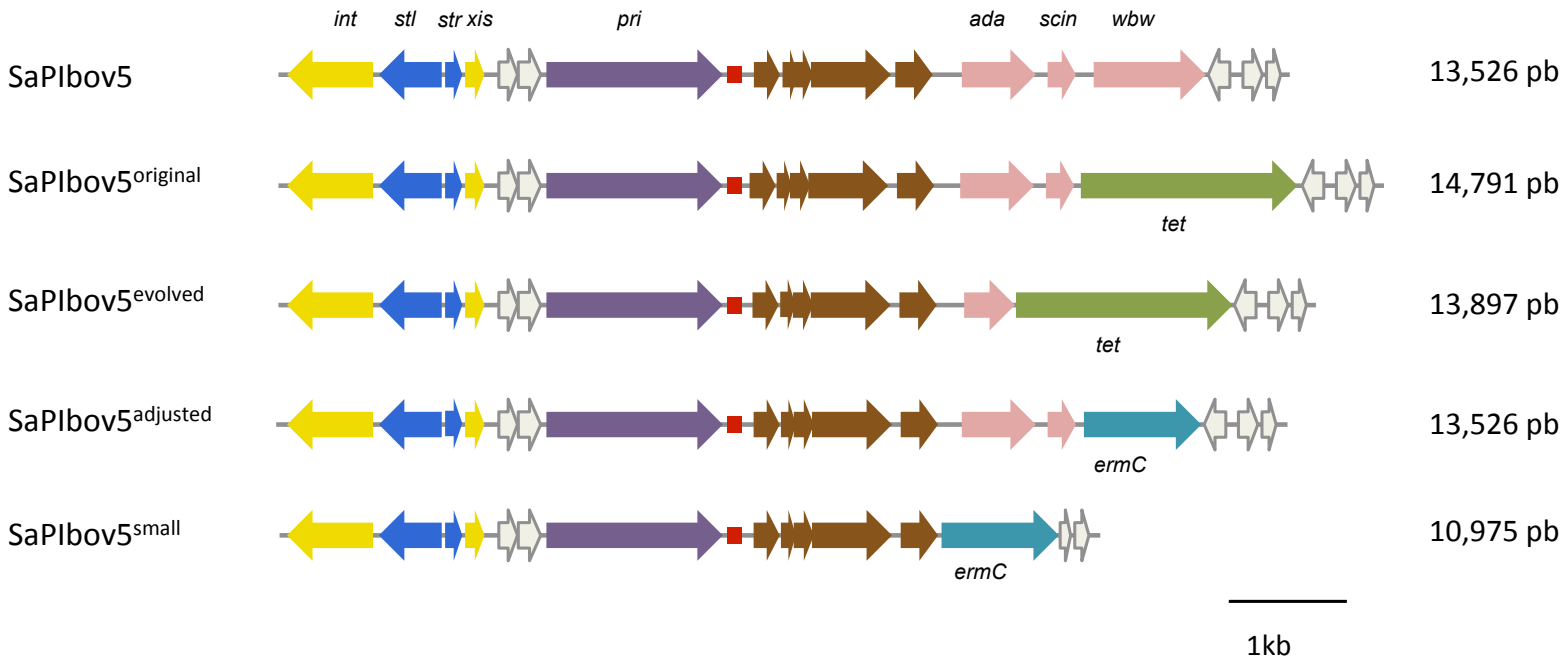


Figure 5.4 Alignment of selected SaPIbov5 size adjustment. Genomes are aligned according to the prophage convention, with the integrase gene (*int*) at the left. Genes are coloured according to their sequence and function: *int* and *xis* (excisionase), yellow; SaPI transcription regulators (*stl, str*), blue; the replication genes primase (*pri*) and replication initiator (*rep*), purple; the replication origin (*ori*), red; genes of the new morphogenetic operon, brown; superantigen and other virulence genes, pink. Genes encoding hypothetical proteins, white. The tetracycline cassette gene, green; the erythromycin cassette gene, turquoise.

Table 5.1 Effect of SaPIbov5 size on phage and SaPI titres[a].

Donor strain			
Phage	**SaPI**	**Phage titre[b]**	**SaPI titre[c]**
ϕ12	-	1.9×10^{10}	-
ϕ12	SaPIbov5^{original}	2.02×10^{5}	3.09×10^{5}
ϕ12	SaPIbov5^{evolved}	1.94×10^{6}	6.8×10^{6}
ϕ12	SaPIbov5^{adjusted}	3.3×10^{6}	1.33×10^{7}*
ϕ12	SaPIbov5^{small}	1.5×10^{6}	1.8×10^{6}
ϕ12	SaPIbov5 ^{adjusted} Δ*ccm*	8.35×10^{7}	3.35×10^{7}
ϕ SLT	-	6.6×10^{7}	-
ϕ SLT	SaPIbov5^{original}	8.2×10^{5}	7.5×10^{3}
ϕ SLT	SaPIbov5^{evolved}	1.03×10^{6}	1.16×10^{6}*
ϕ SLT	SaPIbov5^{adjusted}	1.1×10^{6}	2.7×10^{6}*
ϕ SLT	SaPIbov5^{small}	8.2×10^{5}	1.7×10^{5}

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]PFU/ml induced culture, using RN4220 as recipient strain
[c]No. of transductants/ml induced culture, using RN4220 as recipient strain.
* Significant at $p<0.05$, using Kruskal-Wallis with Dunn's Multiple Comparison Test probabilities.

To further explore this capsid reduction scenario, the SaPIbov5 island with the *tet*M marker (SaPIbov5^{original}) and one of the evolved island (SaPIbov5^{evolved}) were used. Additionally, an adjusted SaPIbov5 was generated to address the size genome requisite (SaPIbov5^{adjusted}). In SaPIbov5^{adjusted}, a fragment of the *vwb* gene was replaced by an *erm*C marker, adjusting the island genome to its original size (see scheme in Figure 5.4). Note that the *vwb* gene encodes the von Willebrand binding protein, a virulence factor with no function in the ERP cycle of the SaPIs[238].

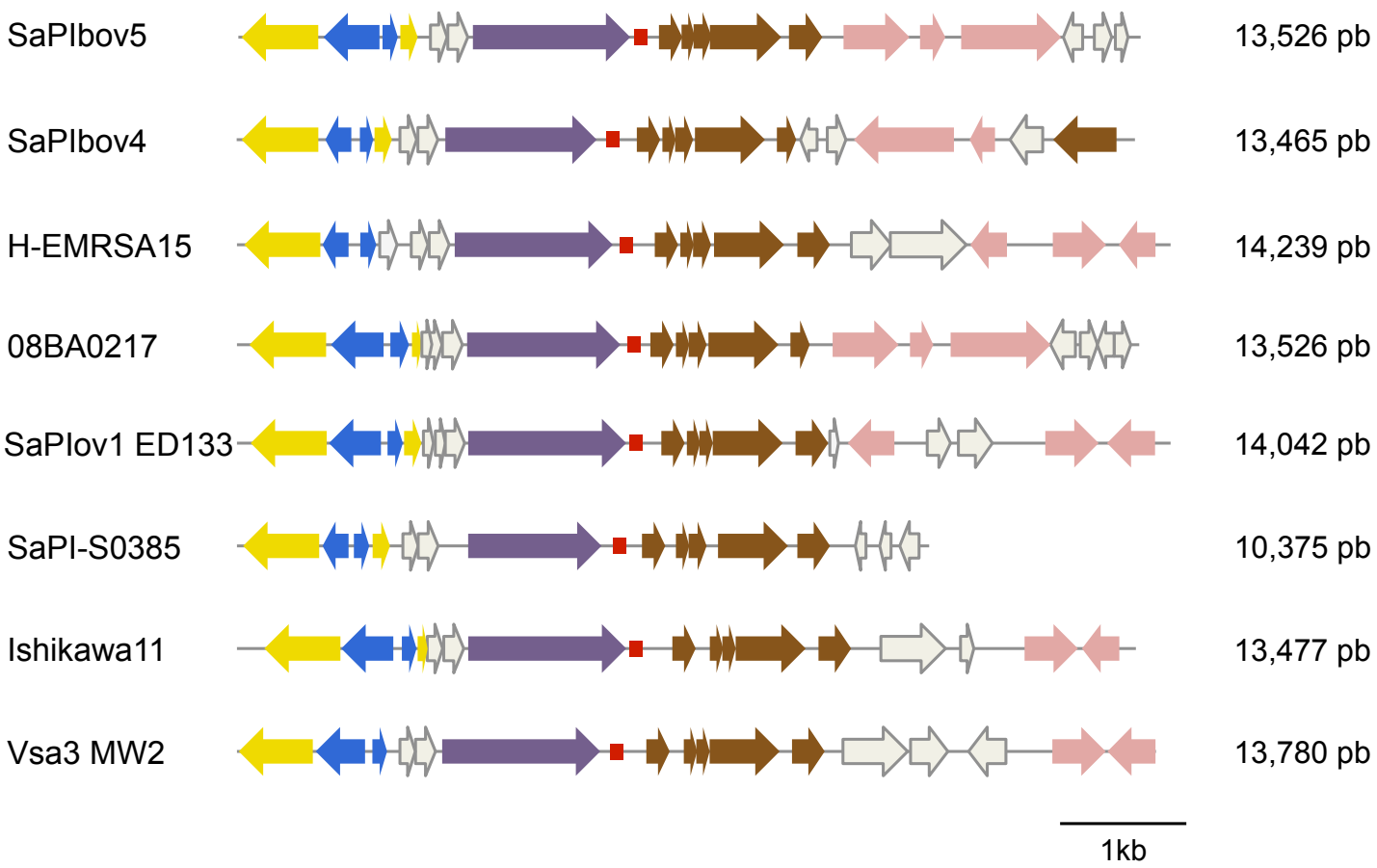


Figure 5.5 Comparison of a subset of phage-inducible chromosomal island genomes. Genomes are aligned according to the prophage convention, with the integrase gene (*int*) at the left. Genes are coloured according to their sequence and function: *int* and *xis* (excisionase), yellow; SaPI transcription regulators (*stl, str*), blue; the replication genes primase (*pri*) and replication initiator (*rep*), purple; the replication origin (*ori*), red; genes of the new morphogenetic operon, brown; superantigen and other virulence genes, pink. Genes encoding hypothetical proteins, white.

Derivatives of all these islands were introduced into lysogenic strains for *cos* phages, ϕSLT and ϕ12, and the SaPIbov5 and phage cycles were SOS-induced (MC). After induction by phages ϕ12 and ϕSLT, the evolved and size-adjusted SaPIbov5 strains generated the characteristic small sized SaPI-specific band, which indicates that SaPI monomers were packaged into small-capsid, while SaPIbov5^{original} did not (Figure 5.2). Mobilisation of SaPIbov5^{adjusted} following induction showed a high transfer ratio similar to that of the SaPIbov5^{evolved} strain, and up to ~ 10^2 times higher compared with the transfer of SaPIbov5^{original} (Table 5.1). These results confirm that a fixed SaPI size is required for a high SaPI mobilisation.

Overall these results suggested that SaPIbov5 was packaged in small capsids. To further support this hypothesis, electron microscopy (EM) images were obtained of the particles produced by ϕ12 in the presence and absence of SaPIbov5. The virion particles for ϕ12 phage had the characteristic size and shape as previously observed in Chapter 3, typical of the bacteriophages of this family[270]. As can be seen in Figure 5.6, the ϕ12 prolate head is 45 nm wide and 100 nm long and presents a 325 nm long flexuous tail. In contrast, smaller capsid particles are produced in the presence of SaPIbov5, having a diameter of about 42-45 nm isometric heads and being attached to the same ϕ12 tail (Figure 5.6). This data confirmed that SaPIbov5 forms small-size virions in order to exclusively accommodate the smaller SaPI genome.

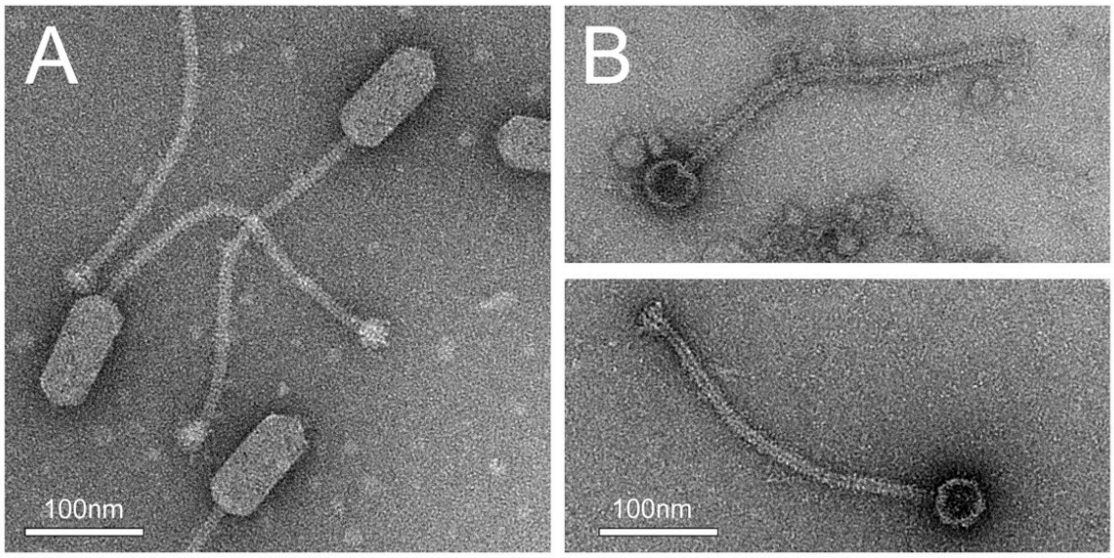


Figure 5.6 Electron microscopy of ϕ12 and SaPIbov5 particles. Electron micrographs of negatively stained wt ϕ12 virions (helper phage lysate) (A), and particles produced after induction of a ϕ12 lysogen containing SaPIbov5^{adjusted} (mixed helper phage–SaPI lysate; only SaPIbov5^{adjusted} particles are shown) (B). Scale bars are 100 nm.

As previously described, SaPI *pac* islands accomplish the production of small capsids by encoding the *cpm*AB morphogenetic genes responsible for size redirection[256,257]. Nevertheless, SaPIbov5 does not encode any of these *pac* SaPI genes. From these selective studies, it can be established that *cos* SaPIs are packaged in small capsids via an uncharacterised mechanism. Both *pac* and *cos* SaPIs share the same convergent evolutionary strategy: to be packaged in small capsids, although using distinctive mechanisms, ultimately interfering with phage reproduction.

5.2.2 Identification of the SaPIbov5-encoded Cpm-like protein

SaPI gene organisation shows synteny and is arranged in genomic clusters. Thus, the uncharacterised genome of SaPIbov5 was compared to one of the *pac* SaPIs to identify the region that could be potentially involved in the formation of the SaPI-sized small capsids. The morphogenetic module in *pac* SaPIs is always located between the SaPI *ori* site and the virulence genes. Therefore, it was hypothesised that in *cos* SaPIs the packaging genes would be placed in the same genome position than in *pac* SaPIs. The analysis of this putative region in SaPIbov5 revealed five genes of interest that include genes

from ORF 8 to 12, which could function as analogues to the ones reported for *pac* SaPIs operon I (Figure 5.1). The only gene with an established role for the protein was SaPIbov5 ORF8, which encodes Ppi. As previously described, Ppi is a protein responsible for blocking the *pac* phage terminase small subunit, ultimately interfering with phage packaging[255].

To identify the gene(s) required for the formation of the SaPIbov5 small capsids, single punctual mutants were generated in the five putative genes by introducing a stop codon (ochre mutation) in the middle of the specific coding sequence. By using this strategy, the SaPIbov5 size remained unchanged, and the packaging process would not be affected by genome length variations. The effects of these mutations were analysed for the three definable stages of the SaPI ERP cycle. The different strains were introduced into ϕ12 lysogens and induced by SOS response, lysates were analysed by gel electrophoresis, and tested for SaPI reproduction and transfer titres. It was expected that if a gene(s) was involved in capsid production, the small band representing the SaPI monomers packaged into small SaPI particles would not be present. As can be seen in Figure 5.7, only the mutant in SaPIbov5 ORF11 did not generate the characteristic SaPI band. These results suggest that this protein could be responsible for the capsid size redirection of the *cos* phage, and will be referred to as Ccm (for *cos* capsid morphogenesis) from here on.

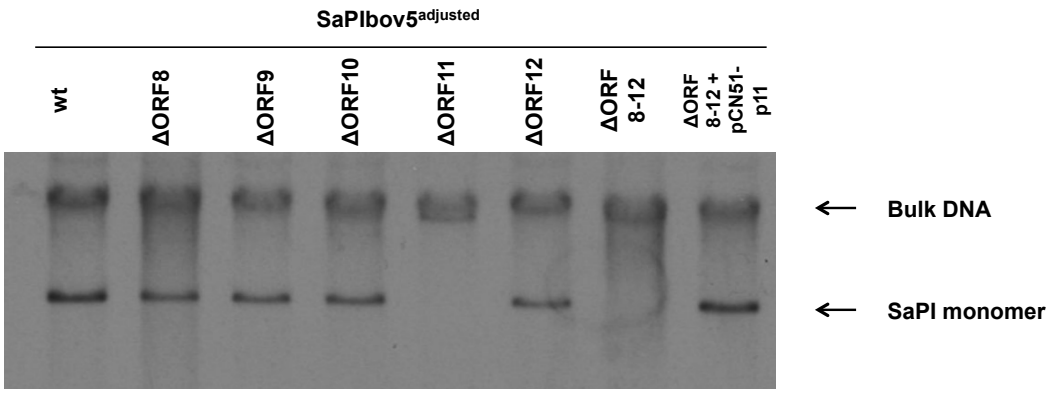


Figure 5.7 Replication analysis of SaPIbov5 mutants. Southern blot of ϕ12 lysates, from strains carrying the wt or the different SaPIbov5 mutants (carrying ochre mutations in the SaPIbov5 genes 8 to 12). Samples were isolated 0 or 90 min after induction with MC, separated on agarose and blotted with a SaPIbov5-specific probe. The upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids. SaPIbov5 ORF11 corresponds to *ccm*.

To ensure that the Ccm protein was indeed responsible for the capsid size remodelling, and to test whether other proteins could act together with Ccm in the packaging process, different gene mutations were combined. Furthermore, a mutant was generated carrying stop codons in all of the genes located in this new *ccm*-morphogenetic operon, named operon I-like. After induction by ϕ12, all the derivatives that contained the Ccm mutation did not generate the small characteristic band (Figure 5.8). To determine whether Ccm alone was able to restore the production of small capsids, a derivative of the island mutated in ORFs 8 to 12 was complemented with a pCN51 derivative plasmid expressing the *ccm* gene. Complementation with the *ccm* gene completely restored the production of small capsids in SaPIbov5 (Figure 5.7).

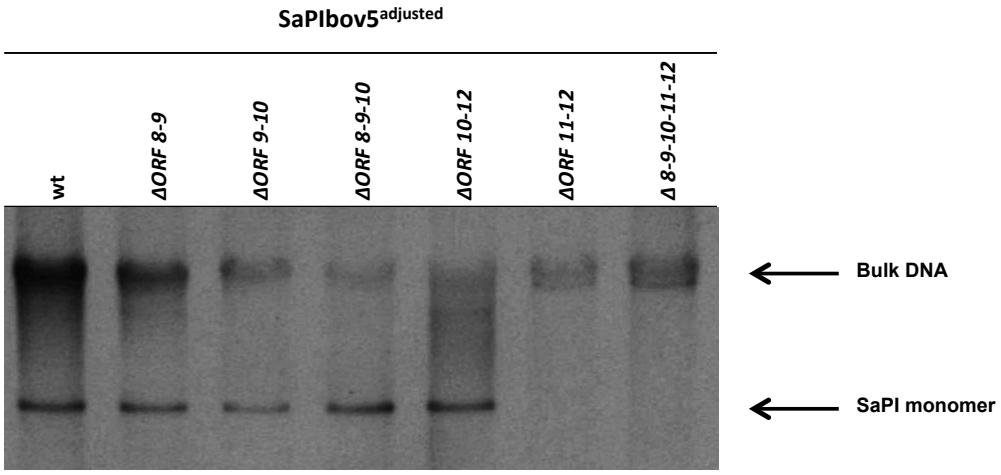


Figure 5.8 Replication analysis of double SaPIbov5 mutants. Southern blot of ϕ12 lysates from strains carrying the wt or the different combinations of SaPIbov5 mutants (carrying ochre mutations). Samples were isolated 0 or 90 min after induction with MC, separated on agarose and blotted with a SaPIbov5-specific probe. The upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids.

When analysing the transduction efficiency of the Ccm mutant, it was observed that this island was mobilised by ϕ12 at levels that were comparable to the one of the wt island (Table 5.1). This high transfer rate can be explained by the packaging of the SaPIbov5 genome into phage large-particles, at a frequency comparable to their packaging into reduced size capsids, as observed for other SaPIs[255-257]. This result highlights the possibility that the production of small capsids, mediated through Ccm, is rather a mechanism of interference with the host phage than a mechanism to maintain

high transduction rates. This hypothesis is further supported by the fact that the phage interference in the Ccm mutant was less pronounced, compered to wt SaPIbov5^{adjusted} (Table 5.1).

5.2.3 Ccm blocks ϕ12 reproduction

As previously established in Chapter 3, SaPIbov5 interferes with ϕ12 reproduction through an unknown mechanism[270]. As *ccm* mediates formation of small capsids, it was hypothesised that it could also be involved in the phage interference process. Since interference occurs during plaque formation, the ability of the *cos* ϕ12 to form plaques was tested using induction and infection assays in two different models.

In the first approach, the SaPIbov5 derivatives with single point mutations, as well as the SaPIbov5 ORFs 8-12 mutant, were used as recipient strains. These strains were first tested for their ability to block plaque formation after ϕ12 infection. ϕ12 ability to produce plaques in the presence of the SaPIbov5 was severely inhibited, with a reduction of 10^6–10^7-fold (Figure 5.9A). Interestingly, both mutants affecting the *ccm* gene, the single SaPIbov5 *ccm* mutant and the SaPIbov5 mutant in ORFs 8-12, slightly interfered with ϕ12 titre, indicating that Ccm has a relevant role in phage interference (Figure 5.9A). However, the size of the phage plaques was not completely restored, especially in the *ccm* mutant, suggesting that other proteins could be involved in the phage interference mechanism (Figure 5.10). This complementary effect was not observed when the different single mutated genes were evaluated individually (Figure 5.9A).

Along with evaluating the infection process, packaging after SOS induction of a resident prophage was assessed. The aim was to evaluate if any of the different SaPIbov5 mutants were able to prevent phage reproduction after induction. For that purpose, ϕ12 lysogen carrying the different SaPIbov5 point mutation islands were SOS-induced (MC) and the phage titre measured in RN4220, a non-lysogenic strain. As can be seen in Figure 5.9B, the number of plaques slightly increased in the Ccm and ORF12 mutant proteins, and the titre was partially restored for the mutant of all the genes (ORFs 8-12). These

results support the existence of other mechanisms being involved in the phage interference process.

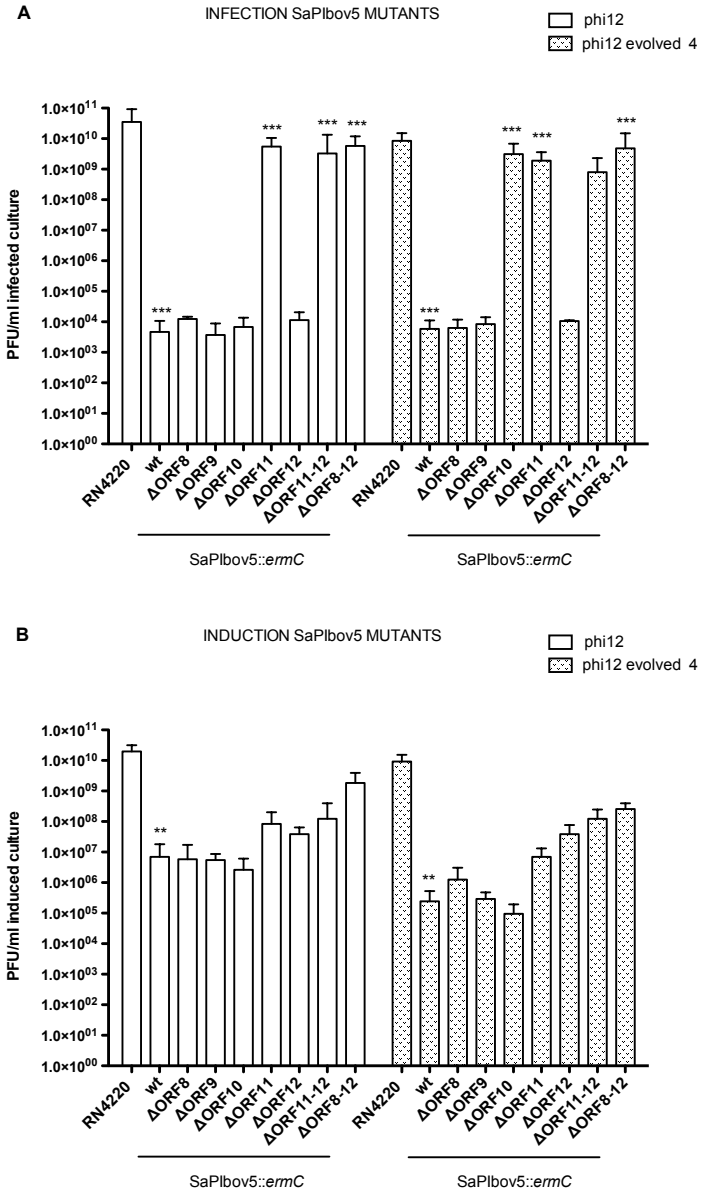


Figure 5.9 Interference of SaPIbov5 proteins ORF8 to ORF12 with ϕ12. (A) Strain RN4220 containing wt or the different SaPIbov5 mutants were infected with ϕ12 or ϕ12$^{evolved\ 4}$, plated on phage bottom agar, and incubated for 48 h at 32°C. (B) Lysogenic strains for the ϕ12 and ϕ12$^{evolved\ 4}$ with the different SaPIbov5 mutants were induced after MC and plated on phage bottom agar for 48 h at 32 °C. The mean values of results from ten independent experiments are shown. Variation was within ±5% in all cases. PFU/ml induced culture, using RN4220 as recipient strain. Statistical significance was determined by using the statistical Kruskal-Wallis test with Dunn's Multiple Comparison Test to determine significant grouping compared with the controls. RN4220 vs. SaPIbov5 ($p < 0,001$) and SaPIbov5 vs. SaPIbov5ΔORF11, SaPIbov5ΔORF11-12 or SaPIbov5ΔORF8-9-10-11-12 ($p < 0,0001$) in both ϕ12 and ϕ12$^{evolved\ 4}$ after infection. Additionally, SaPIbov5 vs. SaPIbov5ΔORF10 ($p < 0,0001$) after ϕ12^{evolved4} infection.

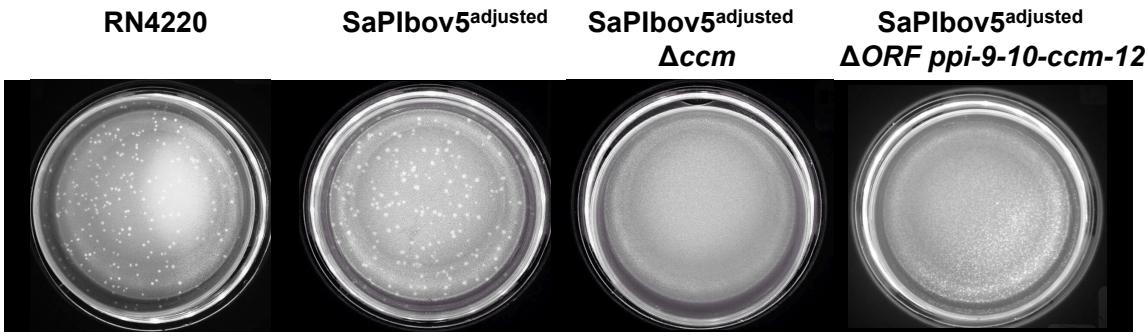


Figure 5.10 Effects of Ccm protein deletion on SaPIbov5 interference with ϕ12 (*cos* phage). Strain RN4220 alone or containing WT SaPIbov5^{adjusted}, SaPIbov5^{adjusted}Δ*ccm* or SaPIbov5^{adjusted}Δ*ppi*-9-10-*ccm*-12 was infected with ϕ12, plated on phage bottom agar and incubated for 48 h at 32 °C. Plates were stained with 0.1% TTC in TSB and photographed. Dilution of 10^{-6} was used in RN4220, 10^{-4} for SaPIbov5Δ*ccm* or SaPIbov5*Δppi-9-10-ccm*-12, and without dilution for the SaPIbov5^{adjusted} wild type.

The second strategy used was to individually express each of the single genes (SaPIbov5 ORFs 8 to 12) in the pCN51 vector[297], under the control of a cadmium-inducible promoter (*P*cad), in order to identify the roles of the operon I-like encoded genes. The same approach used in the first strategy was used to evaluate the two most relevant pathways of the phage life cycle, infection and induction. In accordance with the aforementioned results, only the Ccm protein was found to block ϕ12 reproduction both in the infection and induction processes, with a reduction of 10^3–10^4-fold in the phage titre (Figure 5.11A/B).

Furthermore, ORF12 expressed in the presence of ORF11 (*ccm*) was found to slightly reduce the Ccm interference activity (Figure 5.11C). A similarity in sequence was found between the ORF12 (42%) and the previously described protein PtiM, which modulates the expression of late phage gene transcription regulators LtrC[259]. Besides the similarity in sequence, when comparing the predicted structures of ORF12 and PtiM using the three-dimensional modelling server I-TASSER, a similar folding was found (Appendix 3). Additionally, the spore coat domain, CoatF (COG5577), was identified by BLAST protein research. COG5577 is associated with different roles such as cell wall and membrane/envelope biogenesis. Together, these comparison studies support the hypothesis of ORF12 being homologous to PtiM. Hence, the ORF12 role could involve modulation of the interference blockage by SaPIbo

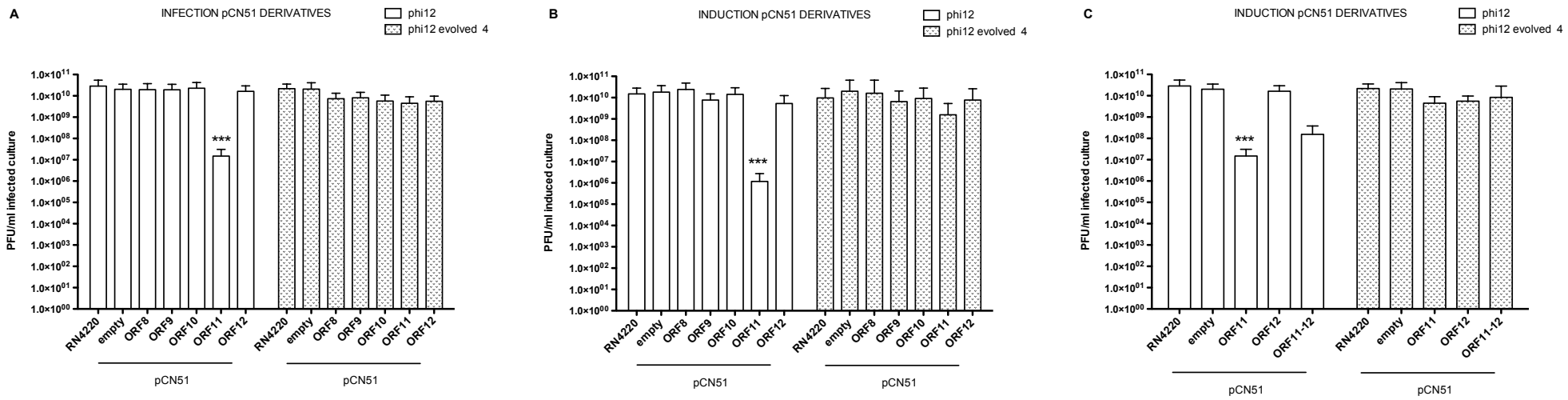


Figure 5.11 Interference of SaPIbov5 proteins ORF8 to ORF12 with ϕ12. (A) Phage interference mediated by cloned SaPIbov5 genes. The indicated genes were cloned into plasmid pCN51. Strain RN4220 containing the indicated plasmids was infected with phages 12 or $\phi12^{evolved4}$, plated on phage bottom agar containing 5 μM CdCl2 (induces the expression of the cloned genes) and incubated for 48 h at 32°C. (B) Effect of the different pCN51 cloned genes in phage reproduction. The lysogenic strains for ϕ12 or $\phi12^{evolved4}$, containing the different pCN51 derivative plasmids, were SOS-induced and the lysates plated on phage bottom agar for 48 h at 32 °C. (C) Effect of the pCN51 cloned with ORF11-12 genes in phage reproduction. The mean values of results from ten independent experiments are shown. Variation was within ±5% in all cases. PFU/ml induced culture, using RN4220 as recipient strain. Statistical significance was determined by using the statistical Kruskal-Wallis test with Dunn's Multiple Comparison Test to determine significant grouping compared with the controls. RN4220 and RN4220 + pCN51-empty vector vs. RN4220 + pCN51-ORF11 ($p < 0,0001$), in both ϕ12 and $\phi12^{evolved\ 4}$ after infection and induction.

Following the results of Ccm-ORF12 overexpression by pCN51, a *ccm*-ORF12 double mutant was tested for the occurrence of a comparable phenotype. Engaging the same strategy, phage infection and induction ability were tested to characterise this derivative strain. As can be seen in Figure 5.9B, after induction of the double mutant with the wt phage, ϕ12 blockage is less pronounced when compared to the effects of the individual *ccm* mutant. The role of ORF12 to the process of SaPI-mediated interference is still unclear and further studies will be needed to identify the function of this protein in the SaPI cycle.

5.2.4 Ccm-mediated interference

Evolutionary experiments were performed to identify the phage protein(s) interacting with the SaPI encoded Ccm and to characterise the mechanism underlying SaPIbov5 interference. A ϕ12 lysogen strain was evolved in the presence of the plasmid pCN51-*ccm*, which expresses *ccm* under control of a cadmium-inducible promoter (*Pcad*). Using this strategy, only those phages carrying mutations in the protein(s) that interact or are affected by Ccm, would become insensitive to the Ccm-mediated blockage.

Four of these evolved phages were found to be resistant to the Ccm-mediated interference and were sequenced to identify the mutations. All of the sequenced ϕ12 evolved phages were found to have mutations in gp33, corresponding to the major capsid protein component of the head ($\phi12_{gp33}$)[270]. In several of the evolved phages, but not in all, mutations were also found in the gp45 gene, which has a structural role in the phage encapsidation process (Table 5.2; Appendix 4).

Table 5.2 ϕ12 mutants insensitive to the Ccm-mediated interference.

Phage	gp33	gp45
$\phi12^{evolved1}$	G356E; T357S	S13R
$\phi12^{evolved2}$	T323P	I53S
$\phi12^{evolved3}$	E236K	E203K
$\phi12^{evolved4}$	E236K	-

These evolved phages were tested to confirm that the Ccm-mediated blockage had been overcome, and additionally, to confirm that $\phi12_{gp33}$ was the target protein for the SaPIbov5 Ccm. For those purposes, the evolved phage $\phi12^{evolved4}$ was used as it only had a single mutation affecting the capsid protein (E236K; Table 5.2; Appendix 4). A lysogenic RN4220 derivative carrying this phage was generated and the $SaPIbov5^{adjusted}$ island introduced. As can be seen in Figure 5.12 and Table 5.3, while the replication and transfer of $SaPIbov5^{adjusted}$ after SOS induction of the strains were unaffected, no SaPI band was produced, with the SaPI DNA being packaged in full-sized phage capsids.

Engaging the same strategy as for the wt phage ϕ12, phage infection and induction ability were tested to characterise the $\phi12^{evolved4}$ phage. Each of the SaPIbov5 operon I-like genes were expressed separately using pCN51, and it was found that the evolved phage was able to form plaques to the same extent as with the wt with the empty vector (Figure 5.11A/B). A minor effect on $\phi12^{evolved4}$ reproduction was observed for the SaPIbov5 encoded Ccm protein (Figure 5.11A/B). Taken together these results clearly confirm the $\phi12_{gp33}$ protein as the target of the SaPI Ccm.

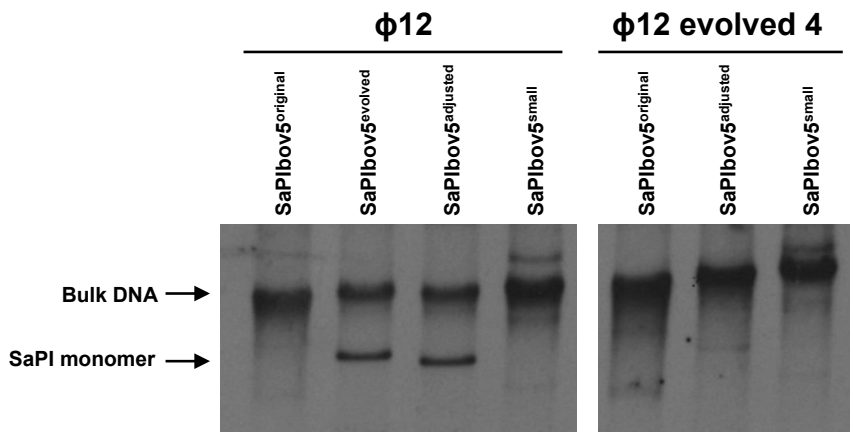


Figure 5.12 Replication analysis of the different sized SaPIbov5 islands induced by ϕ12 or ϕ12^{evolved4}. Southern blot of ϕ12 and ϕ12^{evolved4} lysates, from strains carrying SaPIbov5^{original}, SaPIbov5^{adjusted}, and SaPIbov5^{small} as indicated. Samples were isolated 0 or 90 min after induction with MC, separated on agarose and blotted with a SaPIbov5-specific probe. The upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids.

Table 5.3 Effects of ϕ12^{evolved4} in the transfer of SaPIbov5.

Phage	SaPI	Phage titre[b]	SaPI titre[c]
ϕ12^{evolved4}	-	9.24 x 10^{9}	-
ϕ12^{evolved4}	SaPIbov5^{original}	1.2 x 10^{6}	2.4 x 10^{4}
ϕ12^{evolved4}	SaPIbov5^{adjusted}	4.7 x 10^{6}	1.1 x 10^{7}*
ϕ12^{evolved4}	SaPIbov5^{small}	5.4 x 10^{6}	1.3 x 10^{6}

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]PFU/ml induced culture, using RN4220 as recipient strain
[c]No. of transductants/ml induced culture, using RN4220 as recipient strain.
* Significant at p<0.05, using Kruskal-Wallis with Dunn's Multiple Comparison Test probabilities.

To confirm that this evolved phage was insensitive to the blocking effect of the island, bacteria carrying the wt SaPIbov5 island and the mutant in ORFs 8 to 12 were infected with the ϕ12^{evolved4}. These derivatives were tested individually for plaque formation. As can be seen in Figure 5.9A, and contrary to what was expected for this evolved phage, SaPIbov5^{adjusted} still blocked phage reproduction. This result supports the hypothesis of other SaPIbov5 encoded proteins being involved in additional mechanisms of interference. In addition, and as observed previously for the wt phage, the infection ability of the evolved phage was recovered in the SaPIbov5 *ccm* mutant (Figure 5.9A).

Remarkably, it was found that $\phi12^{evolved4}$ was also able to overcome the interference blockage in the SaPIbov5 ORF10 mutant strain (Figure 5.9A), implying that this gene could be associated with a SaPIbov5-mediated interference mechanism. However, neither recognisable domains nor similar genes could be found in the GenBank database, with ORF10 only present in the identified *cos* SaPIs (Figure 5.5). Thus, the function of this protein remains to be characterised.

When analysing the effect on the evolved phage after SOS induction, a blockage of phage production could still be seen after induction in the SaPIbov5 wild-type strain (Figure 5.9B). Furthermore, a partial blockage of the phage by the SaPIbov5 *ccm* mutant was observed, although this island mutant was not expected to interfere with the evolved virion (Figure 5.9B). Interestingly, the ORF12 mutation and the double mutant (Ccm-ORF12) led to a phage titre recovery, with a slightly higher titre in comparison to the *ccm* mutant. $\phi12^{evolved4}$ moderately recovered the titre only with the island mutant in all operon I-like proteins (Figure 5.9B). All these results support the hypothesis that the other genes encoded in SaPIbov5 operon I-like could have a function related to the interference mechanisms of *cos* SaPIs, as it is the case for *pac* SaPIs operon I. The fact that the operon I-like gene cluster is highly conserved among the entire *cos* SaPIs further indicates a major role of these proteins in the SaPI life cycle (Figure 5.5).

5.2.5 SaPIbov5 Ccm and $\phi12_{gp33}$ are homologs in sequence but not in function

In silico analyses were performed to further characterise the Ccm protein. These studies revealed that the protein has no putative conserved domain in its sequence. Although Ccm and the $\phi12_{gp33}$ were distantly related in sequence, computational modelling analysis with I-TASSER[314], RaptorX[317] and Phyre2[318] revealed that the structural architecture of Ccm displays the same HK97-fold as the capsid protein gp5 of phage HK97[174,345] (Figure 5.13; Figure 5.14).

All capsid proteins that fold in this shape have the common characteristic of sharing only 10-15% or less sequence identity. This standard HK97 folding has some structural features that can be found in the Ccm protein (Figure 5.13B), such as the N-arm in the N-terminal region with α-helical content; an E-loop with two stranded anti-parallel β-sheet with a P-domain composed of a spine helix and a long β-sheet; an A-domain with a central β-sheet and finally a Δ domain at the N-terminal, with a high α-helical content. The 102 residues of the Δ domain are removed by a protease[346] at the end of capsid formation and act as a scaffolding protein limiting the conformational possibilities to accomplish the precise size of the capsids[172] (Figure 5.13).

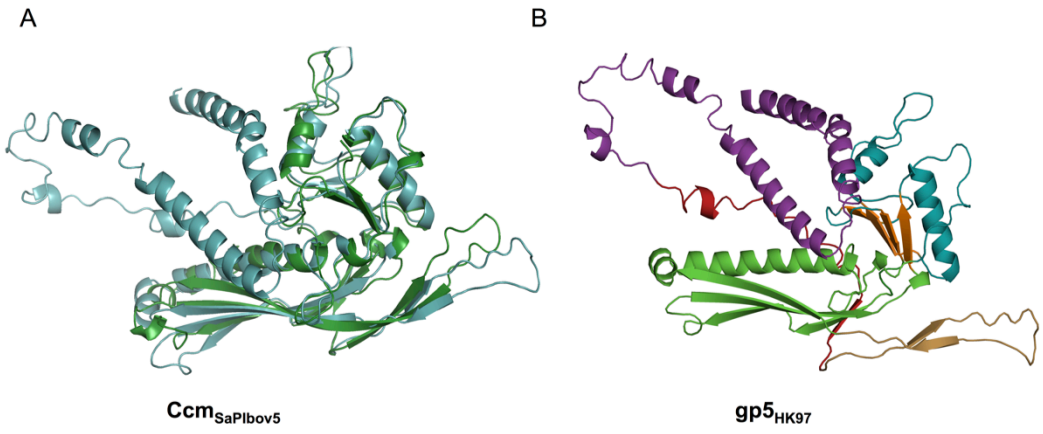


Figure 5.13 SaPIbov5 Ccm shows the characteristic HK97 fold. (A) Superimposed modelling of Ccm (cyan) generated with I-TASSER compared with the experimental structures of the crystal structure of gp5 HK97 Prohead II (green). Model with higher C-score (-0.88 for Ccm) was selected for further structural analysis using PyMOL. (B) The common structural features in Ccm coloured correspondingly; N-arm (red); E-loop (sand); P-domain (green); A-domain (cyan); β-hinge (orange) and Δ domain (purple).

RaptorX and Phyre2 servers modelled the C-terminal portions of gp33 (residues 127-402) and Ccm (residues 83-355) proteins with high confidence (Appendix 5- Table 1 and 2). A high structural similarity to the modelled HK97-fold domains was found between Ccm (RMSD < 1.5 Å for 240 residues) and HK97 capsid protein (RMSD < 2 Å for 210 residues) (Figure 5.14). On the other hand, when modelling the N-terminal portion of Ccm and gp33 proteins (residues 1-82 and 1-126, respectively), a low confidence in the folding models was found (Appendix 5 - Table 1 and 2). Nonetheless, a highly conserved α-helical fold was present in the predicted modelling of the Ccm

and gp33 proteins (Figure 5.15; Figure 5.16), although the helical fold was shorter in the case of Ccm. As previously highlighted, the α-helical N-terminal region is a Δ-domain present in several phages, such as HK97 or T5, which works as a scaffolding protein supporting capsid assembly[174,347].

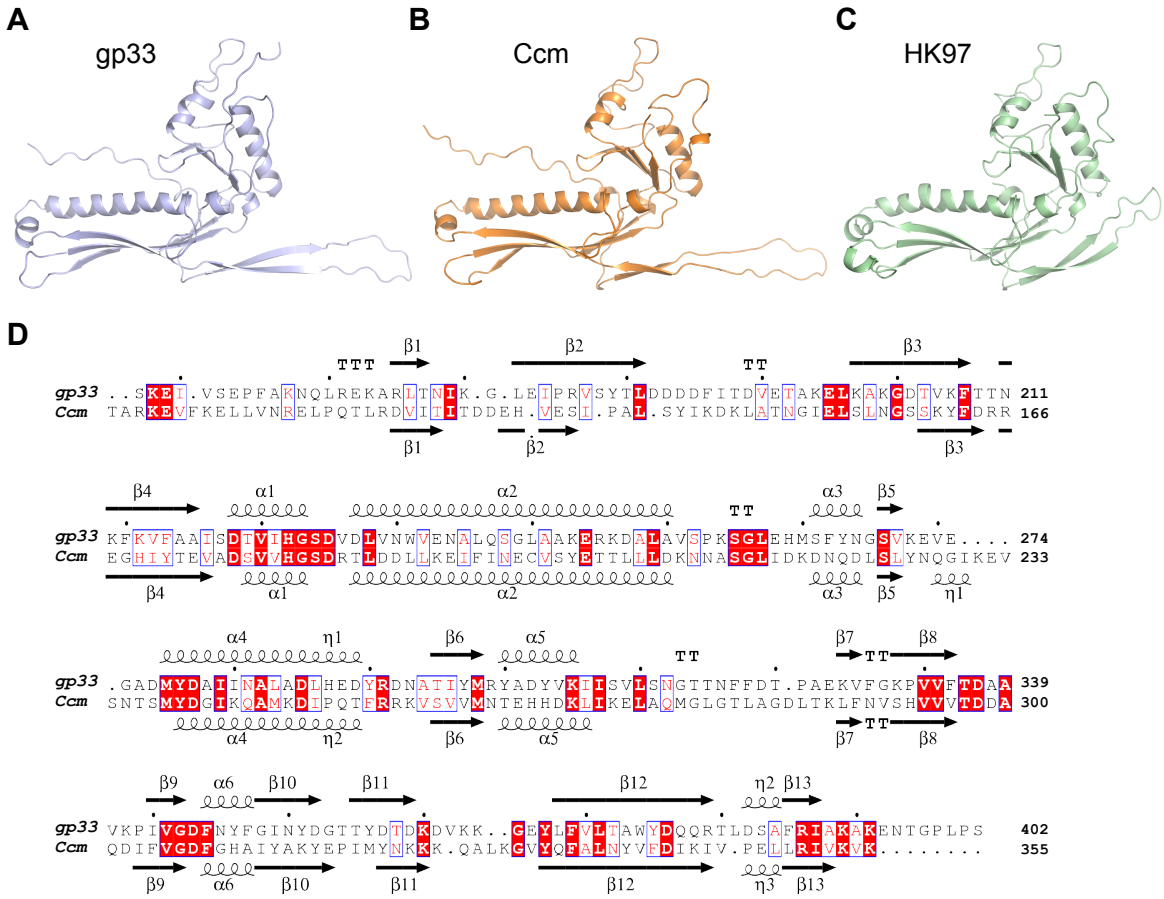


Figure 5.14 The C-terminal portion of gp33 and Ccm proteins are predicted to adopt the characteristic HK97-fold of phage coat proteins. Cartoon representation of C-terminal portion of (A) $\phi12_{gp33}$ (residues 127-402) and (B) SaPIbov5 Ccm (residues 83-355) generated by RaptorX[317]. Both proteins show similar folding to the prototypical coat protein from phage HK97 (C; PDB 1OHG). (D) Structural alignment of $\phi12_{gp33}$ (A) and SaPIbov5 Ccm (B) models carried out with Mustang[320]. Identical residues are highlighted on a red background, and conserved residues are in a blue box with red text. The elements of secondary structure for each model are shown above (gp33) or below (Ccm) the corresponding sequence.

