Rapid Screening for Antimicrobial Genes in Novel Nocardiophages

Youtaro Shibayama

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, for the degree of Master of Science.
September 2006

DECLARATION

I declare that this thesis is my own unaided work. It is being submitted for the degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other university.

Youtaro Shibayama

Date

ABSTRACT

There has been an increase in number of human infections by mycobacteria and opportunistic pathogens of the closely related nocardioform bacteria. Frequent multiple drug resistance in these organisms makes it desirable to identify novel targets for antimicrobial agents. Bacteriophages offer one way to do this as analysis of their DNA reveals great diversity in their genetic makeup, suggesting variety in the way they interfere with host cells. Four novel nocardiophages were therefore isolated from soil and characterized. Libraries of their nucleic acid were constructed and screened for clones inhibitory to a nocardioform of the genus *Rhodococcus*. Nine clones were characterized, and minimum necessary DNA for inhibitory activity sequenced. Of 18 ORFs predicted on these DNAs, 13 could not be assigned a function. Genes similar to ones in databases apparently interfered with DNA metabolism, protein synthesis, or integrity of plasma membrane. This genetic approach may be an efficient and effective way to discover novel targets for antibiotics.

ACKNOWLEDGEMENTS

I would like to express my appreciation to all who supported me throughout this work.

To Professor Eric Dabbs, for his guidance and the opportunity to learn under his supervision.

To my parents, for supporting me throughout this work so that I could spend more hours in the lab.

The results presented here have been used as the basis for a manuscript submitted for publication.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS & ACRONYMS

A	adenine
ADP	adenosine diphosphate
AIDS	Acquired Immune Deficiency Syndrome
Ala	alanine
AMP	antimicrobial peptide
Amp	ampicillin
ATCC	American Type Culture Collection
bp	base pairs
C	Celsius or cytosine
cfu	colony forming units
CHEF	clamped homogenous electric fields
Cm	chloramphenicol
cm	centimeter
d	distilled
DHPPP	dihydropterin pyrophosphate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DR	dynamic regulation
ds	double stranded
DSM	Deutsche Sammlung von Mikroorganismen
dTMP	deoxythymidine monophosphate
dUMP	deoxyuridine monophosphate
EF	elongation factor
eop	efficiency of plaquing
EtBr	ethidium bromide
G	guanine
g	gram
GDP	guanosine diphosphate
GTP	guanosine triphosphate
HIV	Human Immunodeficiency Virus
HNH	histidine-asparagine-histidine
hr	hour
IFM	Institute of Fermentation, Osaka
kb	kilobase
kV	kilovolts
LB	Luria-Bertani broth
LA	Luria-Bertani agar
M	molar
µg	microgram
mg	milligram

min	minute
ml	millilitre
MOI	multiplicity of infection
mM	millimolar
mm	millimetre
MTF	methylenetetrahydrofolate
MW	molecular weight marker
ng	nanogram
nm	nanometer
ORF	open reading frame
pABA	*para*-aminobenzoic acid
PEG	polyethylene glycol
PFGE	pulsed field gel electrophoresis
pfu	plaque forming units
Rep	replication
rRNA	ribosomal ribonucleic acid
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
SC	sodium-calcium
SDS	sodium dodecyl sulfate
sec	second
ss	single stranded
T	thymine
tRNA	transfer ribonucleic acid
TBE	Tris-borate-EDTA
TE	Tris-EDTA
Tris	Tris(hydroxymethyl)-aminomethane
UV	ultraviolet
V	volts
v	volume
w	weight

1. INTRODUCTION

1.1 Nocardioforms

Nocardioforms are a group of Gram-positive bacteria which form a mycelium that breaks up into rods or coccoid elements. They are a member of the aerobic actinomycetes and possess GC-rich DNA. Their cell walls contain mycolic acids and have a type IV composition, recognized by the presence of *meso*-diaminopimelic acid, arabinose, and galactose. In the ninth edition of Bergey's Manual of Determinative Bacteriology, the genera *Nocardia*, *Gordonia*, *Rhodococcus*, and *Tsukamurella* are included in the nocardioforms (Holt *et al*., 1994). Currently, taxonomy within these genera is changing rapidly, as modern molecular biological techniques contribute to the recognition of new species and reassignment of old ones (Brown-Elliot *et al*., 2006). Closely related organisms include *Mycobacterium* and *Corynebacterium*. The nocardioforms are usually saprophytic and abundant in the soil worldwide.