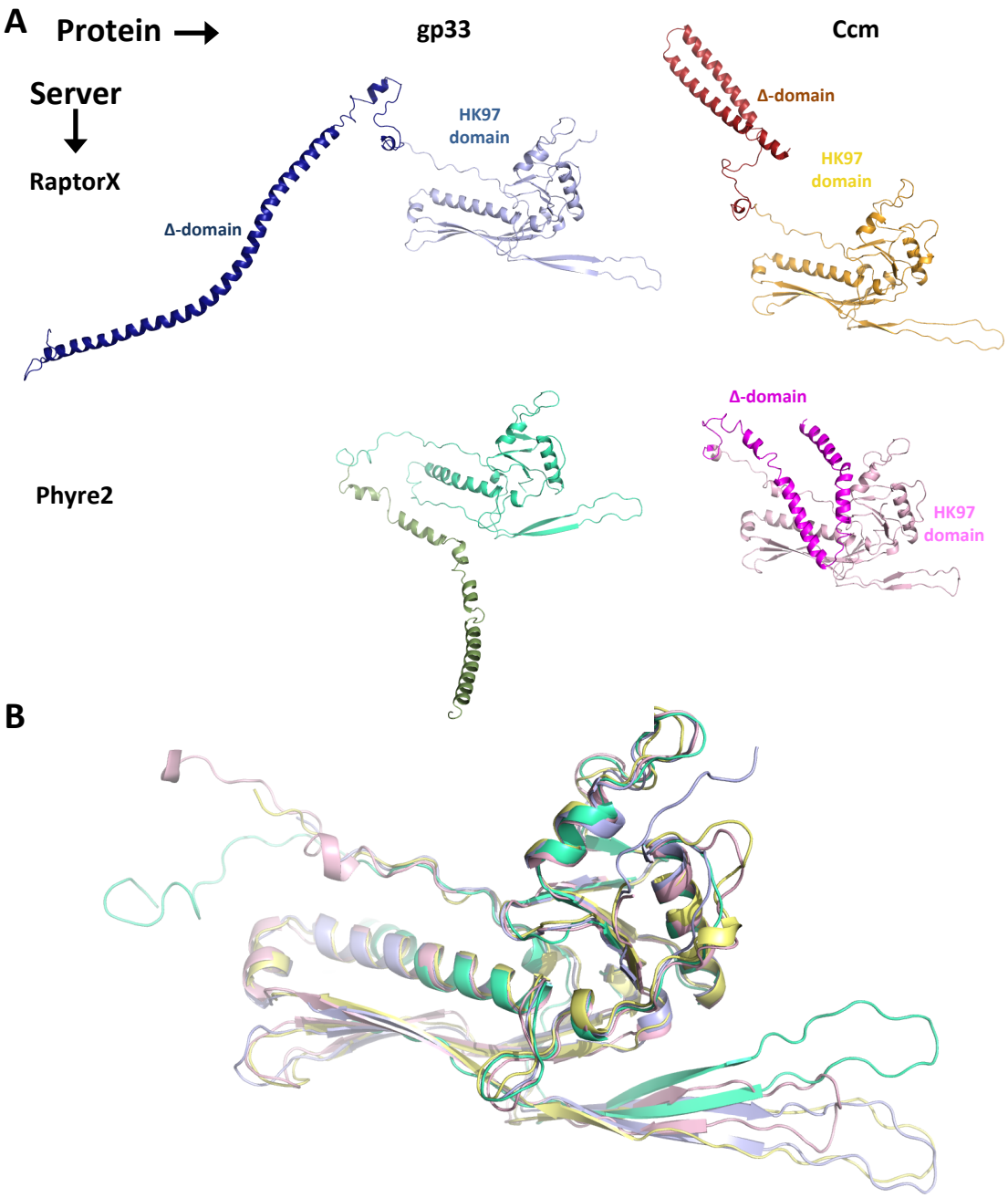


Figure 5.15 Comparison of gp33 and Ccm structural models generated by RaptorX and Phyre2 servers. (A) A cartoon representation of the structural models of gp33 and Ccm proteins produced by RaptorX and Phyre2 servers. Both servers divided the proteins into two portions: an N-terminal portion (highlighted in a dark hue) with a high α-helical content characteristic of coat protein Δ-domains and a C-terminal portion (light hues) showing the HK97-fold. (B) Superimposition of C-terminal portions of the four models (two for each protein). The four models (identical colours to A) are consistent with gp33 and Ccm presenting a structurally similar C-terminal with HK97-fold.

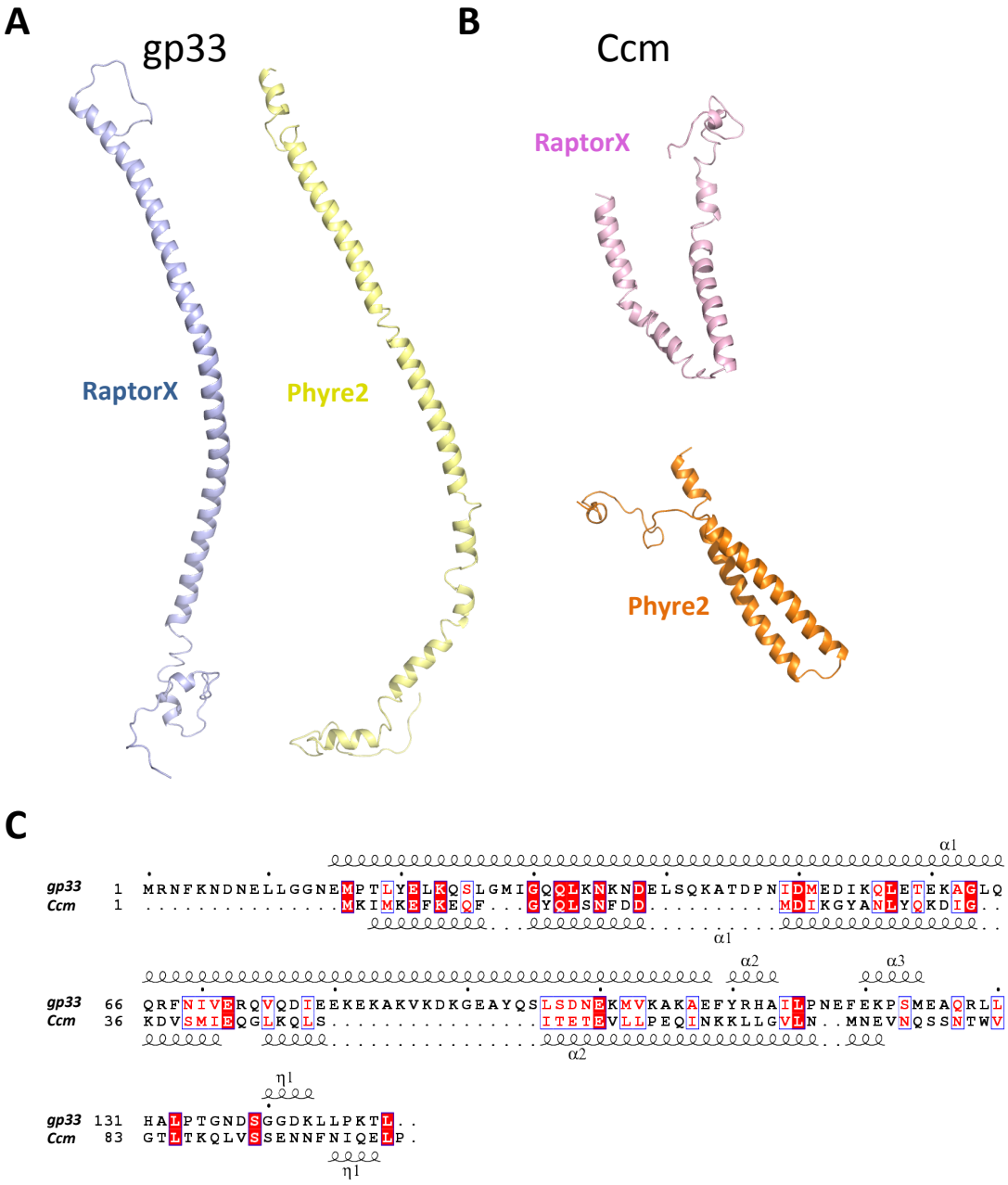


Figure 5.16 The N-terminal portion of gp33 and Ccm are predicted to adopt an α-helical fold. Cartoon representation of the predicted models by RaptorX and Phyre2 server for the N-terminal portion of (A) gp33 (residues 1-126) and (B) Ccm (residues 1-82). (C) Structural alignment of gp33 and Ccm models generated by RaptorX carried out with Mustang. Identical residues are highlighted on a red background, and conserved residues are in a blue box with red text. The elements of the secondary structures for each model are shown above (gp33) or below (Ccm) the corresponding sequence.

The putative structural similarity of Ccm suggested that it could be a capsid protein and raised the question whether the Ccm protein would be able to form capsids in the absence of $\phi12_{gp33}$. This, in turn, would suggest that SaPIbov5 could have a complementary mechanism to interfere with the phage reproduction and at the same time support SaPIbov5 transfer. To determine whether this hypothesis was accurate, a previously generated ϕSLT mutant in the capsid protein $\phi SLT_{gp42}{}^{270}$ was used. The ϕSLT_{gp42} protein belongs to the same family as the HK97-fold coat proteins and is essentially identical to the one encoded by the $\phi12_{gp33}$. The SaPIbov5^{adjusted} island was introduced into this mutant strain and the ability of this mutant to generate both phage and SaPI particles was analysed. As can be seen in Table 5.4, neither the phage nor the island could be mobilised, implying that ϕSLT_{gp42} is essential for both phage and SaPI transfer. This result undoubtedly excludes the Ccm protein from being capable of forming small capsids by itself. From these studies, it can be concluded that both *cos* and *pac* SaPIs use a convergent strategy to redirect capsid size via a scaffolding mechanism.

Table 5.4 Effect of phage mutations on phage and SaPI titre[a].

Donor strain			
Phage	**SaPI**	**Phage titre[b]**	**SaPI titre[c]**
ϕ SLT*pvl*::*tet*M	-	5.0 x 10^6	-
ϕ SLT*pvl*::*tet*M ΔORF42	-	<10	-
ϕ SLT*pvl*::*tet*M	SaPIbov5^{adjusted}	1.74 x 10^6	1.72 x 10^6
ϕ SLT*pvl*::*tet*M ΔORF42	SaPIbov5^{adjusted}	<10	<10

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]PFU/ml induced culture, using RN4220 as recipient strain.
[c]No. of transductants/ml induced culture, using RN4220 as recipient strain.

5.2.6 Ccm blocks *cos* but not *pac* phages

Cos and *pac* phages from *S. aureus* utilise similar packaging mechanisms, although using different families of proteins in order to encapsidate their own dsDNA. Hence, it was further investigated to see if the Ccm mechanism employed by SaPIbov5 for *cos* helper phages could also interfere packaging of the *pac* phages. This was analysed by using the pCN51 plasmid with the cloned

ccm gene, testing if this plasmid could block infection by *pac* phages when the *ccm* gene was expressed. As can be seen in Table 5.5, the expression of Ccm did not block the reproduction of the ϕ11 or ϕ80α *pac* phages, indicating that this mechanism of Ccm-mediated interference is specific for *cos* phages.

Table 5.5 Role of Ccm in phage interference.

Phage infection

Phage	No. phage plaques	
	RN4220[b]	**RN4220 + pCN51-*ccm***[b]
ϕ80α	2×10^{12}	8×10^{12}
ϕ11	5.26×10^{11}	1.6×10^{12}
ϕ12	2.88×10^{10}	1.4×10^{7}
ϕSLT*pvl*::*tet*M	8.8×10^{6}	1.05×10^{2}

Phage induction

Donor Strain		Phage titre
Lysogen	**Plasmid**	**RN4220**[b]
ϕ80α	pCN51	6.8×10^{11}
ϕ80α	pCN51-*ccm*	1.12×10^{11}
ϕ11	pCN51	4.15×10^{10}
ϕ11	pCN51-*ccm*	2.77×10^{10}
ϕ12	pCN51	1.6×10^{10}
ϕ12	pCN51-*ccm*	1.17×10^{6}
ϕSLT*pvl*::*tet*M	pCN51	1.72×10^{6}
ϕSLT*pvl*::*tet*M	pCN51-*ccm*	6×10^{2}

[a]The mean values of results from three independent experiments are shown.
Variation was within ±5% in all cases.
[b]PFU/ml induced culture, using RN4220 as recipient strain.

5.2.7 *Cos* SaPIs reserve space for virulence gene carriage

It is well established that *pac* SaPI genomes can be encapsidated into full phage particles. For instance, SaPIbov2, one of the prototypical *pac* SaPIs[236], packages its genome in full-size phage capsids. The large SaPIbov2 genome size, ~ 27 kb, and the absence of a CpmB homolog protein, causes its encapsidation into large phage capsids[245]. Therefore, besides evaluating the

mechanisms by which *cos* SaPIs are packaged in small capsids, the possibility of *cos* SaPIs having an augmented genome size or lacking the *ccm* gene was also examined.

Thus, GenBank sequence analyses of the whole operon was performed and found that the Ccm gene was preserved in the entire family of *cos* SaPIs. Remarkably, one of the *cos* islands, SaPI-S0385, was found to have a reduced size (10.3 kb) compared to the common *cos* SaPIs (Figure 5.5). This small SaPI featured all genes that were predicted to be required for the SaPI cycle. As described in this chapter, the variation in SaPI size directly affects the efficiency of *cos* SaPI packaging and transfer. For this reason, and in order confirm whether this small SaPI was able to produce small capsids, a small island was generated (SaPIbov5^{small}) containing only the predicted minimal elements believed to be involved in the packaging process. In this SaPIbov5^{small} derivative, the von Willebrand binding protein (*vwb*), the adenosine deaminase (*ada*) and the staphylococcal complement inhibitor (*scn*) were removed, and an erythromycin cassette was inserted giving a total size of 10.9 kb (Figure 5.4).

The SaPIbov5^{small} ERP cycle was tested to analyse its functionality and packaging phenotype. As can be seen in Figure 5.12, replication of SaPIbov5^{small} after SOS induction was unaffected, but since the *cos* unit-length was changed, no SaPI-sized small capsid band was observed. The transfer ability, mediated by ϕ12, was slightly reduced when compared to SaPIbov5^{adjusted}, although as expected this island still blocked ϕ12 reproduction (Table 5.1), reaffirming that this island still expresses the Ccm protein. This result suggests that *cos* SaPIs have evolved to reserve ~2 kb of space for the carriage of virulence genes to complete the ~15 kb *cos* SaPI genome to be efficiently packaged. Carriage of these virulence genes not only allows *cos* SaPIs to be packaged into SaPI-reduced capsids but also confers upon these elements a selective advantage to be maintained in nature.

5.3 Discussion

The overall aim of this work was to determine the mechanism by which SaPIbov5, and by extension *cos* SaPIs, interfere with the life cycles of their helper phages via production of SaPI-sized small capsids. Although little is known about *cos* SaPIs, many studies have been conducted with the related *pac* SaPI family members. Three mechanisms of phage interference have been identified to date. They involve capsid size redirection[257], interference with the late operon transcription phage module[259] or inhibition of the phage-encoded TerS[255]. While the genes responsible for the first two mechanisms are carried in the characteristic operon I present in most of *pac* SaPIs, the *ppi* gene is outside of this operon I. Remarkably, a variant of the *ppi* gene is also present in the *cos* SaPI elements. However, the SaPIbov5 *ppi* gene does not seem to be involved in *cos* phage ϕ12 interference. This result can be easily explained. The Ppi protein blocks the phage-encoded TerSϕ. Since this protein is also required for packaging of both *cos* phage and *cos* SaPI[270], inhibition of the activity of the TerSϕ will in turn block SaPI transfer. Thus, the function of this gene in the SaPIbov5 ERP cycle remains unsolved.

More clear is the role of the *ccm* gene. This gene, present in all analysed *cos* SaPIs, encodes Ccm, which redirects the production of SaPI sized capsids, this being the first mechanism of phage interference that specifically blocks *cos* phage reproduction. The *ccm* gene was found to encode a HK97-like capsid protein (Figure 5.14). Structural analysis of this protein predicted that Ccm could be participating in the assembly process of the small capsids, remodelling the full-size phage shell or even being part of the capsid structure along with the phage-encoded capsid protein. Despite the fact that Ccm by itself was not capable of forming small capsids, the similarity of the coat fold to gp33 indicates that Ccm could form dimers in the presence of the phage capsid protein. Other studies have suggested that the length of the N-terminal Δ-domain is associated with capsid size[348,349]. Our models demonstrate that there is a difference in length between the Δ-domain of Ccm and gp33, with Ccm being 44 residues shorter (Figure 5.15; Figure 5.16). Hence, these modifications in length could be driving capsid size redirection.

This proposed model of action could also clarify why Ccm does not affect *pac* phage reproduction. ϕ11 and ϕ80α *pac* phages do not encode a Δ-domain within the capsid protein; instead, *pac* phages synthesise different scaffolding proteins expressed separately to perform the same function[200]. Therefore, the *pac* SaPI capsid-size remodelling mechanism consists of CpmB, a SaPI encoded alternative scaffolding protein, which redirects phage capsid assembly to form SaPI-size capsids[200,280]. Hence, both proteins CpmB and Ccm, by employing scaffolding domains, can modify the phage capsid assembly process to benefit SaPI transfer and maintenance in nature. Nevertheless, the mechanism behind this production remains unclear.

The Ccm protein was further investigated for its interference activity, severely blocking ϕ12 and ϕSLT reproduction. After evolutionary experiments with ϕ12, $\phi12_{gp33}$ was revealed to be the target protein for the Ccm protein. $\phi12^{evolved4}$ overtook the interference through a punctual mutation in the gp33, being able to escape the blockage of *ccm* gene. Hence, it was suggested that this interference would be as a consequence of the Ccm-mediated modulation redirecting the phage capsid size into small particles. Several lines were studied to characterise this interference mechanism. Overexpression of Ccm blocked the wt phage, but had no effect on $\phi12^{evolved4}$. Furthermore, when infecting the generated SaPI mutants with the wt phage, a full recovery was attained for the mutant in Ccm. As was expected, mutating the whole SaPI operon did abolish interference. However, although showing similar titres to the wt phage, the phage plaque size was impaired in both Ccm and the whole operon mutants, suggesting that other proteins could be involved in other interference mechanisms. In addition, the results obtained by induction did not achieve full recovery of wt phage titre with the *ccm* mutant or the whole operon mutant.

Supporting this theory, the evolved phage was also able to escape to this interference in the ORF10 mutant, generating plaques following infection. This effect could imply a role for this ORF10 in the SaPIbov5 mediated interference mechanism. However, overexpression with the pCN51 in the induction experiments with the ORF10 mutant did not produce any effect on

phage reproduction either with the wt ϕ12 or the ϕ12^{evolved4}. Note that the SaPIbov5 ORF10 is only present in *cos* SaPIs and that this gene has no characterised domains in its sequence. The role of this protein in SaPIbov5 interference process could involve ORF10 somehow modulating the production of the small capsids proteins with Ccm. Further studies need to be completed to identify the precise role of this protein in the SaPI cycle.

Furthermore, when analysing the function of ORF12, the last encoded protein of this novel morphogenetic operon, it was found that overexpression in the presence of the Ccm protein reduced the hindrance triggered by Ccm alone. *In silico* analysis of this protein revealed a similar sequence to the *pac* SaPI protein PtiM. Structural modelling of ORF12 uncovered an overall fold that closely resembled the SaPI2 encoded PtiM protein. PtiM had been described as a modulator of PtiA activity, controlling the required level of transcription of the late operon phage module[259]. This evidence was only obtained after overexpression of both proteins, but less precise results have been achieved from the infecting and inducing analysis. In the induction studies, different patterns can be found when comparing the wt and the evolved phage. In the wt phage, a reduction of the phage titre after induction was observed in the ORF12 mutant. On the other hand, with ϕ12^{evolved4}, this mutation triggered an increment of the phage infection ability. Further studies may arise with more evidence that will clarify the specific role of ORF12 and its putative modulatory effect. Co-expression of any of the hypothetical proteins, ORF10 and ORF12, Ccm and ϕ12$_{gp33}$, should be performed for a better understanding of the capsid redirection mechanism and to confirm the possibility of complexes between any of the proteins (Figure 5.17).

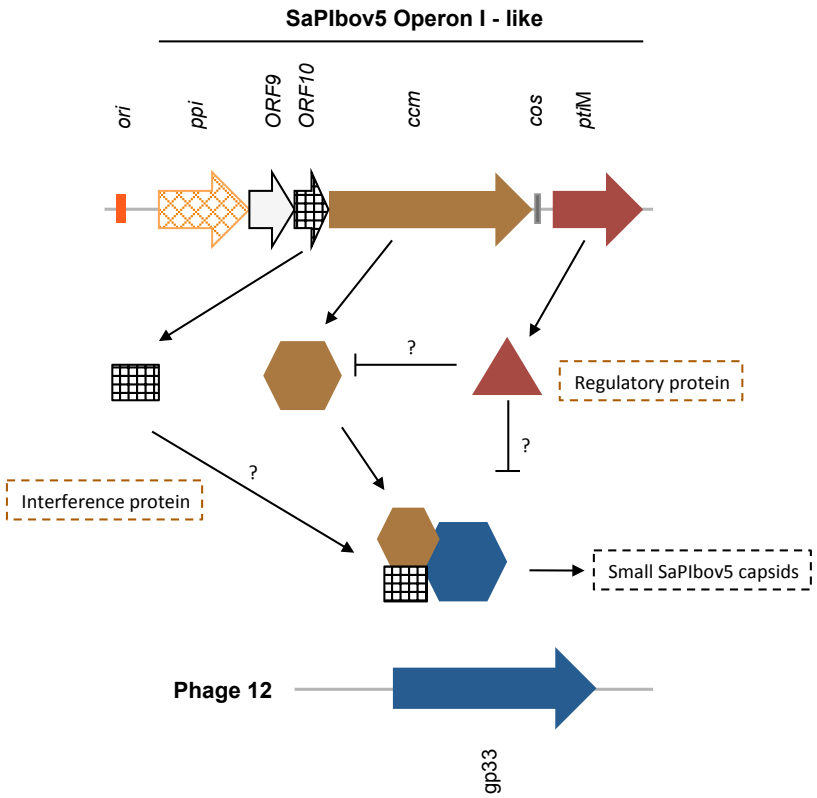


Figure 5.17 Summary of the SaPIbov5-mediated interference with phage 12. SaPIbov5 operon I-like gene cluster coding for five proteins, Ppi, ORF 9, ORF10, Ccm and PtiM, which are somehow involved in the mechanism of ϕ12 interference. Ccm binds to the $\phi12_{gp33}$ and redirects the size of the particle to produce small capsids in which only SaPI genomes can be packaged. ORF10 could exert some interference function together with the Ccm protein, somehow binding to the same protein, gp33, in ϕ12. ORF12, a homolog of PtiM, could be exerting some modulation of the ORF10-Ccm complex, but this mechanism remains to be clarified.

All *cos* SaPIs that were identified in this study encoded the full operon I-like gene cluster, which includes ORFs 8 - 12 for SaPIbov5 (Figure 5.5). As can be seen in Figure 5.18, *cos* SaPIs with a uniform genome size, coding for all the proteins included in the operon-I-like are clustered together. Conservation of all the proteins suggests that they are all required and involved in the interference and packaging process of the SaPIbov5 cycle.

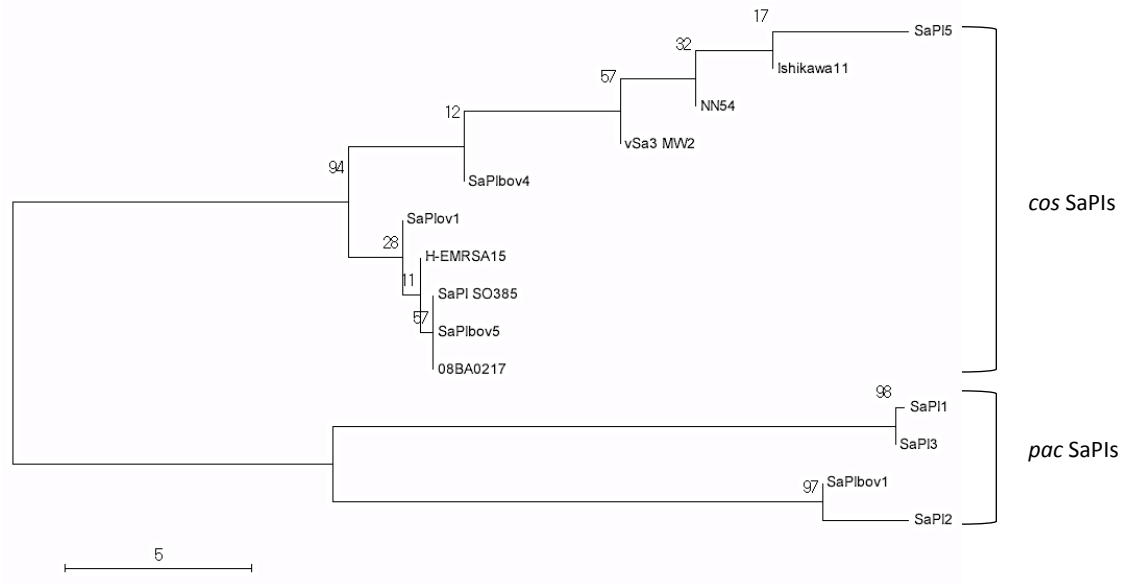


Figure 5.18 Neighbour-Joining tree of different *cos* and *pac* SaPIs. The tree was generated using MEGA7 version with MUSCLE sequence alignment.

Small capsid production is a convergent mechanism widespread in SaPI biology, contributing to phage interference and reduction of phage reproduction. Other PICI elements can be found utilising the same packaging strategy. This is the case for the *E. faecalis* EfCIV583 element, which modifies the phage ϕp1 capsid structure by an unknown mechanism to produce PICI-sized capsids[239]. Furthermore, this strategy is found in the *E. coli* P4 plasmid-ϕP2 model. The relationship between P2-P4 is analogous to the existent between SaPIs and their helper phages, and both involve phage induction to initiate their life cycle[242,275]. P4 uses a similar strategy to that studied in *cos* SaPIs, carrying the phage P2 *cos* sequence in its genome to enable packaging by the phage terminase complex. Moreover, both P4 and SaPIs code scaffolding proteins to remodel the phage-like capsid into small particles. Contrary to SaPIs, the P4 plasmid encodes Sid, an external scaffolding protein that alters the HK97-fold ϕP2-capsid[350]. However, the proteins involved in this interference mechanism have no homology to the ones described for SaPIs, implying that this convergent evolutionary strategy offers an essential advantage for PICI and MGEs maintenance in nature.

Furthermore, to exert the phage interference mechanism and be packaged in small capsids, SaPIs have been forced to maintain a constant genome size.

Within the genome limitation, SaPIs have evolved only to preserve necessary genes related to mobility and phage blockage. Remarkably, inside that reserved and limited space, SaPIs have adapted so that they can carry several virulence genes, which ultimately exert an advantage competing with other SaPIs to be selected and maintained in nature. As established in this chapter, *cos* SaPI size requires a specific unit-length to be encapsidated. Therefore, the numbers of genes that can be included are limited, and any increment of size will be detrimental for SaPI transfer and phage interference.

SaPIs are widespread in nature because of their unique variety of encoded virulence and fitness factors for competing in different bacterium niches[344,351]. Since the numbers of chromosomal (att_C) sites where SaPIs can integrate are reduced, having been identified to date only 5, one could consider that the differences in virulence factors encoded by *cos* and *pac* SaPIs could be an important limiting factor for bacterial niche competition. On the contrary, analysis of the virulence genes encoded by the SaPIs establishes that there are no large differences in the arsenal of virulence genes present in the *cos* or *pac* SaPIs. Thus, genes encoding the TSST-1, SEL or SEC toxins, the staphylococcal complement inhibitor SCIN or the vWbp can be located in both types of islands[89]. Within this context, there is a huge competition between SaPIs to persist in nature. Furthermore, several studies have shown that certain proteins encoded by these elements, such as Bap and vWbp, exert a major role in the pathogenesis of *S. aureus* in the animal hosts they infect[236,238]. Identifying the specific SaPI-encoded virulence genes that are involved or related to the survival of the pathogen in different niches will allow us to recognise novel approaches to fight *S. aureus* infections in a more efficient manner. This same scenario could be playing a major role for other PICI elements. This hypothesis is supported by several studies that indicate how important phages and SaPIs are in bacterial evolution[261,273].

5.4 Conclusions and future work

Of the family of PICIs, SaPIs are the best-studied phage satellites that with a limited number of encoded genes can efficiently exploit helper phages for

their own benefit. *Pac* SaPIs use diverse mechanisms to package their own genome and ultimately interfere with phage reproduction. Amongst all the strategies that *pac* SaPIs employ to manipulate the phage cycle, assembly in small SaPI-sized capsid using SaPI encoded *cpm*A and *cpm*B genes is a significant step in order to interfere with the phage reproducibility.

A novel convergent evolutionary mechanism is described here, by which *cos* SaPIs are packaged in a similar way to their related *pac* family members. Although the genes involved are rather distant to those employed by *pac* SaPIs, these elements have evolved to achieve the same outcome and to target essential phages-functions using unrelated proteins.

However, as described here, it was not possible to decipher the complete basis by which *cos* SaPIs block *cos* phage reproduction. Nevertheless, the gene responsible for the formation of small capsids in *cos* SaPIs was identified and characterised. Hence, this is the first time that a mechanism of interference has been described for *cos* SaPIs and their cognate *cos* phages. This mechanism depends on the *ccm* gene, which encodes a protein with a HK97-fold with a low sequence identity to its related capsid proteins. This novel *cos* SaPI strategy indicates that the production of small capsids is widespread in nature and that it has been conserved during SaPI evolution. Future work will be focused on evaluating the role(s) of the other genes present in this new operon, and in identifying other mechanisms of interference of *cos* SaPIs.

Chapter 6 Characterisation of the putative morphogenetic cluster in EfCIV583

Disclaimer on work performed:

The group of José R Penadés kindly provided some strains in this chapter, and full details are shown in Table 2.1 of Chapter 2.

Dr Renata C. Matos kindly provided some strains mentioned in this chapter, and full details are shown in Table 2.1 of Chapter 2.

6.1 Background

It is well known that PICIs are widely distributed in the *Staphylococcus* genera[32,36]. However, these elements are not genus-restricted, and PICI-like elements have been found already in other bacterial species such as *Streptococcus pyogenes* (SpyCIM1), *Enterococcus faecalis* V583 (EfCIV583) and *Lactococcus lactis* (LlCIbIL312)[89,239,240]. Therefore, the PICIs seem to be widely distributed, suggesting that they have an enormous importance in bacterial evolution[261,273].

As previously described, *E. faecalis* V583 isolate is part of the high-risk clonal complex two, which is particularly well adapted to hospital settings. The genome of V583 strain carries seven prophage-like elements, named as V583-pp1 to - pp7. The EfCIV583 pathogenicity island, also described as pp7+, was analysed by Matos *et al.*[239]. The authors identified this island as a PICI element and demonstrated that ϕp1 (also named pp1+) is the helper phage of this element. Further, Matos *et al.*[239] established that this pathogenicity island produces small capsids, where the island is specifically packaged. This suggested that the EfCIV583 was extremely similar in function to SaPIs. However, they suggested that the EfCIV583 island is induced by the SOS response and that this element only requires the presence of the phage for its own packaging, rather than being needed for the induction of the PICI cycle, in clear contradiction to the SaPI cycle. However, recent results from our group demonstrated that the helper phage is needed both for induction and packaging of the EfCIV583 island, confirming EfCIV583 as a member of the recently identified PICI family[240] (Figure 6.1).

Therefore, in view of the previous results, the aim of this study was to determine the relationship between this novel PICI element and its helper phage ϕp1. Thus, this work attempted to decipher the uncharacterised mechanism by which this island produces small capsids using the ϕp1machinery. In addition, the role of the phage ϕp1-encoded TerS and TerL proteins in the EfCIV583 cycle was also analysed.

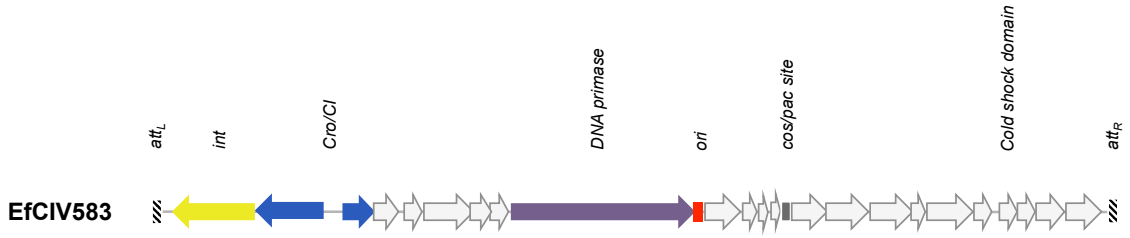


Figure 6.1 Genomic organisation of *E. faecalis* EfCIV583 island. The genome is aligned according to the prophage convention with the integrase gene at the left end. Gene colour code: *int*, yellow; transcription regulators, blue; replication gene, purple; the replication origin (*ori*), red; Genes encoding hypothetical proteins, white. The *cos/pac* site is shown in grey.

6.2 Results

6.2.1 Phage ϕp1 packaging mechanism and the implication for the EfCIV583 cycle

As previously highlighted, PICI induction depends on a phage-encoded protein that acts as an anti-repressor. Martinez-Rubio *et al.*[240] established that induction by the helper phage ϕp1 was required for the activation of the ERC cycle of the EfCIV583 element[240] (Figure 6.2). In addition, it was shown that the deletion of the phage-encoded inducer, ϕp1 EF0309, completely eliminated induction and dramatically decreased EfCIV583 transfer[240]. This study revealed that the product of the phage ϕp1 EF0309 gene blocked the function of EfCIV583 Rpr protein (a Stl-like repressor) by a protein-protein interaction[240]. Overall these results suggested that indeed ϕp1 was the helper phage of EfCIV583 and that the relationship between the phage ϕp1 and EfCIV583 resembled the one described for the staphylococcal helper phages and SaPIs[109,110,242]. Likewise, EfCIV583 hijacks the formation of phage capsids, which leads to capsid size remodelling, as also occurs with the interference shown by the classical SaPI1, SaPIbov1 and SaPIbov5 islands[239].

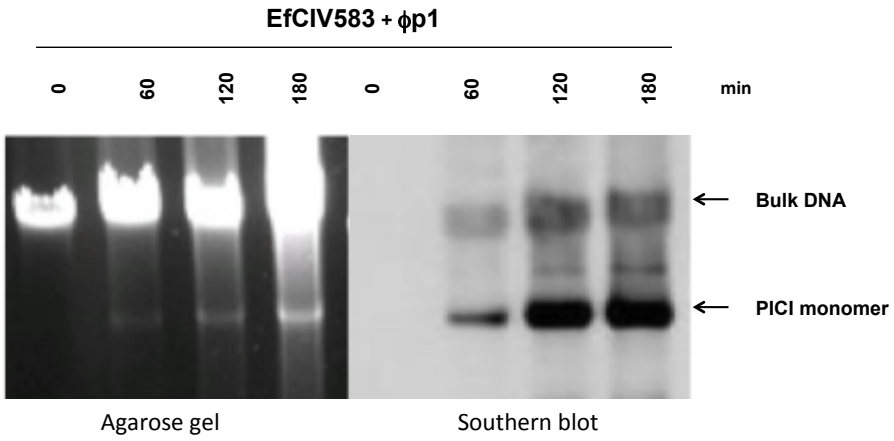


Figure 6.2 Study of the ERC cycle of EfCIV583. Agarose gel (left) and Southern blot (right) of strain V583 lysogenic for the phage ϕp1 after induction with MC. In the Southern blot, a probe specific for the island was used. The characteristic banding pattern of PICIs is shown.

In order to decipher the mechanism behind the mobilisation of EfCIV583 after induction by its helper phage ϕp1, initially the putative locations of the genes encoding the holoterminase complex (TerS and TerL) in the ϕp1 phage were investigated. *In silico* analyses were performed to localise and characterise the hypothetical terminase proteins in the genome of the phage ϕp1. BLAST gene and protein sequence analysis revealed a high degree of similarity between EF0332 and EF0333 in ϕp1 and the phiFL4A TerS and TerL, respectively, suggesting that they might feature the terminase activity. EF0332 and the TerS (gp30) of the *Enterococcus* phage phiFL4A feature 68% similarity (Figure 6.3A). Likewise, TerL (gp31) of phiFL4A was 97% similar compared to EF0333 of ϕp1 (Figure 6.3B). Note that there were no characteristic domains that could be identified in the hypothetical ϕp1 encoded TerS and TerL, suggesting these proteins belong to new classes of terminases.

Furthermore, 3-dimensional structural models were generated to corroborate that EF0332 and EF0333 fold like the classical TerS and TerL proteins. For that purpose, an iterative threading assembly refinement I-TASSER server was used[314]. Despite relatively low sequence identity, TerS ϕp1 EF0332 displayed a high degree of structural similarity to the crystal structure of the TerS protein (gp1) of *Shigella flexneri* bacteriophage Sf6 (PDB 3HEF) (Figure 6.4A). Likewise, the secondary structure of ϕp1 EF0333 resembled the same folding

as the crystal structure of the TerL (gp12) protein of Enterobacteria phage T4 (PDB 3CPE) (Figure 6.4B).

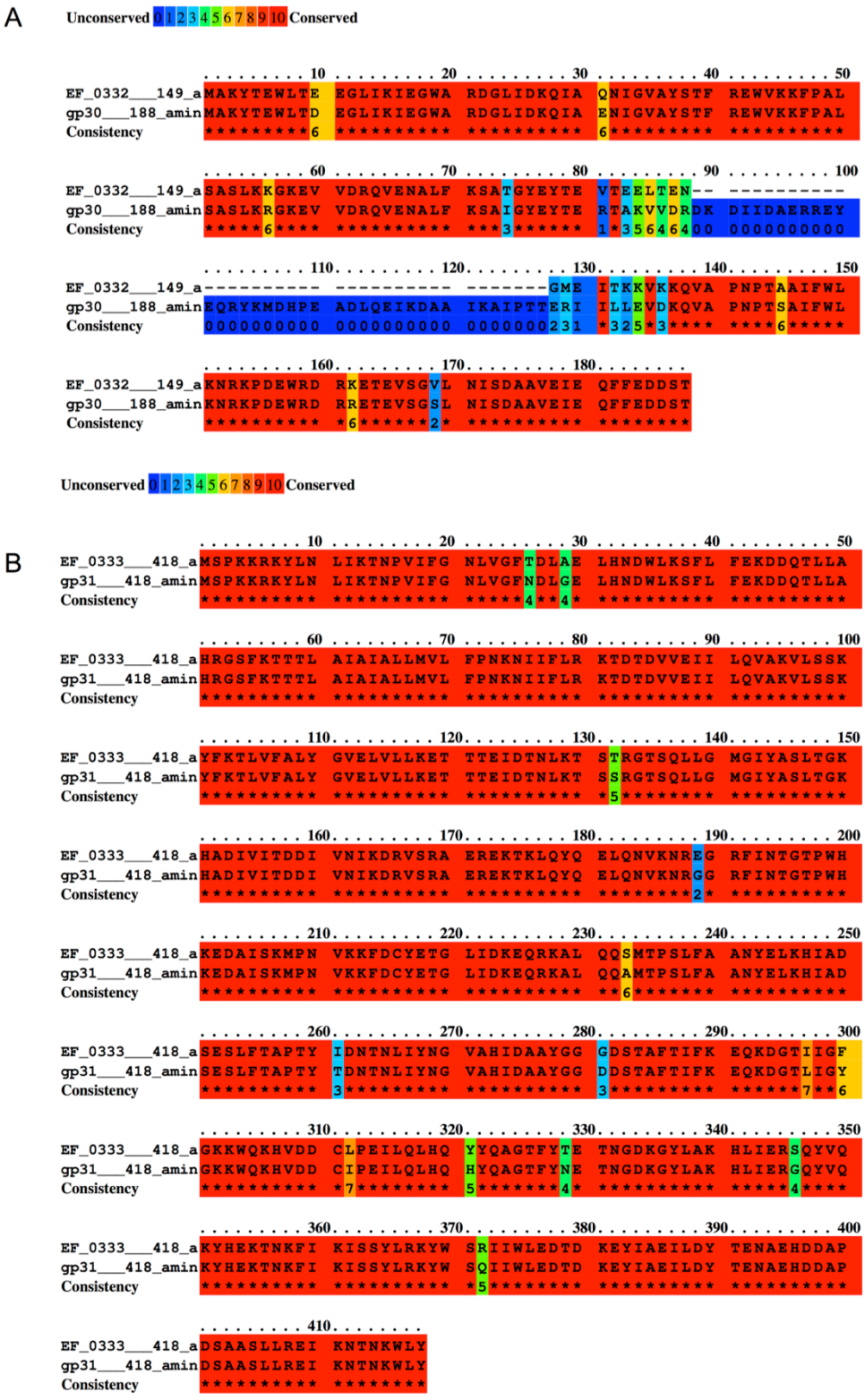


Figure 6.3 Alignment of the putative TerS and TerL proteins from ϕp1 and phiFL4A. (A) EF0332 (TerS) and gp30 (TerS) and (B) EF0333 (TerL) and gp31 (TerS) from phage ϕp1 and phiFL4A, respectively, were aligned with using PRALINE multiple sequence alignment program.

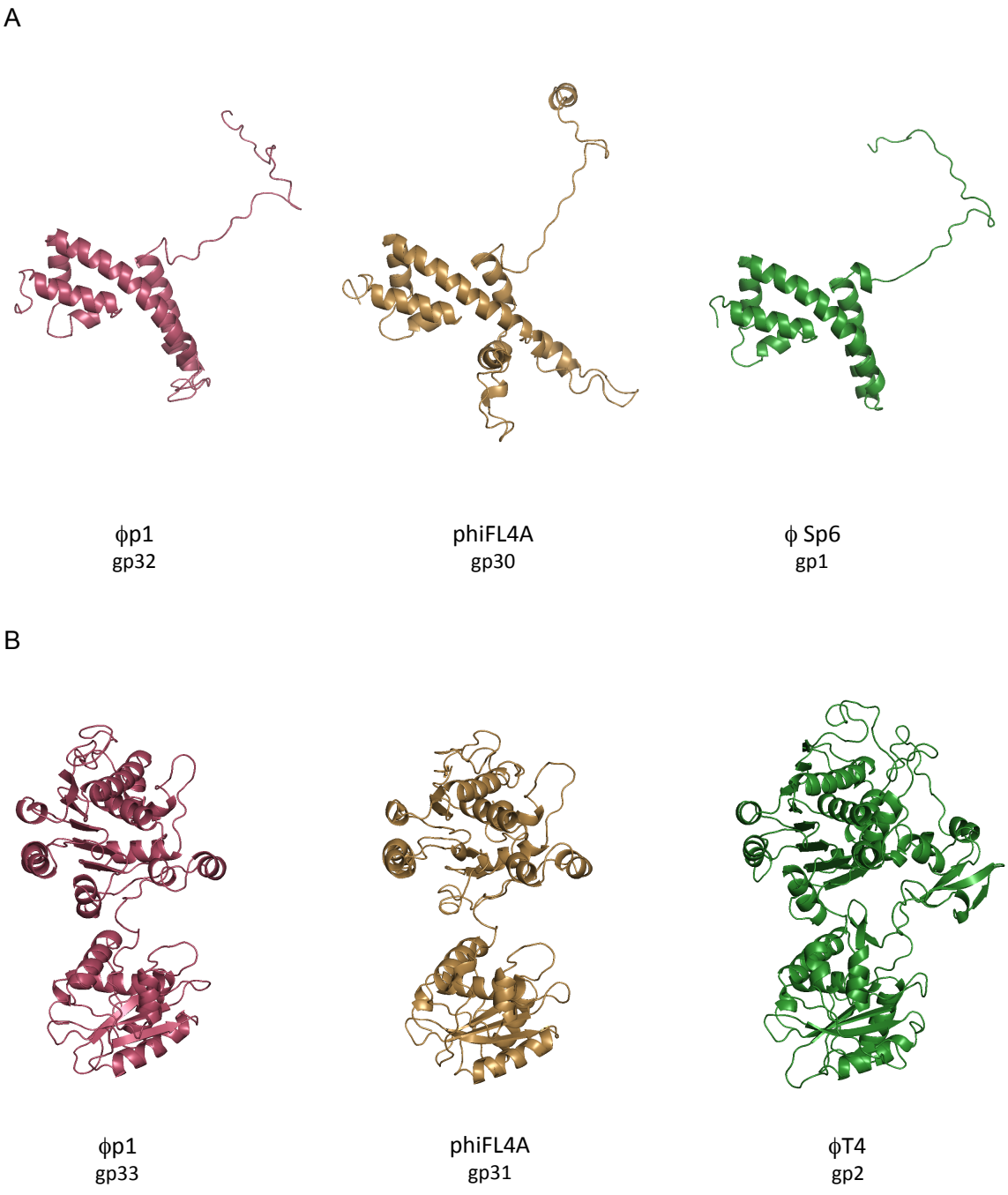


Figure 6.4 Secondary structure of the putative ϕp1 TerS and TerL proteins. (A) ϕp1 TerS (EF0332; pink) and *Enterococcus phage* phiFL4A (gp30; yellow) structural models are compared with the experimental structure of *Shigella flexneri* Sf6 phage (PDB 3HEF; green) small terminase complex. (B) ϕp1 TerL (EF0333; pink) and *Enterococcus phage* phiFL4A (gp31; yellow) are compared with the experimental structure of *E. coli bacteriophage* T4 (green; PDB 3CPE) large terminase complex. Structural models were generated with I-TASSER. Models with higher C-score (-1.79 for EF0332 and -0.77 for EF033) for ϕp1 and (-3.56 for gp30 and -0.70 for gp31) for phiFL4A were selected for further structural analysis using PyMOL.

To confirm the results of these simulations experimentally, deletion mutants were generated for both genes to determine whether EF0332 and EF0333 performed as TerS and TerL in ϕp1. Hence, the terminase mutants were analysed for their ability to produce phage particles. Neither EF0332 nor EF0333 mutants were able to produce infective particles, in accordance with the proposed function. Complementation using a pCU1 plasmid with a cadmium-inducible promoter (*pCad*), encoding TerS and TerL did not recover the titre of these phage mutants (Table 6.1). Since the *pac* site is usually embedded in the *ter*S, it is expected that the *ter*S mutants cannot be complemented in trans. The results, however, were surprising in respect to the TerL. In fact, we are unable to explain it. In conclusion, EF0332 and EF0333, from now designated as TerS$_{\phi p1}$ and TerL$_{\phi p1}$, could be strong candidates for providing the terminase function in the ϕp1 phage.

Thus, it was of interest to know whether the EfCIV583 element encodes its own *ter*S gene in its morphogenetic module, as *pac* SaPIs have[256], or if it uses that from the helper phage. If the EfCIV583 element encodes its own TerS protein, mutation of the *ter*S gene in the helper phage would have no effect on the transfer of the EfCIV583 island. A derivative of EfCIV583 carrying a *tet*M selection marker was generated to facilitate transfer studies and with the purpose of analysing the three definable stages of the PICI ERC cycle for this element. The derivative island was then introduced into lysogenic strains containing either ϕp1 wt or the ϕp1 mutants. As can be seen in Table 6.1, the EfCIV583$_{tetM}$ island was only transferred when the wt helper phage was induced. However, the island could not be mobilised by the phage mutant derivatives. Complementation of the mutants with plasmids encoding the TerS and TerL genes slightly recovered the island titre. Together this data confirms that TerS$_{\phi p1}$ and TerL$_{\phi p1}$ are essential for the EfCIV583 island transfer and that this element does not encode a TerS itself, being dependent on the phage for its activation and transfer. These results also suggest that the EfCIV583 element encodes a *pac* or *cos* site in its sequence, which will be used to initiate packaging of the element.

Table 6.1 Effects of phage mutants on EfCIV583 transfer[a].

Donor strain			
Phage	**EfCIV583**	**Phage titre[b]**	**EfCIV583 titre[c]**
ϕp1	-	1.9×10^9	-
ϕp1	+	9×10^6	1.9×10^4
ϕp1ΔORF32 (TerS)	+	<10	<10
ϕp1ΔORF32 (TerS) + pCU1-Cd-EF0332	+	<10	2.3×10^2
ϕp1ΔORF33 (TerL)	+	<10	<10
ϕp1ΔORF33 (TerL) + pCU1-Cd-EF0333	+	<10	1.3×10^2

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]PFU/ml of lysate, using the non-lysogenic VE18590 as recipient strain.
[c]No. of transductants/ml induced culture, using VE18590 as recipient strain.

6.2.2 Identification of the putative packaging recognition site sequence

The fact that complementation with the plasmid encoding TerS and TerL in the respective mutants could not restore the infectious ability of $TerS_{\phi p1}$ and $TerL_{\phi p1}$ deficient phages suggested that one of these genes might contain a *pac* or *cos* site within the gene coding sequence. At this point, however, it was not known whether the packaging of ϕp1 is dependent on a *pac* or a *cos* sequence. Further, the lack of an identifiable *ter*S gene and the dependence of EfCIV583 on the host TerS and TerL suggested that the island might follow a transfer strategy similar to one of *cos* SaPIs, previously described in this thesis. Thus, packaging of EfCIV583 could be entirely reliant on the machinery of ϕp1. In conclusion, it was hypothesised that EfCIV583 would feature a homologue of the ϕp1 packaging recognition site.

Under the presumption that ϕp1 and EfCIV583 feature the same packaging recognition site, nucleotide identity analyses were performed comparing the ϕp1 and EfCIV583 genomes to identify this site. Searching parameters were based on properties of functional *pac* or *cos* sites that were previously described in other phages, such as the high presence of G and C nucleotides for the *cos* phages, or the location within the phage genome, which normally resides in the module of morphogenesis near the terminase

genes[155,270,288,321,352]. In this context, different repetitive homology sequences from ϕp1 and EfCIV583 were analysed. Only one sequence was found in both genomes matching these characteristics (Figure 6.5). This potential *pac* or *cos* site sequence was located within the phage ϕp1 *ter*S gene and behind the primase gene and the *ori* site, in the morphogenetic module of the island (Figure 6.1; Figure 6.6).

```
EfCIV583        AAAAAAGT--CGGCACTTTCGGCAGCC-TAAAAA
ϕp1             AAAAAATTTCCGGCACTTTCGGCATCCTTAAAAA
                ****** *  ************** ** ******
```

Figure 6.5 Putative packaging recognition site alignments. The EfCIV583 and phage cleavage sites and their flanking sequences were aligned using MUSCLE. The putative cleavage sites are shaded in yellow.

To support the previous assumption and to determine whether the potential packaging recognition site of phage ϕp1 was located within the *ter*S gene, the previously generated pCU1-TerS, pCU1-TerL and pCU1-empty plasmids were used to evaluate the ability of the phage to promote its mobilisation[299]. With this strategy, it was possible to assess whether the *Enterococcus* phage ϕp1 was involved in general transduction enabling horizontal transfer of an empty pCU1 plasmid. If that was the case, ϕp1 should be considered as a *pac* phage. However, if ϕp1 was a *cos* phage, only the plasmid with the encoded packaging recognition site should be transferred.

For that purpose, a ϕp1 lysogen strain carrying the different plasmid derivatives with chloramphenicol markers was used for the transduction studies of the plasmid. After SOS induction by MC, the plasmids were transferred to the VE18590 (pp-) strain, which is a host acceptor cured of any MGE. As can be seen in Table 6.2, general transduction of the empty control plasmid was not observed in the lysogenic strains for ϕp1, but the plasmid containing the TerS gene was transferred. Interestingly, the plasmid containing the putative TerL of ϕp1 was also transduced. This could be explained by the presence of an additional packaging recognition site on the TerL protein, and could also clarify the inability of the mutant of TerL to be restored with the complemented plasmid. Another possibility that may

explain the transfer of the TerL plasmid could involve a recombination process between the sequences of the encoded TerL with ϕp1. This last scenario would suggest ϕp1 as a *pac* phage, since extra packaging will not interfere with the packaging ending, and additional genome could be packaged into the phage particle. This packaging has been observed for *pac* phages with more affinity to horizontally transfer plasmids coding for phage sequences[301]. Thus, further experiments must be done to really clarify whether this sequence is the packaging recognition site sequence and if it is a *cos* or *pac* site.

Table 6.2 The transfer efficiency of the different cloned TerS and TerL in pCU1[a].

Donor strain		
Phage	**Phage titre[b]**	**Plasmid titre[c]**
ϕp1	1.9×10^9	-
ϕp1+ pCU1-Cd	1.7×10^8	<10
ϕp1+ pCU1-Cd-ORF32 ΔORF33	2×10^8	8×10^4
ϕp1+ pCU1-Cd-ORF33	1.8×10^8	4.4×10^4

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]PFU/ml of lysate, using the non-lysogenic VE18590 as recipient strain.
[c]No. of transductants/ml induced culture, using VE18590 as recipient strain.

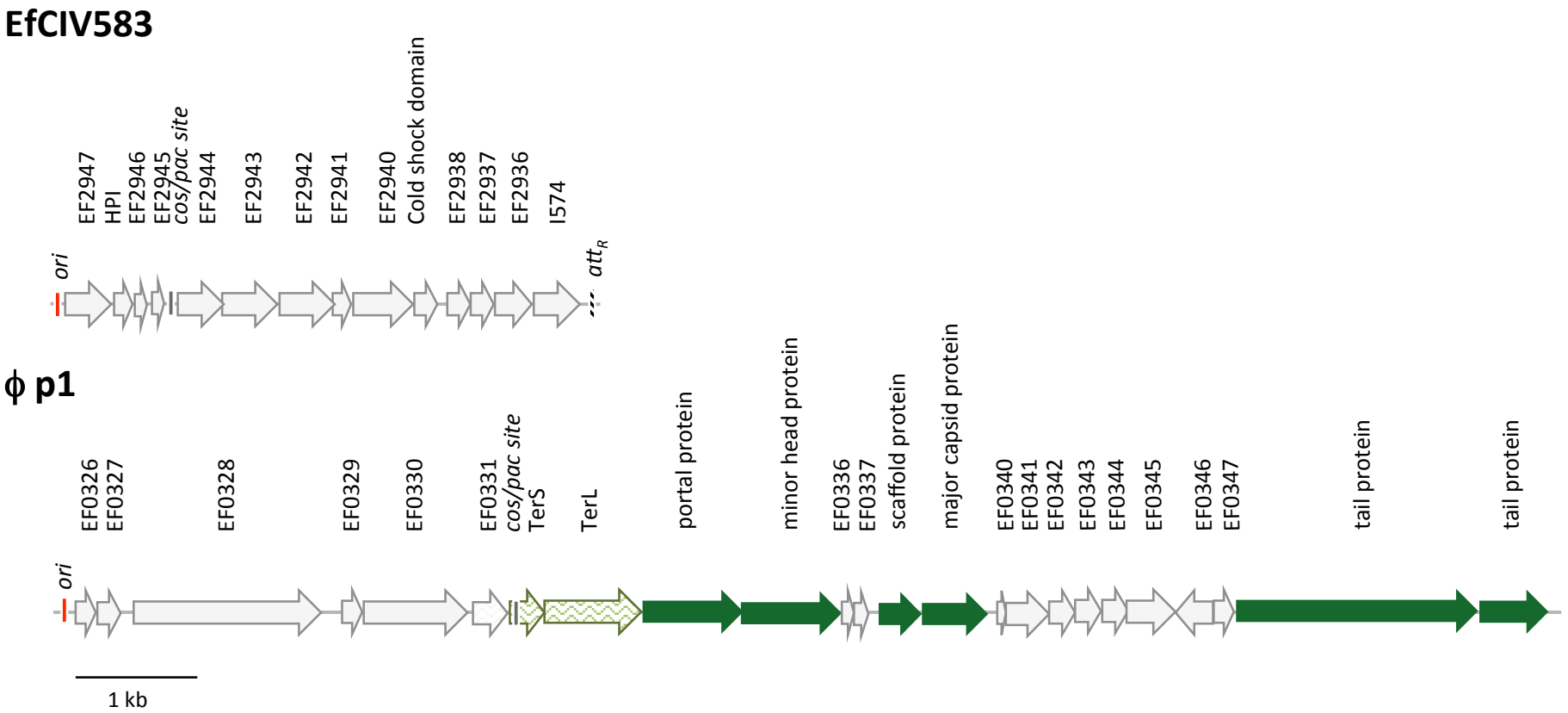


Figure 6.6 Genomic organisation of *E. faecalis* V583 prophage and PICI element. Genes of the putative morphogenetic modules of ϕp1 and EfCI583. Arrows indicate open-reading frames. Only genes encoding predicted genes functions are annotated.

6.2.3 EfCIV583 production of small capsids

Since PICIs show synteny, and because the genes involved in phage morphogenesis are always located between the PICI *ori* site and the virulence genes, it was supposed that the *cpm*-like gene(s) would be situated in a similar position in the EfCIV583 genome. To identify the protein(s) involved in the production of the EfCIV583-sized capsid, using the same strategy previously used for SaPIbov5, deletion mutants for each of the genes present in the putative morphogenetic module of the EfCIV583 island were generated. This putative region comprises 14 genes (Figure 6.6), of so far unknown function.

The EfCIV583 mutants were then introduced into the ϕp1 lysogen, and the ERC cycle was analysed after SOS induction. As shown in Figure 6.7, all the EfCIV583 mutants, except the mutants in the EF2942 gene and the EF2940 gene, generated the distinctive PICI band. The result for EF2942 suggests that this protein could be responsible for the capsid size redirection of EfCIV583, but also could be involved in another process of the ERC cycle since the 'bulk' band corresponding to the replication of the island was also affected. The EF2940 mutant, however, only was affected in the production of the small sized band, which corresponds to the monomer packaged into the small capsids. The transfer ability of these mutant strains compared to the wt did not reveal any difference in transfer titre except for the mutant in EF2942 where the island transfer was reduced (Table 6.3). However, the phage was blocked to the same extent in all cases compared to the wt phage (Table 6.3).

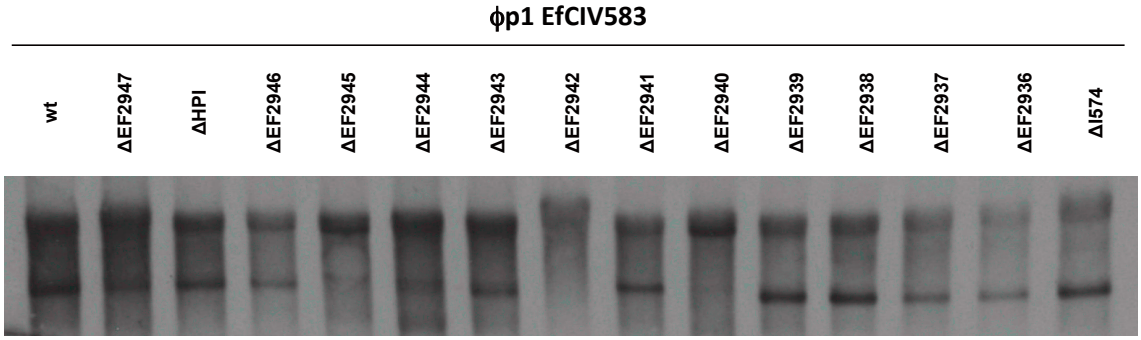


Figure 6.7 Replication analysis of EfCIV583 mutants. Southern blot of ϕp1 lysates, from strains carrying the wt or the different EfCIV583 mutants (carrying allelic mutations in the EfCIV583 genes EF2947 to EF2939). Samples were isolated 0 or 90 min after induction with MC, separated on agarose and blotted with an EfCIV583-specific probe. The upper band is 'bulk' DNA, and represents replicating EfCIV583. PICI monomer represents EfCIV583 DNA packaged in small capsids. EfCIV583 EF2940 corresponds to *cpm*E.

Table 6.3 The effect of EfCIV583 mutations on phage and PICI titres[a].

Donor strain			
Phage	**EfCIV583**	**Phage titre[b]**	**EfCIV583 titre[b]**
ϕp1	-	1.1×10^9	-
ϕp1	EfCIV583::*tet*M	6.94×10^6	3.3×10^4
ϕp1	EfCIV583::*tet*M ΔEF2947	5.50×10^6	1.75×10^4
ϕp1	EfCIV583::*tet*M ΔHPI	3.7×10^6	1.63×10^4
ϕp1	EfCIV583::*tet*M ΔEF2946	3.31×10^6	1.47×10^4
ϕp1	EfCIV583::*tet*M ΔEF2945	3.46×10^6	1.53×10^5
ϕp1	EfCIV583::*tet*M ΔEF2944	3.74×10^6	1.45×10^4
ϕp1	EfCIV583::*tet*M ΔEF2943	2.73×10^6	1.31×10^4
ϕp1	EfCIV583::*tet*M ΔEF2942	6.56×10^6	1.33×10^2
ϕp1	EfCIV583::*tet*M ΔEF2941	3.1×10^6	3.13×10^4
ϕp1	EfCIV583::*tet*M ΔEF2940	2.1×10^6	2.36×10^4
ϕp1	EfCIV583::*tet*M ΔEF2939	2.24×10^6	2.92×10^4
ϕp1	EfCIV583::*tet*M ΔEF2938	4.2×10^6	5.94×10^4
ϕp1	EfCIV583::*tet*M ΔEF2937	3.52×10^6	1.76×10^4
ϕp1	EfCIV583::*tet*M ΔEF2936	4.17×10^6	1.7×10^4
ϕp1	EfCIV583::*tet*M ΔI574	3.18×10^6	6.86×10^4

[a]The mean values of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]PFU/ml of lysate, using the non-lysogenic VE18590 as recipient strain.
[c]No. of transductants/ml induced culture, using VE18590 as recipient strain.

PCR analyses of excision and circularisation on EF2942 and EF2940 mutants were performed to clearly demonstrate that no other stages of the ERC cycle were affected. As can be seen in Figure 6.8, both mutants can excise from the host chromosome. This result is expected since EfCIV583 excises spontaneously from the chromosome in the absence of an inducing phage, as previously reported for other SaPIs[236]. The results obtained with the circularisation analyses were surprising. While they confirm that both mutants are circularised after the excision, an unexpected band was observed in the EF2942 mutant. The presence of this band, with a larger size, suggests an aberrant excision of this mutant, which could explain the reduction in replication observed in Figure 6.7 and Appendix 6. Given the previous results, it was hypothesised that just EF2940 is responsible for the remodelling of the capsid size of the ϕp1phage. This gene was renamed hereafter as CpmE (for capsid protein morphogenesis from EfCIV583).

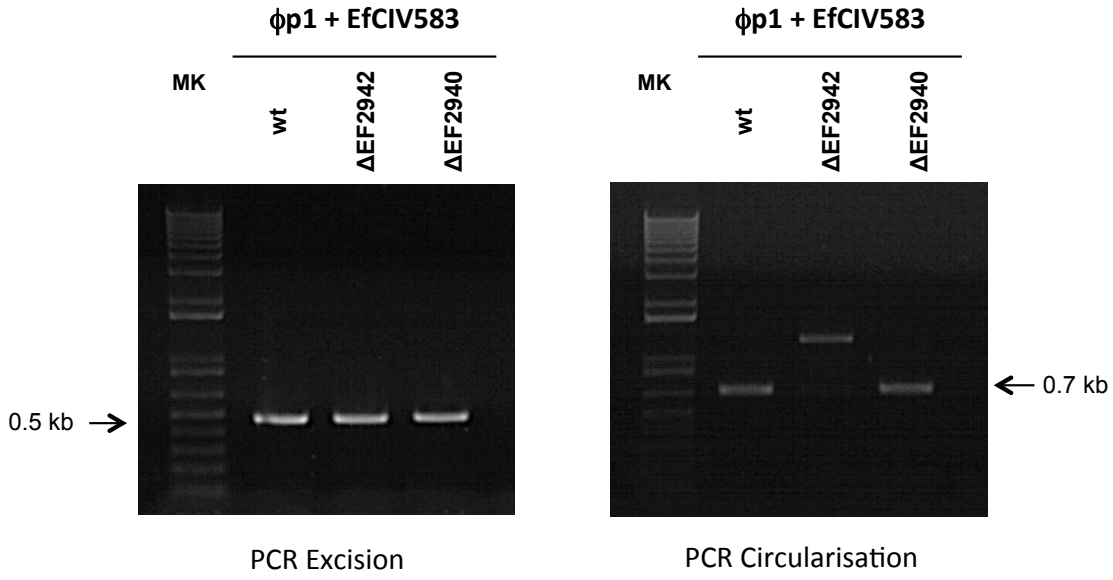


Figure 6.8 Detection of EfCIV583 excision and circularisation of EfCIV583 EF2942 and EF2940 mutants. Samples were isolated 0 or 160 min after induction with MC, and DNA from *E. faecalis* V583 and EF2942 and EF2940 mutants were extracted and PCR-amplified using specific primers recognising the flanking sequences of the island (excision) or PCR-amplified using a pair of primers set divergently at both termini of the island (circularisation).

In an attempt to determine the mechanisms for the size particle modifications, *in silico* analyses were performed to characterise the CpmE protein. Computational modelling analysis with I-TASSER[314] revealed that the

structural architecture of CpmE displays a similar fold to the scaffolding protein EF0338 of ϕp1 phage, despite the low sequence similarity (Figure 6.9).

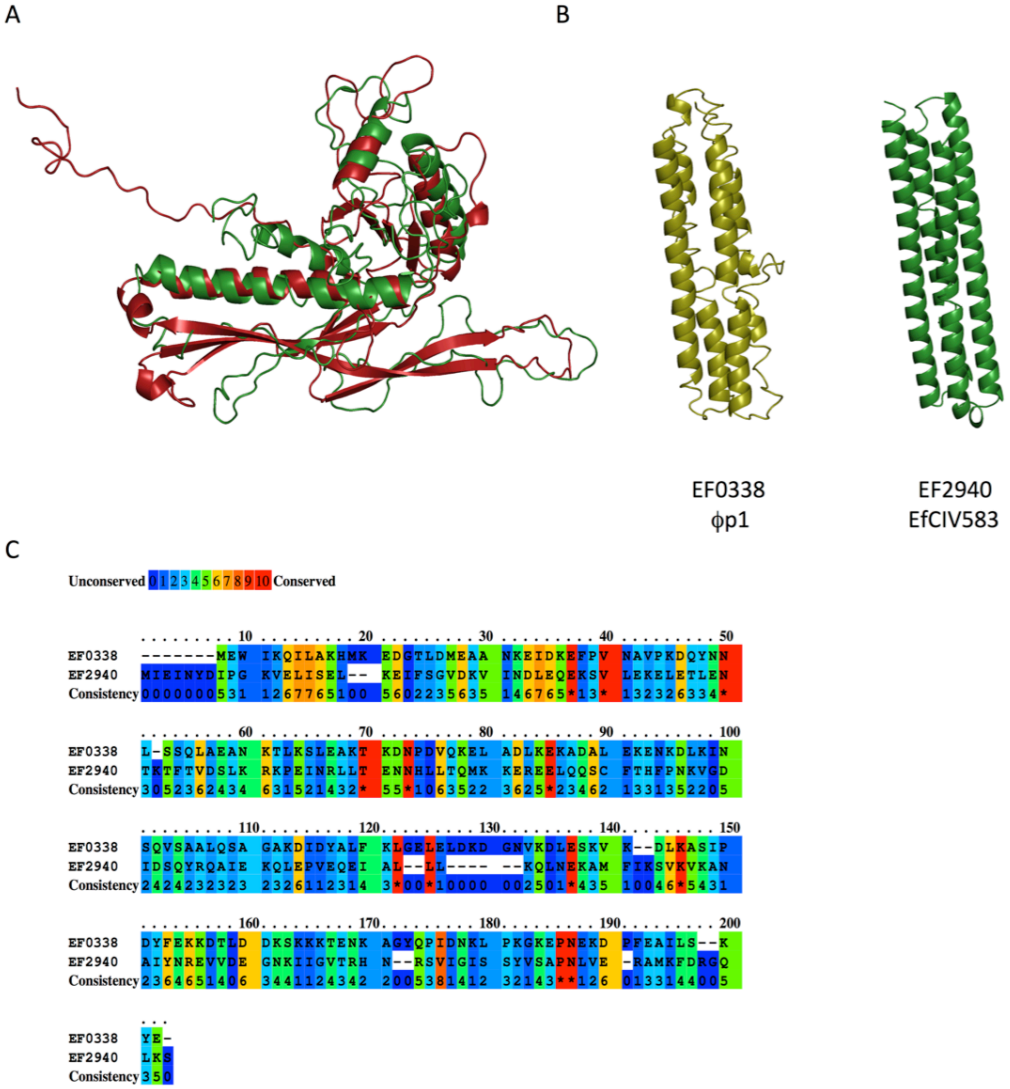


Figure 6.9 EfCIV583 remodels ϕp1 HK97-like fold capsid using an external scaffolding protein. (A) Superimposed modelling of the experimental structure of EF0339 (green) generated with I-TASSER is compared with the crystal structure of gp5 HK97 Prohead I (PDB2FS3; red). (B) Modelling of the experimental structures with I-TASSER of the scaffolding protein of ϕp1, EF0338 (yellow) and the proposed scaffolding protein of the EfCIV583 (EF2940). Models with higher C-score (-1.97 for EF0339; -3.02 for EF0338; -2.10 for EF2940) were selected for further structural analysis using PyMOL. (C) Alignment of the putative scaffolding proteins from ϕp1 (EF0339) and EfCIV583 (EF2940) using PRALINE multiple sequence alignment program.

As previously described for the *pac* and *cos* SaPIs, the production of small capsids is accomplished via coding morphogenetic genes responsible for the capsid size redirection[256,257]. Nevertheless, the EfCIV583 element does not

encode any homologs of these *pac* nor *cos* SaPI genes. Hence, one possible scenario could involve CpmE acting as a scaffolding protein, modifying the phage capsid protein (EF0339) into reduced ones. The phage-capsid of ϕp1 ($\phi p1_{EF0339}$) displays the same HK97-fold as the capsid protein gp5 of phage HK97 (Figure 6.9). In contrast, however, $\phi p1_{EF0339}$ does not encode a Δ-domain inside the capsid protein but instead encodes a protein expressed separately to achieve the scaffolding function. One possible model of action could be similar to the one established for *pac* SaPIs, where CpmB, an internal scaffolding protein mimicking the *pac* phage one, redirects the phage capsid assembly to form SaPI-size capsids[200,280]. From these studies, it could be suggested that EfCIV583, using CpmE, is packaged in small capsids via a scaffolding remodelling mechanism. As described here, however, it was not possible to determine the mechanisms by which these two proteins are performing the capsid size modification.

6.2.4 EfCIV583 interference with ϕp1

Related to the ϕp1-EfCIV583 interaction, it was of interest to identify the EfCIV583 encoded protein(s) responsible for the phage interference ability. This was addressed using the previously generated mutants for each gene present in the putative morphogenetic and virulence modules in the EfCIV583 island. These mutants were infected with ϕp1 lysate, and the capacity of the different mutants to block the formation of viable ϕp1 virion particles was assessed. Phage ϕp1 was not able to recover the ability to produce phage particles in the presence of the wt island and the different EfCIV583 mutants (Figure 6.10). This observation suggests that there are many mechanisms of interference present in the EfCIV583 island. Thus, the mutation of individual genes is not sufficient for the phage to escape from the EfCIV583 blockage.

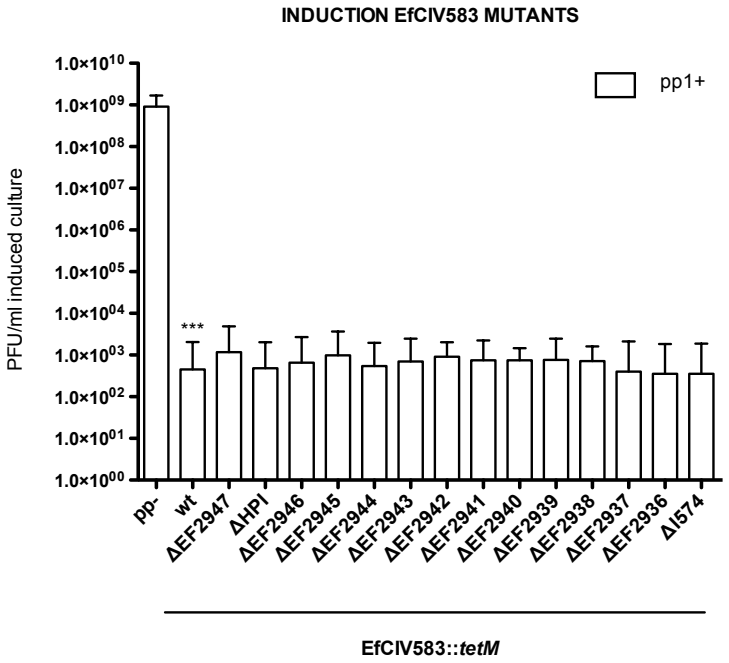


Figure 6.10 Interference of EfCIV583 mutations EF2947 to I574 with phage ϕp1 (pp1+). Strain VE18590 (pp-) containing wt or the different EfCIV583 mutants were infected with ϕp1, plated on phage bottom agar, and incubated for 48 h at 32°C. The mean values of results from ten independent experiments are shown. Variation was within ±5% in all cases. PFU/ml induced culture, using VE18590 as recipient strain. Statistical significance was determined by using the statistical Kruskal-Wallis test with Dunn's Multiple Comparison Test to determine significant grouping compared with the controls. VE18590 (pp-) vs. EfCIV583 wt ($p < 0,001$).

6.3 Discussion

The aim of these studies was to attempt to characterise the mechanism by which EfCIV583 exploits its helper phage ϕp1 to produce reduced size capsids, where the EfCIV583 reduced genome will be specifically packaged. The PICI family widely uses this convergent strategy with proteins with no homology between them, yet with a consistent capsid remodelling function. Within this context, it was suggested here that EfCIV583 could promote small capsid formation using the product of the self-encoded *cpm*E gene (EF2940). CpmE folds in a similar pattern as the scaffolding protein EF0338 of ϕp1.

These data suggest that a scaffolding protein could be driving the capsid size-reduction. This mechanism resembles the one used by *pac* SaPIs, where a distinctive internal scaffolding protein remodels the helper phage capsid. Several studies have elucidated the mechanism by which *pac* SaPIs achieve

the transformation of the phage capsid shell by two SaPI encoded proteins, CpmA and CpmB. In particular, CpmB acts as a scaffolding protein, which is assisted by CpmA protein, defining the structure of the capsid. In this thesis, it was further studied how *cos* SaPIs encode a pseudo-capsid protein coined as Ccm, with a shorter internal scaffolding Δ-domain, that could reshape the size and the form of the phage capsid. A novel convergent evolutionary mechanism is described here by which EfCIV538 package their reduced genome in small capsids via a putative scaffolding protein. Hence, PICIs have maintained this mechanism of phage capsid assembly remodelling process employing different scaffolding strategies, which will be beneficial for the PICI transfer and maintenance in nature.

It was further observed that mutation of the EF02942 resulted in an aberrant excision of this island, showing a different band size compared to the wt island. The exact role of this gene is not clear, and although it could be related to the formation of small capsids, since the small band in the Sothern blot studies was not present, further experiments will have to be conducted to decipher the precise function of this protein in the EfCIV583 cycle.

Further, the roles of the ϕp1 genes EF0332 and EF0333 in the packaging process of the phage and the island were investigated. Sequence identity analysis showed that these two phage genes could encode TerS and TerL proteins, which are components of the phage holoterminase complex. Structural analysis of these putative TerS and TerL showed that both proteins displayed a high degree of structural similarity to the classical phage terminases. Deletion of these ϕp1 genes not only eliminated the phage ability to form plaques but also prevented the EfCIV583 transfer. Interestingly, the infectious ability was not restored by their complementation, but did complement the island transfer. It was shown that EfCIV583 does not encode a *ter*S gene itself, but is likely to be mobilised by a putative *cos* or *pac* sequence homolog in its genome utilising the hosts' terminases. Sequence identity analysis of EfCIV583 and ϕp1 revealed that both feature an identical 13 bp nucleotide sequence. The sequence location and its high content in G and C nucleotides further indicate that this sequence could be a potential

packaging recognition site. This hypothesis is supported by the fact that ϕp1 was able to mobilise a plasmid encoding the TerS, which contained the 13 bp putative cleavage motif. The proposed packaging recognition site sequence, included in the *ter*S gene, could explain the inability of the phage to be complemented, as the phage needs this sequence to initiate cleavage and packaging of its genome. A less clear reason can be argued to explain the inability of the TerL mutant to be complemented. Moreover, ϕp1 was able to mobilise the TerL expressing plasmid but not the empty plasmid. Although other sequences involved in the packaging process could be included in the sequence of the *ter*L gene, a recombining process could better clarify the transfer of the TerL encoded plasmid[301]. It could not definitely be established that EF0332 and EF0333 are the TerL and TerS of phage ϕp1. Nevertheless, the results in this chapter strongly indicate that the proteins encoded by those two genes fulfil this function. Further studies have to be performed to determine if the proposed phage terminases are the phage TerS and TerL and to explain the TerL mutant phenotype.

Overall these results confirm that the PICI element EfCIV583 not only requires the helper phage ϕp1 for the production of a reduced capsid, but also for the activation of the ERC cycle. Moreover, this element does not encode a putative TerS, relying on the helper phage ϕp1 holoterminase complex for its mobilisation. This mechanism was previously described in this thesis for *cos* SaPIs. SaPIbov5 was shown to be packed and mobilised by encoding highly conserved cleavage sequences, both *pac* and *cos* sites, in its genome[270,330]. The results described in this chapter strongly indicate that EfCIV583 might utilise a similar mechanism.

Enterococcus phages have already been acknowledged for being active participants of virulence gene transfer, and having a major role in the adaptation of bacteria to different niches[353–355]. These novel PICI elements use similar convergent strategies with an evolutionary impact on host bacteria plasticity, becoming more prevalent in specific niches with the acquirement of virulence and fitness factors. Not surprisingly, PICI-like elements are not unique to staphylococci or enterococcus genera. Thus, those elements have

not been yet recognised as such and have been designated as defective prophages. In these studies, the *E. faecalis* EfCIV583 island, a novel member of the PICI family was further characterised, proving that this element uses the whole phage machinery to be encapsidated in small phage particles.

6.4 Conclusions and future work

The family of the PICIs contributes substantively to HGT, host adaptation and virulence[240,261,273]. PICIs are widespread all through bacterial biome, evolving separately but parallel to their helper phages. In this study it was reported that *E. faecalis* EfCIV583 island specifically packages its dsDNA into small phage-like infectious particles through CpmE, resulting in high transfer frequencies. A novel convergent evolutionary mechanism is described here, by which the EfCIV583 is packaged in a similar way as their related family members SaPIs. Although the genes involved are rather distant to the ones employed by *pac* or *cos* SaPIs, this element has evolved to achieve the same outcome, encoding unrelated scaffolding proteins. Future work will be focused on evaluating the role of the putative scaffoldings protein CpmE in order to identify its role in the interference and packaging of the EfCIV583 element. Further studies will address the role(s) of the other genes present in EfCIV583, and identification of other mechanisms of interference. Overall, these findings, along with SaPIs, illustrate the diverse mechanisms of piracy that these MGEs employ, which ultimately have an enormous impact on lateral gene transfer of virulence genes within bacteria.

Chapter 7 Concluding Remarks

7.1 SaPI mobilisation by *S. aureus cos* phages

Commensal and pathogenic bacteria evolve rapidly under selective pressures to adapt to different niches. Key players in this adaptation process are fitness and virulence factors provided by MGEs, which can be horizontally transferred between bacterial species. In *S. aureus,* transfer of MGEs is carried out at high frequencies[32,33,356]. In previous studies of *S. aureus* clonal strains, it was estimated that MGEs and non-mobile GIs comprise 25% of the *S. aureus* genome, improving its adaptability and ability to survive in different hosts[35,36]. This interchangeable and variable 'genome library' provides the bacteria with a selection of resistance and virulence genes that will ultimately be advantageous for invasion and survival within different niches and for adaptation to ecological pressures[33,273,351]. The different prevalence of virulence and fitness factors can define different clonal lineages.

In recent years, further insights into the family of SaPIs have been obtained. SaPIs were the first member of the PICIs to be characterised[13] and feature a high content of self-encoded virulence factors. Among these virulence factors are superantigens such as enterotoxins B and C, the toxic shock syndrome toxin-1 or antibiotic resistance genes. SaPIs are phage satellites, exploiting the phage life cycle for their own replication and transfer. The vast majority of characterised SaPIs depends on *pac* phages as their helper phages. These phages use a headful mechanism, which is relatively unrestrictive in terms of the properties of the encapsidated dsDNA, to package their genome. By this mechanism, over 100% of the actual phage genome size can be packaged into the procapsid and it requires only the presence of a single *pac* cleavage site to start the process[159,188–190]. By those means, any fragment of DNA containing this *pac* site or a pseudo-*pac* site, either phage, SaPI or foreign DNA, can be encapsidated by this mechanism.

To fully exploit this packaging process, *pac* SaPIs encode several proteins that contribute to the packaging specificity, also providing interference mechanisms against the helper phage. One of these mechanisms is SaPI encoded homolog of the phage TerS, $TerS_{SP}$, which specifically recognises the

SaPI *pac* site. The $TerS_{SP}$ forms a complex with the phage-encoded TerL and exclusively packages the island genome[241,254]. Besides, the SaPI encoded Ppi will be responsible for hijacking the phage-encoded TerS to facilitate $TerS_{SP}$ binding[255]. Likewise, *pac* SaPIs enables the formation of a size-reduced capsid via the scaffolding morphogenesis proteins, CpmA and CpmB, where *pac* SaPIs reduced genome will be encapsidated[256,257]. Finally, the recently described *pti*A, *pti*B, *pti*M genes modulate a novel interference mechanism through blocking expression of the helper phage late morphogenetic operon[259].

Among the first sequenced SaPIs, several islands present in strains from bovine, ovine and equine hosts were found not to encode a TerS or a CpmAB homologs, and the same setup was found in other SaPI-like elements from other species[239,240]. These SaPIs were long believed to be defective. However, as described in Chapter 3, SaPIbov5, the prototype of these SaPIs, was found to use *pac* and *cos*-like mechanisms to enable its own packaging. This element does not encode homologues of TerS or Cpm, yet is completely functional, being able to be induced, packaged and mobilised at high frequencies by different *pac* and *cos* helper phages. SaPIbov5 encodes homologues of phage *cos* and *pac* cleavage sequences that allow the SaPI to exploit the packaging machinery of phages. When induced by *pac* phages, such as ϕ80α or ϕ11, SaPIbov5 is packaged in full-sized phage particles by the *pac* phage terminase complex[270]. Interestingly, the putative *cos* site present in SaPIbov5 island was located and was found to be identical to the one present in *S. aureus* phage ϕ12. Not surprisingly, *cos* phages induced and packaged SaPIbov5, indicating that this island and probably other elements present in Gram-positive and Gram-negative bacteria could be packaged by *cos* phages. In support of this, several PICI-like elements were found to encode a *cos* site homologue, or to lack homologs of the *ter*S and *cpm* genes, suggesting that this *cos*-packaging strategy might be widespread among the PICIs universe.

Another significant finding of the present study is the requirement of an HNH endonuclease protein for the ϕSLT and ϕ12 *cos* phages, and hence SaPIbov5-*cos*-packaging. HNH endonucleases were found in several *cos* bacteriophages, infecting Gram-positive and Gram-negative bacteria[192,193]. In the studied

phages, the genes encoding the HNH proteins were located near the phage packaging module and were regulated by the same master regulator, RinA, which controls the expression of the late morphogenetic operon genes[153]. HNH overexpression confirmed that these *cos* phage HNH nucleases feature a nonspecific nuclease activity, contributing to the cleavage of the *cos* site during the packaging process. HNH and TerL were found to function together, suggesting that they could separately introduce single strands breaks to achieve *cos*-site cleavage. Supporting this theory, a single-strand nicking activity was established for another member of the HNH endonucleases[193]. Further, the occurrences of HNH proteins is 90% associated with the family of phage-TerL Terminase_1 (PF03354.10), and the combination of HNH-TerL interaction has also been established for the *cos* phage HK97, where a efficient cleavage occurs only when both proteins are present[194]. In addition, for the SPP1 phage model a TerS-TerL complex is vital for the phage nuclease function and viral maturation, protecting the concatemer to be degraded before packaging[329]. Our data confirmed that both ends of the linear ϕSLT DNA were unaffected by the host nucleases, and it can be hypothesised that the HNH–TerS–TerL complex could also be involved in this process.

These results establish that the HNH proteins are required for an efficient packaging of *cos* phages and in extension for *cos* SaPIs. A new model for phage packaging was proposed here, in which HNH proteins perform the classical nuclease activity in coordination with TerL, usually only associated with TerL. Besides the HNH protein, a supramolecular complex, composed of proteins of the morphogenesis module (TerS, TerL, capsid protein, portal, scaffolding, and others), is also required to perform phage dsDNA cleavage reaction. This mechanism was observed in further Gram-negative HNH nucleases encoding *cos* phages. Analysis of the cleavage process in *E. coli* coliphage ϕP27 confirmed that a supramolecular complex of the essential packaging proteins was required to accomplish viral maturation[270].

As previously described, *pac* phages were believed to be the sole participant in the transfer of MGEs and virulence-encoding genes, excluding *cos* phages from this process. During *cos* packaging a pair of specific *cos* site sequences

have to be situated a precise unit-length distance from the phage monomer for the terminase cleavage reaction to ocurr[160]. These requirements, more stringent than those for *pac* phage packaging, led to the false conclusion that *cos* phages could not engage in transduction. The findings described in Chapter 4 revealed that *cos* phages indeed not only participate in the horizontal transfer of self-encoded virulence genes but also mediate mobilisation and spread of other MGE, the *cos* SaPIs. The characterised *cos* dependent packaging is a novel transfer mechanism for *cos* SaPI that could implicate these elements in the transmission of fitness and viral factors to other bacteria. Several studies established the mobilisation of *pac* SaPIs by *pac* phages, not only among *S. aureus* species but also to other staphylococci and *Listeria monocytogenes*[245,250]. Likewise, this novel family of *cos* SaPIs characterised in this study was found to be also transferred from *S. aureus* to other staphylococci (*S. epidermidis* and *S. xylosus*) and also to *L. monocytogenes*, implicating *cos* phages in the transmission of MGEs for the first time.

The concept of 'silent transfer' was reviewed for the novel HNH-*cos*-phages. This mechanism relates to the ability of phages to transfer MGEs into target cells in which they cannot grow themselves. The host range of a phage is generally determined by its ability to lyse the particular bacteria, visualised by the formation of plaques. Nevertheless, the capacity to inject the dsDNA from the capsid into a cell is sufficient to transfer SaPIs, which might be able to integrate into host genomes that are not accessible for phage integration. Hence, the formation of plaques is no accurate method to determine the host range of a phage regarding the transfer of SaPIs and other MGEs. This silent transfer was shown for *pac* phages, transferring *pac* SaPIs to bacteria hosts the phages themselves were not able to parasitize[245,250]. *Cos*-phages, which have always been acknowledged to be poorly lytic and transducing phages, were able to participate in the transfer of virulence factors encoded in MGEs, even to host bacteria they were not able to lyse. The role of non-lysogen phages promoting the transduction of *cos* SaPIs to different bacterial hosts was never acknowledged. However, could indeed have a major impact on bacterial pathogen evolution.

Further, an evolutionary convergent assembling mechanism involving *cos* SaPIs and the packaging of small capsids was assessed. This capsid remodelling process was different but similar in concept to the one previously reviewed for *pac* SaPIs, which include the participation of CpmA and CpmB proteins[241,257,284]. CpmB acts as an internal scaffolding protein, responsible for the redirection of the phage capsid morphogenesis process towards the formation of small virion particles. After sequence and structural analyses, it was revealed that the product of the SaPIbov5 *ccm* gene is a HK97 capsid protein homologue. This capsid homologue (Ccm) is thought to participate somehow in the process of shell formation. *In silico* analysis of the Ccm protein revealed that the Δ-domain is 44 residues shorter than the one present in the helper phage $\phi12_{gp33}$ (capsid protein). Δ-domains are only found in HK97-fold capsid proteins of phages that do not encode a distinctive scaffolding protein. Several studies showed that the length of the N-terminal Δ-domain is related to the final size of the virion particle[348,349]. Moreover, it was established that Ccm alone was not able to form SaPI-sized capsids in the absence of the $\phi12_{gp33}$, being the phage capsid required for this purpose. Hence, it was proposed here that Ccm could form dimers with the phage capsid to promote this *cos* SaPI capsid assembling mechanism. Supporting this data, evolutionary experiments revealed $\phi12_{gp33}$ as the Ccm target. An evolved φ12 phage ($\phi12^{evolved4}$) surpassed the interference through a punctual mutation in the gp33, being able to escape the blockage of *ccm* gene product. Further, this Ccm-mediated mechanism only affected *cos* phages reproduction, since *pac* phages φ11 and φ80α capsid proteins lack this Δ-domain and use a distinctive scaffolding protein[200]. These results established that these two unrelated proteins, CpmB and Ccm, drive the formation of reduced size capsids, simulating the scaffolding functions of their helper phages. The Ccm-mediated mechanism of capsid size-reduction is a perfect example of a convergent mechanism that has been conserved during SaPI evolution to adapt its helper phage and reduce the phage infectivity.

Additionally, the data presented in Chapter 5 describes the mechanism employed by SaPIbov5 to interfere with the *cos* helper phages, severely blocking φ12 and φSLT reproduction. The phage interference was

predominantly mediated by the aforementioned Ccm protein, redirecting the assembling of the phage capsid into smaller sized particles. Other proteins encoded in the operon-I-like were found to be involved in the SaPIbov5 mediated-interference mechanisms. Mutation of SaPIbov5 ORF10 was found to restore $\phi12^{evolved4}$ titre, yet had no effect on the wt phage. However, the role of this protein remains obscure, and further studies will have to be undertaken. Moreover, SaPIbov5 ORF12 was found to be a homologue for PtiM. PtiM is encoded in *pac* SaPIs, where it modulates the blocking effect of PtiA towards the helper phage late gene expression operon[255,259]. However, the function of ORF12 in this system is unknown, and only minor effects were observed, suggesting a regulatory function when overexpressing this protein in the presence of Ccm. Further, the gene product of SaPIbov5 ORF8 encoded a homologue of *pac* SaPIs Ppi. In *pac* SaPIs, Ppi hijacks the phage TerS favouring the SaPI-encoded TerS packaging of SaPI dsDNA. However, no negative effect on the ORF8 expression on the reproduction of *cos* phage $\phi12$ could be established. This result is easily explained since *cos* SaPIs do not encode a $TerS_{SP}$ and therefore do need the helper phage holoterminase complex. The role in SaPIbov5 cycle of this Ppi homolog remains unknown. It was further observed that the SaPIbov5 operon I-like (SaPIbov5 ORFs 8 to 12) was conserved among *cos* SaPIs, suggesting that a major complex with all the proteins could be involved in the interference mechanism and the SaPIbov5 cycle.

The conserved assembly mechanism of reducing phage capsid size is not restricted to SaPIs, and is widespread in nature. For an example, *E. coli* P4 plasmid exploits the phage P2 capsid formation by encoding an external scaffolding protein Sid, which will remodel the capsid size to form small virion particles[350]. Furthermore, P4 enables its packaging into the produced small capsids via a homolog *cos* site of the P2 in its genome[288], an evolutionary convergent strategy similar to the one found here for *cos* SaPIs[270,330]. The novel PICI element found in *E. faecalis*, EfCIV583 island, also performs remodelling of the phage capsid protein, promoting the transfer of its reduced genome[239].

7.2 EfCIV583 mobilisation by *E. faecalis* ϕp1 phage

Up to 25% of the genome of the *E. faecalis* V583 strain is composed of MGEs, contributing significantly to the pathogenicity of the bacteria. The *E. faecalis* strain V583 was the first enterococci described to be vancomycin-resistant, mediated by a transposon-encoded gene[357,358]. This conjugative transposon was later found to be horizontally transferred to an isolate of *S. aureus*. Further, a large GI is also present encoding aggregation substances, cytolysin surface proteins and other proteins of unknown function related with fitness and virulence traits[358–360]. Additionally, the genome of V583 strain carries seven prophage-like elements, named as V583 -pp1 to -pp7. One of those elements encoded in the V583 strain was the EfCIV583 island. The EfCIV583 pathogenicity island, also described as pp7+[239], was proven to confer *E. faecalis* V583 strain with fitness factors providing an advantage in the competition with neighbouring strains in the intestine of mice depleted of normal microbiota[291].

Several studies established that EfCIV583 island exploits its helper phage ϕp1 to produce reduced size capsids[239], and proposed the product of the ϕp1 EF0309 gene to be the phage-encode antirepressor of the EfCIV583 *rpr* gene[240]. Likewise, EfCIV583 hijacks the formation of phage capsids, driving the assembling of small capsids. Due to those similarities to the relationship between SaPIs and staphylococcal helper phages, EfCIV583 was considered to be a PICI element[109,110,240,242]. The data presented in Chapter 6 proposed the EfCIV583-encoded CpmE as the protein responsible for the formation of reduced size capsids. Furthermore, in order to understand how the island is packaged, the morphogenetic genes in the helper phage ϕp1 were located. Among those encoded genes, the product of the $\phi p1_{EF0338}$ was positioned and identified as a scaffolding protein. The major capsid protein $\phi p1_{EF0339}$ was downstream the encoded $\phi p1_{EF0338}$. This putative capsid protein folds into a HK97-like pattern, and an external scaffolding protein assists the packaging process for correct capsid maturation. This data related with the putative assembling mechanism of the phage-encoded capsid proteins suggested the possibly that EfCIV583 was using a self-encoded scaffolding protein to perform

the small capsid remodelling function. It was proposed that CpmE has a scaffolding function in the capsid remodelling process. CpmE was found to fold in a similar fashion to the simulated structure of the scaffolding protein of ϕp1. This strategy could be comparable to the one observed in *pac* SaPIs, where CpmB remodels the *pac* phage structure as an internal scaffolding protein. However, the exact mechanism oh how the EfCIV58 CpmE drives the production of small capsids remains unclear.

Furthermore, it was proposed that the phage-encoded $\phi p1_{EF0332}$ and $\phi p1_{EF0333}$ proteins could be the functional homologues of the TerS and TerL holoterminase complex, respectively. It was found that EfCIV583 does not encode a TerS, and relies on the ϕp1 terminase complex for its mobilisation. Hence, it was suggested that the transfer of this island was enabled by a putative phage terminase cleavage sequence in the EfCIV583 genome. Once again the same mechanism was observed in the transfer of non-*ter*S-*cos* SaPIs, suggesting that this mechanism is widespread among PICI elements enabling for intra- and inter-generic mobilisation[270,330]. Overall these results established that the proteins encoded by the helper phage ϕp1 are not only required for the production of EfCIV58 reduced capsid particles, but also for the genome packaging process of this PICI-like element.

7.3 PICIs: a diverse family of mobile genetic elements

PICIs are a family of MGEs, which are widespread in nature. Although originally identified in *S. aureus*, novel PICI-like elements were lately characterised in other Gram-positive bacteria[240]. The PICI family co-evolved with their helper phages developing unique mechanisms of interference, which ultimately allow their transfer blocking phage reproduction. Several mechanisms have been elucidated in this thesis, illustrating how these elements interact with the helper phages. Remarkably, by reducing the virus load they are providing a significant benefit to their host bacteria. Moreover, PICIs contribute substantively to HGT, host adaptation and virulence.

In the case of the well characterised SaPIs, a vast and a diverse number of virulence and fitness factors were described, such as SEL or SEC toxins, SCIN,

TSST-1 or the vWbp, and are found in *cos* or *pac* islands. This diversity of virulence genes imposes a huge competition among *cos* or a *pac* SaPIs to be maintained in the host bacteria. This bacterial gene reservoir is highly mobile, yet the number of integration attachment sites (att_C) for SaPIs is limited and only 5 have been described to date[89,279]. However, the sequence specificity of SaPI integrases is less stringent than the ones of phage integrases, allowing *in vitro* SaPI integration at alternative att_C variants. This feature could help the mobilisation of SaPIs into different non-*aureus* bacteria, where the exact att_C site is not present[250]. The absorption of SaPI genomes is further assisted by presence of similar cell wall structures such as wall teichoic acids, which acts as an anchor receptor for the SaPI virion to inject their dsDNA[331]. These receptors lead to horizontal gene transfer of MGEs, helping bacteria to evolve with the implementation of fitness and virulence features, which are distributed in some cases even in distant phylogenetic genera.

Furthermore, the number of genes that can be encoded in the SaPI reduced genome is limited. Larger SaPI genomes cannot be packaged into small capsids, thus being encapsidated in phage particles and losing one of the most important mechanisms of phage-interference. As it was established in Chapter 5, the size limitation is more restrictive in *cos* SaPIs since a unit-length has to be maintained to complete the packaging process and be packaged into SaPI-size capsids. Thus, *cos* SaPIs have evolved to reserve ~2 kb of space to carry virulence genes. The carriage of virulence genes provides SaPIs with an essential advantage to compete and to be maintained in nature. The examination of the encoded fitness and virulence genes could be of clinical relevance in the combat against super resistant *S. aureus* clones.

The work conducted in this thesis characterised the participation of *cos* SaPIs, a novel family of horizontally mobile elements, in spreading of virulence and other significant accessory genes among bacteria. It was demonstrated that *cos* phages are capable of mobilising *cos* SaPIs via phage-mediated transfer both intra- and inter-genera. PICIs use a conserved strategy of skewing the phage reproduction machinery towards the production of small capsids, which can accommodate the small PICI genome, but are too small for the larger

phage genome. The mechanisms by which the subclass of *cos* SaPIs and the *Enterococcus* island EfCIV583 are packaged are described in this thesis for the first time. PICIs target the same outcome, the production of size-reduced particles, by targeting the scaffolding process during phage capsid assembling. Considering the extraordinary variety of bacteriophages in the biosphere and the diversity of the mechanisms by which PICIs disseminate, one can assume that many other processes by which PICIs exploit and interfere with phages have not yet been discovered. Further studies will improve our understanding of the impact PICI-mediated horizontal gene transfer has on the evolution of pathogen bacteria.

7.4 Future work

The data generated in this project and presented in this thesis demonstrated that SaPIbov5 and EfCIV583 procure the formation of small capsid shells via the action of unrelated scaffolding-like proteins. This evolutionary convergent mechanism allows PICIs not only to be exclusively package, but also to ultimately interfere with phage reproduction, reducing the phages infectivity. The genes encoding the proteins involved in these processes were identified, and the mechanisms behind the remodelling processes were hypothesised. However, how these proteins achieve the small capsid structure is yet obscure. Further experiments, following up on the data presented in this thesis, and leading towards a better mechanistic understanding of these morphogenetic proteins and the PICI-mediated interference mechanism should include:

- To investigate of the role of the ORF12 SaPIbov5 in the SaPIbov5 ERC cycle, a homolog of the previously characterised PtiM, which modulates of the interference mechanism through blockage of helper phage late gene expression operon[255,259], a pull-down assay could identify target proteins with the helper phage ϕ12 or SaPIbov5.

- To determine the role of the ORF10, which seems to engage some interference action with the helper phage ϕ12 in the SaPIbov5 ERP cycle. This interaction appears to be related to the $\phi12_{gp33}$ protein since

the $\phi12^{evolved4}$ was able to escape the blockage in SaPIbov5 ΔORF10. This result suggested that the target protein of the ORF10 could be the $\phi12_{gp33}$, the same as the Ccm so that could be involved in the modelling process acting as a decoration protein. Overexpression two-hybrid assays with Histidine-tag plasmids of the $\phi12_{gp33}$ and ORF10 should be done to corroborate this interaction. Furthermore, assembling of the Ccm with the presence and absence of the ORF10 will provide data to confirm the morphogenetic role in the capsid formation of SaPIbov5.

- To establish the mechanisms by which Ccm-mediates the redirection of the phage capsid and whether other factors might be involved, *in vitro* models following the different stages of the capsid maturation can be performed (Figure 1.4). HK79-fold capsids can be assembled in an *E. coli* model through the expression of the viral protease and the capsid subunit. Hence, maturation can be triggered and analysed *in vitro*, assessing the effect of individual proteins on the process[361]. Prohead I particles can be assembled without the protease and can be dissociated *in vitro* into single capsomers[48]. Co-expression of SaPIbov5 Ccm, $\phi12_{gp33}$ and the phage protease would result in the digestion of residues 1-102 of the Δ-domain of SaPI and phage capsid subunits. With this approach, the disposition of both proteins in the final virion particle could be achieved via crystallographic data and compared between SaPIbov5 and wt phage.

- The same strategy should be performed to identify the mechanism behind the putative scaffolding protein CpmE of the EfCIV583, which remodels the Prohead I of φp1. In this case, expression of three viral gene products, the protease, the capsid subunit and the scaffolding protein, will trigger maturation *in vitro*[361], and the remodelling process could be assessed and characterised.

- Analysis of the activity of the holoterminase complex of φp1 cleavage reaction and identification of the putative terminase cleavage site

should be performed to further characterise the TerS and TerL function in this ϕp1 phage.

Appendices

Appendix 1 - List of primers used in this study.

***Cos*-site packaging and phage-encoded HNH endonucleases**		
Plasmid	**Oligonucleotides**	**Sequence (5'-3')**
pJP1510	phi12-1mB	CGCGGATCCGCATCATACGATATTAAGCCA
	phiSLTp37-1c	GTATTGATATGACTTACGACC
	phiSLTp37-2m	GGTCGTAAGTCATATCAATACCTAATGTCAGTTTGTTATAGC
	phiSLTp37-3cE	CCGGAATTCTTCATCAAATACCCATTTACC
pJP1076	phiSLT-p36-5mB	CGCGGATCCTTGCAGAATAAAGAACTAACG
	phiSLTp37-1c	GTATTGATATGACTTACGACC
	phiSLTp37-2m	GGTCGTAAGTCATATCAATACCTAATGTCAGTTTGTTATAGC
	phiSLTp37-3cE	CCGGAATTCTTCATCAAATACCCATTTACC
pJP1240	phiSLTp37-4mB	CGCGGATCCTAAGTAAACGAGGCACATCGC
	phiSLTp36-10c	TAACTTCATATAAAGACCCCC
	phiSLTp38-4m	GGGGGTCTTTATATGAAGTTACAAGAAGAAGGTGGTTTTGGT
	phiSLTp38-5cE	CCGGAATTCCAACTTATCACTGTCTGAAGC
pJP1241	phiSLTp37-4mB	CGCGGATCCTAAGTAAACGAGGCACATCGC
	phiSLTp39-1c	CATTTAAAACTTTAATAGTCACC
	phiSLTp39-2m	GGTGACTATTAAAGTTTTAAATGAGTGGTGAAGGAAACATAGAG
	phiSLTp39-3cE	CCGGAATTCCACATCAATCGGACTAATGCC
pJP1242	phiSLTp40-1mB	CGCGGATCCTTCGCTAATAACGACGAAATG
	phiSLTp40-2c	AATTTTTTCTTTATGCGTGTGAC
	phiSLTp40-3m	GTCACACGCATAAAGAAAAAATTTTATACCCAATTGACACGCCAC
	phiSLTp40-4cE	CCGGAATTCTTTAGGTGTTTCAACCAATTC
pJP1243	phiSLTp41-1mB	CGCGGATCCTTTGATAATGCAGTAAGAACC
	phiSLTp41-2c	CCTTTACTTTTTGATTTTCTTTTC
	phiSLTp41-3m	GAAAAGAAAATCAAAAAGTAAAGGGAAAAAATTAAACGCGAATGC
	phiSLTp41-4cE	CCGGAATTCTACTAAATCTACATCTGATCC
pJP1244	phiSLTp41-5mB	CGCGGATCCTTATACCCAATTGACACGCCAC
	phiSLTp42-1c	TTCATATAATGTCGGCATTTC
	phiSLTp42-2m	GAAATGCCGACATTATATGAAGATCAGCAACGTACATTAGAC
	phiSLTp42-3cE	CCGGAATTCTAAAGCTCTATCACTCTTAGC
pJP1509	phiSLTp47-1mB	CGCGGATCCTTTAAAGATACGGGTGCTAGC
	phiSLTp47-2c	ATAAGAACCTTGTCCTTCTGC
	phiSLTp47-3m	CAGAAGGACAAGGTTCTTATGACAGTGAAGATCATTCAGAG
	phiSLTp47-4cE	CCGGAATTCTGCTTCGCCCATTGTCACATC
pJP1254	phiSLT-p36-5mB	CGCGGATCCTTGCAGAATAAAGAACTAACG
	phiSLTp37-14m	GATGCAAACATTGTAGCACACATTATTTATGTCGACGAAGATTTT
	phiSLTp37-15c	AAAATCTTCGTCGACATAAATAATGTGTGCTACAATGTTTGCATC
	phiSLTp37-3cE	CCGGAATTCTTCATCAAATACCCATTTACC
pJP1255	phiSLT-p36-5mB	CGCGGATCCTTGCAGAATAAAGAACTAACG
	phiSLTp37-16m	GATGCAAACATTGTACATGCAATTATTTATGTCGACGAAGATTTT
	phiSLTp37-17c	AAAATCTTCGTCGACATAAATAATTGCATGTACAATGTTTGCATC
	phiSLTp37-3cE	CCGGAATTCTTCATCAAATACCCATTTACC

*Underlined is shown the sequence recognised by the restriction enzymes used

Plasmid	Oligonucleotides	Sequence (5'-3')
pJP1512	phiSLTp39-26mS	ACGCGTCGACTCTATGGGTTTAACTGCAGCAAACTGCAGCA
	phiSLTp39-22m	AAATACACATATGGGTATTTTTGATGCAATTCATGAATTCAAAGATTATAAA
	phiSLTp39-23c	GAAATCAATTTATAATCTTTGAATTCATGAATTGCATCAAAAATACCCATAT
	phiSLTp39-27cB	CGCGGATCCCTCTTTTGCGATAACATTCACAACATTCAC
pJP1513	phiSLTp39-4mB	CGCGGATCCTCTATGGGTTTAACTGCAGCA
	phiSLTp39-24m	ATTTCCTTAGATGAGTTGGAAGGTCGACCATGTACTATAGGTTATGCATTA TCAGAAACAGAGGACTTT
	phiSLTp39-25c	TAAAGTCCTCTGTTTCTGATAATGCATAACCTATAGTACATGGTCGACCTT CCAACTCATCTAAGGAAAT
	phiSLTp39-5cE	CCGGAATTCCTCTTTTGCGATAACATTCAC
pJP1557	SaPIbov5-45mB	CGCGGATCCGAGGGACATATCTATACAGAG
	SaPIbov5-46c	CGCGTTGCAAGCGAAGGGTTTTTTTTTTTACCCGGCGAAAAAACATTTTA AGCCCATGGGCAGGGGGCTATATTTTTTAT
	SaPIbov5-47m	ATAAAAAATATAGCCCCCTGCCCATGGGCTTAAAATGTTTTTTCGCCGGG TAAAAAAAAAAACCCTTCGCTTGCAACGCG
	SaPIbov5-48cE	CCGGAATTCATCATCTCCGTCCCATTCACC
pJP1514	phi12p28-1mB	CGCGGATCCGGCACATCGCTATGCGGTGTG
	phi12p28-2cE	CCGGAATTCTAGGGGGCTATAAAAATAAAATTA
pJP1077	phiSLTp37-4mB	CGCGGATCCTAAGTAAACGAGGCACATCGC
	phiSLTp36-4cE	CCGGAATTCAAACATTTTAAGCCGATGGGC
pJP1245	phiSLTp38-6mB	CGCGGATCCGAGAGGCCCAAACGCTAGC
	phiSLTp38-7cE	CCGGAATTCGTTGTTAATAGTTTTGGTGAAGG
pJP1246	phiSLTp39-4mB	CGCGGATCCTCTATGGGTTTAACTGCAGCA
	phiSLTp39-5cE	CCGGAATTCCTCTTTTGCGATAACATTCAC
pJP1247	phiSLTp40-5mB	CGCGGATCCAGTGGTGAAGGAAACATAGAG
	phiSLTp40-6cE	CCGGAATTCCCTTTACTTTTTGATTTTCTTTTC
pJP1248	phiSLTp41-5mB	CGCGGATCCTTATACCCAATTGACACGCCAC
	phiSLTp41-6cE	CCGGAATTCTTCATATAATGTCGGCATTTC
pJP1249	phiSLTp42-4mB	CGCGGATCCCCTAAAGAAAGTATGTCACTAG
	phiSLTp42-5cE	CCGGAATTCGCTGATTTAGCCTTAGCTG
pJP1520	phiSLTp47-5mS	ACGCGTCGACCAGCGCATCGAATAGGTGTG
	phiSLTp47-6cB	CGCGGATCCCTGCTACTTCAACATTTTGGG
pJP1521	phiSLTp39-4mB	CGCGGATCCTCTATGGGTTTAACTGCAGCA
	phiSLTp39-18m	CACATATGGGTATTTTTGATGCAATTCATGAATTTAAAGATTA
	phiSLTp39-19c	TAATCTTTAAATTCATGAATTGCATCAAAAATACCCATATGTG
	phiSLTp39-5cE	CCGGAATTCCTCTTTTGCGATAACATTCAC
pJP1522	phiSLTp39-4mB	CGCGGATCCTCTATGGGTTTAACTGCAGCA
	phiSLTp39-20m	GACCATGTACTATAGGTTATGCATTATCAGAAACAGAGGACTT
	phiSLTp39-21c	AAGTCCTCTGTTTCTGATAATGCATAACCTATAGTACATGGTC
	phiSLTp39-5cE	CCGGAATTCCTCTTTTGCGATAACATTCAC
pJP1537	phiSLTp37-4mB	CGCGGATCCTAAGTAAACGAGGCACATCGC
	phiSLTp36-4cE	CCGGAATTCAAACATTTTAAGCCGATGGGC

*Underlined is shown the sequence recognised by the restriction enzymes used.

Plasmid	Oligonucleotides	Sequence (5'-3')
pJP1538	phiSLTp37-4mB	CGCGGATCCTAAGTAAACGAGGCACATCGC
	phiSLTp36-4cE	CCGGAATTCAAACATTTTAAGCCGATGGGC
pJP1523	phi12p27-1mS	ACGCGTCGACGGGCACATTATTGTTTGGT
	phi12p27-4cB	CGCGGATCCTCTTCACGTAAACACATTTGAC
pJP1524	phi12p27-1mS	ACGCGTCGACGGGCACATTATTGTTTGGT
	phi12p27-2c	AGCTCTTTCGTTGGGTTAA
	phi12p27-3m	TTAACCCAACGAAAGAGCTGAACTTTGTTAAAGCGGTAG
	phi12p27-4cB	CGCGGATCCTCTTCACGTAAACACATTTGAC
pJP836	phiSLTp36-7mB	CGCGGATCCGACATTAAGTTGCTTATAGCG
	phiSLTp36-8cE	CCGGAATTCCTCTCTTAACTTCTTCCATGC
pJP837	phiSLTp36-7mB	CGCGGATCCGACATTAAGTTGCTTATAGCG
	phiSLTp36-2c	GATATCATATATTGTGTTCCC
	phiSLTp36-3m	GGAACACAATATATGATATCAACTTTGTTAAAGCGGTAGCG
	phiSLTp36-8cE	CCGGAATTCCTCTCTTAACTTCTTCCATGC
pJP1539	phiP27-HNH-15mXbaI	GCTCTAGAAGTAACAGGCATTACAGCAGC
	phiP27-HNH-16cH	CCCAAGCTTTCACCTGATTCGTTCGCGCG
pJP1540	phiP27-terS-15mXbaI	GCTCTAGAAATTTTTACACCCGCGAAATT
	phiP27-terS-16cH	CCCAAGCTTTCATAGGTTTTTAAATGGATTATCGC
pJP1541	phiP27-terL-15mXbaI	GCTCTAGAGGGAGCGATAATCCATTTAAAA
	phiP27-terL-16cH	CCCAAGCTTTTAAAGCGAACGGATCCCGTA
pJP1542	phiP27p38-5mXbaI	GCTCTAGAGCACGATATCTCGACCGTACA
	phiP27p38-6cH	CCCAAGCTTTCAAGACGCTGTTTTGTCTGC
pJP1543	phiP27p39-6mXbaI	GCTCTAGAGATGAACATGACCACCAGTCC
	phiP27p39-5cH	CCCAAGCTTTTAAAATTTAAGATTTTTCAGTGCATTC
pJP1544	phiP27p40-5mXbaI	GCTCTAGACGCACTGAATGCACTGAAAAA
	phiP27p40-6cH	CCCAAGCTTTTACGCAGCGGCTTTCTGG
pJP1545	phiP27p47-5mXbaI	GCTCTAGATATTTCGCAGTTTACAGAACTG
	phiP27p47-6cH	CCCAAGCTTTTAACGCCTGAAAGGTGAATA
pJP1525	phiSLT-45mH	CCCAAGCTTTTTTATGCCCCCCTGCC
	phiSLT-46cB	CGCGGATCCGACCCCCTTTCATGAAAAAT
pJP1526	phiSLT-50mE	CCGGAATTCCCCCCCTGCCCATCGGCTTAAAATGTTTTTTCGCCGGGT AAAAAAAAAAACCAAACGCTAGCAACGCGGA
	phiSLT-51cB	CGCGGATCCCCCCTTTCATGAAAAATTTATCCGCGTTGCTAGCGTTTGG TTTTTTTTTTTACCCGGCGAAAAAACATTT
pJP1527	phi12-7mE	CCGGAATTCGCCCCCTACCCATCGGCTTAAAATGTTTTTTCGACGGGTA CCGGCGGGGGCCCTTCGCTTGCAACGCGGA
	phi12-8cB	CGCGGATCCCCCCCTTTCATAAAAGTTTATCCGCGTTGCAAGCGAAGG GCCCCCGCCGGTACCCGTCGAAAAAACATTT
pJP1528	phi12-9mE	CCGGAATTCGCCCCCTACCCATCGGCTTAAAATGTTTTTTCGACGGGTA AAAAAAAAAACCCTTCGCTTGCAACGCGGA
	phi12-10mB	CGCGGATCCCCCCCTTTCATAAAAGTTTATCCGCGTTGCAAGCGAAGG GTTTTTTTTTTTACCCGTCGAAAAAACATTT

*Underlined is shown the sequence recognised by the restriction enzymes used

Plasmid	Oligonucleotides	Sequence (5'-3')
pJP1529	SaPIbov5-41mE	CCGGAATTCGCCCCCTGCCCATTGGCTTAAAATGTTTTTTCGCCGGGTAC CGGCGGGGGCCCTTCGCTTGCAACGCGGA
	SaPIbov5-42cB	CGCGGATCCCCCCCTTTCATAAAAGTTTATCCGCGTTGCAAGCGAAGGG CCCCCGCCGGTACCCGGCGAAAAAACATTT
pJP1530	SaPIbov5-43mE	CCGGAATTCGCCCCCTGCCCATTGGCTTAAAATGTTTTTTCGCCGGGTAA AAAAAAAAACCCTTCGCTTGCAACGCGGA
	SaPIbov5-44cB	CGCGGATCCCCCCCTTTCATAAAAGTTTATCCGCGTTGCAAGCGAAGGG TTTTTTTTTTTACCCGGCGAAAAAACATTT
pJP1531	phip27-18mE	CCGGAATTCAGGGGCGGGTCAAATCCCTGCAACCCTGGCTGTCCGGGA CCGCCCGCCCCGTCAAATTTTTACACCCGCG
	phip27-19cB	CGCGGATCCAATCCTGAAATTTTTAATTTCGCGGGTGTAAAAATTTGACG GGGCGGGCGGTCCCGGACAGCCAGGGTTG
pJP1532	phip27-16mE	CCGGAATTCAGGGGCGGGTCAAATCCCTGCAACCCTGGCTGTCCGGGA CAAAAAAAAAAGTCAAATTTTTACACCCGCG
	phip27-17cB	CGCGGATCCAATCCTGAAATTTTTAATTTCGCGGGTGTAAAAATTTGACTT TTTTTTTTGTCCCGGACAGCCAGGGTTG
pJP1533	phiSLTp37-9mB	CGCGGATCCGATGACCAAGCATAATAACATT
	phiSLTp37-10cE	CCGGAATTCTTAAATTTTTAGAACTCTAATTTTC
pJP1534	phiSLTp37-9mB	CGCGGATCCGATGACCAAGCATAATAACATT
	phiSLTp37-10cE	CCGGAATTCTTAAATTTTTAGAACTCTAATTTTC
pJP1535	phiSLTp37-9mB	CGCGGATCCGATGACCAAGCATAATAACATT
	phiSLTp37-10cE	CCGGAATTCTTAAATTTTTAGAACTCTAATTTTC
pJP1253	phiSLTp39-7mB	CGCGGATCCGGTGGTTTTGGTGACTATTAAAG
	phiSLTp39-8cE	CCGGAATTCTTAACGCATTATGTCTTTAATAC
pJP1570	orf29phi11-21mB	CGCGGATCCTGGGTTGGCTGATTATAGCC
	orf29phi11-22cE	CCGGAATTCGTTAAAGTTTAATTTAACTTTCG

*Underlined is shown the sequence recognised by the restriction enzymes used

ϕP27 Mutagenesis	Oligonucleotides	Sequence (5'-3')
ϕP27 *stx*::*tet*A	phip27-1m	GGGTCTGGTACTGATTACCTTAGCCAAAAGGAATATATGTATA TGAAGTGTATATTGTTACGCTGTTAACTACTTTACTT
	phip27-2c	TGCGTCCAGAAACAAAAGACGCGCATAAATAAACCGTAGATT CTCAGTTAAACTTCACCTGGTTATCAAGAGGGTCATTA
ϕP27 *stx*::*tet*A *hnh*::*clor*	phip27-HNH-1m	TTCAGTGAGGGGCTGCGATAATGCCGGTATTAAGGAGATTCC AATGCCATCACGAATACCTGTGTAGGCTGGAGCTGCTTC
	phip27-HNH-2c	ATTTGACCCGCCCCTCCCCACAAGTGAGAATAATTATCACCT GATTCGTTCGCGCGCTGTCATATGAATATCCTCCTTA
ϕP27 *stx*::*tet*A *ter*S::*clor*	phip27-terS-5m	CCGCGAAATTAAAAATTTCAGGATTTGACATGTCAGGAAAATC TGTTGCGCCCGGAAGAGTGTGTAGGCTGGAGCTGCTTC
	phip27-terS-6c	TTTTTTCGTGTCATAGGTTTTTAAATGGATTATCGCTCCCCTG CTTCGGCGTCATAAGCCCATATGAATATCCTCCTTA
ϕP27 *stx*::*tet*A *ter*L::*clor*	phip27-terL-6m	CACGAAAAAATTACGTGAACGTTAACGCCGCAAATCAGTATG CACGTGACGTGGTTCGCGTGTGTAGGCTGGAGCTGCTTC
	phip27-terL-7c	GCGCGAGAATCAGCATGATCATAATTACCTCAGTTAAAGCGA ACGGATCCCGTAGGACTCCATATGAATATCCTCCTTA
ϕP27 *stx*::*tet*A *orf38*::*clor*	phip27-p38-1m	ACACGGCAGTCTGTCGGCGGAGGTAAATAGTGTTCTTTTCGG GGTTATTTCAACGAAAAATGTGTAGGCTGGAGCTGCTTC
	phip27-p38-2c	ACTGACGGATTTCAGGTTCAGCGGTATATCAAGACGCTGTTT TGTCTGCATCTCCACTCTCCATATGAATATCCTCCTTA

ϕP27 Mutagenesis	Oligonucleotides	Sequence (5'-3')
ϕP27 *stx*::*tet*A *orf39*::*clor*	phip27-p39-1m	TGACAACGGTAAGAAAAAGGAGAGTGGAGATGCAGACAAAA CAGCGTCTTGATATACCGCTGTGTAGGCTGGAGCTGCTTC
	phip27-p39-2c	TCTTTAACATCAGCCATTATTTTCTCCTGGTTAAAATTTAAGAT TTTTCAGTGCATTCAGCATATGAATATCCTCCTTA
ϕP27 *stx*::*tet*A *orf40*::*clor*	phip27-p40-1m	TGAAAAATCTTAAATTTTAACCAGGAGAAAATAATGGCTGATG TTAAAGATGTGGAACAGTGTGTAGGCTGGAGCTGCTTC
	phip27-p40-2c	GCCCGCCATTACGCGGGCACAAAAAAAACCGCATTACGCAG CGGCTTTCTGGCGGGTTGCCATATGAATATCCTCCTTA
ϕP27 *stx*::*tet*A *orf47*::*clor*	phip27-p47-1m	GAACTGGAAGCCGAGGCCGTCCGCATTAACGAGGAGATGAA GCATGGCTGACAGTTTTCATGTGTAGGCTGGAGCTGCTTC
	phip27-p47-2c	AAGGTAAGGAAGGAAAATCCATTCTGATACCCCATTATTAAC GCCTGAAAGGTGAATACCCATATGAATATCCTCCTTA

Intra- and inter-generic SaPI transfer by cos phages		
PCR/plasmid	**Oligonucleotides**	**Sequence (5'-3')**
Integration	Ipr-13mP	AAACTGCAGGAAGTTCAAGACAGATTAGGTCACGG
	Ipr-2c	TTGGTGGATTACCAGAAGACATGG
stl	SaPIbov att-9cE	CCGGAATTCATTGAGTGGGAATAATTATATATAGC
	SaPIbov1-203mB	CGCGGATCCCATGGAAGGAGCTGGTCAAATGGC
pri	SaPIbov1-215cB	CGCGGATCCAATTGTGCCGTAATGTGTTGG
	SaPIhor2-12c	CTTCATCAAATTGATACATAGC
ori	SaPIhor2-7m	ATAGTATTAATGGTTACAGAGC
	SaPIbov5-22c	ATACCCAAATTGTTCTTTAAACTC
cos	SaPIbov5-54mB	CGCGGATCCTTTCTTCATAACATTTCGTTCTATGA
	SaPIbov5-14c	TGTCATAATTATTCTCCTATC
scn	SaPIbov5-28mB	CGCGGATCCAGATAAGCATGATATTAACGGG
	SaPIbov5-38cS	ACGCGTCGACATTTTTCATTACCATTAAATAAATTTATATACCTG
vwb	SaPIbov5-18m	ATAGGAGAATAATTATGACATACTAAAAACACCCATTTGGG
	SaPIbov att-8mH	CCCAAGCTTTATATATAATTATTCCCACTCAATG
pJP1511	phi12p29-1mB	CGCGGATCCGGACCGTTGCAAAAAGG
	phiSLTp29-2c	CCTCTATATATTCTTTCAGCTGTTT
	phi12p29-3m	AAACAGCTGAAAGAATATATAGAGGAGAAGAAGGTGGATTCGG
	phi12p29-4cS	ACGCGTCGACATGAATTTCATCAAAAATGCCC

*****Underlined is shown the sequence recognised by the restriction enzymes used.

SaPIbov5-mediated phage interference via small capsid production		
Plasmid	**Oligonucleotides**	**Sequence (5'-3')**
pJP1712	SaPIbov5-29mE	CCGGAATTCGCATAATGACTACCGCATTAG
	SaPIbov5-30cB	CGCGGATCCATTTTTCATTACCATTAAATAAATTTATATACCTG
	SaPIbov5-31mPstI	AACTGCAGAAATTAATAAGGTGTATAACGGTAAATAG
	SaPIbov5-32sEcoRV	AAGATATCGTAATCATAACATAAAAGCACCCC
pJP1761	SaPIbov5-57mE	CCGGAATTCTATGTAAAGGTGTGGTTATTTGG
	SaPIbov55-58cB	CGCGGATCCCTATGAAAGTTCTTTTTTAACTTGT
	SaPIbov5-59mPst1	AACTGCAGGTTATGATATGTTATACATGCATCA
	SaPIbov5-60cEcoRV	AAGATATCAATGAAATAAGTATTTTTCGTTAGTTTG
pJP1757	SaPIbov5-90mB	CGCGGATCCTTCAGTCACACCTATAAACACTG
	SaPIbov5-91c	TTTGAATTCTTGTTCTTAAAATACATT
	SaPIbov5-92m	AATGTATTTTAAGAACAAGAATTCAAA
	SaPIbov5-93cS	ACGCGTCGACTCATTTGTATGATCTAGCACGTC
pJP1758	SaPIbov5-78mB	CGCGGATCCGAAGAACTTGCAGTTACTCAAAA
	SaPIbov5-75c	GATTATTTGGAATTCGATTCTTTAATGATT
	SaPIbov5-76m	AATCATTAAAGAATCGAATTCCAAATAATC
	SaPIbov5-77cS	ACGCGTCGACAAAGTTAGATAGTTGGTACCCAA
pJP1759	SaPIbov5-81mB	CGCGGATCCAAGGTCTAAGGCTACCAATTGT
	SaPIbov5-94c	TACGTTCGTTGTTTTGAATTCTTCTTAGAACGAAATGTTA
	SaPIbov5-95m	TAACATTTCGTTCTAAGAAGAATTCAAAACAACGAACGTA
	SaPIbov5-96cS	ACGCGTCGACGATAGCTCAATACCATTAGTAGCTA
pJP1717	SaPIbov5-97mB	CGCGGATCCGTAAGGCGCTATACAAAAAAGC
	SaPIbov5-98c	TCTCACTACTCACTAAGAATTCCTGTTTATGTACCTACCCATG
	SaPIbov5-99m	CATGGGTAGGTACATAAACAGGAATTCTTAGTGAGTAGTGAGA
	SaPIbov5-100cS	ACGCGTCGACCAACAGATACTTTTCGTCTAAACG
pJP1760	SaPIbov5-65mB	CGCGGATCCCGTTTAGACGAAAAGTATCTGT
	SaPIbov5-66c	ATTGTAGAATTCTAAAAGTTCTACTTAAGTGATTTTATC
	SaPIbov5-67m	GATAAAATCACTTAAGTAGAACTTTTAGAATTCTACAAT
	SaPIbov5-68cS	ACGCGTCGACTAACTCCTCATCATCTCCGTCC
pJP1749	SaPIbov5-78mB	CGCGGATCCGAAGAACTTGCAGTTACTCAAAA
	SaPIbov5-87cXmaI	CCCCCCGGGTTAATTTACTAAACTAAATGTTACTG
pJP1748	SaPIbov5-81mB	CGCGGATCCAAGGTCTAAGGCTACCAATTGT
	SaPIbov5-82cE	CCGGAATTCTCATTTGTATGATCTAGCACGTC
pJP1747	SaPIvov5-80mB	CGCGGATCCGTATAGATATGGGAGTGTGTATATT
	SaPiBov5-88cXmaI	CCCCCCGGGTTAAACTCTTTCATAATCTTCATCTC
pJP1730	SaPIbov5-54mB	CGCGGATCCTTTCTTCATAACATTTCGTTCTATGA
	SaPIbov5-55cE	CCGGAATTCTTTCATAAAAGTTTATCCGCGTT
pJP1745	SaPIbov5-64mB	CGCGGATCCGCAACGCGGATAAACTTTTATG
	SaPIbov5-56cE	CCGGAATTCCTATGAAAGTTCTTTTTTAACTTGT

*Underlined is shown the sequence recognised by the restriction enzymes used.

Characterisation of the putative morphogenetic cluster in EfCIV583		
Plasmid	**Oligonucleotides**	**Sequence (5'-3')**
pJP1744	SaPIbov5-54mB	CGCGGATCCTTTCTTCATAACATTTCGTTCTATGA
	SaPIbov5-56cE	CCGGAATTCCTATGAAAGTTCTTTTTTAACTTGT
pJP1750	SaPIbov5-78mB	CGCGGATCCGAAGAACTTGCAGTTACTCAAAA
	SaPIbov5-103cE	CCGGAATTCCTATGAAAGTTCTTTTTTAACTTGT
pJP1769	SaPIbov5-54mB	CGCGGATCCTTTCTTCATAACATTTCGTTCTATGA
	SaPIbov5-55cE	CCGGAATTCTTTCATAAAAGTTTATCCGCGTT
pJP1719	EF2947-1mB	CGCGGATCCTATTTATTAGGGGATAGTTGGATAA
	EF2947-2c	CAAAATTCGGTTCTTTTTCTG
	EF2947-3m	CAGAAAAAGAACCGAATTTTGGTATCTTGATAGTCAACCCAA
	EF2947-4cS	ACGCGTCGACGTATTCATACTTTTTCACACCTAAT
pJP1718	HPI-1mB	CGCGGATCCCAACAGAAAAAGAACCGAATTTT
	EF2946-2c	CTGGTTTTTTCTCTCTAATCAT
	EF2946-3m	ATGATTAGAGAGAAAAAACCAGAAAAAGATTGAAGAAGAAATAAC
	EF2946-4cS	ACGCGTCGACCATCAACATGAGCTAAAACTCC
pJP1731	HPI-1mB	CGCGGATCC CAACAGAAAAAGAACCGAATTTT
	HPI-2c	CTTTGAAAAATTTTAACTTTCAT
	HPI-3m	ATGAAAGTTAAAATTTTTCAAAGTATTTGGTATGAGGAGGAATAA
	EF2946-4cS	ACGCGTCGACCATCAACATGAGCTAAAACTCC
pJP1733	EF2945-7mS	ACGCGTCGACGGAAAGGAGGAGATCGTTATTAT
	EF2945-2c	TTATCAATGTATACTTTATCCAT
	EF2945-3m	ATGGATAAAGTATACATTGATAAGCATGTTTTACCTAAGAATTAGC
	EF2945-5cSphI	CATGCATGCCCGCTCAAAATGTTTTTGTCG
pJP1734	EF2944-1mS	ACGCGTCGACCTAAAGAGCTACTTAACAACGTG
	EF2944-2c	ACACTAACGGAAATGAGAGATA
	EF2946-3m	TATCTCTCATTTCCGTTAGTGTGAGCTTTCCGACGATGTAGA
	EF2946-5cSphI	CATGCATGCATTTCCAAGAACAAACTGTTCT
pJP1735	EF2943-1mB	CGCGGATCCGTGTTGCAAAAGTTGCTTTCAA
	EF2943-2c	TATATCCAGCGTGCAAGATGA
	EF2943-3m	TCATCTTGCACGCTGGATATAGACACAATTAGAAAGTTTATG
	EF2943-4cS	ACGCGTCGACCAATAAAACTTAGCTAGCCTGT
pJP1736	EF2942-1mB	CGCGGATCCAAATTTAACCGACAAACAGGA
	EF2942-2c	CTTTTTTTACTCCTACTAAACCG
	EF2942-3m	CGGTTTAGTAGGAGTAAAAAAAGAAGAAATATGACAGCGGAA
	EF2941-4cS	ACGCGTCGACTTATGGCGAGTAACTCCGATGA
pJP1737	EF2941-1mB	CGCGGATCCAGTCCAATATCAATAGCATTAG
	EF2941-7c	TTCATCTACAAATTCCTCTAAAACTTTT
	EF2941-8m	AAAAGTTTTAGAGGAATTTGTAGATGAAGAAAGAAGATAAACCTGATCCATTTAAAA
	EF2941-4cS	ACGCGTCGACTTATGGCGAGTAACTCCGATGA

*Underlined is shown the sequence recognised by the restriction enzymes used.

Plasmid	Oligonucleotides	Sequence (5'-3')
pJP1738	EF2941-1mB	CGCGGATCCAGTCCAATATCAATAGCATTAG
	EF2940-1c	GCTCAACTTTTCCCGGAATATC
	EF2940-2m	GATATTCCGGGAAAAGTTGAGC AAAGAGCTATGAAATTTGATC
	EF2940-3cS	ACGCGTCGACGAATAGGACACTCGCAACAACT
pJP1739	EF2939-1mB	CGCGGATCCAGCTTATTTCCGAGCTAAAAGA
	EF2939-2c	TTTTAACTGTTCCAGTTTTCAC
	EF2939-3m	GTGAAAACTGGAACAGTTAAAAAATGTTACTGTAATTGAATAG
	EF2939-4cS	ACGCGTCGACACATAAAAAACCTCCAATCTCAC
pJP1740	EF2938-1mB	CGCGGATCCAAAGAGCTATGAAATTTGATCGT
	EF2938-2c	CAACAACTGCTGATCGATTAT
	EF2938-3m	ATAATCGATCAGCAGTTGTTGTTTATTTTGGCAATGTCTTTAGG
	EF2938-4cS	ACGCGTCGACCTGTTAACAGAATAGTCCAGAAT
pJP1741	EF2938-1mB	CGCGGATCCAAAGAGCTATGAAATTTGATCGT
	EF2937-1c	AAATCTATCCCTATAGGTAACAT
	EF2937-2m	ATGTTACCTATAGGGATAGATTTAGAATGAAAAAGGGGGATTA
	EF2937-3cS	ACGCGTCGACAAACTATCACACTGACTAACCAC
pJP1742	EF2936-1mB	CGCGGATCCTTTTTCTACAATCCCTATCATGG
	EF2936-2c	AGTTCATCAGGAACCTTCGAC
	EF2936-3m	GTCGAAGGTTCCTGATGAACTATTAAAAATAAACTTTTGGAGAGC
	EF2936-7cS	ACGCGTCGACCGCCAATAGTTACATATTACGT
pJP1743	I574-1mB	CGCGGATCCTGTTATTACAGAGGCAATGATT
	I574-2c	ACTTGACATAAAGTGTTAGATCG
	I574-3m	TCTAACACTTTATGTCAAGTGGTATAACTATTCTTTTCCC
	I574-4cS	ACGCGTCGACAGAGACTGAATAAGCAAAGAAT
pJP1710	phi1-p32-1mB	CGCGGATCCGCTATCCGCATTATTAACTCTAATTC
	phi1-p32-2c	CACTCTGTGTACTTCGCCAT
	phi1-p32-3m	ATGGCGAAGTACACAGAGTGTCGAGGATGATTCTACATGA
	phi1-p32-4cS	ACGCGTCGACGAAATTGCATCTTCCTTATGCCA
pJP1711	phi1-p33-5mB	CGCGGATCCGGCTATCGTTAAAACAAATTGCGG
	phi1-p33-6c	AGTATTTTCGCTTTTTAGGGCTCAT
	phi1-p33-7m	ATGAGCCCTAAAAAGCGAAAATACTACACAAATAAATGGCTATACTAG
	phi1-p33-8cS	ACGCGTCGACCCTGTTTATTATTTGATAACCGAT
pJP1775	phi1-p32-5mS	ACGCGTCGACTTCCCATTGATAAATCGTGATAT
	phi1-p32-6cB	CGCGGATCCTCATGTAGAATCATCCTCGAAAA
pJP1774	phi1-p33-1mS	ACGCGTCGACGTTGAAAAATAGAAAGCCGGA
	phi1-p33-2cB	CGCGGATCCCTAGTATAGCCATTTATTTGTGT

*Underlined is shown the sequence recognised by the restriction enzymes used.

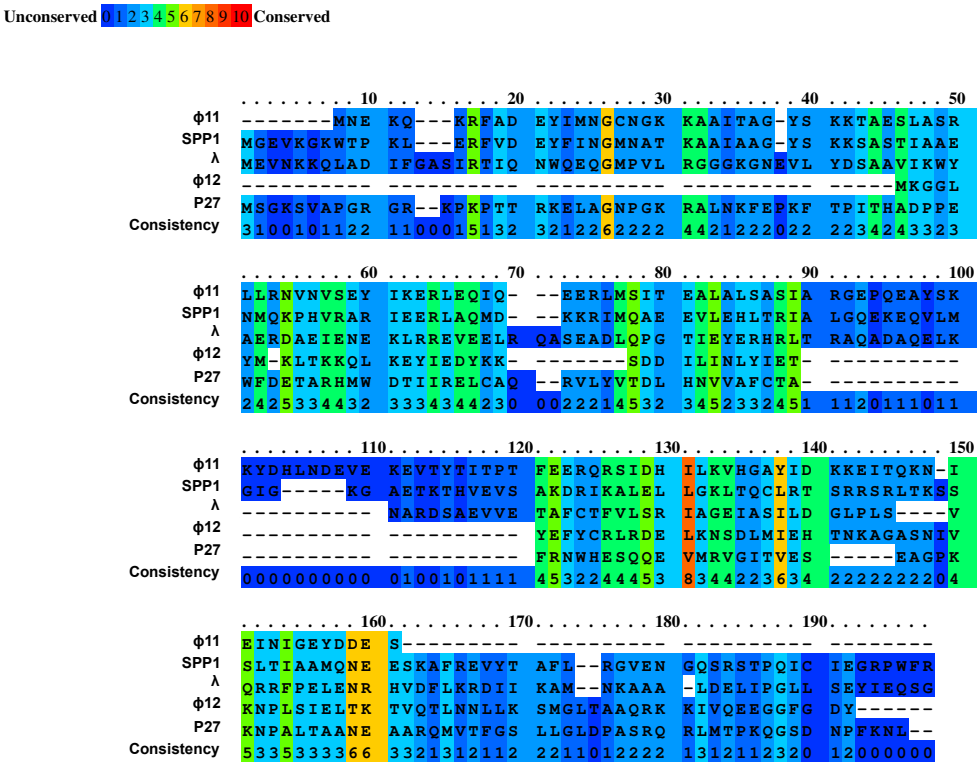
PCR	Oligonucleotides	Sequence (5'-3')
Excision PCR	EfPI-1m	TAAAAACAGCGCCTTCGTCC
	EfPI-2c	AATCGAGTAGTAGCTGAAACG
Circularisation PCR	EfPI-8m	CTTCTTCAATCAGGAGTGCC
	EfPI-12c	TATGGTTGGTACTGATAGGCG

Primers used for sequencing plasmid constructions		
Plasmid	**Oligonucleotides**	**Sequence (5'-3')**
pCN51	pCN51-1m	TAGGTGATGAACATATCAGGC
	pCN51-2c	CAGTATTTATTATGCATTTAGAATA
pCN42	pCN42-1m	TAGGTGATGAACATATCAGGC
	pCN42-2c	CAGTATTTATTATGCATTTAGAATA
pBAD-18	pBAD-18-1m	CTGTTTCTCCATACCCGTT
	pBAD-18-2c	CTCATC CGCCAAAACAG
pCU1	Direct	GTAAAACGACGGCCAG
	Reverse	CAGGAAACAGCTATGAC
pPROEX HTa	pPROEX HTa-1m	GTATTTTCTCCTTACGCATCTGTGC
	pPROEX HTa-2c	TCACGTAGCGATAGCGGA
pMAD	pMAD-2m	AAGCGAGAAGAATCATAATGGG
	pMAD -4c	CTAGCTAATGTTACGTTACAC
pBT_2-βgal	βgal-2c	TGGCCATTGCTCTGGGTTATAAT
	pBT_2-βgal -2c	GGTGCCGAGGATGACGATGAGCGC

Primers used for Southern blot probes		
Southern	**Oligonucleotides**	**Sequence (5'-3')**
SaPIbov1/5 probe	SaPIbov1-112mE	CCGGAATTCAATTGCTGAGGCAAAACTTC
	SaPIbov1-113cB	CGCGGATCCTAATTCTCCACGTCTAAAGC
ϕSLT probe (*rin*A)	phiSLTp36-1mB	CGCGGATCCATCAGAGCTGAAGTTTCATGG
	phiSLTp36-2c	GATATCATATATTGTGTTCCC
ϕSLT probe (*ter*S)	phiSLTp38-9mB	CGCGGATCCGATGAAAGGGGGTCTTTATATG
	phiSLTp37-5c	AATAAAATGTTGCCAAGGCTG
ϕP27 probe (*ter*S-*ter*L)	phip27-terL-2c	CGCGTTGGGATTACGGTAGAG
	phiP27-terS-7c	CACTCCCCTTTGGTGT
EfCIV583 probe (*int*)	EfCI- EF2954-1mB	CGCGGATCCTATAGAACGAGTATTGCTTGTGTCC
	EfCI- EF2954-2c	GAAAAAAAACGTAAATCAAAAGCT

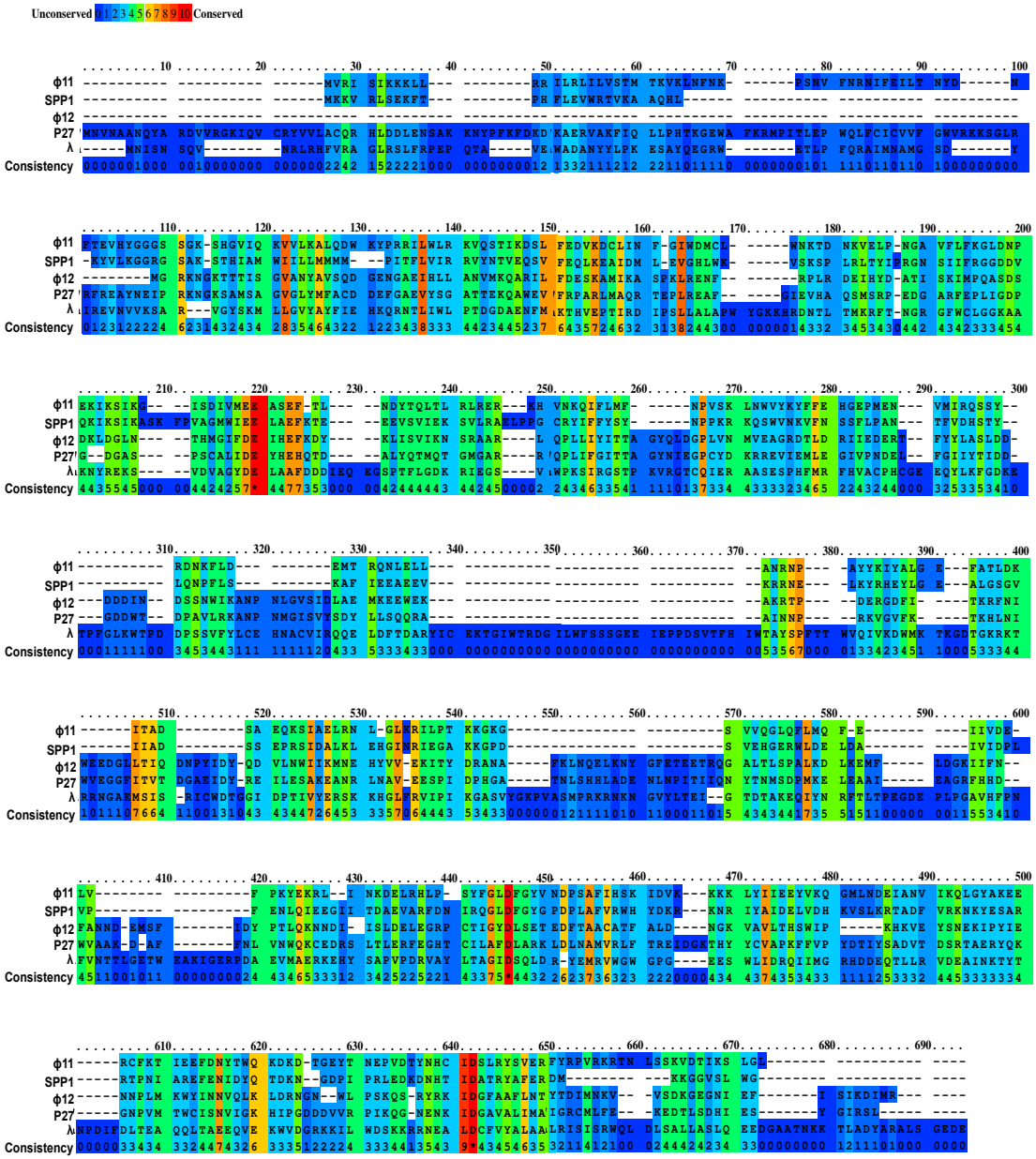
Appendix 2 - Alignment of TerS, TerL or portal protein amino acid sequences from phages ϕ11, SPP1, λ, ϕ12 and ϕP27.

A



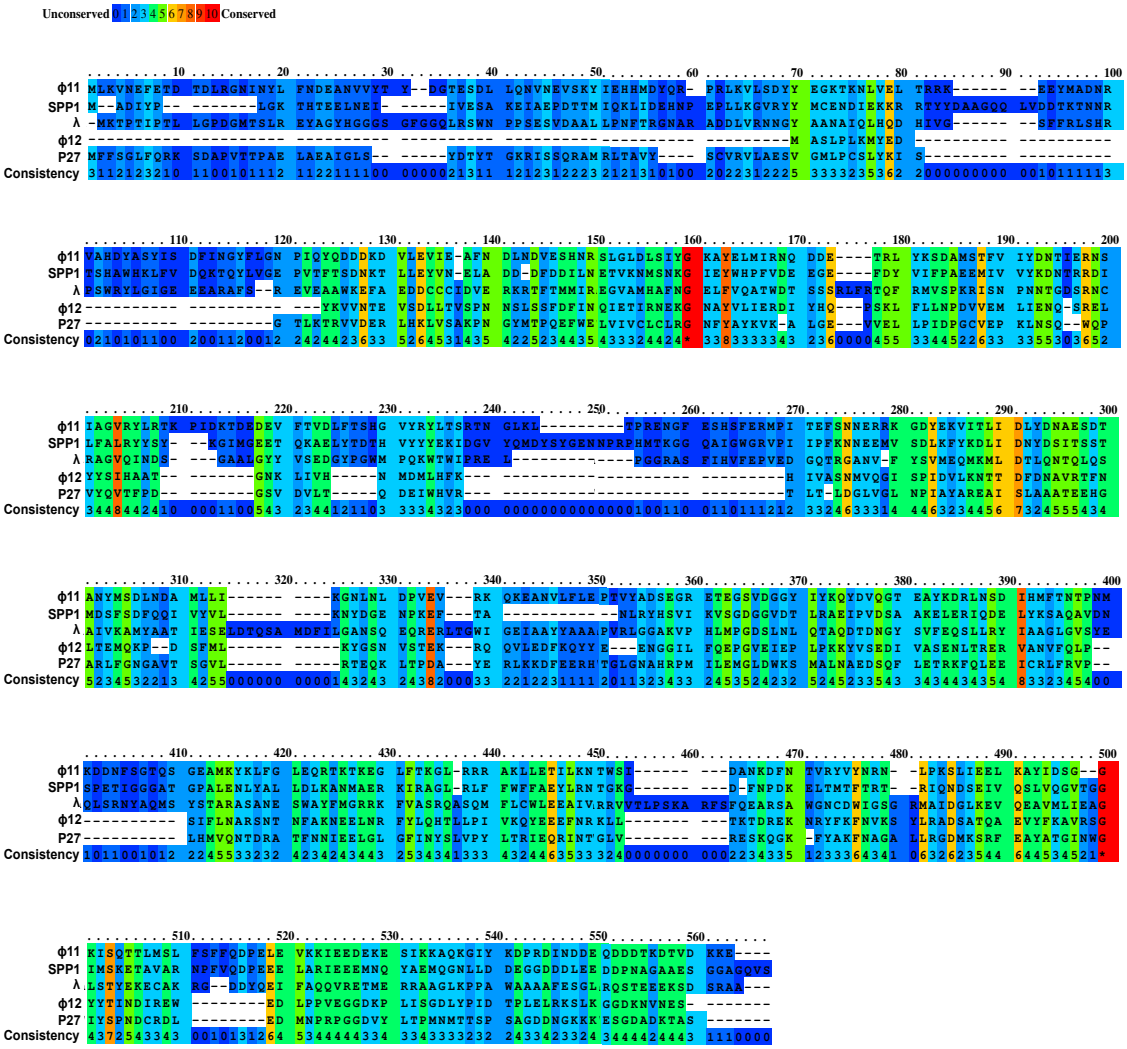
TerS

B


Unconserved 1 1 2 3 4 5 6 7 8 9 10 Conserved
ϕ11
SPP1
ϕ12
P27
λ
Consistency

TerL

C

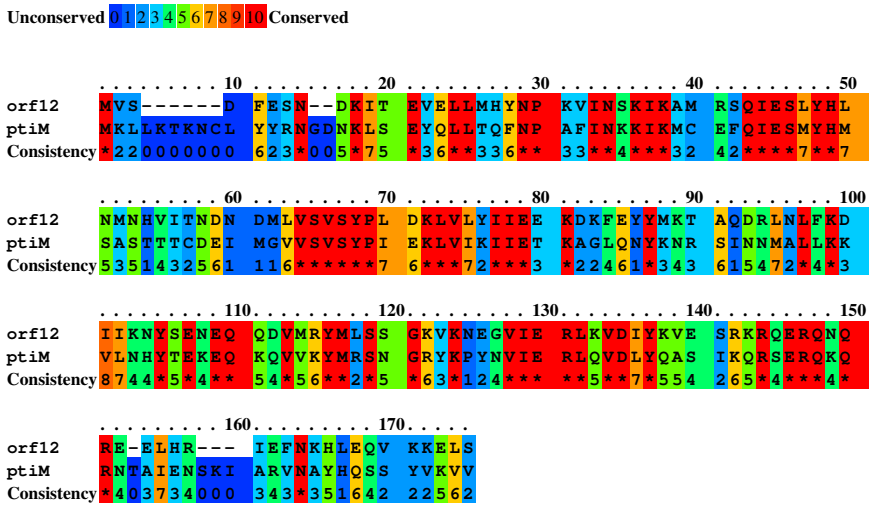


Portal

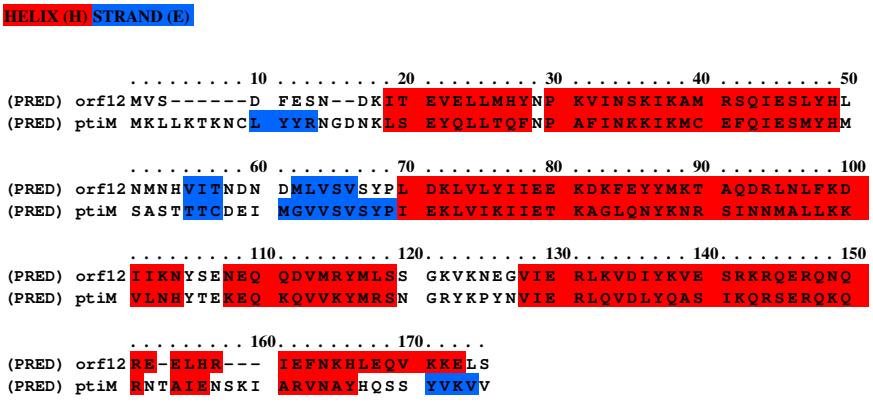
Figure 2.1 Alignment of TerS (A), TerL (B) or portal protein (C) amino acid sequences from phages ϕ11, SPP1, λ, ϕ12 and ϕP27, coloured according to relative sequence conservation at each position. Adapted from alignment generated by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position.

Appendix 3 - Alignment and modelled three-dimensional structures of ORF12 from SaPIbov5 and PtiM from SaPI2.

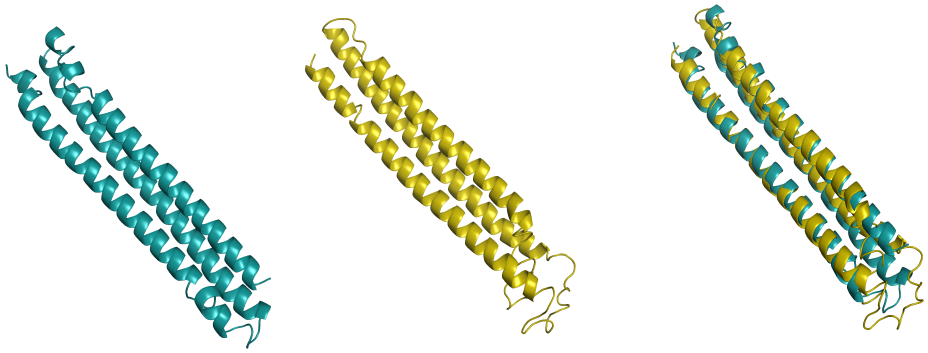
A



B



C

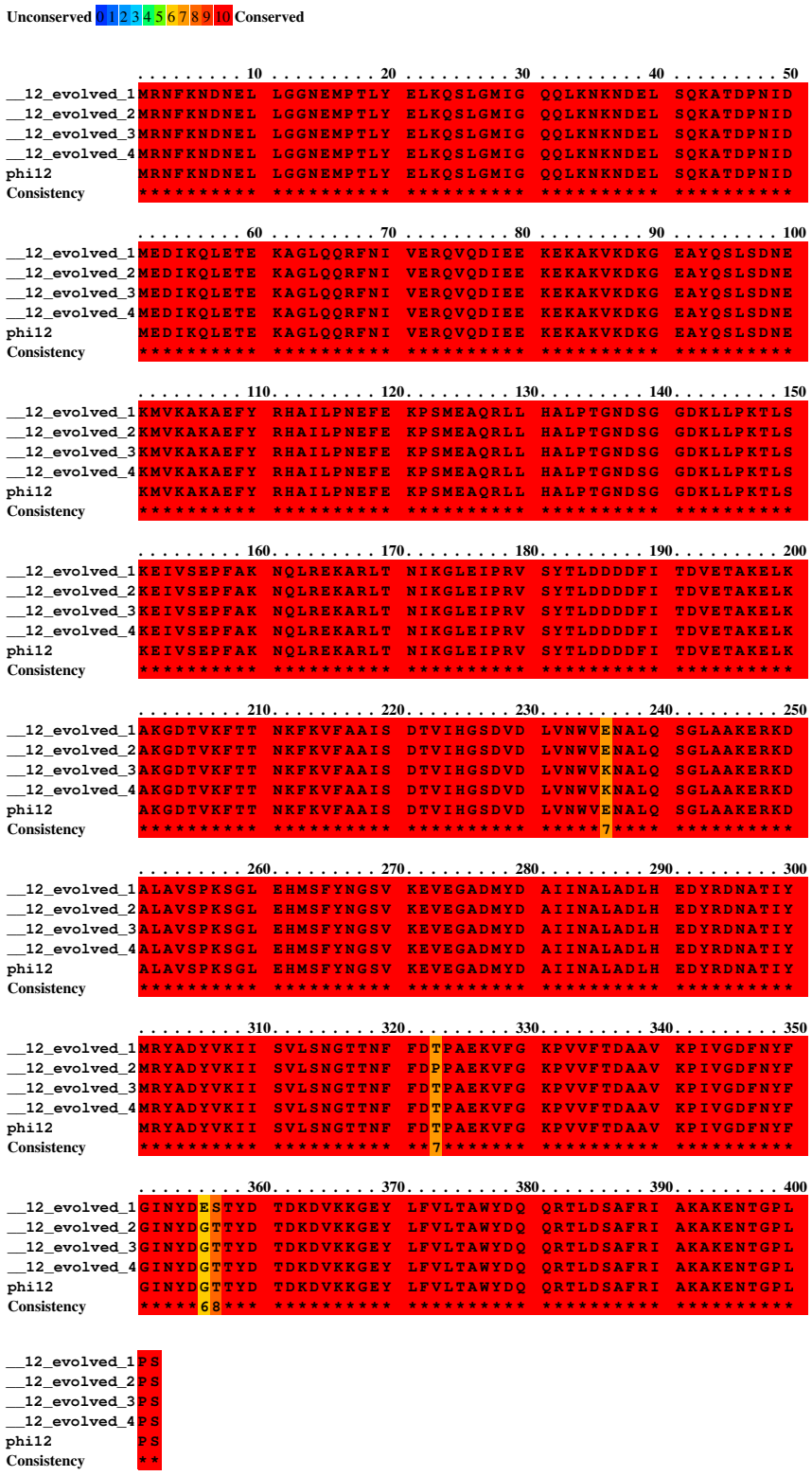


SaPIbov5 ORF12 **SaPI2 PtiM**

Figure 3.1 Alignment of ORF12 (A) or PtiM (B) amino acid sequences from SaPIbov5 and SaPI2 respectively, coloured according to relative sequence conservation at each position and overall modelled three-dimensional structure of ORF12 or PtiM (C). (A-B) Adapted from alignment generated by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position. For the structural prediction, the current colour scheme of the alignment is for secondary structure type is red for helix (H) and blue for strand (E). (C) The three-dimensional structure of SaPIbov5 ORF12 modelled with I-TASSER is shown in ribbon representation coloured in cyan. ORF12 SaPIbov5 showed an overall fold that closely resembles to SaPI2 PtiM (yellow). Models with higher C-score (-3.07F for PtiM and -2.84 for ORF12) were selected for further structural analysis using PyMOL.

Appendix 4 - Alignment of ORF33 and ORF45 from wt phage ϕ12 and evolved mutants.

A



B

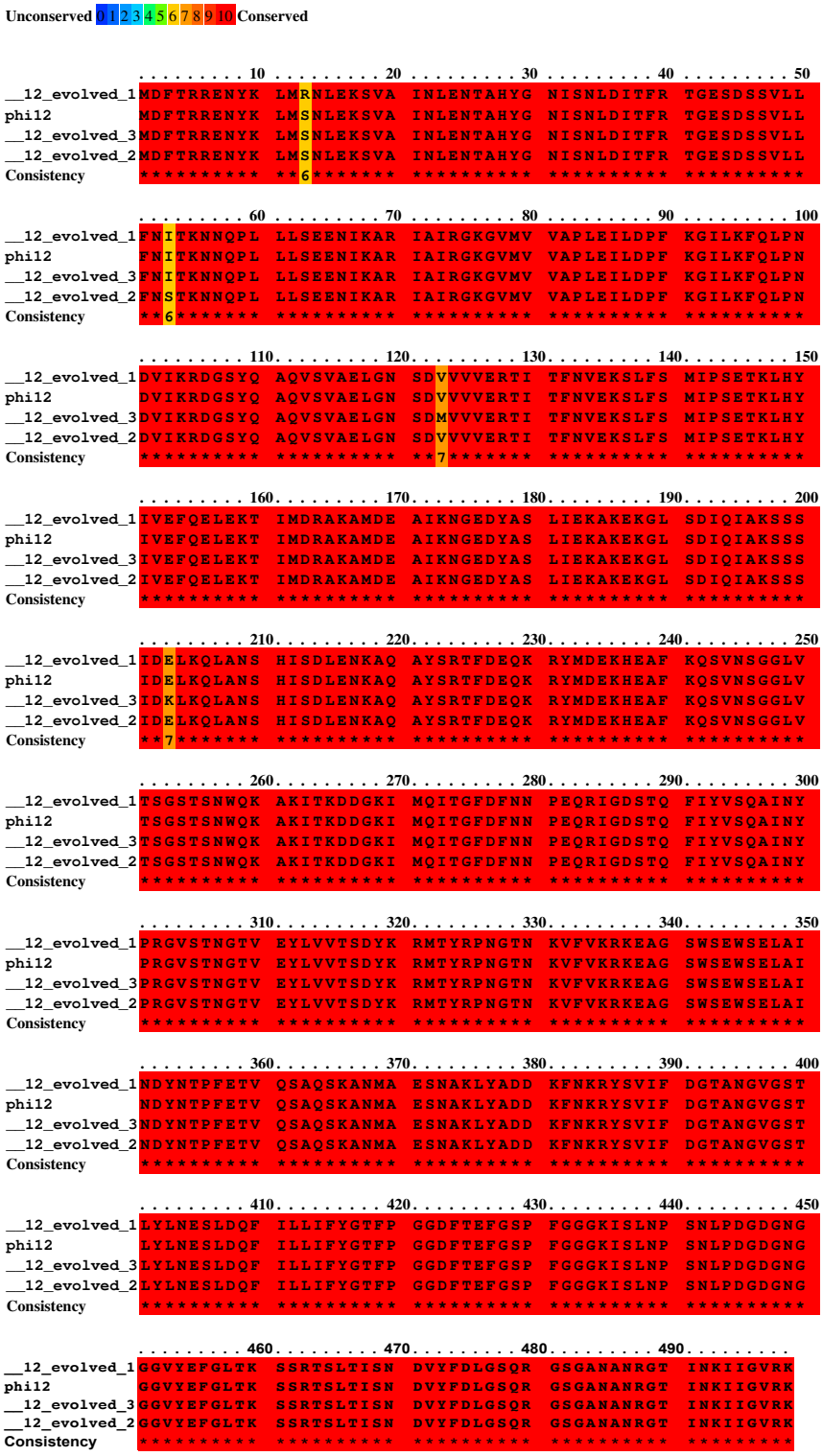


Figure 4.1 Alignment of ORF33 (A) or ORF45 (B) amino acid sequences from phages ϕ12, and ϕ12 evolved mutants, coloured according to relative sequence conservation at each position. Adapted from alignment generated by PRALINE. The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position.

Appendix 5 - N-terminal portion of gp33 and Ccm are predicted to adopt an α-helical fold.

Table 1. Templates and confidence values in gp33 and Ccm models generated by Phyre2 server[a].

	Protein	Template information	PDB code	Aligned Residues[a]	Confidence[b] (%)	Identity (%)
C-terminal portion	gp33	Capsid protein form bacteriophage HK97	2FT1	132-388	100	18
		Capsid protein form bacteriophage CW02	3J1A	211-388	99.6	25
		Capsid protein form bacteriophage BPP-1	3J4U	148-338	98.2	17
		Thermotoga maritima encapsulin protein (TM-0786)	3DKT	147-392	97.9	15
		Capsid protein form bacteriophage T7	3J7W	135-387	97.8	15
		Myxococcus xanthus encpasulin protein (EncA)	4PT2	147-388	97.6	11
		Capsid protein form marine bacteriophage Syn5	4BML	144-387	97.6	15
		Virus-like particule from *Pyrococcus furiosus*	2E0Z	147-392	97.4	9
	Ccm	Capsid protein form bacteriophage HK9	2FT1	153-349	99	15
		Capsid protein form bacteriophage HK97	1IF0	159-349	98.9	14
		Capsid protein form bacteriophage HK97	3P8Q	153-341	97.9	15
		Capsid protein form bacteriophage CW02	3J1A	186-309	76.8	26
		Capsid protein form bacteriophage BPP-1	3J4U	234-299	74.7	20
		Capsid protein form bacteriophage HK97	2FT1	153-349	99	15
N-terminal portion	gp33	Coiled-Coil Protein of Unknown Function from *Eubacterium eligens*	3HNW	19-84	93.3	12
		Human Fibrinogen beta chain	3GHG	17-87	84.4	19
		Anopheles Plasmodium-responsive Leucine-rich repeat protein 1 (LRR1)	3OJA	17-90	84.3	12
		Bovine Fibrinogen beta chain	1DEQ	20-91	84.1	13
		Human cytoplasmic dynein 2 heavy chain	4RH7	17-119	83.6	10
		PcsB protein from *Streptococcus pneumoniae*	4CGK	17-82	79.4	13
		Seryl-tRNA synthetase N-terminal domain from *Aquifex aeolicus*	2DQ3	14-91	78.8	13
		dynein motor heavy chain	3VKG	1-84	67.5	5
	Ccm	Mit domain-containing protein 1	2YMB	109-157	14.9	20
		Human Striatin-3 coiled coil domain	4N6J	31-47	14.8	35

[a]See reference [318] for details.
[b]Aligned residues. Part of protein sequence that is aligned with the template sequence.
[c]Confidence represents the probability (from 0 to 100) that the match between the query sequence and the corresponding template is a true homology. A match with confidence >90%, generally should indicate that the query sequence adopts the overall fold shown by the template and that the core of the protein is modelled at high accuracy.

Table 2. Templates and confidence values in gp33 and Ccm models generated by RaptorX server[a].

	Protein	Template information	PDB code	P-value[b]	Score[c]	uGDT/ GDT[d]	uSeqID/ SeqID[e]
C-terminal portion	gp33	Capsid protein form bacteriophage HK97	1OHG	3.1 e-08	149	147/57	42/16
		Capsid protein form bacteriophage HK97	3QPR	2.1 e-05	99	121/51	31/13
		Thermotoga maritima encapsulin protein (TM-0786)	3DKT	1.5 e-04	83	72/31	27/12
		Myxococcus xanthus encpasulin protein (EncA)	4PT2	3.0 e-04	78	51/29	10/6
		Virus-like particule from *Pyrococcus furiosus*	2E0Z	1.0 e-04	86	60/27	18/8
	Ccm	Capsid protein form bacteriophage HK97	1OHG	9.1 e-08	133	144/55	29/11
		Capsid protein form bacteriophage HK97	3QPR	2.2 e-05	93	94/47	18/9
		Thermotoga maritima encapsulin protein (TM-0786)	3DKT	5.9 e-05	85	72/30	26/11
		Myxococcus xanthus encpasulin protein (EncA)	4PT2	2.3 e-04	75	50/28	12/7
		Virus-like particule from Pyrococcus furiosus	2E0Z	6.7 e-05	84	56/26	22/10
N-terminal portion	gp33	Structural maintenance of chromosomes protein 4 (Smc4) from Yeast	4RSI	1.1 e-02	61	52/47	13/12
		CT398 protein from *Chlamydia trachomatis*	4ILO	1.0 e-02	62	52/47	12/11
		HP0958 Protein from *Helicobacter pylori*	3NA7	1.1 e-02	60	52/47	18/16
		Colicin IA from *Escharichia coli*	1CII	8.0 e-03	64	41/43	8/9
		Cytoplasmic dynein 1 heavy chain 1 from *Mus musculus*	3WUQ	9.9 e-03	62	49/44	15/14
	Ccm	iSH2 domain of human phosphatidylinositol 3-kinase p85beta	3MTT	2.7 e-02	28	42/53	10/12
		Vacuolar protein sorting-associated protein 30	5DFZ	1.8 e-02	30	42/52	8/10
		PcsB protein from *Streptococcus pneumoniae*	4CGK	2.9 e-02	27	42/52	10/12
		HP0958 Protein from *Helicobacter pylori*	3NA7	2.9 e-02	28	42/52	8/10
		Autophagy protein Apg17	4HPQ	3.1 e-02	27	41/53	5/6

[a]See reference [317] for details.
[b]P-value. Likelihood of a predicted model being worse than the best of a set of randomly generated models. For mainly alpha proteins, P-value < e-3 and e-4 are good indicator for mainly alpha and beta proteins, respectively. [c]Score. The alignment score falling between 0 and the (domain) sequence length, with 0 indicating the worst. [d]uSeqID and SeqID. uSeqID is the number of identical residues in the alignment. SeqID is uSeqID normalized by the protein (or domain) sequence length and multiplied by 100. [e]uGDT and GDT. uGDT is the unnormalized GDT (Global Distance Test) score defined as 1*N(1) + 0.75 * N(2) + 0.5 * N(4) + 0.25 * N(8), where N(x) is the number of residues with estimated modelling error (in Å) smaller than x. GDT is calculated as uGDT divided by the protein (or domain) length and multiplied by a 100.

Appendix 6 - Sequence analysis of EfCIV583 and EfCIV583 mutant in EF2942 circularisation.

EfCIV583 wt Circularised:

GCATTTATAAACACCGACGAAGTGAGCGAATTTATCTGCTGTTTTTTCTTTTTCTTCGGGCA
TAACGTGAGAAGAGATTACTCTCTGTAAGGGCATGGTTGGTAAGTCCTAATCCTTTAACCTA
GGAATTTTCTCTTTTTACAATTGTAGTGCCTATTAAAGAAGTTGCTGAAAGATTAGGACATAC
TAGCACTGCTATTACTGATCGTATCTATTCTCATGTTATGCCTGAAGAAAAAGAAAAAACAG
CAGATAAGTTCGCTCAATTCGTCGGTTTTTAATAAATGCACAGACAAAAGCACAGATCTTTA
CTATTTTTATGTAAAAAAAAGAGCTAGAATCCTTATAAATAAAGGTTTCTAACTGGCATTAGT
AACGCTATTAATGAAACAACGTGCTTCCTTGAGTGGATATTATTAAGCACGAATAGCCCCAT
TTTACTTGATACAATAGGAATTTGCAACAAAACATAATTAGTTAAATACACATAGTTTTGATC
AAAAGCACAGACCAAAGCACAGACCTTTTTTAATCCATCACGCCAATAGTTACATATTACGT
TTACTGTTTTTTAAACCCAACGGGAAAAGAATAGTTATACCTGATATACCTATACTGATAAAA
AAATAAATGAGGTAACTTGTCCAGTCAGCGGTACCCCCATCTATATAACCTGAAATTGAAAA
ATATGAAAACAAAACTATCACACTGACTAACCACTGATATCTTAACCTTAATTTTAATTTA

EfCIV583 mutant EF2942 Circularised:

ATGATTAAATGGAGGAGTTGAGAAAAATATCTATTAGAGAAACGGAAAGGGTTTTACATTGA
AGAAACAGCAACTCTCACTTTACGGCACCATTTTCTTTCTTTGCTTTTTCCTAATTGTGGGCT
TCTTCTTCTCAAGTCGTGGCGAAGATTATGCCAAAGAATCTGAACAAAAAGTTACAATTGAT
TCCGCCAAACATGAGAAGCATACGAAAGACAAGGAAGAAAATAATTCAGCAAATACAGTAT
TTTTTGATAAAATAAATGACCTGCTCGTCGCTTCTGTCAAAGAATTTGAAGGAACGGTCGGG
ATTAGTTATTTGGATTTAGAAACTGGCGAACAGCGTTCCGTGAATGGCCAACATGAATTTTA
TACAGCTAGTACTATTAAAGTGCCCCTGACGATGCTAGTTGCCGACACGGTTGCTTCTGGT
CAAAAAAAATGGACAGACCTGATTCCCTATAACGCCGAAGAAGATTATGAAGAAGGAACTG
GGATCATTGCCTATAATATCCAACCAGAATATCCATTAAAAACGTTACAAGAATATGCCATC
ACCTATTCTGATAATATCGCTAAAAATATGCTCTATGACACATTGGGGTGGGTGATGCAAAA
GCAAAACGAGAAATGTATCAGCGTTATTTGCACAAGACACCTTCGATTGAAGAGCCACAAT
TTTCTTCTGAAGATGCGCTTGTTATTCTACAAAAATTATATACTGAAAAAGCAACAAAACCAG
ATTACCAAGCGATTTATGATTCTATGAAACAACGTGCTTCCTTGAGTGGATATTATTAAGCA
CGAATAGCCCCATTTTACTTGATACAATAGGAATTTGCAACAAAACATAATTAGTTAAATACA
CATAGTTTTGATCAAAAGCACAGACCAAAGCACAGACCTTTTTTAATCCATCACGCCAATAG
TTACATATTACGTTTACTGTTTTTTAAACCCAACGGGAAAAGAATAGTTATACCTGATATACC
TATACTGATAAAAAAATAAATGAGGTAACTTGTCCAGTCAGCGGTACCCCCATCTATATAAC
CTGAAATTGAAAAATATGAAAACAAAACTATCACACTGACTAACCACTGATATCTTAACCTTA
ATTTTAATTTATGTC

att_S

EfCIV583

Partial att_S

E. faecalis V584 (1457538-1456760)

Figure 6.1 Detection of EfCIV583 circularisation. DNA from *E. faecalis* EfCIV583 and EF2942 mutant were extracted and PCR-amplified using a pair of primers set divergently at both termini of the island (circularisation). PCR were sequenced and BLAST sequence analysis was performed to identify *E. faecalis* V583 and EfCIV583 genome.

Glossary

Prophage: phage integrated into the bacterial chromosome.

Temperate phage: a phage that can sustain either a lytic or lysogenic life cycle.

Lysogen: a bacterium harbouring a prophage.

Horizontal gene transfer: The transfer of foreign DNA among unicellular or multicellular organisms that are not parents and offspring.

Mobile genetic element (MGE): a DNA sequence that is able to promote its transfer between organisms.

Genomic island (GI): discrete sections of DNA, which are found inserted into bacterial genomes and is assumingly acquired via horizontal gene transfer.

Staphylococcal pathogenicity island (SaPI): a GI present in *S. aureus* encoding phage-related genes that is mobilised by induction of a helper phage. The SaPIs were the first PICIs identified.

Phage-inducible chromosomal island (PICI): a GI encoding phage-related genes equivalent to those of the SaPIs and that is mobilised after induction by a helper phage.

Satellite virus: a genetic element whose life cycle depends on a helper virus, which parasitizes.

Terminase complex: an enzyme complex composed of two subunits, small (TerS) and large (TerL) terminase, responsible for the DNA packaging in phages.

***pac* site:** a specific DNA sequence recognised by the terminase complex of a *pac* phage to initiate packaging.

***cos* site:** a cohesive end site is a DNA sequence recognised by the terminase complex of a *cos* phage to initiate packaging.

Headful packaging: DNA packaging mechanism of *pac* phages in which concatemeric DNA is cleaved at the *pac* site and encapsidated into a preformed prohead. The full fill of the prohead is the signal for a nonspecific cleavage of the DNA, which finishes the packaging process.

***Cos* unit-length packaging:** DNA packaging mechanism of *cos* phages in which concatemeric DNA is cleaved at the *cos* site and encapsidated into a preformed prohead. The recognition of the *cos* site at a precise unit-length distance is the signal for the second cleavage, which finishes the packaging process.

Molecular piracy: a biological process in which a replicon takes over another replicon structural proteins' to package and spread its own genome.

SOS: bacterial response to DNA damage, detaining the cell cycle and triggering DNA repair. The SOS response includes the RecA-mediated cleavage of LexA repressor and derepression of repair response genes.

Campbell mechanism: an integrase mediated site-specific integration of a circularised genetic element into a bacterial chromosome by a specific reciprocal crossover between shared sequences.

Concatemeric DNA: a large, non-interrupted molecule of DNA, which encompasses numerous copies of the same genome linked end to end in tandem.

Procapsid: the empty precursor of the capsid particle comprised of several copies of the capsid protein subunits, which is normally constituted by internal scaffolding proteins.

Capsid: a polyhedral protein complex, or shell of a virus, which will harbour the virus genome; usually icosahedral in bacteriophages.

Virion: a virus infective particle.

List of References

1. Kloos, W. E. Natural populations of the genus *Staphylococcus*. *Annu. Rev. Microbiol.* **34,** 559–92 (1980).
2. Peacock, S. J., de Silva, I. & Lowy, F. D. What determines nasal carriage of *Staphylococcus aureus*? *Trends Microbiol.* **9,** 605–10 (2001).
3. Yu, V. L. *et al. Staphylococcus aureus* nasal carriage and infection in patients on hemodialysis. Efficacy of antibiotic prophylaxis. *N. Engl. J. Med.* **315,** 91–6 (1986).
4. Wertheim, H. F. L. *et al.* The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet. Infect. Dis.* **5,** 751–62 (2005).
5. Nienaber, J. J. C. *et al.* Methicillin-susceptible *Staphylococcus aureus* endocarditis isolates are associated with clonal complex 30 genotype and a distinct repertoire of enterotoxins and adhesins. *J. Infect. Dis.* **204,** 704–13 (2011).
6. Rocha, L. A., Marques Ribas, R., da Costa Darini, A. L. & Gontijo Filho, P. P. Relationship between nasal colonization and ventilator-associated pneumonia and the role of the environment in transmission of *Staphylococcus aureus* in intensive care units. *Am. J. Infect. Control* **41,** 1236–1240 (2013).
7. Holland, T. L., Arnold, C. & Fowler, V. G. Clinical management of *Staphylococcus aureus* bacteremia: A review. *JAMA* **312,** 1330–41 (2014).
8. Lowy, F. D. *Staphylococcus aureus* Infections. *N. Engl. J. Med.* **339,** 520–532 (1998).
9. Labandeira-Rey, M. *et al. Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. *Science* **315,** 1130–3 (2007).
10. Gillet, Y. *et al.* Association between *Staphylococcus aureus* strains carrying gene for Panton-Valentine leukocidin and highly lethal necrotising pneumonia in young immunocompetent patients. *Lancet* **359,** 753–759 (2002).
11. Baba, T. *et al.* Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet (London, England)* **359,** 1819–27 (2002).
12. Kreiswirth, B. N., Projan, S. J., Schlievert, P. M. & Novick, R. P. Toxic shock syndrome toxin 1 is encoded by a variable genetic element. *Rev. Infect. Dis.* **1,** S83-8-9 (1989).
13. Lindsay, J. A., Ruzin, A., Ross, H. F., Kurepina, N. & Novick, R. P. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29,** 527–43 (1998).
14. Kuroda, M. *et al.* Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet (London, England)* **357,** 1225–40 (2001).
15. Yamaguchi, T. *et al.* Identification of the *Staphylococcus aureus* etd pathogenicity island which encodes a novel exfoliative toxin, ETD, and EDIN-B. *Infect. Immun.* **70,** 5835–45 (2002).
16. Bergonier, D., de Crémoux, R., Rupp, R., Lagriffoul, G. & Berthelot, X. Mastitis of dairy small ruminants. *Vet. Res.* **34,** 689–716 (2003).
17. Hermans, K., Devriese, L. A. & Haesebrouck, F. Rabbit staphylococcosis: Difficult solutions for serious problems. *Vet. Microbiol.* **91,** 57–64 (2003).
18. McNamee, P. T. *et al.* Study of leg weakness in two commercial broiler flocks. *Vet. Rec.* **143,** 131–5 (1998).
19. Chambers, H. F. & Deleo, F. R. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. *Nat. Rev. Microbiol.* **7,** 629–41 (2009).
20. Dinges, M. M., Orwin, P. M. & Schlievert, P. M. Exotoxins of *Staphylococcus aureus*. *Clin. Microbiol. Rev.* **13,** 16–34 (2000).
21. Chang, S. *et al.* Infection with vancomycin-resistant *Staphylococcus aureus* containing the *van*A resistance gene. *N. Engl. J. Med.* **348,** 1342–7 (2003).

22. Diekema, D. J. *et al.* Survey of infections due to *Staphylococcus* species: Frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific Region for the SENTRY antimicrobial surveillance. *Clin. Infect. Dis.* **32,** S114–S132 (2001).

23. Fluit, A. C., Verhoef, J. & Schmitz, F. J. Frequency of isolation and antimicrobial resistance of Gram-negative and Gram-positive bacteria from patients in intensive care units of 25 European university hospitals participating in the European arm of the SENTRY Antimicrobial Surveillance Program 19. *Eur. J. Clin. Microbiol. Infect. Dis.* **20,** 617–25 (2001).

24. Katayama, Y., Ito, T. & Hiramatsu, K. A new class of genetic element, *Staphylococcus* cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **44,** 1549–1555 (2000).

25. Chambers, H. F. The changing epidemiology of *Staphylococcus aureus*? *Emerg. Infect. Dis.* **7,** 178–82 (2001).

26. Fridkin, S. K. *et al.* Methicillin-resistant *Staphylococcus aureus* disease in three communities. *N. Engl. J. Med.* **352,** 1436–44 (2005).

27. Köck, R. *et al.* Methicillin-resistant *Staphylococcus aureus* (MRSA): Burden of disease and control challenges in Europe. *Euro surveillance : Bulletin européen sur les maladies transmissibles = European communicable disease bulletin* **15,** 19688 (2010).

28. Knight, G. M. *et al.* Shift in dominant hospital-associated methicillin-resistant *Staphylococcus aureus* (HA-MRSA) clones over time. *J. Antimicrob. Chemother.* **67,** 2514–2522 (2012).

29. Vandenesch, F. *et al.* Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: Worldwide emergence. *Emerg. Infect. Dis.* **9,** 978–84 (2003).

30. Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **3,** 722–32 (2005).

31. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3,** 711–21 (2005).

32. Lindsay, J. A. Genomic variation and evolution of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **300,** 98–103 (2010).

33. McCarthy, A. J., Witney, A. A. & Lindsay, J. A. *Staphylococcus aureus* temperate bacteriophage: Carriage and horizontal gene transfer is lineage associated. *Front. Cell. Infect. Microbiol.* **2,** 6 (2012).

34. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19,** 2226–2238 (2002).

35. Lindsay, J. A. & Holden, M. T. G. *Staphylococcus aureus*: Superbug, super genome? *Trends Microbiol.* **12,** 378–85 (2004).

36. Malachowa, N. & DeLeo, F. R. Mobile genetic elements of *Staphylococcus aureus*. *Cell. Mol. Life Sci.* **67,** 3057–71 (2010).

37. Alibayov, B., Baba-Moussa, L., Sina, H., Zdenkova, K. & Demnerova, K. *Staphylococcus aureus* mobile genetic elements. *Mol Biol Rep* **41,** 5005–5018 (2014).

38. Fitzgerald, J. R., Sturdevant, D. E., Mackie, S. M., Gill, S. R. & Musser, J. M. Evolutionary genomics of *Staphylococcus aureus*: Insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8821–6 (2001).

39. Sulakvelidze, A. Bacteriophage: A new journal for the most ubiquitous organisms on Earth. *Bacteriophage* **1,** 1–2 (2011).

40. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nat. Rev. Microbiol.* **3,** 504–10 (2005).

41. Rohwer, F. & Thurber, R. V. Viruses manipulate the marine environment. *Nature* **459,** 207–12 (2009).

42. Suttle, C. A. Viruses in the sea. *Nature* **437,** 356–61 (2005).

43. Suttle, C. A. Marine viruses--major players in the global ecosystem. *Nat. Rev. Microbiol.* **5,** 801–12 (2007).

44. Wommack, K. E. & Colwell, R. R. Virioplankton: Viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64,** 69–114 (2000).

45. Bergh, O., Børsheim, K. Y., Bratbak, G. & Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **340,** 467–8 (1989).

46. Marsh, P. & Wellington, E. M. H. Phage-host interactions in soil. *FEMS Microbiol. Ecol.* **15,** 99–107 (1994).

47. Ashelford, K. E., Day, M. J. & Fry, J. C. Elevated abundance of bacteriophage infecting bacteria in soil. *Appl. Environ. Microbiol.* **69,** 285–9 (2003).

48. Bratbak, G., Heldal, M., Norland, S. & Thingstad, T. F. Viruses as partners in spring bloom microbial trophodynamics. *Appl. Environ. Microbiol.* **56,** 1400–1405 (1990).

49. Sano, E., Carlson, S., Wegley, L. & Rohwer, F. Movement of viruses between biomes. *Appl. Environ. Microbiol.* **70,** 5842–6 (2004).

50. Brüssow, H., Canchaya, C. & Hardt, W.-D. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* **68,** 560–602 (2004).

51. Stern, A. & Sorek, R. The phage-host arms race: Shaping the evolution of microbes. *Bioessays* **33,** 43–51 (2011).

52. Canchaya, C., Fournous, G. & Brüssow, H. The impact of prophages on bacterial chromosomes. *Mol. Microbiol.* **53,** 9–18 (2004).

53. Coulthurst, S. J. The Type VI secretion system - a widespread and versatile cell targeting system. *Res. Microbiol.* **164,** 640–654 (2013).

54. Shikuma, N. J. *et al.* Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* **343,** 529–33 (2014).

55. Holden, M. T. G. *et al.* Complete genomes of two clinical *Staphylococcus aureus* strains: Evidence for the rapid evolution of virulence and drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9786–9791 (2004).

56. Kwan, T., Liu, J., DuBow, M., Gros, P. & Pelletier, J. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 5174–9 (2005).

57. Brown, N. F., Wickham, M. E., Coombes, B. K. & Finlay, B. B. Crossing the line: Selection and evolution of virulence traits. *PLoS Pathog.* **2,** e42 (2006).

58. Nanda, A. M., Thormann, K. & Frunzke, J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.* **197,** 410–419 (2015).

59. Sumby, P. & Waldor, M. K. Transcription of the toxin genes present within the staphylococcal phage phiSa3ms is intimately linked with the phage's life cycle. *J. Bacteriol.* **185,** 6841–51 (2003).

60. Waldor, M. K. & Friedman, D. I. Phage regulatory circuits and virulence gene expression. *Curr. Opin. Microbiol.* **8,** 459–65 (2005).

61. Novick, R. P. Mobile genetic elements and bacterial toxinoses: The superantigen-encoding pathogenicity islands of *Staphylococcus aureus*. *Plasmid* **49,** 93–105 (2003).

62. Betley, M. J. & Mekalanos, J. J. Staphylococcal enterotoxin A is encoded by phage. *Science* **229,** 185–7 (1985).

63. Kaneko, J., Kimura, T., Narita, S., Tomita, T. & Kamio, Y. Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton-Valentine leukocidin genes. *Gene* **215,** 57–67 (1998).

64. Narita, S. *et al.* Phage conversion of Panton-Valentine leukocidin in *Staphylococcus*

aureus: Molecular analysisi of a PVL-converting phage, phiSLT. (2001).

65. van Wamel, W. J. B., Rooijakkers, S. H. M., Ruyken, M., van Kessel, K. P. M. & van Strijp, J. A. G. The innate immune modulators staphylococcal complement inhibitor and chemotaxis inhibitory protein of *Staphylococcus aureus* are located on beta-hemolysin-converting bacteriophages. *J. Bacteriol.* **188,** 1310–5 (2006).

66. Muniesa, M., Recktenwald, J., Bielaszewska, M., Karch, H. & Schmidt, H. Characterization of a shiga toxin 2e-converting bacteriophage from an *Escherichia coli* strain of human origin. *Infect. Immun.* **68,** 4850–5 (2000).

67. Zhang, X. *et al.* Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J. Infect. Dis.* **181,** 664–70 (2000).

68. Boyd, E. F. Bacteriophage-encoded bacterial virulence factors and phage-pathogenicity island interactions. *Adv. Virus Res.* **82,** 91–118 (2012).

69. Casjens, S. Prophages and bacterial genomics: What have we learned so far? *Mol. Microbiol.* **49,** 277–300 (2003).

70. Canchaya, C., Proux, C., Fournous, G., Bruttin, A. & Brüssow, H. Prophage genomics. *Microbiol. Mol. Biol. Rev.* **67,** 238–76, table of contents (2003).

71. Brüssow, H. & Hendrix, R. W. Phage Genomics. *Cell* **108,** 13–16 (2002).

72. Clokie, M. R., Millard, A. D., Letarov, A. V & Heaphy, S. Phages in nature. *Bacteriophage* **1,** 31–45 (2011).

73. Coleman, D. C. *et al. Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: Molecular mechanism of triple conversion. *J. Gen. Microbiol.* **135,** 1679–97 (1989).

74. Lee, C. Y. & Iandolo, J. J. Mechanism of bacteriophage conversion of lipase activity in *Staphylococcus aureus*. *J. Bacteriol.* **164,** 288–93 (1985).

75. Day, W. A., Fernández, R. E. & Maurelli, A. T. Pathoadaptive mutations that enhance virulence: Genetic organization of the *cad*A regions of *Shigella spp*. *Infect. Immun.* **69,** 7471–80 (2001).

76. Berngruber, T. W., Weissing, F. J. & Gandon, S. Inhibition of superinfection and the evolution of viral latency. *J. Virol.* **84,** 10200–8 (2010).

77. Kliem, M. & Dreiseikelmann, B. The superimmunity gene *sim* of bacteriophage P1 causes superinfection exclusion. *Virology* **171,** 350–5 (1989).

78. Susskind, M. M., Botstein, D. & Wright, A. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. III. Failure of superinfecting phage DNA to enter *sie*A+ lysogens. *Virology* **62,** 350–66 (1974).

79. Selva, L. *et al.* Killing niche competitors by remote-control bacteriophage induction. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 1234–8 (2009).

80. Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10771–6 (2013).

81. Lennox, E. S. Transduction of linked genetic characters of the host by bacteriophage P1. *Virology* **1,** 190–206 (1955).

82. Zinder, N. D. & Lederberg, J. Genetic exchange in *Salmonella*. *J. Bacteriol.* **64,** 679–99 (1952).

83. Sandri, R. M. & Berger, H. Bacteriophage P1-mediated generalized transduction in *Escherichia coli*: Fate of transduced DNA in *rec*+ and *rec*A- recipients. *Virology* **106,** 14–29 (1980).

84. Ebel-Tsipis, J., Botstein, D. & Fox, M. S. Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. *J. Mol. Biol.* **71,** 433–48 (1972).

85. Ikeda, H. & Tomizawa, J. I. Transducing fragments in generalized transduction by phage P1. I. Molecular origin of the fragments. *J. Mol. Biol.* **14,** 85–109 (1965).

86. Mašlaňová, I. *et al.* Bacteriophages of *Staphylococcus aureu*s efficiently package various bacterial genes and mobile genetic elements including SCC*mec* with different frequencies. *Environ. Microbiol. Rep.* **5,** 66–73 (2013).

87. O'Shea, Y. A. & Boyd, E. F. Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction. *FEMS Microbiol. Lett.* **214,** 153–7 (2002).

88. Lindqvist, B. H., Dehò, G. & Calendar, R. Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol. Rev.* **57,** 683–702 (1993).

89. Novick, R. P., Christie, G. E. & Penades, J. R. The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol* **8,** 541–551 (2010).

90. Faruque, S. M. *et al.* RS1 element of *Vibrio cholerae* can propagate horizontally as a filamentous phage exploiting the morphogenesis genes of CTXphi. *Infect. Immun.* **70,** 163–70 (2002).

91. Del Campillo-Campbell, A., Kayajanian, G., Campbell, A. & Adhya, S. Biotin-requiring mutants of *Escherichia coli* K-12. *J. Bacteriol.* **94,** 2065–6 (1967).

92. Morse, M. L., Lederberg, E. M. & Lederberg, J. Transduction in Escherichia Coli K-12. *Genetics* **41,** 142–56 (1956).

93. Kenzaka, T., Tani, K. & Nasu, M. High-frequency phage-mediated gene transfer in freshwater environments determined at single-cell level. *ISME J.* **4,** 648–59 (2010).

94. Modi, S. R., Lee, H. H., Spina, C. S. & Collins, J. J. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499,** 219–22 (2013).

95. Minot, S. *et al.* The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Res.* **21,** 1616–25 (2011).

96. Quirós, P. *et al.* Antibiotic resistance genes in the bacteriophage DNA fraction of human fecal samples. *Antimicrob. Agents Chemother.* **58,** 606–9 (2014).

97. Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289,** 2129–33 (2000).

98. International Committee on Taxonomy of Viruses. Virus taxonomy: Classification and nomenclature of viruses. *Ninth report of the International Committee on Taxonomy of Viruses* 686–713 (2012).

99. Ackermann, H.-W. Phage classification and characterization. *Methods Mol. Biol.* **501,** 127–40 (2009).

100. Nelson, D. Phage taxonomy: We agree to disagree. *J. Bacteriol.* **186,** 7029–31 (2004).

101. Rohwer, F. & Edwards, R. The phage proteomic tree: A genome-based taxonomy for phage. *J. Bacteriol.* **184,** 4529–4535 (2002).

102. Lawrence, J. G., Hatfull, G. F. & Hendrix, R. W. Imbroglios of viral taxonomy: Genetic exchange and failings of phenetic approaches. *J. Bacteriol.* **184,** 4891–4905 (2002).

103. Black, L. W. DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* **43,** 267–92 (1989).

104. Johnson, J. E. & Chiu, W. DNA packaging and delivery machines in tailed bacteriophages. *Curr. Opin. Struct. Biol.* **17,** 237–43 (2007).

105. Aksyuk, A. A. & Rossmann, M. G. Bacteriophage assembly. *Viruses* **3,** 172–203 (2011).

106. Ackermann, H. W. Tailed bacteriophages: The order caudovirales. *Adv. Virus Res.* **51,** 135–201 (1998).

107. Deghorain, M. & Van Melderen, L. The staphylococci phages family: An overview. *Viruses* **4,** 3316–3335 (2012).

108. Trun, N. & Trempy, J. in *Fundamental Bacterial Genetics* (John Wiley & Sons, 2009).

109. Maiques, E. *et al.* Beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J. Bacteriol.* **188,** 2726–9 (2006).

110. Ubeda, C. *et al.* Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.* **56,** 836–44 (2005).

111. Miller, C. *et al.* SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. *Science* **305,** 1629–31 (2004).

112. Phillips, I., Culebras, E., Moreno, F. & Baquero, F. Induction of the SOS response by new 4-quinolones. *J. Antimicrob. Chemother.* **20,** 631–8 (1987).

113. Favre, A., Chams, V. & Caldeira de Araujo, A. Photosensitized UVA light induction of the SOS response in *Escherichia coli*. *Biochimie* **68,** 857–64 (1986).

114. Łoś, M. & Węgrzyn, G. Pseudolysogeny. *Adv. Virus Res.* **82,** 339–49 (2012).

115. Fortier, L.-C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4,** 354–65 (2013).

116. Campbell, A. Sensitive mutants of bacteriophage lambda. *Virology* **14,** 22–32 (1961).

117. Hendrix, R. W. in *The Bacteriophage Lambda (Volume 02)* (1971).

118. Hershey, A. D., Burgi, E. & Ingraham, L. Cohesion of DNA molecules isolated from phage lamda. *Proc. Natl. Acad. Sci. U. S. A.* **49,** 748–55 (1963).

119. Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. Nucleotide sequence of bacteriophage lambda DNA. *J. Mol. Biol.* **162,** 729–73 (1982).

120. Friedman, D. I. & Gottesman, M. Lytic mode of lambda development. *Lambda II - NLM Catalog - NCBI* (1983).

121. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Evolution (N. Y).* **96,** 2192–2197 (1999).

122. Botstein, D. A theory of modular evolution for bacteriophages. *Ann. N. Y. Acad. Sci.* **354,** 484–90 (1980).

123. Clark, A. J., Inwood, W., Cloutier, T. & Dhillon, T. . Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J. Mol. Biol.* **311,** 657–679 (2001).

124. Mirold, S. *et al.* Salmonella host cell invasion emerged by acquisition of a mosaic of separate genetic elements, including *Salmonella* pathogenicity island 1 (SPI1), SPI5, and sopE2. *J. Bacteriol.* **183,** 2348–58 (2001).

125. De Paepe, M. *et al.* Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: The role of Rad52-like recombinases. *PLoS Genet.* **10,** e1004181 (2014).

126. Martinsohn, J. T., Radman, M. & Petit, M. A. The lambda red proteins promote efficient recombination between diverged sequences: Implications for bacteriophage genome mosaicism. *PLoS Genet.* **4,** e1000065 (2008).

127. Swenson, K. M., Guertin, P., Deschênes, H. & Bergeron, A. Reconstructing the modular recombination history of *Staphylococcus aureus* phages. *BMC Bioinformatics* S17 (2013).

128. Baker, J., Limberger, R., Schneider, S. J. & Campbell, A. Recombination and modular exchange in the genesis of new lambdoid phages. *New Biol.* **3,** 297–308 (1991).

129. Casjens, S. R. Comparative genomics and evolution of the tailed-bacteriophages. *Curr. Opin. Microbiol.* **8,** 451–8 (2005).

130. Ptashne, M. Phage repressors. In: *Strategy of the viral genom*e. *Ciba Found. Symp.* 155–68 (1971).

131. Benson, N. & Youderian, P. Phage lambda Cro protein and cI repressor use two different patterns of specific protein-DNA interactions to achieve sequence specificity *in vivo*. *Genetics* **121,** 5–12 (1989).

132. Little, J. W. Autodigestion of LexA and phage lambda repressors. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 1375–9 (1984).

133. Kim, B. & Little, J. W. LexA and lambda Cl repressors as enzymes: Specific cleavage in an intermolecular reaction. *Cell* **73,** 1165–73 (1993).

134. Campbell, A. M. Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* **174,** 7495–9 (1992).

135. Robert A. Weisberg & Arthur Landy. In *Lambda II - NLM Catalog - NCBI* (1983).

136. Gottesman, M. E. & Weisberg, R. A. In *The Bacteriophage Lambda (*Chapter 6*, Volume 02)* (1971).

137. Gottesman, M. E. & Yarmolinsky, M. B. Integration-negative mutants of bacteriophage lambda. *J. Mol. Biol.* **31,** 487–505 (1968).

138. Seah, N. E. *et al.* Nucleoprotein architectures regulating the directionality of viral integration and excision. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 12372–7 (2014).

139. Miller, H. I. *et al.* Regulation of the integration-excision reaction by bacteriophage lambda. *Cold Spring Harb. Symp. Quant. Biol.* **45 Pt 1,** 439–45 (1981).

140. Guarneros, G. & Echols, H. New mutants of bacteriophage lambda with a specific defect in excision from the host chromosome. *J. Mol. Biol.* **47,** 565–74 (1970).

141. Schindler, D. & Echols, H. Retroregulation of the int gene of bacteriophage lambda: Control of translation completion. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4475–9 (1981).

142. Kobiler, O., Oppenheim, A. B. & Herman, C. Recruitment of host ATP-dependent proteases by bacteriophage lambda. *J. Struct. Biol.* **146,** 72–8

143. Ball, C. A. & Johnson, R. C. Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. *J. Bacteriol.* **173,** 4027–31 (1991).

144. Craig, N. L. & Nash, H. A. *E. coli* integration host factor binds to specific sites in DNA. *Cell* **39,** 707–16 (1984).

145. Skalka, A., Poonian, M. & Bartl, P. Concatemers in DNA replication: Electron microscopic studies of partially denatured intracellular lambda DNA. *J. Mol. Biol.* **64,** 541–50 (1972).

146. Wake, R. G., Kaiser, A. D. & Inman, R. B. Isolation and structure of phage lambda head-mutant DNA. *J. Mol. Biol.* **64,** 519–40 (1972).

147. Dodson, M. *et al.* Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda: Localized unwinding of duplex DNA by a six-protein reaction. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 7638–42 (1986).

148. Bouché, J. P., Rowen, L. & Kornberg, A. The RNA primer synthesized by primase to initiate phage G4 DNA replication. *J. Biol. Chem.* **253,** 765–9 (1978).

149. Espeli, O. & Marians, K. J. Untangling intracellular DNA topology. *Molecular Microbiology* **52,** 925–931 (2004).

150. Weigel, C. *et al.* Bacteriophage replication modules. *FEMS Microbiol. Rev.* **30,** 321–81 (2006).

151. Takahashi, S. The starting point and direction of rolling-circle replicative intermediates of coliphage lambda DNA. *Mol. Gen. Genet.* **142,** 137–53 (1975).

152. Yarnell, W. S. & Roberts, J. W. Mechanism of intrinsic transcription termination and antitermination. *Science* **284,** 611–5 (1999).

153. Ferrer, M. D. *et al.* RinA controls phage-mediated packaging and transfer of virulence genes in Gram-positive bacteria. *Nucleic Acids Res* **39,** 5866–5878 (2011).

154. Quiles-Puchalt, N. *et al.* A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. *Nucleic Acids Res* **41,** 7260–75 (2013).

155. Desiere, F., Pridmore, R. D. & Brüssow, H. Comparative genomics of the late gene cluster from *Lactobacillus* phages. *Virology* **275,** 294–305 (2000).

156. Brussow, H., Desiere, F., Brüssow, H. & Desiere, F. Comparative phage genomics and the evolution of *Siphoviridae*: Insights from dairy phages. *Mol. Microbiol.* **39,** 213–22

(2001).

157. Rixon, F. J. & Schmid, M. F. Structural similarities in DNA packaging and delivery apparatuses in Herpesvirus and dsDNA bacteriophages. *Curr. Opin. Virol.* **5,** 105–10 (2014).

158. Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V, Arisaka, F. & Rossmann, M. G. Structure and morphogenesis of bacteriophage T4. *Cell. Mol. Life Sci.* **60,** 2356–70 (2003).

159. Feiss, M. & Rao, V. B. The bacteriophage DNA packaging machine. *Adv Exp Med Biol* **726,** 489–509 (2012).

160. Rao, V. B. & Feiss, M. Mechanisms of DNA packaging by large double-stranded DNA Viruses. *Annu. Rev. Virol.* **2,** 351–78 (2015).

161. Kaiser, D., Syvanen, M. & Masuda, T. DNA packaging steps in bacteriophage lambda head assembly. *J. Mol. Biol.* **91,** 175–86 (1975).

162. Valpuesta, J. M. & Carrascosa, J. L. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Q. Rev. Biophys.* **27,** 107–155 (1994).

163. Agirrezabala, X. *et al.* Structure of the connector of bacteriophage T7 at 8A resolution: Structural homologies of a basic component of a DNA translocating machinery. *J. Mol. Biol.* **347,** 895–902 (2005).

164. Lander, G. C. *et al.* The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* **312,** 1791–5 (2006).

165. Doan, D. N. P. & Dokland, T. The gpQ portal protein of bacteriophage P2 forms dodecameric connectors in crystals. *J. Struct. Biol.* **157,** 432–6 (2007).

166. Morais, M. C. *et al.* Defining molecular and domain boundaries in the bacteriophage phi29 DNA packaging motor. *Structure* **16,** 1267–74 (2008).

167. Duda, R. L. *et al.* Structural transitions during bacteriophage HK97 head assembly. *J. Mol. Biol.* **247,** 618–35 (1995).

168. Huet, A., Conway, J. F., Letellier, L. & Boulanger, P. *In vitro* assembly of the T=13 procapsid of bacteriophage T5 with its scaffolding domain. *J. Virol.* **84,** 9350–8 (2010).

169. Weigele, P. R., Sampson, L., Winn-Stapley, D. & Casjens, S. R. Molecular genetics of bacteriophage P22 scaffolding protein's functional domains. *J. Mol. Biol.* **348,** 831–44 (2005).

170. Black, L.W., Showe, M.K., and Steven, A. In *Morphogenesis of the T4 head: Karam, J. D. Ed.* (1994).

171. Rossmann, M. G., Mesyanzhinov, V. V, Arisaka, F. & Leiman, P. G. The bacteriophage T4 DNA injection machine. *Curr. Opin. Struct. Biol.* **14,** 171–80 (2004).

172. Dokland, T. Scaffolding proteins and their role in viral assembly. *Cell. Mol. Life Sci.* **56,** 580–603 (1999).

173. Rizzo, A. A. *et al.* Multiple functional roles of the accessory I-domain of bacteriophage P22 coat protein revealed by NMR structure and CryoEM modeling. *Structure* **22,** 830–41 (2014).

174. Suhanovsky, M. M. & Teschke, C. M. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* **479–480,** 487–97 (2015).

175. Mousset, S. & Thomas, R. Ter, a function which generates the ends of the mature λ chromosome. *Nature* **221,** 242–244 (1969).

176. Mitchell, M. S., Matsuzaki, S., Imai, S. & Rao, V. B. Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. *Nucleic Acids Res.* **30,** 4009–21 (2002).

177. Hwang, Y., Hang, J. Q., Neagle, J., Duffy, C. & Feiss, M. Endonuclease and helicase activities of bacteriophage lambda terminase: Changing nearby residue 515 restores activity to the gpA K497D mutant enzyme. *Virology* **277,** 204–14 (2000).

178. Oliveira, L., Henriques, A. O. & Tavares, P. Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency. *J. Biol. Chem.* **281,** 21914–23 (2006).

179. Goetzinger, K. R. & Rao, V. B. Defining the ATPase center of bacteriophage T4 DNA packaging machine: Requirement for a catalytic glutamate residue in the large terminase protein gp17. *J. Mol. Biol.* **331,** 139–54 (2003).

180. Hwang, Y. & Feiss, M. Mutations affecting the high affinity ATPase center of gpA, the large subunit of bacteriophage lambda terminase, inactivate the endonuclease activity of terminase. *J. Mol. Biol.* **261,** 524–35 (1996).

181. Nemecek, D. *et al.* Subunit conformations and assembly states of a DNA-translocating motor: Tthe terminase of bacteriophage P22. *J. Mol. Biol.* **374,** 817–36 (2007).

182. Casjens, S. R. *et al.* The generalized transducing *Salmonella* bacteriophage ES18: Complete genome sequence and DNA packaging strategy. *J. Bacteriol.* **187,** 1091–104 (2005).

183. Camacho, A. G., Gual, A., Lurz, R., Tavares, P. & Alonso, J. C. Bacillus subtilis bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. *J. Biol. Chem.* **278,** 23251–9 (2003).

184. Ghosh-Kumar, M., Alam, T. I., Draper, B., Stack, J. D. & Rao, V. B. Regulation by interdomain communication of a headful packaging nuclease from bacteriophage T4. *Nucleic Acids Res.* **39,** 2742–55 (2011).

185. de Beer, T. *et al.* Assignment of the 1H, 13C, and 15N resonances of the DNA binding domain of gpNu1, a genome packaging protein from bacteriophage lambda. *J. Biomol. NMR* **18,** 69–70 (2000).

186. Sun, S. *et al.* The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* **135,** 1251–62 (2008).

187. Roy, A. *et al.* Small terminase couples viral DNA binding to genome-packaging ATPase activity. *Structure* **20,** 1403–1413 (2012).

188. Rao, V. B. & Feiss, M. The bacteriophage DNA packaging motor. *Annu Rev Genet* **42,** 647–681 (2008).

189. Oliveira, L., Tavares, P. & Alonso, J. C. Headful DNA packaging: Bacteriophage SPP1 as a model system. *Virus Res* **173,** 247–259 (2013).

190. Wu, H., Sampson, L., Parr, R. & Casjens, S. The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. *Mol. Microbiol.* **45,** 1631–46 (2002).

191. Huang, H. & Yuan, H. S. The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. *J. Mol. Biol.* **368,** 812–21 (2007).

192. Moodley, S., Maxwell, K. L. & Kanelis, V. The protein gp74 from the bacteriophage HK97 functions as a HNH endonuclease. *Protein Sci.* **21,** 809–18 (2012).

193. Xu, S. & Gupta, Y. K. Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease. *Nucleic Acids Res.* **41,** 378–90 (2013).

194. Kala, S. *et al.* HNH proteins are a widespread component of phage DNA packaging machines. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6022–7 (2014).

195. Heming, J. D., Huffman, J. B., Jones, L. M. & Homa, F. L. Isolation and characterization of the herpes simplex virus 1 terminase complex. *J. Virol.* **88,** 225–36 (2014).

196. Borst, E. M. *et al.* The human cytomegalovirus UL51 protein is essential for viral genome cleavage-packaging and interacts with the terminase subunits pUL56 and pUL89. *J. Virol.* **87,** 1720–32 (2013).

197. Chen, D.-H. *et al.* Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 1355–60 (2011).

198. King, J. & Casjens, S. Catalytic head assembling protein in virus morphogenesis. *Nature* **251,** 112–9 (1974).

199. Fane, B. A. & Prevelige, P. E. Mechanism of scaffolding-assisted viral assembly. *Adv.*

Protein Chem. **64,** 259–99 (2003).

200. Spilman, M. S. *et al.* Assembly of bacteriophage 80α capsids in a *Staphylococcus aureus* expression system. *Virology* **434,** 242–50 (2012).

201. Steven, A. C., Heymann, J. B., Cheng, N., Trus, B. L. & Conway, J. F. Virus maturation: Dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr. Opin. Struct. Biol.* **15,** 227–36 (2005).

202. Gilcrease, E. B., Winn-Stapley, D. A., Hewitt, F. C., Joss, L. & Casjens, S. R. Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. *J. Bacteriol.* **187,** 2050–7 (2005).

203. Tang, L., Gilcrease, E. B., Casjens, S. R. & Johnson, J. E. Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. *Structure* **14,** 837–45 (2006).

204. Qin, L., Fokine, A., O'Donnell, E., Rao, V. B. & Rossmann, M. G. Structure of the small outer capsid protein, Soc: A clamp for stabilizing capsids of T4-like phages. *J. Mol. Biol.* **395,** 728–41 (2010).

205. King, J., Lenk, E. V & Botstein, D. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. II. Morphogenetic pathway. *J. Mol. Biol.* **80,** 697–731 (1973).

206. Orlova, E. V *et al.* Structure of a viral DNA gatekeeper at 10 A resolution by cryo-electron microscopy. *EMBO J.* **22,** 1255–62 (2003).

207. Lhuillier, S. *et al.* Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 8507–12 (2009).

208. Casjens S, H. R. in *The Bacteriophages* 15–75. (1988).

209. Camacho, A., Jiménez, F., Viñuela, E. & Salas, M. Order of assembly of the lower collar and the tail proteins of *Bacillus subtilis* bacteriophage phi 29. *J. Virol.* **29,** 540–5 (1979).

210. Lander, G. C. *et al.* The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit. *Structure* **17,** 789–99 (2009).

211. Sciara, G. *et al.* Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 6852–7 (2010).

212. Katsura, I. Mechanism of length determination in bacteriophage lambda tails. *Adv. Biophys.* **26,** 1–18 (1990).

213. Auzat, I., Dröge, A., Weise, F., Lurz, R. & Tavares, P. Origin and function of the two major tail proteins of bacteriophage SPP1. *Mol. Microbiol.* **70,** 557–69 (2008).

214. Aksyuk, A. A. *et al.* The tail sheath structure of bacteriophage T4: A molecular machine for infecting bacteria. *EMBO J.* **28,** 821–9 (2009).

215. Vianelli, A. *et al.* Bacteriophage T4 self-assembly: Localization of gp3 and its role in determining tail length. *J. Bacteriol.* **182,** 680–8 (2000).

216. Zhao, L., Kanamaru, S., Chaidirek, C. & Arisaka, F. P15 and P3, the tail completion proteins of bacteriophage T4, both form hexameric rings. *J. Bacteriol.* **185,** 1693–700 (2003).

217. Heller, K. J. Identification of the phage gene for host receptor specificity by analyzing hybrid phages of T5 and BF23. *Virology* **139,** 11–21 (1984).

218. Wang, J., Hofnung, M. & Charbit, A. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12. *J. Bacteriol.* **182,** 508–12 (2000).

219. Coombs, D.H., and Arisaka, F. In *Molecular Biology of Bacteriophage T4.* 259–281 (1994).

220. Young, R. Bacteriophage lysis: Mechanism and regulation. *Microbiol. Rev.* **56,** 430–81 (1992).

221. Young, R. & Bläsi, U. Holins: Form and function in bacteriophage lysis. *FEMS Microbiol. Rev.* **17,** 191–205 (1995).

222. Schmidt, H. & Hensel, M. Pathogenicity islands in bacterial pathogenesis. *Clin. Microbiol.*

Rev. **17,** 14–56 (2004).

223. Hacker, J. *et al.* Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. *Microb. Pathog.* **8,** 213–25 (1990).

224. Blum, G. *et al.* Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immun.* **62,** 606–14 (1994).

225. Groisman, E. A. & Ochman, H. Pathogenicity islands: Bacterial evolution in quantum leaps. *Cell* **87,** 791–4 (1996).

226. Hacker, J. *et al.* Pathogenomics of mobile genetic elements of toxigenic bacteria. *Int. J. Med. Microbiol.* **293,** 453–61 (2004).

227. Dobrindt, U. *et al.* Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J. Bacteriol.* **185,** 1831–40 (2003).

228. Hacker, J. & Kaper, J. B. Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54,** 641–79 (2000).

229. Sullivan, J. T. *et al.* Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J. Bacteriol.* **184,** 3086–95 (2002).

230. Gaillard, M. *et al.* The clc element of *Pseudomonas sp*. strain B13, a genomic island with various catabolic properties. *J. Bacteriol.* **188,** 1999–2013 (2006).

231. Larbig, K. D. *et al.* Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J. Bacteriol.* **184,** 6665–80 (2002).

232. Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2,** 414–424 (2004).

233. Gill, S. R. *et al.* Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J. Bacteriol.* **187,** 2426–2438 (2005).

234. Ito, T., Okuma, K., Ma, X. X., Yuzawa, H. & Hiramatsu, K. Insights on antibiotic resistance of *Staphylococcus aureus* from its whole genome: Genomic island SCC. *Drug Resist. Updat.* **6,** 41–52 (2003).

235. Arnold, H. P. *et al.* The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol. Microbiol.* **34,** 217–26 (1999).

236. Úbeda, C. *et al.* Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol. Microbiol.* **49,** 193–210 (2003).

237. O'Neill, A. J., Larsen, A. R., Skov, R., Henriksen, A. S. & Chopra, I. Characterization of the epidemic European fusidic acid-resistant impetigo clone of *Staphylococcus aureus*. *J. Clin. Microbiol.* **45,** 1505–10 (2007).

238. Viana, D. *et al.* Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. *Mol. Microbiol.* **77,** 1583–1594 (2010).

239. Matos, R. C. *et al. Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. *PLoS Genet.* **9,** e1003539 (2013).

240. Martínez-Rubio, R. *et al.* Phage-inducible islands in the coccal sea. *The ISME Journal* (2016) (Epub ahead of print).

241. Úbeda, C. *et al.* SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol. Microbiol.* **67,** 493–503 (2008).

242. Tormo-Más, M. A. *et al.* Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465,** 779–782 (2010).

243. Penades, J. R., Donderis, J., Garcia-Caballer, M., Tormo-Mas, M. A. & Marina, A. dUTPases, the unexplored family of signalling molecules. *Curr Opin Microbiol* **16,** 163–170 (2013).

244. Lane, K. Transcriptional crosstalk between helper bacteriophages and *Staphylococcal aureus* pathogenicity islands. PhD Dissertation. Virginia Commonwealth University (2013).

245. Maiques, E. *et al.* Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. *J. Bacteriol.* **189,** 5608–5616 (2007).

246. Ubeda, C. *et al.* Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Mol. Microbiol.* **72,** 98–108 (2009).

247. Ruzin, A., Lindsay, J. & Novick, R. P. Molecular genetics of SaPI1--a mobile pathogenicity island in *Staphylococcus aureus*. *Mol. Microbiol.* **41,** 365–77 (2001).

248. Mir-Sanchis, I. *et al.* Control of *Staphylococcus aureus* pathogenicity island excision. *Mol. Microbiol.* **85,** 833–45 (2012).

249. Campbell, A. Phage integration and chromosome structure. A personal history. *Annu. Rev. Genet.* **41,** 1–11 (2007).

250. Chen, J. & Novick, R. P. Phage-mediated intergeneric transfer of toxin genes. *Science* **323,** 139–41 (2009).

251. Ubeda, C., Barry, P., Penadés, J. R. & Novick, R. P. A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 14182–8 (2007).

252. Ubeda, C. *et al.* Structure-function analysis of the SaPIbov1 replication origin in *Staphylococcus aureus*. *Plasmid* **67,** 183–90 (2012).

253. Tallent, S. M., Langston, T. B., Moran, R. G. & Christie, G. E. Transducing particles of *Staphylococcus aureus* pathogenicity island SaPI1 are comprised of helper phage-encoded proteins. *J. Bacteriol.* **189,** 7520–4 (2007).

254. Tormo, M. A. *et al. Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *J Bacteriol* **190,** 2434–2440 (2008).

255. Ram, G. *et al.* Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 16300–5 (2012).

256. Úbeda, C. *et al.* SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol. Microbiol.* **65,** 41–50 (2007).

257. Damle, P. K. *et al.* The roles of SaPI1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. *Virology* **432,** 277–82 (2012).

258. Poliakov, A. *et al.* Capsid size determination by *Staphylococcus aureus* pathogenicity island SaPI1 involves specific incorporation of SaPI1 proteins into procapsids. *J. Mol. Biol.* **380,** 465–75 (2008).

259. Ram, G., Chen, J., Ross, H. F. & Novick, R. P. Precisely modulated pathogenicity island interference with late phage gene transcription. *Proc Natl Acad Sci U S A* **111,** 14536–14541 (2014).

260. Bento, J. C., Lane, K. D., Read, E. K., Cerca, N. & Christie, G. E. Sequence determinants for DNA packaging specificity in the *S. aureus* pathogenicity island SaPI1. *Plasmid* **71,** 8–15 (2014).

261. Chen, J., Ram, G., Penadés, J. R., Brown, S. & Novick, R. P. Pathogenicity island-directed transfer of unlinked chromosomal virulence genes. *Mol. Cell* **57,** 138–49 (2015).

262. Harwich, MD. Transcriptional profiling of staphylococcal bacteriophage 80α and regulatory interactions with pathogenicity island SaPI1. PhD Dissertation. Virginia Commonwealth University (2009).

263. Fitzgerald, J. R. *et al.* Characterization of a putative pathogenicity island from bovine *Staphylococcus aureus* encoding multiple superantigens. *J. Bacteriol.* **183,** 63–70

(2001).

264. Subedi, A. *et al.* Sequence analysis reveals genetic exchanges and intraspecific spread of SaPI2, a pathogenicity island involved in menstrual toxic shock. *Microbiology* **153,** 3235–45 (2007).

265. Guinane, C. M. *et al.* Evolutionary genomics of *Staphylococcus aureus* reveals insights into the origin and molecular basis of ruminant host adaptation. *Genome Biol Evol* **2,** 454–466 (2010).

266. Chen, H.-J. *et al.* New structure of phage-related islands carrying *fus*B and a virulence gene in fusidic acid-resistant *Staphylococcus epidermidis*. *Antimicrob. Agents Chemother.* **57,** 5737–9 (2013).

267. Kuroda, M. *et al.* Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13272–7 (2005).

268. Diep, B. A. *et al.* Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet (London, England)* **367,** 731–9 (2006).

269. Tormo-Más, M. Á. *et al.* Phage dUTPases control transfer of virulence genes by a proto-oncogenic G protein-like mechanism. *Mol. Cell* **49,** 947–58 (2013).

270. Quiles-Puchalt, N. & Carpena, N. Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 6016–21 (2014).

271. Dearborn, A. D. & Dokland, T. Mobilization of pathogenicity islands by *Staphylococcus aureus* strain Newman bacteriophages. *Bacteriophage* **2,** 70–78 (2012).

272. Liu, J. *et al.* Antimicrobial drug discovery through bacteriophage genomics. *Nat. Biotechnol.* **22,** 185–91 (2004).

273. Frígols, B. *et al.* Virus satellites drive viral evolution and ecology. *PLOS Genet.* **11,** e1005609 (2015).

274. Six, E. W. The helper dependence of satellite bacteriophage P4: Which gene functions of bacteriophage P2 are needed by P4? *Virology* **67,** 249–263 (1975).

275. Christie, G. E. & Dokland, T. Pirates of the Caudovirales. *Virology* **434,** 210–21 (2012).

276. Chang, J. R., Poliakov, A., Prevelige, P. E., Mobley, J. A. & Dokland, T. Incorporation of scaffolding protein gpO in bacteriophages P2 and P4. *Virology* **370,** 352–61 (2008).

277. Quiles-Puchalt, N., Martinez-Rubio, R., Ram, G., Lasa, I. & Penades, J. R. Unravelling bacteriophage phi11 requirements for packaging and transfer of mobile genetic elements in *Staphylococcus aureus*. *Mol Microbiol* **91,** 423–437 (2014).

278. Spilman, M. S. *et al.* A conformational switch involved in maturation of *Staphylococcus aureus* bacteriophage 80α capsids. *J. Mol. Biol.* **405,** 863–76 (2011).

279. Novick, R. P. & Ram, G. The floating (pathogenicity) island: A genomic dessert. *Trends Genet.* **32,** 114–126 (2016).

280. Dearborn, A. D. *et al.* The *Staphylococcus aureus* pathogenicity island 1 protein gp6 functions as an internal scaffold during capsid size determination. *J. Mol. Biol.* **412,** 710–22 (2011).

281. Christie, G. E. *et al.* The complete genomes of *Staphylococcus aureus* bacteriophages 80 and 80alpha--implications for the specificity of SaPI mobilization. *Virology* **407,** 381–390 (2010).

282. Barrett, K. J., Marsh, M. L. & Calendar, R. Interactions between a satellite bacteriophage and its helper. *J. Mol. Biol.* **106,** 683–707 (1976).

283. Marvik, O. J. *et al.* The capsid size-determining protein Sid forms an external scaffold on phage P4 procapsids. *J. Mol. Biol.* **251,** 59–75 (1995).

284. Dearborn, A. D. *et al.* Structure and size determination of bacteriophage P2 and P4 procapsids: Function of size responsiveness mutations. *J. Struct. Biol.* **178,** 215–24

(2012).

285. Shore, D., Dehò, G., Tsipis, J. & Goldstein, R. Determination of capsid size by satellite bacteriophage P4. *Proc. Natl. Acad. Sci. U. S. A.* **75,** 400–4 (1978).

286. Isaksen, M. L., Dokland, T. & Lindqvist, B. H. Characterization of the capsid associating activity of bacteriophage P4's Psu protein. *Virology* **194,** 674–681 (1993).

287. Ranjan, A., Banerjee, R., Pani, B., Sen, U. & Sen, R. The moonlighting function of bacteriophage P4 capsid protein, Psu, as a transcription antiterminator. *Bacteriophage* **3,** e25657 (2013).

288. Ziermann, R. & Calendar, R. Characterization of the *cos* sites of bacteriophages P2 and P4. *Gene* **96,** 9–15 (1990).

289. Diana, C., Dehò, G., Geisselsoder, J., Tinelli, L. & Goldstein, R. Viral interference at the level of capsid size determination by satellite phage P4. *J. Mol. Biol.* **126,** 433–45 (1978).

290. Six, E. W., Sunshine, M. G., Williams, J., Haggard-Ljungquist, E. & Lindqvist, B. H. Morphopoietic switch mutations of bacteriophage P2. *Virology* **182,** 34–46 (1991).

291. Duerkop, B. A., Clements, C. V, Rollins, D., Rodrigues, J. L. M. & Hooper, L. V. A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17621–6 (2012).

292. Scott, J., Thompson-Mayberry, P., Lahmamsi, S., King, C. J. & McShan, W. M. Phage-associated mutator phenotype in group A streptococcus. *J. Bacteriol.* **190,** 6290–301 (2008).

293. Scott, J., Nguyen, S. V, King, C. J., Hendrickson, C. & McShan, W. M. Phage-like *Streptococcus pyogenes* chromosomal islands (SpyCI) and mutator phenotypes: Control by growth state and rescue by a SpyCI-encoded promoter. *Front. Microbiol.* **3,** 317 (2012).

294. Nguyen, S. V & McShan, W. M. Chromosomal islands of *Streptococcus pyogenes* and related streptococci: Molecular switches for survival and virulence. *Front. Cell. Infect. Microbiol.* **4,** 109 (2014).

295. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494,** 489–91 (2013).

296. Cucarella, C. *et al.* Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *J. Bacteriol.* **183,** 2888–96 (2001).

297. Charpentier, E. *et al.* Novel cassette-based shuttle vector system for Gram-positive bacteria. *Appl. Environ. Microbiol.* **70,** 6076–85 (2004).

298. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177,** 4121–30 (1995).

299. Augustin, J. *et al.* Genetic analysis of epidermin biosynthetic genes and epidermin-negative mutants of *Staphylococcus epidermidis*. *Eur. J. Biochem.* **204,** 1149–54 (1992).

300. Kreiswirth, B. N. *et al.* The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305,** 709–12cd (1983).

301. Novick, R. Properties of a cryptic high-frequency transducing phage in *Staphylococcus aureus*. *Virology* **33,** 155–166 (1967).

302. Arnaud, M., Chastanet, A. & Débarbouillé, M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, Gram-positive bacteria. *Appl. Environ. Microbiol.* **70,** 6887–91 (2004).

303. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97,** 6640–6645 (2000).

304. Brückner, R., Zyprian, E. & Matzura, H. Expression of a chloramphenicol-resistance determinant carried on hybrid plasmids in Gram-positive and Gram-negative bacteria.

Gene **32,** 151–60 (1984).

305. Sambrook, J. Molecular cloning: A laboratory manual. Second edition. Volumes 1, 2, and 3. Current protocols in molecular biology. *Cell* **61,** 17–18 (1989).

306. Ausubel, F.M., Brent, R.M, Kingston, R.E. *et al*. Current Protocols in Molecular Biology. Wiley: Ausubel, F (1990)

307. Chien, Y., Manna, A. C., Projan, S. J. & Cheung, A. L. SarA, a global regulator of virulence determinants in *Staphylococcus aureus*, binds to a conserved motif essential for *sar*-dependent gene regulation. *J. Biol. Chem.* **274,** 37169–76 (1999).

308. Straus, D. & Ausubel, F. M. Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 1889–93 (1990).

309. Brückner, R. Gene replacement in *Staphylococcus carnosus* and *Staphylococcus xylosus*. *FEMS Microbiol. Lett.* **151,** 1–8 (1997).

310. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–402 (1997).

311. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–9 (2011).

312. Simossis, V. A. & Heringa, J. PRALINE: A multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* **33,** W289-94 (2005).

313. Karplus, K., Barrett, C. & Hughey, R. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* **14,** 846–56 (1998).

314. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–38 (2010).

315. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67,** 235–42 (2011).

316. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66,** 486–501 (2010).

317. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7,** 1511–22 (2012).

318. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–58 (2015).

319. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320-4 (2014).

320. Konagurthu, A. S. *et al.* MUSTANG-MR structural sieving server: Applications in protein structural analysis and crystallography. *PLoS One* **5,** e10048 (2010).

321. García, P. *et al.* Functional genomic analysis of two *Staphylococcus aureu*s phages isolated from the dairy environment. *Appl. Environ. Microbiol.* **75,** 7663–73 (2009).

322. Shen, B. W., Landthaler, M., Shub, D. A. & Stoddard, B. L. DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **342,** 43–56 (2004).

323. Xu, S. *et al.* Structure determination and biochemical characterization of a putative HNH endonuclease from *Geobacter metallireducens* GS-15. *PLoS One* **8,** e72114 (2013).

324. Shen, B. W. *et al.* Unusual target site disruption by the rare-cutting HNH restriction endonuclease *Pac*I. *Structure* **18,** 734–43 (2010).

325. Zhao, H., Christensen, T. E., Kamau, Y. N. & Tang, L. Structures of the phage Sf6 large terminase provide new insights into DNA translocation and cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 8075–80 (2013).

326. Rentas, F. J. & Rao, V. B. Defining the bacteriophage T4 DNA packaging machine: Evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17. *J. Mol. Biol.* **334,** 37–52 (2003).

327. Smits, C. *et al.* Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep.* **10,** 592–8 (2009).

328. Sippy, J. & Feiss, M. Initial *cos* cleavage of bacteriophage lambda concatemers requires proheads and gpFI in vivo. *Mol. Microbiol.* **52,** 501–13 (2004).

329. Cornilleau, C. *et al.* The nuclease domain of the SPP1 packaging motor coordinates DNA cleavage and encapsidation. *Nucleic Acids Res.* **41,** 340–354 (2013).

330. Chen, J. *et al.* Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by *cos* phages. *ISME J.* **9,** 1260–1263 (2014).

331. Winstel, V. *et al.* Wall teichoic acid structure governs horizontal gene transfer between major bacterial pathogens. *Nat. Commun.* **4,** 2345 (2013).

332. Rooijakkers, S. H. M. *et al.* Early expression of SCIN and CHIPS drives instant immune evasion by *Staphylococcus aureus*. *Cell. Microbiol.* **8,** 1282–93 (2006).

333. Hyman, P. & Abedon, S. T. In *Advances in Applied Microbiology*. **70,** 217–248 (2010).

334. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8,** 317–327 (2010).

335. Sass, P. & Bierbaum, G. Lytic activity of recombinant bacteriophage phi11 and phi12 endolysins on whole cells and biofilms of *Staphylococcus aureus*. *Appl. Environ. Microbiol.* **73,** 347–52 (2007).

336. Matsuzaki, S. *et al.* Experimental protection of mice against lethal *Staphylococcus aureus* infection by novel bacteriophage phi MR11. *J. Infect. Dis.* **187,** 613–24 (2003).

337. Wills, Q. F., Kerrigan, C. & Soothill, J. S. Experimental bacteriophage protection against *Staphylococcus aureus* abscesses in a rabbit model. *Antimicrob. Agents Chemother.* **49,** 1220–1 (2005).

338. Bruttin, A. & Brüssow, H. Human volunteers receiving *Escherichia coli* phage T4 orally: A safety test of phage therapy. *Antimicrob. Agents Chemother.* **49,** 2874–8 (2005).

339. Kishor, C. *et al.* Phage therapy of staphylococcal chronic osteomyelitis in experimental animal model. *Indian J. Med. Res.* **143,** 87–94 (2016).

340. Pincus, N. B. *et al.* Strain specific phage treatment for *Staphylococcus aureu*s infection is influenced by host immunity and site of infection. *PLoS One* **10,** e0124280 (2015).

341. Kaźmierczak, Z., Górski, A. & Dąbrowska, K. Facing antibiotic resistance: *Staphylococcus aureus* phages as a medical tool. *Viruses* **6,** 2551–70 (2014).

342. Hejnowicz, M. S. *et al.* In *Advances in Virus Research* **83,** 143–216 (2012).

343. Uchiyama, J. *et al.* Intragenus generalized transduction in *Staphylococcus spp.* by a novel giant phage. *ISME J.* **8,** 1949–52 (2014).

344. Penadés, J. R. & Christie, G. E. The phage-inducible chromosomal islands: A family of highly evolved molecular parasites. *Annu. Rev. Virol.* **2,** 181–201 (2015).

345. Helgstrand, C. *et al.* The refined structure of a protein catenane: The HK97 bacteriophage capsid at 3.44Å resolution. *J. Mol. Biol.* **334,** 885–899 (2003).

346. Conway, J. F., Duda, R. L., Cheng, N., Hendrix, R. W. & Steven, A. C. Proteolytic and conformational control of virus capsid maturation: The bacteriophage HK97 system. *J. Mol. Biol.* **253,** 86–99 (1995).

347. Oh, B., Moyer, C. L., Hendrix, R. W. & Duda, R. L. The delta domain of the HK97 major capsid protein is essential for assembly. *Virology* **456–457,** 171–8 (2014).

348. Chang, J. R., Spilman, M. S. & Dokland, T. Assembly of bacteriophage P2 capsids from capsid protein fused to internal scaffolding protein. *Virus Genes* **40,** 298–306 (2010).

349. Huet, A., Duda, R. L., Hendrix, R. W., Boulanger, P. & Conway, J. F. Correct assembly of the bacteriophage T5 procapsid requires both the maturation protease and the portal complex. *J. Mol. Biol.* **428,** 165–181 (2016).

350. Wang, S., Chang, J. R. & Dokland, T. Assembly of bacteriophage P2 and P4 procapsids with internal scaffolding protein. *Virology* **348,** 133–40 (2006).

351. Penadés, J. R., Chen, J., Quiles-Puchalt, N., Carpena, N. & Novick, R. P. Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* **23C,** 171–178 (2014).

352. Iandolo, J. J. *et al.* Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of *Staphylococcus aureus* 8325. *Gene* **289,** 109–18 (2002).

353. Yasmin, A. *et al.* Comparative genomics and transduction potential of *Enterococcus faecalis* temperate bacteriophages. *J. Bacteriol.* **192,** 1122–1130 (2010).

354. Mazaheri Nezhad Fard, R., Barton, M. D. & Heuzenroeder, M. W. Bacteriophage-mediated transduction of antibiotic resistance in enterococci. *Lett. Appl. Microbiol.* **52,** 559–564 (2011).

355. Bossi, L., Fuentes, J. A., Mora, G. & Figueroa-Bossi, N. Prophage contribution to bacterial population dynamics. *J. Bacteriol.* **185,** 6467–6471 (2003).

356. (IWG-SCC), I. W. G. on the C. of S. C. C. E. Classification of staphylococcal cassette chromosome *mec* (SCC*mec*): Guidelines for reporting novel SCC*mec* elements. *Antimicrob. Agents Chemother.* **53,** 4961–4967 (2009).

357. Sahm, D. F. *et al. In vitro* susceptibility studies of vancomycin-resistant *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **33,** 1588–91 (1989).

358. Paulsen, I. T. *et al.* Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* **299,** 2071–4 (2003).

359. Shankar, N., Baghdayan, A. S. & Gilmore, M. S. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. *Nature* **417,** 746–750 (2002).

360. McBride, S. M. *et al.* Genetic variation and evolution of the pathogenicity island of *Enterococcus faecalis*. *J. Bacteriol.* **191,** 3392–3402 (2009).

361. Gan, L. *et al.* Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure* **14,** 1655–65 (2006).

Publications related to this work

Quiles-Puchalt N[1], **Carpena N**[1], Alonso JC, Novick RP, Marina A, Penadés JR. (2014) Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. *Proc Natl Acad Sci U S A*. doi: 10.1073/pnas.1320538111. (*[1]N.Q.-P. and N.C. contributed equally to this work).*

Chen J, **Carpena N**, Quiles-Puchalt N, Ram G, Novick RP, Penadés JR. (2014) Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. *ISME Journal*. doi: 10.1038/ismej.2014.187.

Penadés JR, Chen J, Quiles-Puchalt N, **Carpena N**, Novick RP. (2014) Bacteriophage-mediated spread of bacterial virulence genes. *Curr Opin Microbiol*. doi: 10.1016/j.mib.2014.11.019.

Carpena N, Manning KA, Dokland T, Marina A, Penadés JR. (2016) A remarkable example of convergent evolution involving pathogenicity islands in helper *cos* phages interference. *Phil. Trans. R. Soc. B*. doi: 10.1098/rstb.2015.0505.

Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases

Nuria Quiles-Puchalt[a,b,1], Nuria Carpena[a,b,1], Juan C. Alonso[c], Richard P. Novick[d], Alberto Marina[b], and José R. Penadés[a,b,2]

[a]Institute of Infection, Immunity, and Inflammation, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow G12 8TA, United Kingdom; [b]Instituto de Biomedicina de Valencia–Consejo Superior de Investigaciones Científicas, 46010 Valencia, Spain; [c]Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnología–Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain; and [d]Skirball Institute Program in Molecular Pathogenesis and Departments of Microbiology and Medicine, New York University Medical Center, New York, NY 10016

Edited by Sankar Adhya, National Institutes of Health, National Cancer Institute, Bethesda, MD, and approved March 13, 2014 (received for review November 6, 2013)

Staphylococcal pathogenicity islands (SaPIs) are the prototypical members of a widespread family of chromosomally located mobile genetic elements that contribute substantially to intra- and interspecies gene transfer, host adaptation, and virulence. The key feature of their mobility is the induction of SaPI excision and replication by certain helper phages and their efficient encapsidation into phage-like infectious particles. Most SaPIs use the headful packaging mechanism and encode small terminase subunit (TerS) homologs that recognize the SaPI-specific *pac* site and determine SaPI packaging specificity. Several of the known SaPIs do not encode a recognizable TerS homolog but are nevertheless packaged efficiently by helper phages and transferred at high frequencies. In this report, we have characterized one of the non–*ter*S-coding SaPIs, SaPIbov5, and found that it uses two different, undescribed packaging strategies. SaPIbov5 is packaged in full-sized phage-like particles either by typical *pac*-type helper phages, or by *cos*-type phages—i.e., it has both *pac* and *cos* sites—a configuration that has not hitherto been described for any mobile element, phages included—and uses the two different phage-coded TerSs. To our knowledge, this is the first example of SaPI packaging by a *cos* phage, and in this, it resembles the P4 plasmid of *Escherichia coli*. *Cos*-site packaging in *Staphylococcus aureus* is additionally unique in that it requires the HNH nuclease, carried only by *cos* phages, in addition to the large terminase subunit, for *cos*-site cleavage and melting.

interference | horizontal gene transfer | PICIs | molecular parasitism

Staphylococcal pathogenicity islands (SaPIs) are highly mobile and parasitize temperate helper phages to enable their reproduction and their highly efficient encapsidation in infectious phage-like particles. Previously characterized SaPIs and their helper phages use the *pac* mechanism for DNA encapsidation and encode small terminase subunit (TerS) homologs which complex with the phage-coded large terminase subunit and thus determine SaPI packaging specificity. This process is aided by a SaPI-coded protein, Ppi, which complexes with the phage TerS, blocking its function (1). Many of the SaPIs also encode a capsid morphogenesis module consisting of two proteins, CpmA and -B, which cause the formation of small procapsids that match the smaller size of the SaPI genome (2).

Among the ∼20 SaPIs originally identified were several that lacked a TerS homolog and an identifiable morphogenetic module. At the time, these were thought to be defective and were not studied further. More recently, several newly identified SaPIs of bovine mastitis origin were also found to lack these genes; nevertheless, they were efficiently packaged and transferred by helper phages, suggesting that they are not defective after all (3). The prototype of these recently identified islands is SaPIbov5 (3). In this report we decipher a strategy of molecular piracy by which the non–*ter*S-coding SaPIs can be efficiently packaged and transferred.

Results

Phage-Specific SaPI Packaging and Transfer. SaPIbov1, a prototypical SaPI (4), and SaPIbov5 encode identical SaPI Stl repressors, and therefore, it was predicted that both islands would be induced by the same helper phages (5, 6). A comparison of phage-specific mobilizations of SaPIbov1 and SaPIbov5 revealed that two standard SaPI helper phages, ϕ11 and 80α, induced and mobilized both SaPIs, whereas two additional closely related *cos* phages, ϕ12 and ϕSLT, not previously known to be SaPI helper phages, induced and mobilized SaPIbov5 but not SaPIbov1 (Fig. 1 and Table 1). SaPIbov1 induction by phages ϕ11 and 80α, which use a headful packaging mechanism, generated the characteristic SaPI-specific band, which represents the SaPI monomers packaged into the small-capsid SaPI particles generated by the SaPI-encoded packaging module (2). No SaPI-specific band was seen with SaPIbov5. This was expected because SaPIbov5 lacks the SaPI packaging module and is packaged in full-sized phage particles (see below). As previously demonstrated (7, 8), deletion of ϕ11 *ter*S eliminated the packaging of phage DNA but had no effect on SaPIbov1 transfer (Table 1). This mutation, however, eliminated SaPIbov5 transfer (Table 1), suggesting that SaPIbov5 contains a *pac* site that is identical or very similar to that of ϕ11 and explaining the ability of SaPIbov5

Significance

Staphylococcal pathogenicity islands (SaPIs) are highly mobile and carry and disseminate superantigen and other virulence genes. Here we report a remarkable example of molecular parasitism in which the SaPIs hijack the packaging machinery of the phages they victimize, using two unrelated and complementary mechanisms. Phage packaging starts with the recognition in the phage DNA of a specific sequence, termed "*pac*" or "*cos*" depending on the phage type. The SaPI strategies involve carriage of the helper phage *pac*- or *cos*-like sequences in the SaPI genome, which ensures SaPI packaging in full-sized phage particles, depending on the helper phage machinery. These strategies interfere with phage reproduction, which ultimately is a critical advantage for the bacterial population by reducing the number of phage particles.

Author contributions: J.C.A., R.P.N., A.M., and J.R.P. designed research; N.Q.-P. and N.C. performed research; N.Q.-P., N.C., J.C.A., R.P.N., A.M., and J.R.P. analyzed data; and R.P.N. and J.R.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]N.Q.-P. and N.C. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: JoseR.Penades@glasgow.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1320538111/-/DCSupplemental.

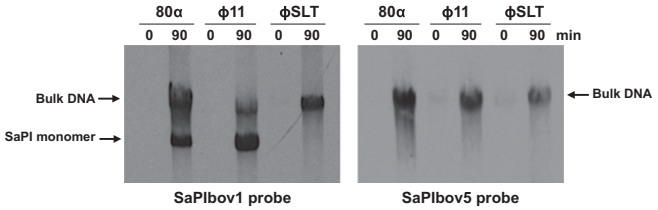


Fig. 1. Replication and encapsidation analysis of SaPIbov5. A Southern blot of the different phage lysates carrying SaPIbov1 (*Left*) or SaPIbov5 (*Right*), obtained with samples taken 0 or 90 min after MC induction, separated on agarose, and blotted with SaPIbov1- or SaPIbov5-specific probe, is shown. The upper band is "bulk" DNA, including chromosomal, phage, and replicating SaPI DNA. The lower band is SaPI linear monomers released from phage heads.

to be mobilized by a helper phage despite its apparent lack of any *ter*S gene. Induction and mobilization of SaPIbov5 by ϕ12 and ϕSLT, however, is not so easily explained. ϕ12 and ϕSLT have a putative *cos* site that is present and shown to be functional in several other related phages (9) (*SI Appendix*, Fig. S1). Examination of the SaPIbov5 sequence revealed a putative *cos* site identical to the putative phage *cos* sites. Several other SaPIs missing the specific packaging module also carry the putative *cos* site (*SI Appendix*, Fig. S1). To test these phage *cos* sites for function, we cloned the ϕ12 and ϕSLT *cos* sites to a plasmid, pCU1 (10), which was not transferrable by ϕ12, and found that the cloned *cos* sites, but not the cloned flanking sequences, enabled transfer of the plasmids by ϕ12 (*SI Appendix*, Table S1) and did not affect their frequency of transfer by ϕ11, which would be by conventional generalized transduction. This result confirmed the identity of these sequences as *cos* sites. We next deleted the *cos* site of SaPIbov5 and found that this eliminated transfer of the island by ϕ12 and ϕSLT but not by ϕ11 (Table 1). These results establish that SaPIbov5, unlike any other SaPI thus far characterized, has both *pac* and *cos* sites and therefore can be mobilized and packaged by either a *pac* or a *cos* helper phage.

Phage-Coded HNH Endonuclease Is Required for Packaging of ϕ12, ϕSLT, and SaPIbov5. Examination of the packaging modules of phages ϕ12 and ϕSLT revealed an uncharacterized ORF between the late gene regulator *rin*A (11, 12) and the terminase genes *ter*S and *ter*L (*SI Appendix*, Fig. S2). The products of these ORFs, ϕ12p28 and ϕSLTp37 (98% identity), contain a region of homology to proteins of the widespread HNH family of endonucleases, and we considered the possibility that these proteins are involved in packaging. If they were involved in packaging, they would presumably be coregulated with the terminase genes—i.e., activated by the late gene regulator RinA (11, 12). Accordingly, we constructed reporter gene fusions with and without *rin*A and tested them for reporter gene expression. As shown in Fig. 2, the HNH ORF is under RinA control. As its proximity to the terminase genes suggested that it might have a role in phage DNA packaging, we deleted it from both ϕ12 and ϕSLT prophages and tested the deletion derivatives for the production of plaque-forming particles following mitomycin C (MC) induction. As shown in Table 1, neither of these deletion derivatives was capable of producing infectious phage, nor were the ϕ12 and ϕSLT mutants able to mobilize SaPIbov5. Complementation of the *hnh* mutants re-sorted both the phage and SaPI titers (Table 1). These results suggest that ϕ12p28 and ϕSLTp37 are essential for ϕ12 and ϕSLT reproduction, respectively. Because proteins of the HNH family have endonuclease activity, usually nonspecific, but in some cases directed by a recognition motif to bind and cleave at a specific site, we considered the possibility that these proteins have nuclease activity and that this activity is required for phage encapsidation.

We first analyzed ϕSLTp37 to determine whether it has the structural features of the HNH nucleases by constructing a 3D structural model of the protein using iterative threading assembly refinement (I-TASSER) server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (13) (Fig. 3). According to this model, the protein folds with similar topology to structures of known HNH nucleases such as that of *Geobacter metallireducen*s GS-15 [Protein Data Bank (PDB) ID code 4H9D] (14) and *Pseudomonas alcaligenes Pac*I endonuclease (PDB ID code 3LDY) (15), despite the low sequence similarity of ϕSLTp37 with these proteins. In the HNH model, the conserved HNH ββα–metal-binding motif comprises residues from 38 to 90 and the catalytic HNH residues are placed in the proper position for catalysis as is observed in structural alignment with other HNH nucleases (*SI Appendix*, Fig. S3). ϕSLTp37 His58 occupies the place of the histidine general base that activates the water nucleophile attacking the sugar phosphate backbone. ϕSLTp37 His83 and His57 correspond to the metal-binding residues that chelate the catalytic divalent cation (Mg^{2+} or Mn^{2+}) and Asn74 to the residue that correctly orientates the histidine general base for catalysis. In addition, four cysteine residues, Cys41, Cys44, Cys79, and Cys82 are placed as expected for the tetrahedral coordination of a Zn^{2+} cation for folding as is observed in previously determined HNH structures. Altogether, the model predicts that ϕSLTp37 and ϕ12p28 are functional HNH nucleases. To confirm this, we expressed the ϕSLTp37 gene in *Escherichia coli* and tested it for its ability to degrade DNA in vitro. As shown in *SI Appendix*, Fig. S4

Table 1. Effect of phage mutations on phage and SaPI titers

Donor strain			Titer	
Phage	SaPI	pCN51*	Phage†	SaPI‡
ϕ11	—	—	6.6×10^9	—
ϕ11 Δ*ter*S	—	—	<10	—
ϕ11	SaPIbov1	—	2×10^7	1×10^7
ϕ11	SaPIbov5	—	2×10^9	1.3×10^7
ϕ11	SaPIbov5 Δ*cos*	—	1.6×10^9	1.1×10^7
ϕ11 Δ*ter*S	SaPIbov1	—	<10	1.9×10^7
ϕ11 Δ*ter*S	SaPIbov5	—	<10	<10
ϕ11 Δ*ter*S	SaPIbov5	pCN51-*terS*$_{ϕ11}$	1.8×10^5	2.3×10^5
80α	—	—	8×10^{10}	—
80α	SaPIbov1	—	3.4×10^9	1×10^7
80α	SaPIbov5	—	1.5×10^{11}	1.2×10^7
ϕ12	—	—	8×10^7	—
ϕ12 Δ*hnh*	—	—	<10	—
ϕ12 Δ*hnh*	—	pCN51-*hnh*$_{ϕ12}$	2.2×10^4	—
ϕ12	SaPIbov1	—	1.1×10^9	<10
ϕ12	SaPIbov5	—	1.1×10^6	1.1×10^6
ϕ12	SaPIbov5 Δ*cos*	—	1.4×10^7	<10
ϕ12 Δ*hnh*	SaPIbov5	—	<10	<10
ϕ12 Δ*hnh*	SaPIbov5	pCN51-*hnh*$_{ϕ12}$	1.4×10^4	8.1×10^4
ϕSLT	—	—	6.2×10^6	—
ϕSLT Δ*hnh*	—	—	<10	—
ϕSLT Δ*hnh*	—	pCN51-*hnh*$_{ϕSLT}$	3.1×10^4	—
ϕSLT	SaPIbov1	—	2.7×10^6	<10
ϕSLT	SaPIbov5	—	8.2×10^5	8.7×10^3
ϕSLT	SaPIbov5 Δ*cos*	—	3.3×10^6	<10
ϕSLT Δ*hnh*	SaPIbov5	—	<10	<10
ϕSLT Δ*hnh*	SaPIbov5	pCN51-*hnh*$_{ϕSLT}$	5.7×10^3	1.3×10^3

The means of results from three independent experiments are shown. Variation was within ±5% in all cases. —, the strain has not SaPIs or plasmids, and consequently, the transfer of the SaPIs can not be analyzed.
*Complemented both the donor and the recipient strains.
†PFU/mL induced culture, using RN4220 as the recipient strain.
‡Number of transductants per milliliter induced culture, using RN4220 as the recipient strain.

MICROBIOLOGY

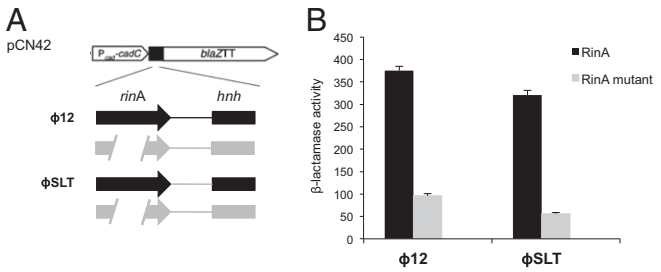


Fig. 2. RinA proteins control *hnh* expression. (*A*) Schematic representation of the different *bla*Z transcriptional fusions. (*B*) Derivatives of strain RN4220 containing each of the indicated plasmids were assayed at the midexponential phase for β-lactamase activity under standard conditions. Samples were normalized for total cell mass. Error bars show SEM.

and as reported for other phage-encoded HNH nucleases (16), ϕSLTp37 degrades DNA. Because this degradation is nonspecific, the protein evidently lacks a sequence-recognition domain. We also tested derivatives of ϕSLTp37 with alanine replacements of the two putative catalytic histidines H57 and H58 and found that these also lacked nuclease activity (*SI Appendix*, Fig. S4). Accordingly, the two proteins are henceforth designated $HNH_{\phi SLT}$ and $HNH_{\phi 12}$ and the genes $hnh_{\phi SLT}$ and $hnh_{\phi 12}$, respectively. These results suggest that it is the nuclease activity of these proteins that is required for DNA packaging, and that *cos*-site specificity must be provided by other protein(s).

We next analyzed the *cos*-site cleavage reaction by means of ϕSLT derivatives with mutations in *ter*S, *ter*L, or *hnh,* using a detoxified derivative of the phage containing a tetracycline-resistance marker (*tet*M) in the Panton–Valentine leukocidin locus. This provided a sensitive test for packaging, as a single packaged molecule could be detected in the recipient strain as a prophage. For TerL, we included alanine replacements of E208 (motor) and D363 (nuclease) on the basis of previous studies (17) and comparative TerL sequences (*SI Appendix*, Figs. S5 and S6); and for HNH we included alanine replacements of H57 and H58 (Fig. 3). For these experiments, the WT and mutant prophages were induced with MC and DNA prepared from a 90-min postinduction culture sample and analyzed for *cos*-site cleavage by means of codigestion with XhoI and SphI and Southern blotting with *rin*A- and *ter*S-specific probes. In the absence of *cos*-site cleavage, a 5.62-Kb fragment was expected; following *cos*-site cleavage, the *rin*A probe would hybridize with a 2.15-Kb fragment and the *ter*S probe with a 3.47-Kb fragment (see the scheme in Fig. 4*A*). As shown in Fig. 4*B*, the smaller fragments produced by cleavage at the *cos* site appeared with the WT phage but not with any of the *hnh, ter*S, or *ter*L mutants. Cleavage was partial with the WT, and a very faint band, corresponding to the processed species, was seen with most of the mutant phages, suggesting that very weak processing could occur in the absence of any one of the three proteins. Note that *cos*-site duplexes cannot account for the appearance of partial cleavage or for the weakness of the bands in the mutants because the samples were prepared under conditions that would ensure *cos*-site melting. This poor processing, however, was not sufficient to enable detectable packaging, which was partially re-sorted after complementation of the different phage mutants (*SI Appendix*, Table S2). Of interest is the fact that the complementation rate of the single point mutants carrying the plasmids expressing the WT proteins was reduced compared with that observed for the complementation of the deletion mutants (*SI Appendix*, Table S2). This suggests that the mutant proteins are expressed, forming nonfunctional complexes with the other phage-encoded proteins and interfering with the WT proteins to generate functional phage particles. Taken together, these results suggest that the packaging of this family of phages (and their parasite SaPIs) requires the nuclease activity of both TerL and HNH proteins. We propose here that these enzymes plus TerS form a cooperative nuclease complex to bring about *cos*-site cleavage, with TerS presumably determining the cleavage site.

The HNH–TerS–TerL Complex Is Not Sufficient to Activate *Cos*-Site Processing. To prevent premature *cos*-site processing, the λ TerS–TerL complex performs its nuclease function only in the presence of a preformed procapsid (18). We tested this with the HNH–TerS–TerL complex in vivo by mutating various proteins involved in ϕSLT capsid formation and assaying for phage packaging and *cos*-site cleavage as above (Fig. 4*A* and *SI Appendix*, Table S2). This included *ter*S, p40 (portal), p41 (protease), p42 (major capsid protein), and p47 (major tail protein). Electron microscopic analysis of the particles present in lysates prepared with the different deletion-carrying phages confirmed the predicted roles for the deleted genes. Thus, although deletions of the *hnh, ter*S, *ter*L,

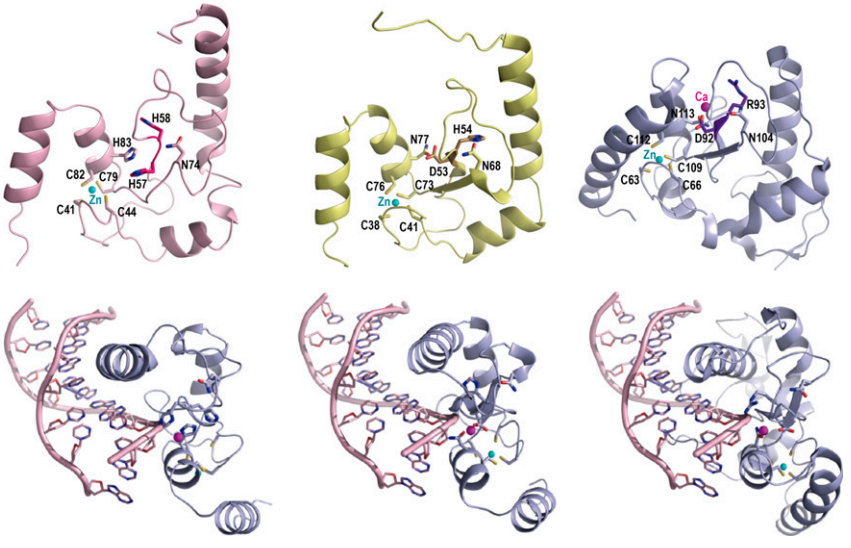


Fig. 3. ϕSLT HNH shows the characteristic HNH nuclease fold. ϕSLT HNH structural model (pink) generated with I-TASSER is compared with the experimental structures of *G. metallireducens* GS-15 (yellow; PDB ID code 4H9D) and *P. alcaligenes* PacI (blue; PDB ID code 3LDY) HNH nucleases. (*Upper*) The ribbon representations of the three structures are shown in the same orientation with catalytic and structural relevant residues represented as sticks and colored by atoms (nitrogen, oxygen, and sulfur in blue, red, and yellow, respectively) with carbon in the same color of the corresponding structure. Residues mutated to Ala in ϕSLT HNH (H57 and H58) and the equivalent residues in GS-15 and PacI are highlighted in darker hues. The structural Zn^{2+} ion is shown as a cyan sphere. (*Lower*) The DNA–HNH complex is modeled for $HNH_{\phi SLT}$ and GS-15 by superimposing these protein structures with PacI HNH in the PacI–DNA experimental complex (PDB ID code 3LDY).

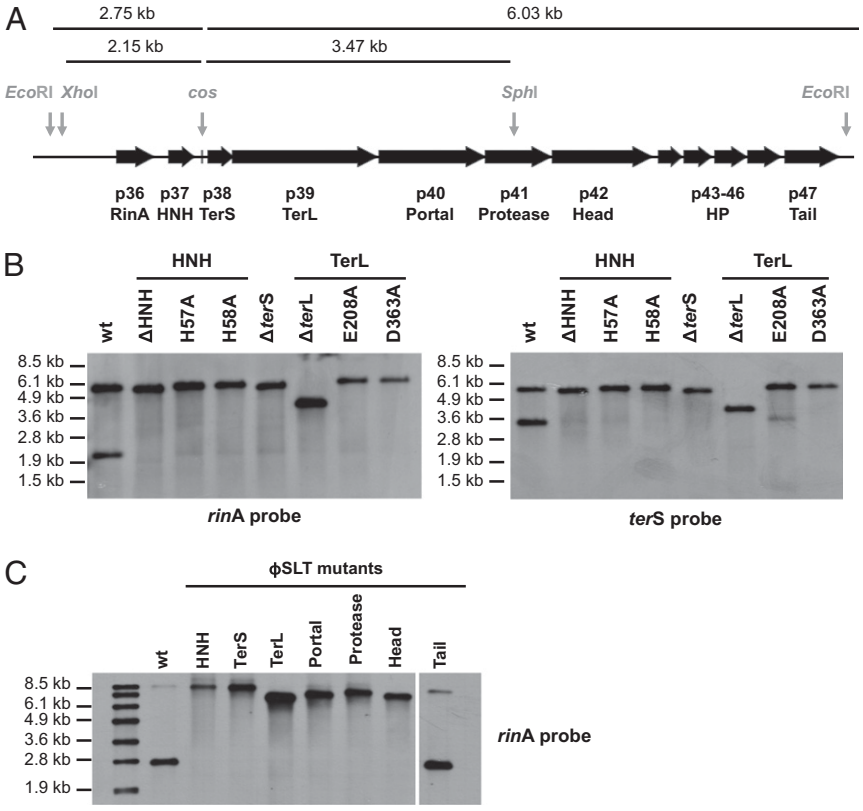


Fig. 4. Role of the different ϕSLT mutants in phage packaging. (*A*) ϕSLT map. The relevant genes and the *cos* site are shown. (*B* and *C*) Lysogenic strains carrying the different phage mutations were exposed to MC, then incubated in broth at 32 °C. Samples were removed after 90 min and used to isolate DNA, which was digested with XhoI/SphI (*B*) or with EcoRI (*C*). DNAs were separated by agarose gel electrophoresis and transferred. Southern blot hybridization patterns of these samples were hybridized overnight with a *rin*A or *ter*S (to the left and right of the *cos* site, respectively in *A*) phage-specific probes. In *C*, all samples were run on the same gel, but four lanes analysing mutants that are not included in the manuscript have been removed. Noncontiguous lanes are divided by a white line.

and portal protein genes resulted in empty capsids and tails, deletions of the protease and capsid genes resulted in only tail structures, whereas the deletion of p47 resulted in capsids only (Fig. 5). Surprisingly, some of the tails observed in the protease and major capsid protein mutants were extremely long, suggesting that the capsid structure also controls the length of the tail. Complementation of the different mutants re-sorted the production of functional phage particles (*SI Appendix*, Table S2). As shown in Fig. 4*C*, each of the deletions affecting capsid formation or DNA packaging eliminated *cos*-site cleavage, whereas deletion of the major tail protein gene did not. This result confirms the prediction that *cos*-site cleavage by the HNH–TerS–TerL complex is activated by the preformed capsid—which is required for *cos* but not for *pac* phages (19).

The Packaging System Involving HNH Proteins Is Widespread in Nature. A recent study has reported that many phages, infecting both Gram-positive and -negative bacteria, encode an HNH protein adjacent to the genes encoding the TerS, TerL, and portal proteins (20). Interestingly, all of the characterized phages carrying an *hnh* gene are *cos* phages. Our results have demonstrated that not only these proteins but also those encoding the major capsid, the portal, and the protease proteins are required for *cos*-site cleavage and therefore for DNA packaging. To demonstrate the diversity of this mechanism, we analyzed *E. coli* phage ϕP27, which is clinically relevant as it encodes the Shiga toxin (Stx) (21). We used a detoxified derivative of phage ϕP27 containing a tetracycline-resistance marker (*tet*A) inserted into the *stx* locus. We deleted from the ϕP27 prophage the genes encoding the HNH, TerS, TerL, portal, protease, and major capsid proteins (*SI Appendix*, Fig. S2). Deletion of these genes did not affect phage DNA replication but completely eliminated phage packaging and infectivity (*SI Appendix*, Table S3). These defects were partially re-sorted by complementation (*SI Appendix*, Table S3). In addition, as previously demonstrated for the staphylococcal phages, *cos*-site cleavage in vivo was observed only after induction of the WT prophage, but not with any of the deletion mutants (Fig. 6), confirming that the supramolecular complex containing HNH–TerS–TerL plus an intact capsid is required for *cos*-site cleavage.

Discussion

Characterization of several of the phage-inducible SaPIs and their helper phages has established that the *pac* (or headful) mechanism is used for encapsidation, which is exactly what would be expected for a transductional mode of transfer. In keeping with this concept, the SaPIs thus far characterized encode a homolog of TerS, which complexes with the phage-coded large terminase subunit TerL to enable packaging of the SaPI DNA in infectious particles composed of phage proteins (2, 7). These also contain a morphogenesis (*cpm*) module that causes the formation of small capsids commensurate with the small SaPI genomes (2). Among the SaPI sequences first characterized several years ago, there were several that did not include either a TerS homolog or a *cpm* homolog, and the same is true of several recently identified SaPIs from bovine sources (3) and for many phage-inducible chromosomal islands from other species. It was assumed, for these several islands, either that they were defective derivatives of elements that originally possessed these genes, or that *ter*S and *cpm* genes were present but not recognized by homology. We have performed an extensive study of one of these islands, SaPIbov5, and found that the element is fully functional—able to be induced, efficiently packaged, and transferred at high frequency by helper phages, and therefore is not defective, and by implication, nor are the others defective. Rather than encoding unrecognizable TerS and Cpm proteins, SaPIbov5 contains neither. Rather, it is encapsidated by two different mechanisms, both different from the classical TerS/Cpm mechanism used by the previously characterized SaPIs. One of these mechanisms involves induction by *pac* type helper

phages (ϕ11 and 80α), followed by packaging in full-sized phage particles, initiated by the phage terminase. This means that SaPIbov5 must have a *pac* site that is recognized by at least these two helper phages (which encode virtually identical TerS). The other mechanism involves induction and packaging by different helper phages, ϕ12 and ϕSLT, that we have shown to be *cos* phages. SaPIbov5 and phages ϕ12 and ϕSLT have similar *cos* sites and flanking sequences, which accounts for SaPIbov5 DNA encapsidation by ϕ12 and ϕSLT. Thus, SaPIbov5 has two unrelated encapsidation sequences to enable independent packaging by *pac* and *cos* phages.

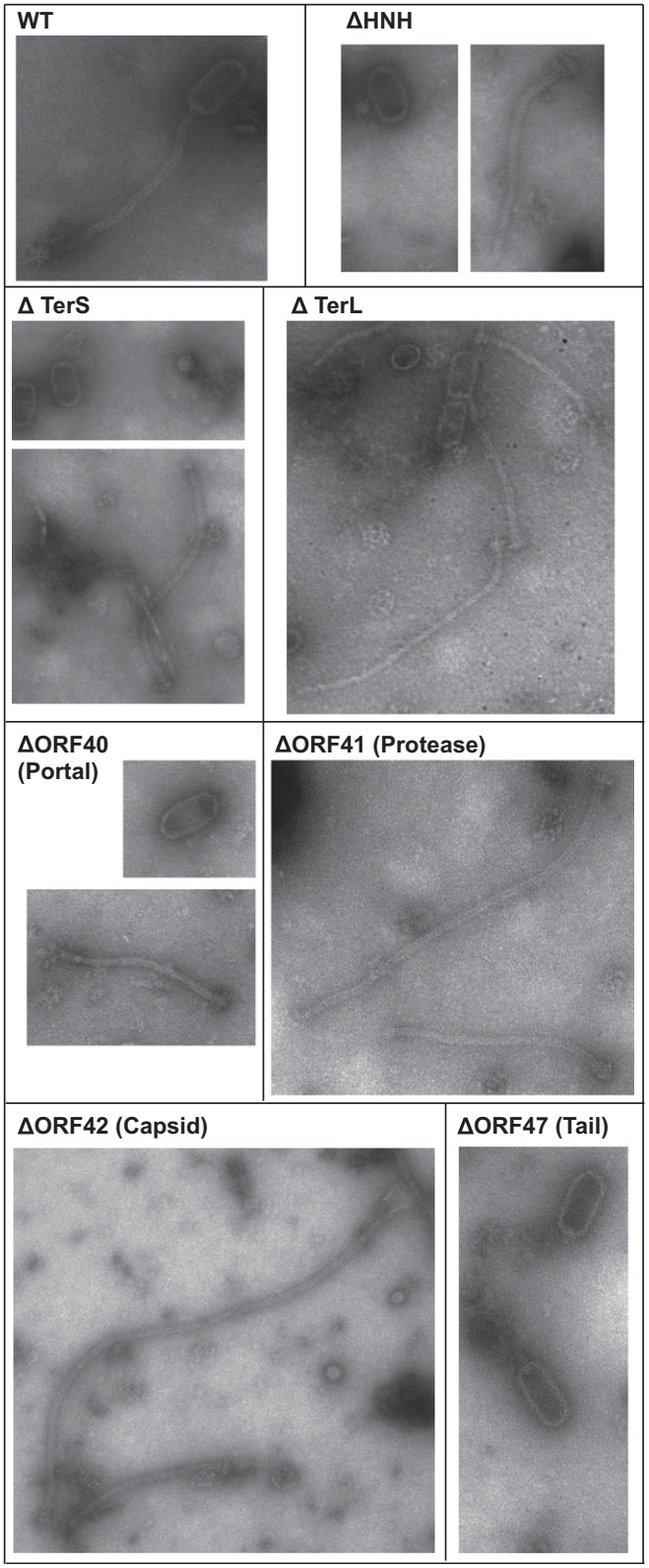


Fig. 5. Electron micrographs of ϕSLT mutant lysates.

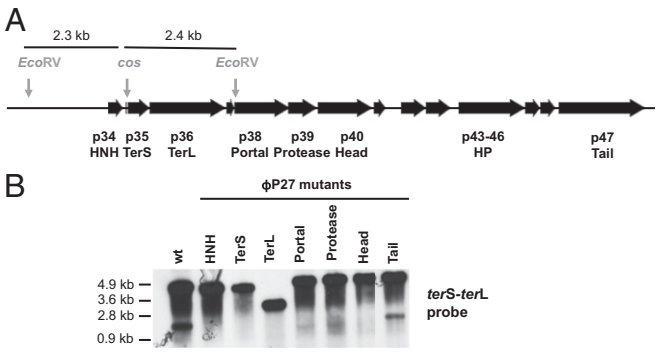


Fig. 6. Role of the different ϕP27 mutants in phage packaging. (*A*) ϕP27 map. The relevant genes and the *cos* site are shown. (*B*) Lysogenic strains carrying the different phage mutations were exposed to MC, then incubated in broth at 32 °C. Samples were removed after 90 min and used to isolate DNA, which was digested with EcoRV. DNAs were separated by agarose gel electrophoresis and transferred. Southern blot hybridization patterns of these samples were hybridized overnight with a *terS–terL* phage-specific probe.

Cos sites were also identified in several other SaPIs that lack *ter*S and *cpm,* suggesting that *cos* packaging in full-sized phage particles is a widespread alternative strategy for the SaPIs, which is reminiscent of plasmid P4 encapsidation by coliphage P2 (22). However, in contrast to the P4–P2 interaction (22), SaPIbov5 does not induce its helper phage ϕ12. Interestingly, an evolutionary tree including SaPIs that use the *cos* and *pac* packaging strategies revealed that the *cos* SaPIs clustered together (*SI Appendix*, Fig. S7). However, some *pac* SaPIs (like SaPI2 or SaPI4) were more related to the *cos* SaPIs than to the other *pac* SaPIs. There are a number of possibilities to explain this. One is that a recombinatorial exchange of packaging modules between *pac* and *cos* SaPIs generated the different subbranches observed (*SI Appendix*, Fig. S7). Another is that some *pac* SaPIs have acquired a *cos* site, presumably from a *cos* phage or a *cos*-type SaPI and have lost the characteristic operon I present in the *pac* islands.

We have observed that an important feature of ϕSLT/SaPIbov5 packaging is the requirement for an HNH nuclease, which is encoded next to the ϕSLT terminase module. Proteins carrying HNH domains are widespread in nature, being present in organisms of all kingdoms. The HNH motif is a degenerate small nucleic acid-binding and cleavage module of about 30–40 aa residues and is bound by a single divalent metal ion (23). The HNH motif has been found in a variety of enzymes playing important roles in many different cellular processes, including bacterial killing; DNA repair, replication, and recombination; and processes related to RNA (23). HNH endonucleases are present in a number of *cos*-site bacteriophages of Gram-positive and -negative bacteria (16, 20), always adjacent to the genes encoding the terminases and other morphogenetic proteins. However, their biological role in the phage life cycle had not been determined.

We have demonstrated that the HNH nucleases encoded by ϕ12 and the closely related ϕSLT have nonspecific nuclease activity and are required for the packaging of these phages and of SaPIbov5. We have shown that HNH and TerL are jointly required for *cos*-site cleavage; as a different member of the HNH family has been shown to have single-stranded nicking activity (20), it is possible that *cos*-site cleavage in these phages involves the nicking of one strand by HNH and the other strand by TerL; because

spontaneous melting of the 100% G + C *cos*-site duplex generated by staggered nicks would be inefficient, the HNH–TerS–TerL complex may also catalyze this melting, as occurs with the λ phage gpNu1-gpA complex (24). Alternatively, melting may be facilitated by a different protein, corresponding to the *E. coli* integration host factor (24).

We have also observed that only *cos* phages of Gram-negative as well as of Gram-positive bacteria encode HNH nucleases, consistent with a special requirement for *cos*-site cleavage as opposed to *pac*-site cleavage, which generates flush-ended products. The demonstration that HNH nuclease activity is required for some but not other *cos* phages suggests that there is a difference between the TerL proteins of the two types of phages—one able to cut both strands and the other needing a second protein to enable the generation of a double-stranded cut. The biological significance of these alternative packaging mechanisms remains to be determined. In addition, previous studies with phage SPP1 have demonstrated that the DNA end not used for encapsidation is degraded in vivo whereas the TerSL–DNA complex protects the DNA end that will be packaged. This protection mechanism, which is operative even in absence of subsequent genome packaging (25), is not yet understood. Because our experiments demonstrated that both ϕSLT DNA ends are resistant to host nucleases, it is tempting to speculate that the HNH–TerS–TerL complex can also be involved in this process.

It has been classically assumed that transfer of mobile genetic elements (MGEs) involves *pac* phages but not *cos* phages. Our results demonstrate that not only *pac* but also *cos* phages can have a relevant role in spreading MGEs with their virulence and resistance genes. Our study also establishes a strategy used by the SaPIs to hijack and interfere with the phage biology. HNH-mediated *cos*-site packaging thus represents a remarkable and novel mechanism of MGE transfer which may have important implications for phage-mediated horizontal gene transfer in general.

Materials and Methods

Bacterial Strains and Growth Conditions. The bacterial strains used in these studies are listed in *SI Appendix*, Table S4. The procedures for preparation and analysis of phage lysates, in addition to transduction and transformation of *Staphylococcus aureus,* were performed essentially as previously described (5, 6, 26).

DNA Methods. General DNA manipulations were performed using standard procedures. DNA samples were heated at 75 °C for 10 min before the electrophoresis to ensure *cos*-site melting. The plasmids and oligonucleotides used in this study are listed in *SI Appendix*, Tables S5 and S6, respectively. The labeling of the probes and DNA hybridization were performed according to the protocol supplied with the PCR-DIG DNA labeling and Chemiluminescent Detection Kit (Roche). To produce the phage and SaPI mutations, we used plasmid pMAD, as previously described (5).

Complementation of the Mutants. The different phage genes under study were PCR amplified using the oligonucleotides listed in *SI Appendix*, Table S6. PCR products were cloned into pCN51 (*S. aureus*) (27) or pBAD18 (*E. coli*) (28) and the resulting plasmids (*SI Appendix*, Table S5) were introduced into the appropriate recipient strains (*SI Appendix*, Table S4).

Enzyme Assays. β-Lactamase assays, using nitrocefin as substrate, were performed as described (5, 6), using a Thermomax (Molecular Devices) microtiter plate reader. Cells were obtained in exponential phase. β-Lactamase units are defined as $(V_{max})/OD_{650}$.

Electron Microscopy. Electron microscopy was performed as previsouly described (1) and performed at Centro de Investigación Príncipe Felipe (Valencia, Spain).

Structural Modeling of $HNH_{\phi STL}$ and $TerL_{\phi STL}$ Nucleases. Models of the 3D structure of $HNH_{\phi STL}$ and $TerL_{\phi STL}$ nucleases were generated using the I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (13). Models with higher a C score (−0.55 and −1.21 for $HNH_{\phi STL}$ and $TerL_{\phi STL}$, respectively) were selected for further structural analysis using Collaborative Computational Project No. 4 suite (29) and Crystallographic Object-Oriented Toolkit software (30).

ACKNOWLEDGMENTS. This work was supported by Grants Consolider-Ingenio CSD2009-00006, BIO2011-30503-C02-01, and Eranet-Pathogenomics PIM2010EPA-00606 (to J.R.P.); Ministerio de Ciencia e Innovación Grant BIO2010-15424 (to A.M.); and National Institutes of Health Grant R01AI022159 (to R.P.N. and J.R.P.).

1. Ram G, et al. (2012) Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc Natl Acad Sci USA* 109(40):16300–16305.
2. Ubeda C, et al. (2007) SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol Microbiol* 65(1):41–50.
3. Viana D, et al. (2010) Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. *Mol Microbiol* 77(6):1583–1594.
4. Ubeda C, et al. (2008) SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol Microbiol* 67(3):493–503.
5. Tormo-Más MÁ, et al. (2010) Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* 465(7299):779–782.
6. Tormo-Más MÁ, et al. (2013) Phage dUTPases control transfer of virulence genes by a proto-oncogenic G protein-like mechanism. *Mol Cell* 49(5):947–958.
7. Tormo MA, et al. (2008) *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *J Bacteriol* 190(7):2434–2440.
8. Quiles-Puchalt N, Martínez-Rubio R, Ram G, Lasa I, Penadés JR (2014) Unravelling bacteriophage ϕ11 requirements for packaging and transfer of mobile genetic elements in *Staphylococcus aureus*. *Mol Microbiol* 91(3):423–437.
9. García P, et al. (2009) Functional genomic analysis of two *Staphylococcus aureus* phages isolated from the dairy environment. *Appl Environ Microbiol* 75(24):7663–7673.
10. Augustin J, et al. (1992) Genetic analysis of epidermin biosynthetic genes and epidermin-negative mutants of *Staphylococcus epidermidis*. *Eur J Biochem* 204(3):1149–1154.
11. Ferrer MD, et al. (2011) RinA controls phage-mediated packaging and transfer of virulence genes in Gram-positive bacteria. *Nucleic Acids Res* 39(14):5866–5878.
12. Quiles-Puchalt N, et al. (2013) A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. *Nucleic Acids Res* 41(15):7260–7275.
13. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5(4):725–738.
14. Xu S-Y, et al. (2013) Structure determination and biochemical characterization of a putative HNH endonuclease from *Geobacter metallireducens* GS-15. *PLoS ONE* 8(9):e72114.
15. Shen BW, et al. (2010) Unusual target site disruption by the rare-cutting HNH restriction endonuclease *Pac*I. *Structure* 18(6):734–743.
16. Moodley S, Maxwell KL, Kanelis V (2012) The protein gp74 from the bacteriophage HK97 functions as a HNH endonuclease. *Protein Sci* 21(6):809–818.
17. Smits C, et al. (2009) Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep* 10(6):592–598.
18. Sippy J, Feiss M (2004) Initial cos cleavage of bacteriophage lambda concatemers requires proheads and gpFI in vivo. *Mol Microbiol* 52(2):501–513.
19. Chai S, et al. (1992) Molecular analysis of the *Bacillus subtilis* bacteriophage SPP1 region encompassing genes 1 to 6. The products of gene 1 and gene 2 are required for *pac* cleavage. *J Mol Biol* 224(1):87–102.
20. Xu S-Y, Gupta YK (2013) Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease. *Nucleic Acids Res* 41(1):378–390.
21. Muniesa M, Recktenwald J, Bielaszewska M, Karch H, Schmidt H (2000) Characterization of a shiga toxin 2e-converting bacteriophage from an *Escherichia coli* strain of human origin. *Infect Immun* 68(9):4850–4855.
22. Lindqvist BH, Dehò G, Calendar R (1993) Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiol Rev* 57(3):683–702.
23. Huang H, Yuan HS (2007) The conserved asparagine in the HNH motif serves an important structural role in metal finger endonucleases. *J Mol Biol* 368(3):812–821.
24. Yang Q, Catalano CE (1997) Kinetic characterization of the strand separation ("helicase") activity of the DNA packaging enzyme from bacteriophage lambda. *Biochemistry* 36(35):10638–10645.
25. Cornilleau C, et al. (2013) The nuclease domain of the SPP1 packaging motor coordinates DNA cleavage and encapsidation. *Nucleic Acids Res* 41(1):340–354.
26. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP (1998) The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol Microbiol* 29(2):527–543.
27. Charpentier E, et al. (2004) Novel cassette-based shuttle vector system for Gram-positive bacteria. *Appl Environ Microbiol* 70(10):6076–6085.
28. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* 177(14):4121–4130.
29. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.
30. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.

The ISME Journal (2015) 9, 1260–1263

www.nature.com/ismej

SHORT COMMUNICATION

Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by *cos* phages

John Chen[1], Nuria Carpena[2,3], Nuria Quiles-Puchalt[3,4], Geeta Ram[1], Richard P Novick[1] and José R Penadés[3]

[1]*Skirball Institute Program in Molecular Pathogenesis and Departments of Microbiology and Medicine, New York University Medical Center, New York, NY, USA;* [2]*Instituto de Biomedicina de Valencia (IBV-CSIC), Valencia, Spain;* [3]*Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK and* [4]*Universidad Cardenal Herrera CEU, Moncada, Valencia, Spain*

Bacteriophage-mediated horizontal gene transfer is one of the primary driving forces of bacterial evolution. The *pac*-type phages are generally thought to facilitate most of the phage-mediated gene transfer between closely related bacteria, including that of mobile genetic elements-encoded virulence genes. In this study, we report that staphylococcal *cos*-type phages transferred the *Staphylococcus aureus* pathogenicity island SaPIbov5 to non-*aureus* staphylococcal species and also to different genera. Our results describe the first intra- and intergeneric transfer of a pathogenicity island by a *cos* phage, and highlight a gene transfer mechanism that may have important implications for pathogen evolution.
The ISME Journal (2015) **9,** 1260–1263; doi:10.1038/ismej.2014.187; published online 14 October 2014

Classically, transducing phages use the *pac* site-headful system for DNA packaging. Packaging is initiated on concatemeric post-replicative DNA by terminase cleavage at the sequence-specific *pac* site, a genome slightly longer than unit length is packaged, and packaging is completed by non-sequence-specific cleavage (reviewed in Rao and Feiss, 2008). Generalized transduction results from the initiation of packaging at *pac* site homologs in host chromosomal or plasmid DNA, and typically represents ~1% of the total number of phage particles. In the alternative *cos* site mechanism packaging is also initiated on concatemeric post-replicative DNA by terminase cleavage at a sequence-specific (*cos*) site. Here, however, packaging is completed by terminase cleavage at the next *cos* site, generating a precise monomer with the cohesive termini used for subsequent circularization (Rao and Feiss, 2008). Although *cos* site homologs may exist in host DNA, it is exceedingly rare that two such sites would be appropriately spaced. Consequently, *cos* phages, of which lambda is the prototype, do not engage in generalized transduction. For this reason, *cos*-site phages have been preferred for possible phage therapy, since they would not introduce adventitious host DNA into target organisms.

The *Staphylococcus aureus* pathogenicity islands (SaPIs) are the best-characterized members of the phage-inducible chromosomal island family of mobile genetic elements (MGEs; Novick *et al.*, 2010). SaPIs are ~15 kb mobile elements that encode virulence factors and are parasitic on specific temperate (helper) phages. Helper phage proteins are required to lift their repression (Tormo-Más *et al.*, 2010, 2013), thereby initiating their excision, circularization and replication. Phage-induced lysis releases vast numbers of infectious SaPI particles, resulting in high frequencies of transfer. Most SaPI helper phages identified to date are *pac* phages, and many well-studied SaPIs are packaged by the headful mechanism (Ruzin *et al.*, 2001; Ubeda *et al.*, 2007). Recently, we have reported that some SaPIs, of which the prototype is SaPIbov5 (Viana *et al.*, 2010), carry phage *cos* sequences in their genomes, and can be efficiently packaged and transferred by *cos* phages to *S. aureus* strains at high frequencies (Quiles-Puchalt *et al.*, 2014). Here we show that this transfer extends to non-*aureus* staphylococci and to *Listeria monocytogenes*.

Since the *pac* phages transfer SaPIs to non-*aureus* staphylococci and to the Gram-positive pathogen

Correspondence: J Chen or RP Novick, Skirball Institute Program in Molecular Pathogenesis and Departments of Microbiology and Medicine, New York University Medical Center, New York, NY 10016, USA or JR Penadés, Institute of Infection, Immunity and Inflammation, University of Glasgow, 120 University Place, Sir Graeme Davies Building, Glasgow G12 8TA, UK.
E-mail: john.chen@med.nyu.edu or Richard.Novick@med.nyu.edu or JoseR.Penades@glasgow.ac.uk
Received 15 May 2014; revised 22 July 2014; accepted 4 August 2014; published online 14 October 2014

Listeria monocytogenes (Maiques *et al.*, 2007; Chen and Novick, 2009), we reasoned that *cos* phages might also be capable of intra- and intergeneric transfer. We tested this with SaPIbov5, into which we had previously inserted a tetracycline resistance (*tetM*) marker to enable selection, and with lysogens of two helper *cos* phages, ϕ12 and ϕSLT, carrying SaPIbov5 (strains JP11010 and JP11194, respectively; Supplementary Table 1). The prophages in these strains were induced with mitomycin C, and the resulting lysates were adjusted to 1 μg ml^{-1} DNase I and RNase A, filter sterilized (0.2 μm pore), and tested for SaPI transfer with tetracycline selection, as previously described (Ubeda *et al.*, 2008). To test for *trans*-specific or *trans*-generic transduction, coagulase-negative staphylococci species and *L. monocytogenes* strains were used as recipients for SaPIbov5 transfer, respectively, as previously described (Maiques *et al.*, 2007; Chen and Novick, 2009). As shown in Table 1, SaPIbov5 was transferred to *S. xylosus, S. epidermidis* and *L. monocytogenes* strains at frequencies only slightly lower than to *S. aureus*. PCR analysis demonstrated that the complete island was transferred to the recipient strains and integrated at the cognate att_B site in the host chromosome (Figure 1 and Supplementary Table 2). In contrast, deletion of the SaPIbov5 *cos* site (strains JP11229 and JP11230) did not affect SaPI replication (Supplementary Figure 1), but completely eliminated SaPIbov5 transfer (Table 1). To rule out the possibility that other mechanisms of gene transfer were involved in this process, we generated a ϕ12 phage mutant in the small terminase (*ter*S) gene (strain JP11012), using plasmid pJP1511 (Supplementary Table 2). The TerS protein is essential for ϕ12 and SaPIbov5 DNA packaging, but not for phage-mediated lysis (Quiles-Puchalt *et al.*, 2014). As expected, this mutation abolished SaPIbov5 transfer (Table 1). Taken together, these results show that intra- and intergeneric transfer of the island was *cos* phage mediated. Furthermore, SaPI proteins, such as integrase (Int) and potentially toxins, can be expressed and functional in non-*aureus* strains.

Because plaque formation is commonly used to determine phage host range, we next determined the ability of phages ϕ12 and ϕSLT to parasitize and form plaques on *S. xylosus, S. epidermidis* and *L. monocytogenes* strains. As shown in Supplementary Figure 2, phages ϕ12 and ϕSLT can parasitize and form plaques on their normal *S. aureus* hosts, but are completely unable to lyse the non-*aureus* strains. Therefore, as previously observed with *pac* phages (Chen and Novick, 2009), these results indicate that the overall host range of a *cos* phage may also be much wider if it includes infection without plaque formation.

Previous studies have demonstrated *pac* phage-mediated transfer of MGEs between *S. aureus* and other bacterial species (Maiques *et al.*, 2007; Chen and Novick, 2009; Uchiyama *et al.*, 2014); however, no previous studies have described the natural intra- or intergeneric transfer of pathogenicity islands by *cos* phages. As bacterial pathogens become increasingly antibiotic resistant, lytic and poorly transducing phages, such as *cos* phages, have been proposed for phage therapy, on the grounds that they would not introduce adventitious host DNA into target organisms and that the phages are so restricted in host range that the resulting progeny are harmless and will not result in dysbiosis of human bacterial flora. Because plaque formation was once thought to determine the host range of a phage, the evolutionary impact of phages on bacterial strains they can transduce, but are unable to parasitize, has remained an unrecognized aspect of phage biology and pathogen evolution. Our results add to the recently recognized concept of 'silent transfer' of pathogenicity factors carried by MGEs (Maiques *et al.*, 2007; Chen and Novick, 2009) by phages that cannot grow on the target organism. They extend this capability to *cos* phages, which have hitherto been unrecognized as mediators of natural genetic transfer.

Table 1 Intra- and intergeneric SaPIbov5 transfer[a]

Donor strain			
Phage	*SaPI*	*Recipient strain*	*SaPI titre*[b]
ϕ12	SaPIbov5	*S. aureus* JP4226	8.3×10^4
		S. epidermidis JP829	2.4×10^4
		S. epidermidis JP830	4.7×10^4
		L. monocytogenes SK1351	6.6×10^3
		L. monocytogenes EGDe	2.1×10^4
		S. xylosus C2a	7.1×10^4
ϕ12	SaPIbov5 Δ*cos*	*S. aureus* JP4226	<10
		S. epidermidis JP829	<10
		S. epidermidis JP830	<10
		L. monocytogenes SK1351	<10
		L. monocytogenes EGDe	<10
		S. xylosus C2a	<10
ϕ12 Δ*ter*S	SaPIbov5	*S. aureus* JP4226	<10
		S. epidermidis JP829	<10
		S. epidermidis JP830	<10
		L. monocytogenes SK1351	<10
		L. monocytogenes EGDe	<10
		S. xylosus C2a	<10
ϕSLT	SaPIbov5	*S. aureus* JP4226	4.1×10^3
		S. epidermidis JP829	1.1×10^3
		S. epidermidis JP830	2.1×10^3
		L. monocytogenes SK1351	3.6×10^2
		L. monocytogenes EGDe	3.1×10^3
		S. xylosus C2a	4.0×10^3
ϕSLT	SaPIbov5 Δ*cos*	*S. aureus* JP4226	<10
		S. epidermidis JP829	<10
		S. epidermidis JP830	<10
		L. monocytogenes SK1351	<10
		L. monocytogenes EGDe	<10
		S. xylosus C2a	<10

Abbreviation: SAPI, *Staphylococcus aureus* pathogenicity island.
[a]The means of results from three independent experiments are shown. Variation was within ±5% in all cases.
[b]No. of transductants per ml induced culture.

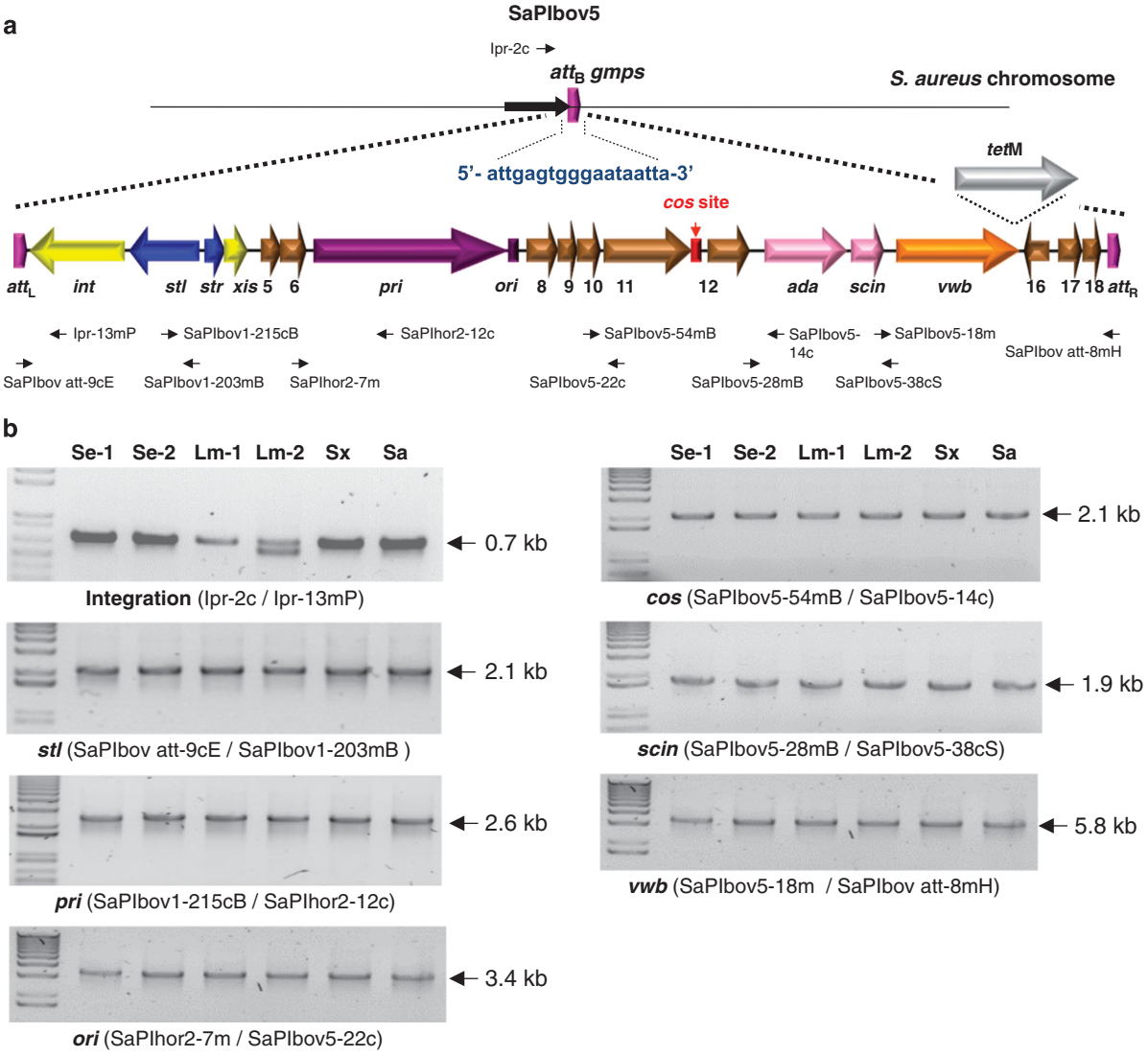


Figure 1 (**a**) Map of SaPIbov5. Arrows represent the localization and orientation of ORFs greater than 50 amino acids in length. Rectangles represent the position of the *ori* (in purple) or *cos* (in red) sites. Positions of different primers described in the text are shown. (**b**) Amplimers generated for detection of SaPIbov5 in the different recipient strains. Supplementary Table 2 lists the sequence of the different primers used. The element was detected in *S. epidermidis* JP829 (Se-1), *S. epidermidis* JP830 (Se-2), *L. monocytogenes* SK1351 (Lm-1), *L. monocytogenes* EGDe (Lm-2), *S. xylosus* C2a (Sx) and *S. aureus* JP4226 (Sa).

The potential for gene transfer of MGEs by this mechanism is limited by the ability of *cos* phages to adsorb and inject DNA into recipient strains, and also by the presence of suitable attachment sites in recipient genomes. However, since different bacterial genera express wall teichoic acid with similar structures, which can act as bacteriophage receptors governing the routes of horizontal gene transfer between major bacterial pathogens, horizontal gene transfer even across long phylogenetic distances is possible (Winstel *et al.*, 2013). In addition, our previous results also demonstrated that the SaPI integrases have much lower sequence specificity than other typical integrases, and SaPIs readily integrate into alternative sites in the absence of the cognate att_C site, such that any bacterium that can adsorb SaPI helper phage is a potential recipient (Chen and Novick, 2009). Thus, we anticipate that *cos* phages can have an important role in spreading MGEs carrying virulence and resistance genes. We also predict that *cos* sites will be found on many other MGEs, enabling *cos* phage-mediated transfer of any such element that can generate post-replicative concatemeric DNA.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank R Calendar for gifts of strains and phages. This work was supported by grants Consolider-Ingenio CSD2009-00006, BIO2011-30503-C02-01 and Eranet-pathogenomics PIM2010EPA-00606 to JRP, from the

Ministerio de Ciencia e Innovación (MICINN, Spain), and grant R01AI022159 to RPN and JRP, from the National Institutes of Health.

References

Chen J, Novick RP. (2009). Phage-mediated intergeneric transfer of toxin genes. *Science* **323**: 139–141.

Maiques E, Ubeda C, Tormo MA, Ferrer MD, Lasa I, Novick RP *et al.* (2007). Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. *J Bacteriol* **189**: 5608–5616.

Novick RP, Christie GE, Penadés JR. (2010). The phage-related chromosomal islands of Gram-positive bacteria. *Nat Rev Microbiol* **8**: 541–551.

Quiles-Puchalt N, Carpena N, Alonso JC, Novick RP, Marina A, Penadés JR. (2014). Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. *Proc Natl Acad Sci USA* **111**: 6016–6021.

Rao VB, Feiss M. (2008). The bacteriophage DNA packaging motor. *Annu Rev Genet* **42**: 647–681.

Ruzin A, Lindsay J, Novick RP. (2001). Molecular genetics of SaPI1–a mobile pathogenicity island in *Staphylococcus aureus*. *Mol Microbiol* **41**: 365–377.

Tormo-Más MÁ, Donderis J, García-Caballer M, Alt A, Mir-Sanchis I, Marina A *et al.* (2013). Phage dUTPases control transfer of virulence genes by a proto-oncogenic G protein-like mechanism. *Mol Cell* **49**: 947–958.

Tormo-Más MÁ, Mir I, Shrestha A, Tallent SM, Campoy S, Lasa I *et al.* (2010). Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465**: 779–782.

Ubeda C, Maiques E, Barry P, Matthews A, Tormo MA, Lasa I *et al.* (2008). SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol Microbiol* **67**: 493–503.

Ubeda C, Maiques E, Tormo MA, Campoy S, Lasa I, Barbé J *et al.* (2007). SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol Microbiol* **65**: 41–50.

Uchiyama J, Takemura-Uchiyama I, Sakaguchi Y, Gamoh K, Kato S-I, Daibata M *et al.* (2014). Intragenus generalized transduction in *Staphylococcus* spp. by a novel giant phage. *ISME J* **8**: 1949–1952.

Viana D, Blanco J, Tormo-Más MÁ, Selva L, Guinane CM, Baselga R *et al.* (2010). Adaptation of Staphylococcus aureus to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. *Mol Microbiol* **77**: 1583–1594.

Winstel V, Liang C, Sanchez-Carballo P, Steglich M, Munar M, Bröker BM *et al.* (2013). Wall teichoic acid structure governs horizontal gene transfer between major bacterial pathogens. *Nat Commun* **4**: 2345.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)

PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Carpena N, Manning KA, Dokland T, Marina A, Penadés JR. 2016 Convergent evolution of pathogenicity islands in helper *cos* phage interference. *Phil. Trans. R. Soc. B* **371**: 20150505. http://dx.doi.org/10.1098/rstb.2015.0505

Accepted: 28 July 2016

One contribution of 15 to a discussion meeting issue 'The new bacteriology'.

Subject Areas:
microbiology, evolution

Keywords:
capsid morphogenesis, SaPIs, PICIs, small capsids, bacteriophage resistance, bacteriophage packaging

Author for correspondence:
José R. Penadés
e-mail: joser.penades@glasgow.ac.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2015.0505 or via http://rstb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Convergent evolution of pathogenicity islands in helper *cos* phage interference

Nuria Carpena[1,2], Keith A. Manning[3], Terje Dokland[3], Alberto Marina[4] and José R. Penadés[1]

[1]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK
[2]Departamento de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad CEU Cardenal Herrera, 46113 Moncada, Valencia, Spain
[3]Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294, USA
[4]Instituto de Biomedicina de Valencia (IBV-CSIC) and CIBER de Enfermedades Raras (CIBERER), 46010 Valencia, Spain

NC, 0000-0002-1017-4499; JRP, 0000-0002-6439-5262

Staphylococcus aureus pathogenicity islands (SaPIs) are phage satellites that exploit the life cycle of their helper phages for their own benefit. Most SaPIs are packaged by their helper phages using a headful (*pac*) packaging mechanism. These SaPIs interfere with *pac* phage reproduction through a variety of strategies, including the redirection of phage capsid assembly to form small capsids, a process that depends on the expression of the SaPI-encoded *cpm*A and *cpm*B genes. Another SaPI subfamily is induced and packaged by *cos*-type phages, and although these *cos* SaPIs also block the life cycle of their inducing phages, the basis for this mechanism of interference remains to be deciphered. Here we have identified and characterized one mechanism by which the SaPIs interfere with *cos* phage reproduction. This mechanism depends on a SaPI-encoded gene, *ccm*, which encodes a protein involved in the production of small isometric capsids, compared with the prolate helper phage capsids. As the Ccm and CpmAB proteins are completely unrelated in sequence, this strategy represents a fascinating example of convergent evolution. Moreover, this result also indicates that the production of SaPI-sized particles is a widespread strategy of phage interference conserved during SaPI evolution.

This article is part of the themed issue 'The new bacteriology'.

1. Introduction

The *Staphylococcus aureus* pathogenicity islands (SaPIs) are the prototypical members of a novel family of mobile genetic elements, the phage-inducible chromosomal islands (PICIs). These elements are intimately related to certain helper phages, whose life cycles they parasitize [1], driving helper phage evolution [2]. Following infection by a helper phage or SOS induction of a helper prophage, the PICI genome excises, using the PICI-encoded integrases (*int*) and excision functions (*xis*) [3,4]. The PICI genome replicates extensively using its replicon [5,6] and is efficiently packaged into infectious particles composed of phage-encoded structural proteins [7,8]. These events, which constitute the excision–replication–packaging (ERP) cycle of the PICIs, allow both the intra- and intergeneric transfer of these elements at extremely high frequencies [9,10]. The hallmark of this parasitism is a key PICI gene that encodes a master repressor (Stl), which controls expression of most of the PICI genome. Contrary to the classical phage repressors, the Stl repressors are not cleaved following activation of the SOS response; rather the repression is lifted by the formation of a complex between the repressor and a specific helper phage protein [11,12], thereby linking PICI replication to the helper phage lytic cycle.

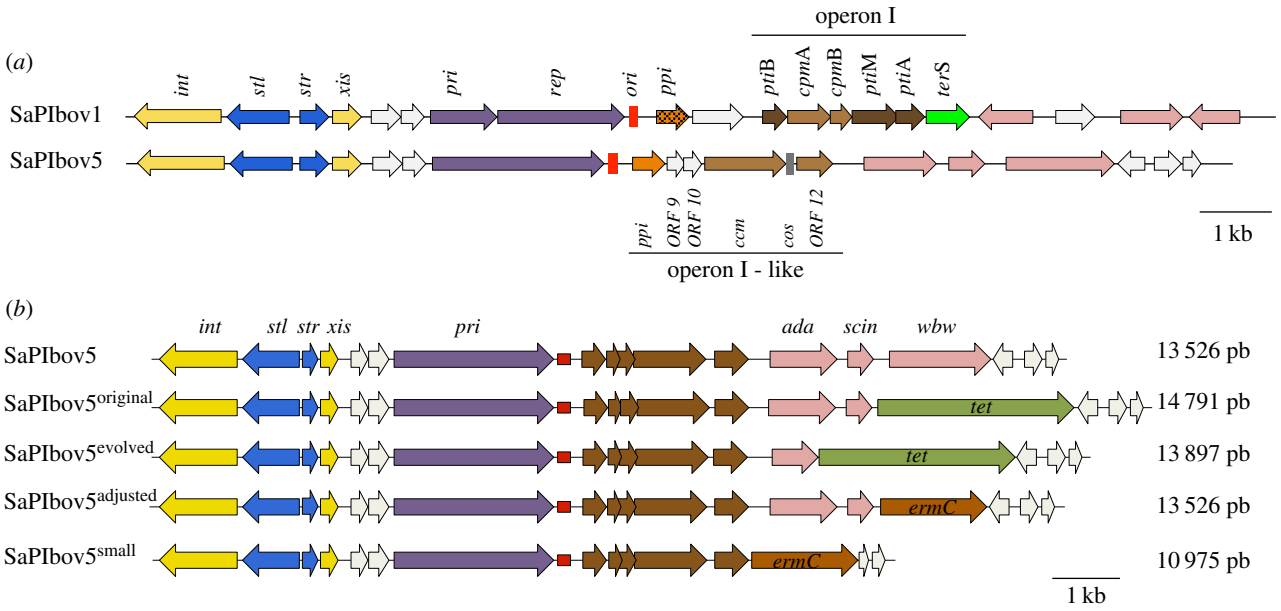


Figure 1. Genomic structure of the *cos* SaPIs. (*a*) Comparison of the *pac* (SaPIbov1) and *cos* (SaPIbov5) SaPIs. (*b*) Alignment of selected SaPIbov5 size adjustment. Genomes are aligned according to the prophage convention with the integrase gene at the left end. Gene colour code: *int* and *xis*, yellow; transcription regulators, blue; replication genes, purple; replication origin, red; genes affecting expression (*pti*) or assembly (*cpm*) of helper phage virion components are dark brown and medium brown, respectively; the terminase small subunit gene (*terS*) is green; *pip* (phage interference) orange, the two variant subsets are distinguished by dark versus light fill; superantigen and other accessory genes, pink. Genes encoding hypothetical proteins, white. In (*a*), the *cos* site is shown in grey. In (*b*), the tetracycline resistance gene is light green, and the erythromycin resistance gene is dark red.

Another key feature of all the analysed PICIs is their capacity to severely interfere with phage reproduction. To date, all described mechanisms of phage interference target key proteins of the phage DNA packaging machinery. Like their helper phages, PICIs can be packaged using two different strategies: a headful (also called *pac*) mechanism, in which DNA packaging continues until the capsid is full; or *cos* site packaging, in which units of DNA delimited by *cos* sites are packaged [13]. Most of the characterized SaPIs (and their helper phages) use the headful packaging mechanism for packaging. The *pac* SaPIs encode a small terminase subunit ($TerS_{SP}$) which interacts with the phage-coded large terminase subunit (TerL), promoting SaPI-specific DNA packaging [14,15]. Additionally, many *pac* SaPIs redirect the helper phage assembly pathway to generate SaPI capsids that are one-third of the size of the helper phage capsids [16,17], commensurate with the smaller size of the SaPI genome. The small SaPI capsids are incapable of accommodating complete helper phage genomes [17–19]. This size redirection depends on the SaPI-encoded *cpm*A and *cpm*B genes [5,20–22]. Like *ter*S_{SP}, the *cpm*AB genes are located in the SaPI packaging module, also termed operon I (figure 1), whose expression is controlled by the SOS-specific repressor LexA [14]. Apparently, the *raison d'être* of this operon is to interfere with phage reproduction. Operon I also contains the *pti*A, *pti*B and *pti*M genes (figure 1) [23]. PtiA and PtiM modulate the function of the late phage gene transcriptional regulator LtrC [23–25], while the mechanism of phage interference depending on PtiB remains unresolved [23]. The remaining known mechanism of interference depends on the *ppi* gene, located between the SaPI *ori* site and the SaPI packaging module (figure 1). The SaPI-coded Ppi protein interacts with the phage TerS, preventing phage DNA packaging [26].

We recently identified a subfamily of SaPIs in which the complete operon I, except the 3′ region of the SaPI *ter*S_{SP} gene, had been replaced by a DNA region, that we have termed 'operon I-like', containing a highly conserved phage *cos* site (electronic supplementary material, figure S1) and a set of conserved genes whose functions remain obscure (figure 1). These variants, represented by SaPIbov4 and SaPIbov5 [27], are induced by certain *cos* phages, such as ϕ12 or ϕSLT, which all share basically the same *cos* site (electronic supplementary material, figure S1), and are efficiently packaged in infectious phage-like particles, leading to high-frequency intra- and intergeneric transfer [9,28]. While these variant islands lack the classical operon I, they also severely interfere with phage reproduction [28], suggesting they encode alternative strategies of phage interference. In this report we characterize the first interference mechanism involving *cos* SaPIs and show that these SaPIs also redirect the capsid assembly of their helpers using a novel mechanism.

2. Material and methods

(a) Bacterial strains and growth conditions

The bacterial strains used in this study are listed in the electronic supplementary material, table S1. The procedures for preparation and analysis of phage lysates, in addition to transduction and transformation of *S. aureus*, were performed essentially as previously described [11,12,18].

(i) DNA methods

General DNA manipulations were performed using standard procedures. DNA samples were heated at 75°C for 10 min prior to the electrophoresis to ensure *cos* site melting. The plasmids and oligonucleotides used in this study are listed in the electronic supplementary material, tables S2 and S3, respectively. The labelling of the probes and DNA hybridization were performed according to the protocol supplied with the PCR-DIG DNA-labelling and Chemiluminescent Detection Kit (Roche).

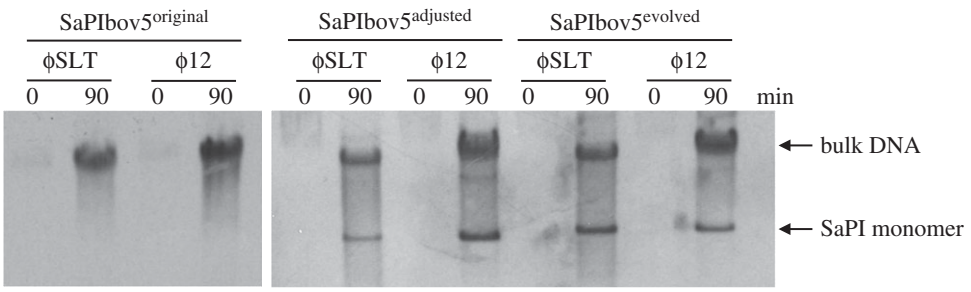


Figure 2. Replication analysis of the different SaPIbov5 derivative islands. Southern blot of ϕ12 and ϕSLT lysates, from strains carrying SaPIbov5^{original}, SaPIbov5^{adjusted} and SaPIbov5^{evolved} as indicated (see text for details). Samples were isolated 0 or 90 min after induction with mitomycin C, separated on agarose gels and blotted with a SaPIbov5-specific probe. Upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids.

To produce the phage and SaPI mutations, we used plasmid pBT2-βgal, as previously described [11].

(ii) Complementation of the mutants

The different phage genes under study were PCR amplified using oligonucleotides listed in the electronic supplementary material, table S3. PCR products were cloned into pCN51 [29] and the resulting plasmids (electronic supplementary material, table S2) were introduced into the appropriate recipient strains (electronic supplementary material, table S1).

(b) Experimental evolution

A ϕSLT lysogen carrying the SaPIbov5 *tet*M island was SOS (mitomycin C) induced and the island transferred to a ϕSLT lysogen. After the transfer, the SaPIbov5-positive strains were recollected and the procedure repeated four more times. After the fifth passage, three individual colonies were isolated, SOS induced and the SaPI titre obtained compared with that obtained with the original SaPIbov5 *tet*M.

(c) Electron microscopy

To produce ϕ12 phage and SaPIbov5 transducing particles, strains JP10435 and JP12419, respectively, were induced with 1 mg l^{-1} mitomycin C at $OD_{600} = 0.5$, and grown for an additional 3 h. As lysis was incomplete, the cell pellets were treated with lysostaphin before collecting lysate supernatants, which were further purified by PEG precipitation and CsCl centrifugation, as previously described [30]. The purified phage and transducing particles were negatively stained with 1% uranyl acetate and observed in an FEI Tecnai F20 electron microscope operated at 200 kV with magnifications of 65 500× or 81 200×. Images were captured on a Gatan Ultrascan 4000 CCD camera.

(i) *In silico* protein modelling and structure comparison

The three-dimensional homology models of ϕ12 gp33 and SaPIbov5 Ccm were constructed using the RaptorX (default mode) [31] and Phyre2 (intensive mode) [32] servers. Both servers generated models with low confidence for the N-terminal portions and high confidence for the C-terminal portions of ϕ12 gp33 and SaPIbov5 Ccm (electronic supplementary material, tables S4 and S5). The models of the C-terminal portions of gp33 and Ccm were structurally aligned with MUSTANG [33] and this alignment was rendered with ESPRIPT v. 3.0 [34].

3. Results

(a) SaPIbov5 is packaged in small capsids

In previous work, we noted that both *cos* phages ϕSLT and ϕ12 induce SaPIbov5 replication to a similar extent, although SaPIbov5 transfer by ϕ12 was approximately 10^2 times higher than that observed for phage ϕSLT [28]. As the SaPIbov5 *cos* site is more similar to that present in ϕ12 (electronic supplementary material, figure S1), we speculated that this would be the reason underlying the different SaPIbov5 packaging efficiency observed with these two phages. Indeed, when SaPIbov5 was evolved through five cycles of induction in the presence of ϕSLT, the transducing titre increased by up to 10^3-fold (electronic supplementary material, table S6), indicating that the evolved SaPIs could be efficiently packaged by phage ϕSLT. However, the SaPI *cos* site sequence remained invariable. Instead, the evolved SaPIs had reduced their size by losing some of the virulence genes contained in the island (figure 1). When we originally introduced *tet*M into SaPIbov5, we had artificially increased the size of the element. The evolved SaPI had been restored to its original size. The increased size caused the reduced transfer observed for SaPIbov5.

This restriction on genome size suggested that the *cos* SaPIs, similar to the previously described *pac* SaPIs [16,17], were packaged into capsids smaller than those normally made by the phage, as the helper phage genomes are about 3× larger (42–45 kb) than the SaPI genomes (≈14 kb, figure 1). This is consistent with the *cos* site packaging mechanism, which packages DNA units delimited by *cos* sites at either end [13].

To test this possibility, we used both the original SaPIbov5 island (SaPIbov5^{original}) and the evolved one (SaPIbov5^{evolved}), each carrying the *tet*M marker. We also generated a third SaPIbov5 that maintained its correct size but in which part of the *vwb* gene was replaced by an *erm*C marker (SaPIbov5^{adjusted}, figure 1). The *vwb* gene encodes the von Willebrand binding protein, a virulence factor with no role in the ERP cycle of the SaPIs [27]. All these islands were introduced into strains LUG1170 and JP10435, lysogenic for the *cos* phages ϕSLT and ϕ12, respectively, and the SaPIbov5 cycle was induced. Remarkably, the evolved and size-adjusted SaPIbov5 islands, but not SaPIbov5^{original}, generated the characteristic SaPI-specific band after induction of these islands by phages ϕSLT and ϕ12 (figure 2). All SaPIs, except for the original SaPIbov5^{original}, were also highly transferred by these phages (electronic supplementary material, table S6), confirming that the limitation of the SaPI genome size to less than around 14 kb was a prerequisite for high-frequency SaPI transfer.

The previous results showed that the length of DNA isolated from capsids produced in the presence of SaPIbov5 was consistent with a single unit of SaPIbov5 DNA, suggestive of formation of small capsids. To confirm that this was the case, we subjected the particles produced by ϕ12 in the absence and presence of SaPIbov5 to electron microscopy (EM). ϕ12 phage particles had the characteristic size and shape of this

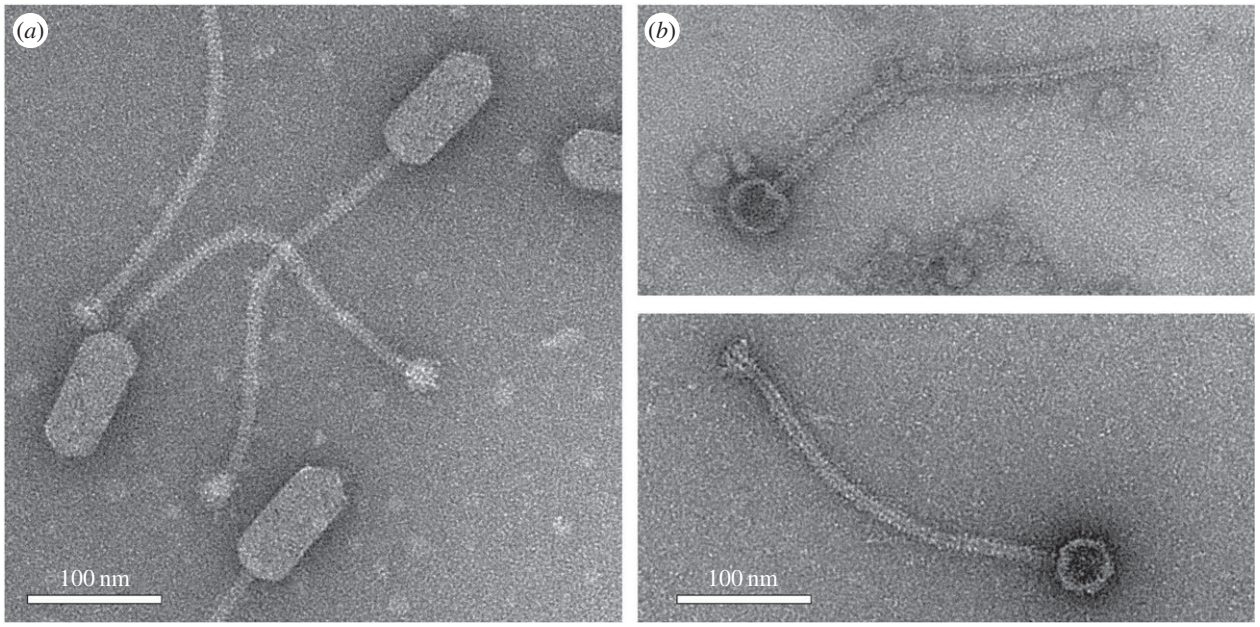


Figure 3. Electron microscopy of ϕ12 and SaPIbov5 particles. Electron micrographs of negatively stained wt ϕ12 virions (*a*), and particles produced by induction of a ϕ12 lysogen containing SaPIbov5^{adjusted} (*b*). Scale bars are 100 nm.

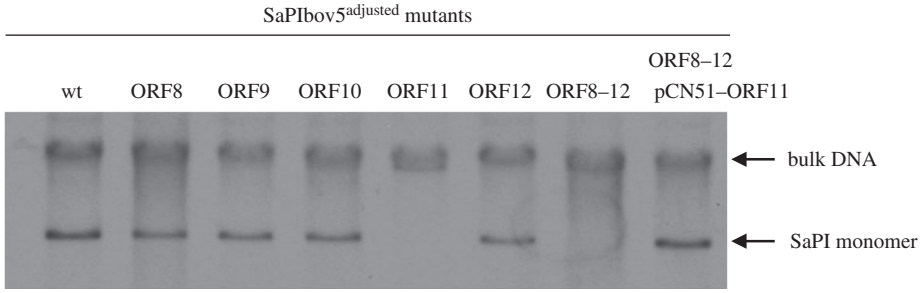


Figure 4. Replication analysis of SaPIbov5 mutants. Southern blot of ϕ12 lysates, from strains carrying the wt or the different SaPIbov5 mutants (carrying ochre mutations in the SaPIbov5 genes 8–12). Samples were isolated 90 min after induction with mitomycin C, separated on agarose and blotted with a SaPIbov5-specific probe. Upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids. SaPIbov5 ORF11 corresponds to *ccm*.

class of bacteriophages [28]: a prolate head, 45 nm wide and 100 nm long, and a 325 nm long, flexuous tail (figure 3). By contrast, virions produced in the presence of SaPIbov5 had small, isometric heads, about 42–45 nm in diameter, attached to a 325 nm tail (figure 3). This result showed that SaPIbov5 caused the formation of small capsids, consistent with its smaller genome size.

(b) Identification of the SaPIbov5-encoded capsid size redirection protein

As mobile genetics elements show synteny, and as in *pac* SaPIs the genes involved in phage interference are located between the SaPI *ori* site and the virulence genes, we speculated that the *cpm*-like gene(s) would be located in a similar position in the SaPIbov5 genome. This putative region comprises five genes (operon I-like genes: open reading frames (ORFs) 8–12; figure 1), including *ppi* (SaPIbov5 ORF8) and SaPIbov5 ORF12, which encodes a highly homologous protein (35% identity) to the SaPIbov1 coded PtiM. Both the Ppi and the PtiM have been previously involved in phage interference [23–26]. To identify the gene(s) involved in the formation of the SaPIbov5 small capsids, we generated individual mutants in all the aforementioned five genes by introducing a stop codon (ochre mutation) in the middle of their coding sequences. This strategy does not change the SaPIbov5 size. The different SaPIbov5 mutant islands were then introduced into the ϕ12 lysogen and the SaPIbov5 ERP cycle analysed after SOS induction of the different strains. As shown in figure 4, all the SaPIbov5 mutants except that for SaPIbov5 ORF11 generated the characteristic SaPI-size DNA band on an agarose gel.

We also generated a SaPIbov5 mutant carrying stop codons in all the genes from ORFs 8–12. As expected, this mutant did not generate the characteristic SaPI band when induced by phage ϕ12 (figure 4). However, complementation of this strain with a plasmid expressing ORF11 restored the production of the SaPI characteristic band, confirming the role of ORF11 in capsid size redirection. As the protein encoded by ORF11 seemed to remodel the capsid size of the helper phage, it was renamed Ccm for *cos* capsid morphogenesis.

(c) Ccm blocks ϕ12 reproduction

In previous work, we had demonstrated that SaPIbov5 interferes with ϕ12 reproduction [28]. To test whether this interference was mediated by Ccm, we used two complementary strategies: first, we introduced into the non-lysogenic

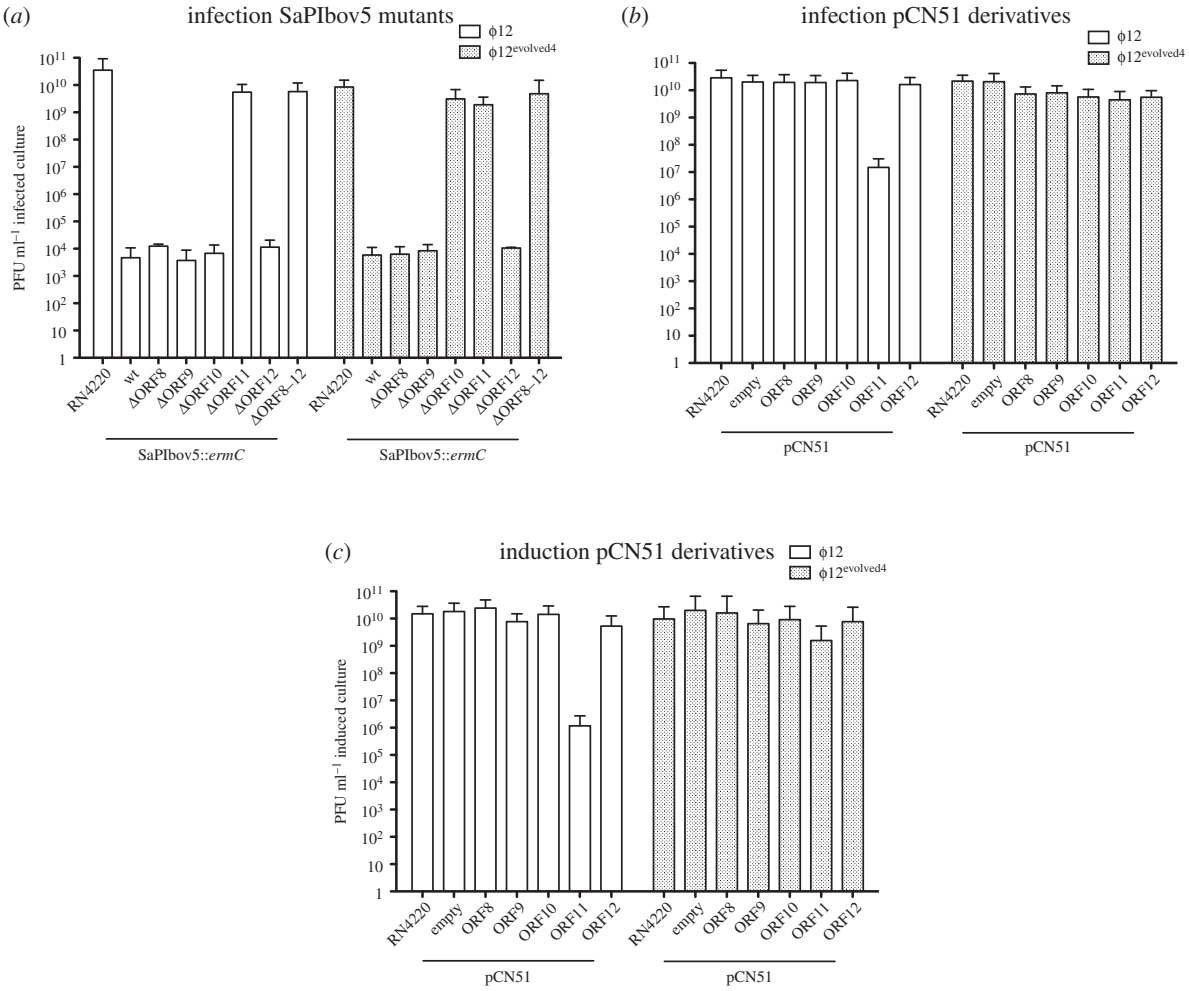


Figure 5. SaPIbov5 Ccm-mediated interference. (*a*) Strain RN4220 containing wt or the different SaPIbov5 mutants were infected with ϕ12 or ϕ12^{evolved4}, plated on phage bottom agar, and incubated for 48 h at 32°C. (*b*) Phage interference mediated by cloned SaPIbov5 genes. The indicated genes were cloned into plasmid pCN51. Strain RN4220 containing the indicated plasmids was infected with phages 12 or ϕ12^{evolved4}, plated on phage bottom agar containing 5 µM CdCl2 (induces the expression of the cloned genes) and incubated for 48 h at 32°C. (*c*) Effect of the different pCN51 cloned genes in phage reproduction. The lysogenic strains for ϕ12 or ϕ12^{evolved4}, containing the different pCN51 derivative plasmids, were SOS induced and the lysates plated on phage bottom agar for 48 h at 32°C.

RN4220 strain the SaPIbov5 mutants described above, including mutants in ORFs 8–11 (*ccm*) and 12 individually and all (ORFs 8–12) together. Then, the capacity of these strains to block plaque formation by phage ϕ12 infection was tested. As shown in figure 5*a*, all mutants except those in the *ccm* gene led to a 10^6- to 10^7-fold reduction in ϕ12 titre, showing that Ccm was primarily responsible for the SaPIbov5-mediated interference. Although the number of plaques obtained in the *ccm* mutant was basically the same as in the SaPIbov5-negative strain, the size of the plaques was reduced. This result suggested that some of the other genes may also be involved in phage interference, although this residual effect was not observed when the different genes were analysed individually (figure 5*a*).

Second, SaPIbov5 genes ORFs 8–12 were expressed from the vector pCN51 [29] under control of the exogenous cadmium-inducible promoter P*cad* in the non-lysogenic strain RN4420, followed by infection with ϕ12, or in the ϕ12 lysogen JP10435, followed by SOS induction. In either case, the resulting titres were reduced 10^3- to 10^4-fold only upon expression of *ccm* (figure 5*b*,*c*).

(d) Target for Ccm-mediated interference

To identify the ϕ12 gene(s) targeted by Ccm, we isolated ϕ12 mutants insensitive to the Ccm-mediated interference. Four of the mutants were sequenced. All had point mutations in gp33, which corresponds to the ϕ12 major capsid protein (CP) [28], although some of the mutants also had mutations in other genes (table 1). To clearly establish whether ϕ12 gp33 was the target gene of the SaPIbov5 Ccm, we generated a lysogenic RN4220 derivative carrying the phage ϕ12^{evolved4} and the SaPIbov5^{adjusted} island. SOS induction of this strain induced SaPIbov5 replication and transfer (electronic supplementary material, table S6), but not the production of SaPI-sized DNA (figure 6). Moreover, ovexpression of SaPIbov5 Ccm protein from the expression constructs described above caused only a slight reduction of ϕ12^{evolved4} titres (figure 5*b*,*c*). Taken together, these results confirm that the ϕ12 CP (gp33) was the target for SaPIbov5 Ccm.

Table 1. ϕ12 mutants insensitive to the Ccm-mediated interference.

phage	ORF33	ORF45
ϕ12^{evolved1}	G3576E; T357S	S13R
ϕ12^{evolved2}	T323P	I53S
ϕ12^{evolved3}	E236 K	E203 K
ϕ12^{evolved4}	E236 K	—

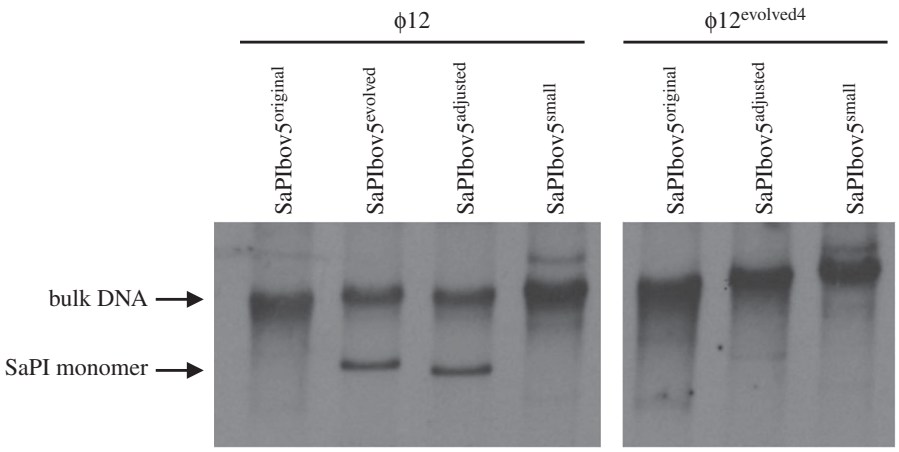


Figure 6. Replication analysis of the different sized SaPIbov5 islands induced by phages ϕ12 or ϕ12^{evolved4}. Southern blot of ϕ12 and ϕ12^{evolved4} lysates, from strains carrying SaPIbov5^{original}, SaPIbov5^{evolved}, SaPIbov5^{adjusted} or SaPIbov5^{small}, as indicated. Samples were taken 90 min after induction with mitomycin C, separated on agarose and blotted with a SaPIbov5-specific probe. Upper band is 'bulk' DNA, and represents replicating SaPIbov5. SaPI monomer represents SaPI DNA packaged in small capsids.

Finally, RN4220 derivatives carrying SaPIbov5 mutants in ORFs 8-12 were infected with ϕ12^{evolved4} and both the phage titre and the plaque sizes were analysed. Based on the results above, we expected this phage to be insensitive to SaPIbov5-mediated interference. However, SaPIbov5 severely blocked ϕ12^{evolved4} reproduction, as it did with the original ϕ12 phage (figure 5*a*), suggesting that other SaPIbov5 genes could have a role in this process, similar to the headful SaPIs described previously [26]. Indeed, the titre of ϕ12^{evolved4} was restored to normal by mutants in either ORF10 or ORF11 (*ccm*) (figure 5*a*), suggesting that ORF10 also plays a role in ϕ12 interference.

(e) SaPIbov5 Ccm and ϕ12 CP are homologues in sequence but not in function

In silico analysis of Ccm revealed that this protein has a HK97 major CP-like fold, similar to that of the ϕ12 CP (gp33). In fact, Ccm and the ϕ12 CP seem to be distantly related, based on sequence similarity (figure 7). *In silico* modelling of Ccm and gp33 with RaptorX [31] and Phyre2 [32] servers predicted with high confidence (electronic supplementary material, table S4 and S5) that the C-terminal portions of gp33 (residues 127–402) and Ccm (residues 83–355) both adopt the prototypical coat protein fold from the phage HK97 (figure 7; electronic supplementary material, figure S2) [35,36]. The modelled HK97-fold domains present a high structural similarity both between Ccm and gp33 (RMSD < 1.5 Å for 240 residues) and with HK97 CP (RMSD < 2 Å for 210 residues) despite the low sequence identity (19.2%) (figure 7). By contrast, models with different folds were predicted with low confidence (electronic supplementary material, tables S4 and S5) for the N-terminal portions of Ccm and gp33 proteins (residues 1–82 and 1–126, respectively). However, in all predictions these regions present high α-helical content (electronic supplementary material, figure S3), consistent with the so-called Δ-domain of HK97-like phages, which works as an internal scaffolding protein that assists in CP assembly and is subsequently removed by a phage-encoded protease [36,37].

This putative structural homology raised the interesting possibility that Ccm would be able to form SaPIbov5 capsids in the absence of the ϕ12 CP, suggesting an alternative mechanism to prevent phage reproduction and favouring SaPIbov5 transfer. To address this possibility, we used a previously generated deletion mutant in the gene encoding the CP of ϕSLT (gp42) [28] which is nearly identical to ϕ12 CP (gp33). Next, we introduced the SaPIbov5^{adjusted} island into this strain and measured the phage and transducing titres after SOS induction of the mutant phage. As shown in table 2, ϕSLT CP was essential both for phage and SaPI transfer, showing that Ccm is unable to take the place of the ϕSLT CP.

(f) Ccm blocks *cos* but not *pac* phages

Although conceptually they perform similar functions, *S. aureus cos* and *pac* phages use different proteins for capsid formation and DNA packaging. Thus, we wanted to test whether the reproduction cycle of the *pac* phages was also blocked by the Ccm protein. This was not the case, and expression of the Ccm from plasmid pJP1730 did not block either ϕ11 or 80α reproduction (electronic supplementary material, table S7).

(g) *Cos* SaPIs reserve space for virulence-gene carriage

SaPIbov2, one of the prototypical *pac* SaPIs [3], is approximately 27 kb in size and cannot redirect the production of small-sized capsids because it does not encode *cpm*B. Consequently, SaPIbov2 is exclusively packaged in large capsids [38]. To know if a similar scenario exists in the *cos* SaPIs, we searched in GenBank for *cos* SaPIs with an increased size and lacking the *ccm* gene. All *cos* SaPIs that were identified encoded Ccm, but one, SaPIS0385, had a reduced size (10.3 kb) compared with the others (electronic supplementary material, figure S4). This island encoded all the genes required for the SaPI cycle, but lacked the classical SaPI-encoded virulence genes. To determine whether a *cos* SaPI with reduced size had a functional ERP cycle, we generated a SaPIbov5 derivative in which the von Willebrand binding protein (*vwb*) and the staphylococcal complement inhibitor (*scn*) genes were deleted (SaPIbov5^{small}) (figure 1). The resulting size was 10.9 kb, similar to SaPIS0385 (electronic supplementary material, figure S4). The ϕ12 mediated transfer of the SaPIbov5^{small} element was only slightly reduced (less than twofold) compared with that observed with

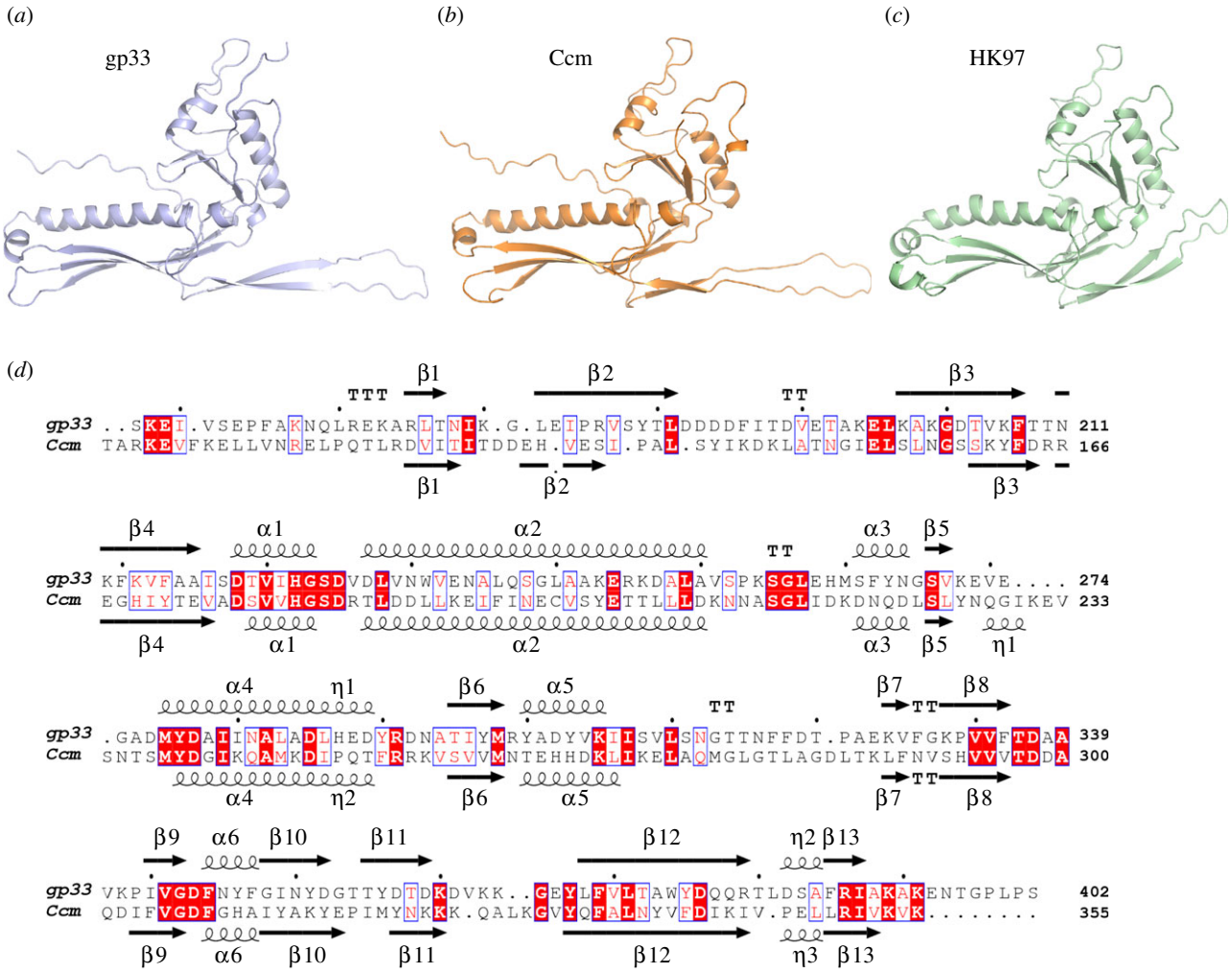


Figure 7. C-terminal portion of gp33 and Ccm proteins are predicted to adopt the characteristic HK97-fold of phage coat proteins. Cartoon representation of the C-terminal portion of (*a*) ϕ12 gp33 (residues 127–402) and (*b*) SaPIbov5 Ccm (residues 83–355), generated by RaptorX [31]. Both proteins show similar folding to the prototypical coat protein from phage HK97 (*c*; PDB 1OHG). (*d*) Structural alignment of ϕ12 gp33 (*a*) and SaPIbov5 Ccm (*b*) models carried out with MUSTANG [33]. Identical residues are highlighted on a red background and conserved residues are in a blue box with red text. The elements of secondary structure for each model are shown above (gp33) or below (Ccm) the corresponding sequence.

Table 2. Effect of phage mutations on phage and SaPI titres. (The means of results from three independent experiments are shown. Variation was within ±5% in all cases.)

donor strain			
phage	SaPI	phage titre[a]	SaPI titre[b]
ϕ SLT*pvl*::*tetM*	—	5.0×10^6	
ϕ SLT*pvl*::*tetM* ΔORF42	—	<10	—
ϕ SLT*pvl*::*tetM*	SaPIbov5^{adjusted}	1.74×10^6	1.72×10^6
ϕ SLT*pvl*::*tetM* ΔORF42	SaPIbov5^{adjusted}	<10	<10

[a]PFU ml^{-1} induced culture, using RN4220 as recipient strain.
[b]Number of transductants ml^{-1} induced culture, using RN4220 as recipient strain.

the wt SaPIbov5. Surprisingly, although the small island expresses the Ccm protein and interferes with ϕ12 reproduction (electronic supplementary material, table S4), it does not produce the characteristic SaPI band (figure 6). Apparently, SaPIbov5^{small} concatemers are packaged more efficiently into the large capsids. This result suggests that during evolution the *cos* SaPIs have reserved approximately 2 kb of DNA space for the carriage of virulence genes.

4. Discussion

In this study, we have described packaging of a family of *cos* SaPIs by *cos* helper phages ϕ12 and ϕSLT, and show that these SaPIs interfere with phage production by forming small capsids that are unable to package complete helper phage genomes. This size redirection process is reminiscent of that found in the previously described *pac* SaPIs, where size redirection depends on the two proteins CpmA and CpmB [14], and CpmB acts as an alternative internal scaffolding protein for the small SaPI capsids [20,39]. Here, we have found that size redirection by SaPIbov5 is dependent on the *ccm* gene, which encodes a HK97-like CP homologue.

How Ccm drives the production of small capsids remains unresolved. The HK97-like CP fold predicted for Ccm raises the interesting possibility that this protein could participate in the capsid assembly or even be part of the capsid shell. Even though our experiments have shown that Ccm is unable to form SaPI-sized capsids by itself, the Ccm fold,

highly similar to gp33, might enable both proteins to be assembled together. It has been suggested that the length of the N-terminal Δ-domain correlates with capsid size [40,41]. Our models indicate that the Δ-domain of Ccm is 44 residues shorter than that in gp33. Therefore, the inclusion of Ccm during formation of procapsids could conceivably drive the formation of smaller capsids. This proposed mechanism of action also explains why Ccm does not block ϕ11 and 80α *pac* phages, whose capsid proteins lack a Δ-domain and require a separately expressed scaffolding protein for capsid assembly [21]. The SaPIs mobilized by these phages use an alternative scaffolding protein, CpmB, to induce small capsid formation [21,42]. Thus, both Ccm and CpmB proteins drive small capsid formation by mimicking the scaffolding function in the assembly process, representing another example of the SaPIs' capacity for adaptation to their helper phages.

The role of SaPIbov5 ORF10 in this process is unclear. Although deletion of ORF10 had no effect on the SaPIbov5-induced suppression of wild-type ϕ12, it restored reproduction of $\phi12^{evolved4}$ (figure 5). However, overexpression of ORF10 alone had no effect on either wild-type or evolved ϕ12. Perhaps, ORF10 and Ccm somehow work together to effect the SaPIbov5-mediated interference, a line of reasoning that we will explore in future research.

The production of small capsids is not just a key feature of SaPI biology, but a widespread mechanism of phage interference. The *Enterococcus faecalis* EfCIV583 element also remodels capsid formation, promoting the formation of small capsids [43]. A similar strategy is used by the *Escherichia coli* P4 plasmid, which remodels helper phage P2 capsid formation by the expression of the P4-encoded external scaffolding protein Sid [44]. The proteins involved in these mechanisms share no homology, suggesting that this is a convergent evolutionary strategy that provides a significant advantage in nature.

All the *cos* SaPIs that we have identified encode proteins basically identical to the SaPIbov5 ORFs 8–12 (electronic supplementary material, figure S4), suggesting that all these proteins are involved in the same biological process, being required together to develop their function in the SaPIbov5 cycle. Many of the *pac* SaPIs also encode a variant of the *ppi* gene that is found in *cos* SaPIs, where they act to suppress helper phage DNA packaging. However, the SaPIbov5 *ppi* gene does seem to be involved in *cos* phage ϕ12 interference, which is not surprising, as the terminase enzymes of the *pac* and *cos* site phages are completely different. The function of *ppi* in the SaPIbov5 ERP cycle thus remains unsolved.

SaPIs are widespread elements in nature. Most *S. aureus* strains carry more than one of these elements. SaPIs carry important virulence factors that are unique to these elements and affect the fitness of their bacterial hosts [1,45]. Interestingly, there appears to be little difference in the virulence genes that are carried by the *pac* and *cos* SaPIs. Thus, the genes encoding the TSST-1, Sel or Sec toxins, the staphylococcal complement inhibitor Scin or the von Willebrand factor-binding protein are found in both types of SaPIs. Because of the limited number of chromosomal (*att*C) sites where the SaPIs can integrate and the high number of circulating SaPIs, there is strong competition for the SaPIs to persist. These elements have evolved to carry all the genes required for their own replication and helper phage exploitation, while 'reserving' a space for the carriage of virulence genes, which at the end will be essential to compete with other SaPIs. However, the number of genes that can be carried in an SaPI is limited, as an increase beyond the size that can be carried within a small capsid would be absolutely detrimental. Thus, the SaPI-encoded virulence genes should be key for the adaption of *S. aureus* to specific niches or hosts. In support of this, two SaPI-coded genes, *bap* and *vwb*, carried in ruminant *S. aureus* strains, play an important role in the pathogenesis of *S. aureus* in these animal hosts [3,27]. Thus, the identification and blockage of the activity of the SaPI-coded virulence genes can provide novel strategies to combat *S. aureus* infections in a more efficient way. The SaPIs have evolved to exploit and interfere with phage reproduction in a multitude of ways. Other PICIs are likely to use similar strategies. We anticipate that there are many additional mechanisms of interference in SaPIs and other PICIs that remain to be uncovered, and that will be of considerable importance to the evolution of virulence in *S. aureus*.

Authors' contributions. N.C. and J.R.P designed research; N.C. and K.A.M. performed research; T.D. and A.M. contributed new reagents/analytic tools; N.C., T.D., A.M. and J.R.P. analysed data; and T.D., A.M and J.R.P wrote the paper.

Competing interests. The authors declare no competing interests.

Funding. This work was supported by grants MR/M003876/1 from the Medical Research Council (UK), BB/N002873/1 from the Biotechnology and Biological Sciences Research Council (BBSRC, UK) and ERC-ADG-2014 proposal no. 670932 Dut-signal (from EU) to J.R.P., and by The National Institutes of Health (USA) grant no. R01 AI083255-06 to T.D., and by grants BIO2013-42619-P from MINECO (Spain) and Prometeo II/2014/029 from the Valencian Government to A.M.

Acknowledgement. We appreciate the assistance of Cynthia Rodenburg (UAB) with some of the EM experiments.

References

1. Penadés JR, Christie GE. 2015 The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. *Annu. Rev. Virol.* **2**, 181–201. (doi:10.1146/annurev-virology-031413-085446)
2. Frígols B, Quiles-Puchalt N, Mir-Sanchis I, Donderis J, Elena SF, Buckling A, Novick RP, Marina A, Penadés JR. 2015 Virus satellites drive viral evolution and ecology. *PLoS Genet.* **11**, e1005609. (doi:10.1371/journal.pgen.1005609)
3. Ubeda C, Tormo MA, Cucarella C, Trotonda P, Foster TJ, Lasa I, Penadés JR. 2003 Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. *Mol. Microbiol.* **49**, 193–210. (doi:10.1046/j.1365-2958.2003.03577.x)
4. Mir-Sanchis I, Martínez-Rubio R, Martí M, Chen J, Lasa I, Novick RP, Tormo-Más MÁ, Penadés JR. 2012 Control of *Staphylococcus aureus* pathogenicity island excision. *Mol. Microbiol.* **85**, 833–845. (doi:10.1111/j.1365-2958.2012.08145.x)
5. Ubeda C, Barry P, Penadés JR, Novick RP. 2007 A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proc. Natl Acad. Sci. USA* **104**, 14 182–14 188. (doi:10.1073/pnas.0705994104)
6. Ubeda C, Tormo-Más MÁ, Penadés JR, Novick RP. 2012 Structure-function analysis of the SaPIbov1 replication origin in *Staphylococcus aureus*. *Plasmid* **67**, 183–190. (doi:10.1016/j.plasmid.2012.01.006)
7. Tormo MA, Ferrer MD, Maiques E, Ubeda C, Selva L, Lasa I, Calvete JJ, Novick RP, Penadés JR. 2008 *Staphylococcus aureus* pathogenicity island DNA is packaged in particles composed of phage proteins. *J. Bacteriol.* **190**, 2434–2440. (doi:10.1128/JB.01349-07)

8. Tallent SM, Christie GE. 2007 Transducing particles of *Staphylococcus aureus* pathogenicity island SaPI1 are comprised of helper phage-encoded proteins. *J. Bacteriol.* **189**, 7520–7524. (doi:10.1128/JB.00738-07)
9. Chen J, Carpena N, Quiles-Puchalt N, Ram G, Novick RP, Penadés JR. 2015 Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by *cos* phages. *ISME J.* **9**, 1260–1263. (doi:10.1038/ismej.2014.187)
10. Chen J, Novick RP. 2009 Phage-mediated intergeneric transfer of toxin genes. *Science* **323**, 139–141. (doi:10.1126/science.1164783)
11. Tormo-Más MÁ *et al.* 2010 Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465**, 779–782. (doi:10.1038/nature09065)
12. Tormo-Más MÁ, Donderis J, García-Caballer M, Alt A, Mir-Sanchis I, Marina A, Penadés JR. 2013 Phage dUTPases control transfer of virulence genes by a proto-oncogenic G protein-like mechanism. *Mol. Cell* **49**, 947–958. (doi:10.1016/j.molcel.2012.12.013)
13. Feiss M, Rao VB. 2012 The bacteriophage DNA packaging machine. *Adv. Exp. Med. Biol.* **726**, 489–509. (doi:10.1007/978-1-4614-0980-9_22)
14. Ubeda C, Maiques E, Tormo MA, Campoy S, Lasa I, Barbé J, Novick RP, Penadés JR. 2007 SaPI operon I is required for SaPI packaging and is controlled by LexA. *Mol. Microbiol.* **65**, 41–50. (doi:10.1111/j.1365-2958.2007.05758.x)
15. Ubeda C, Maiques E, Barry P, Matthews A, Tormo MA, Lasa I, Novick RP, Penadés JR. 2008 SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Mol. Microbiol.* **67**, 493–503. (doi:10.1111/j.1365-2958.2007.06027.x)
16. Ruzin A, Lindsay J, Novick RP. 2001 Molecular genetics of SaPI1: a mobile pathogenicity island in *Staphylococcus aureus*. *Mol. Microbiol.* **41**, 365–377. (doi:10.1046/j.1365-2958.2001.02488.x)
17. Ubeda C, Maiques E, Knecht E, Lasa I, Novick RP, Penadés JR. 2005 Antibiotic-induced SOS response promotes horizontal dissemination of pathogenicity island-encoded virulence factors in staphylococci. *Mol. Microbiol.* **56**, 836–844. (doi:10.1111/j.1365-2958.2005.04584.x)
18. Lindsay JA, Ruzin A, Ross HF, Kurepina N, Novick RP. 1998 The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* **29**, 527–543. (doi:10.1046/j.1365-2958.1998.00947.x)
19. Ubeda C, Olivarez NP, Barry P, Wang H, Kong X, Matthews A, Tallent SM, Christie GE, Novick RP. 2009 Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Mol. Microbiol.* **72**, 98–108. (doi:10.1111/j.1365-2958.2009.06634.x)
20. Damle PK, Wall EA, Spilman MS, Dearborn AD, Ram G, Novick RP, Dokland T, Christie GE. 2012 The roles of SaPI1 proteins gp7 (CpmA) and gp6 (CpmB) in capsid size determination and helper phage interference. *Virology* **432**, 277–282. (doi:10.1016/j.virol.2012.05.026)
21. Spilman MS, Damle PK, Dearborn AD, Rodenburg CM, Chang JR, Wall EA, Christie GE, Dokland T. 2012 Assembly of bacteriophage 80α capsids in a *Staphylococcus aureus* expression system. *Virology* **434**, 242–250. (doi:10.1016/j.virol.2012.08.031)
22. Poliakov A, Chang JR, Spilman MS, Damle PK, Christie GE, Mobley JA, Dokland T. 2008 Capsid size determination by *Staphylococcus aureus* pathogenicity island SaPI1 involves specific incorporation of SaPI1 proteins into procapsids. *J. Mol. Biol.* **380**, 465–475. (doi:10.1016/j.jmb.2008.04.065)
23. Ram G, Chen J, Ross HF, Novick RP. 2014 Precisely modulated pathogenicity island interference with late phage gene transcription. *Proc. Natl Acad. Sci. USA* **111**, 14 536–14 541. (doi:10.1073/pnas.1406749111)
24. Quiles-Puchalt N, Tormo-Más MÁ, Campoy S, Toledo-Arana A, Monedero V, Lasa I, Novick RP, Christie GE, Penadés JR. 2013 A super-family of transcriptional activators regulates bacteriophage packaging and lysis in Gram-positive bacteria. *Nucleic Acids Res.* **41**, 7260–7275. (doi:10.1093/nar/gkt508)
25. Ferrer MD *et al.* 2011 RinA controls phage-mediated packaging and transfer of virulence genes in Gram-positive bacteria. *Nucleic Acids Res.* **39**, 5866–5878. (doi:10.1093/nar/gkr158)
26. Ram G *et al.* 2012 Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl Acad. Sci. USA* **109**, 16 300–16 305. (doi:10.1073/pnas.1204615109)
27. Viana D *et al.* 2010 Adaptation of *Staphylococcus aureus* to ruminant and equine hosts involves SaPI-carried variants of von Willebrand factor-binding protein. *Mol. Microbiol.* **77**, 1583–1594. (doi:10.1111/j.1365-2958.2010.07312.x)
28. Quiles-Puchalt N, Carpena N, Alonso JC, Novick RP, Marina A, Penadés JR. 2014 Staphylococcal pathogenicity island DNA packaging system involving *cos*-site packaging and phage-encoded HNH endonucleases. *Proc. Natl Acad. Sci. USA* **111**, 6016–6021. (doi:10.1073/pnas.1320538111)
29. Charpentier E, Anton AI, Barry P, Alfonso B, Fang Y, Novick RP. 2004 Novel cassette-based shuttle vector system for Gram-positive bacteria. *Appl. Environ. Microbiol.* **70**, 6076–6085. (doi:10.1128/AEM.70.10.6076-6085.2004)
30. Spilman MS, Dearborn AD, Chang JR, Damle PK, Christie GE, Dokland T. 2011 A conformational switch involved in maturation of *Staphylococcus aureus* bacteriophage 80α capsids. *J. Mol. Biol.* **405**, 863–876. (doi:10.1016/j.jmb.2010.11.047)
31. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. 2012 Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522. (doi:10.1038/nprot.2012.085)
32. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015 The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858. (doi:10.1038/nprot.2015.053)
33. Konagurthu AS, Reboul CF, Schmidberger JW, Irving JA, Lesk AM, Stuckey PJ, Whisstock JC, Buckle AM. 2010 MUSTANG-MR structural sieving server: applications in protein structural analysis and crystallography. *PLoS ONE* **5**, e10048. (doi:10.1371/journal.pone.0010048)
34. Robert X, Gouet P. 2014 Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324. (doi:10.1093/nar/gku316)
35. Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. 2003 The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 A resolution. *J. Mol. Biol.* **334**, 885–899. (doi:10.1016/j.jmb.2003.09.035)
36. Suhanovsky MM, Teschke CM. 2015 Nature's favorite building block: deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* **479–480**, 487–497. (doi:10.1016/j.virol.2015.02.055)
37. Oh B, Moyer CL, Hendrix RW, Duda RL. 2014 The delta domain of the HK97 major capsid protein is essential for assembly. *Virology* **456–457**, 171–178. (doi:10.1016/j.virol.2014.03.022)
38. Maiques E, Ubeda C, Tormo MA, Ferrer MD, Lasa I, Novick RP, Penadés JR. 2007 Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. *J. Bacteriol.* **189**, 5608–5616. (doi:10.1128/JB.00619-07)
39. Dearborn AD, Laurinmaki P, Chandramouli P, Rodenburg CM, Wang S, Butcher SJ, Dokland T. 2012 Structure and size determination of bacteriophage P2 and P4 procapsids: function of size responsiveness mutations. *J. Struct. Biol.* **178**, 215–224. (doi:10.1016/j.jsb.2012.04.002)
40. Chang JR, Spilman MS, Dokland T. 2010 Assembly of bacteriophage P2 capsids from capsid protein fused to internal scaffolding protein. *Virus Genes* **40**, 298–306. (doi:10.1007/s11262-009-0442-2)
41. Huet A, Duda RL, Hendrix RW, Boulanger P, Conway JF. 2016 Correct assembly of the bacteriophage T5 procapsid requires both the maturation protease and the portal complex. *J. Mol. Biol.* **428**, 165–181. (doi:10.1016/j.jmb.2015.11.019)
42. Dearborn AD, Spilman MS, Damle PK, Chang JR, Monroe EB, Saad JS, Christie GE, Dokland T. 2011 The *Staphylococcus aureus* pathogenicity island 1 protein gp6 functions as an internal scaffold during capsid size determination. *J. Mol. Biol.* **412**, 710–722. (doi:10.1016/j.jmb.2011.07.036)
43. Matos RC *et al.* 2013 *Enterococcus faecalis* prophage dynamics and contributions to pathogenic traits. *PLoS Genet.* **9**, e1003539. (doi:10.1371/journal.pgen.1003539)
44. Wang S, Chang JR, Dokland T. 2006 Assembly of bacteriophage P2 and P4 procapsids with internal scaffolding protein. *Virology* **348**, 133–140. (doi:10.1016/j.virol.2005.12.021)
45. Penadés JR, Chen J, Quiles-Puchalt N, Carpena N, Novick RP. 2015 Bacteriophage-mediated spread of bacterial virulence genes. *Curr. Opin. Microbiol.* **23**, 171–178. (doi:10.1016/j.mib.2014.11.019)