1.1.1 Nocardioform infections

Many nocardioform species cause infection in animals and humans, both immunocompetent and immunocompromised. Typically, they cause a granulomatous inflammatory reaction, which may progress to abscess formation (McNeil and Brown, 1994). However, symptoms may vary at each of potential infection sites including lungs, brain, bones, pericardium, skin, among others (Salinas-Carmona, 2000). Nocardioform species which cause human infections are listed in Table 1.1.

Table 1.1 Nocardioform species associated with human infections.

Genus	**Species**
Nocardia	*N. asteroides*
	N. brasiliensis
	N. farcinica
	N. nova
	N. otitidiscaviarum
	N. pseudobrasiliensis
Rhodococcus	*R. equi*
Gordonia	*G. terrae*
	G. bronchialis
	G. rubropertincta
Tsukamurella	*T. paurometabola*

Although infrequently encountered in clinical practice, nocardioform infections are challenging to clinicians for several reasons. They are difficult to diagnose because of nonspecific clinical presentations. This problem is made worse by the difficulty in species identification, caused by their slow growth rate and incomplete classification. Recent advances in serological, biochemical, and molecular biological methods have not completely solved this problem. Obtaining biological samples is another problem, as it requires invasive procedures in the infected organs (Corti and Fioti, 2003). In addition, a rapid test for routine use for most nocardioform species does not exist (Lederman and Crum, 2004).

Knowledge of molecular basis of nocardioform virulence is slight, and host immune responses affecting disease progression are not well understood. Main route of acquisition is believed to be through direct inhalation into the lungs or inoculation through the skin.

Incidence of human infections by nocardioforms has increased significantly over the last two decades, probably due to increase in population of the immunocompromised, mostly attributed to HIV/AIDS. For example, autopsies have revealed that as many as 4% of AIDS patients die with evidence of *Nocardia* infection (Lederman and Crum, 2004). There is a significant risk of mortality associated with many nocardioform infections. For example, with *R. equi*, the overall rate of mortality is ~25%: 50 – 55% in HIV-infected patients; 20 – 25% in non-HIV immunocompromised patients; and ~11% in immunocompetent hosts (Kedlaya, *et al*., 2001). With *Nocardia* infections, the rate of mortality in immunocompetent hosts is estimated to be 10 – 50% (Lederman and Crum, 2004).

In addition to their own importance in clinical settings, nocardioforms are phylogenetically close to major human pathogens *Mycobacterium tuberculosis*, causing human tuberculosis, and *Mycobacterium leprae*, causing leprosy. Tuberculosis claims ~2 million lives per year, more than can be attributed to any other bacterial infection (Corbett *et al*., 2003). They are also pathogenically related, as both nocardioforms and mycobacteria may produce similar symptoms. As a result, many *Rhodococcus* and *Nocardia* infections have been mistaken for tuberculosis (Takasugi and Godwin, 1991).

There is no standard treatment for nocardioform infections, due to diverse clinical manifestations, variable levels of host immunocompetence, and variation in susceptibility to drugs. Although surgery has been beneficial depending on the site and symptom of infection, combination antibiotic therapy is most widely in use (McNeil and Brown, 1994). However, multiple resistance to all major classes of drugs by some of these organisms has been reported (Corti and Fioti, 2003; Hsueh *et al*., 1998; McNeil and Brown, 1994; Nordmann *et al*., 1992; Yazawa *et al*., 1993), making it desirable to identify additional targets and develop novel antimicrobial agents.

1.2 Targets for antibiotics

Antibiotics are low molecular weight compounds that kill or inhibit susceptible bacteria and are either produced by microorganisms or artificially synthesized. Currently, there are five major targets for the action of antibiotics: biosynthesis of cell wall, protein, DNA, RNA, and folate coenzyme. Major classes of antibiotics inhibiting these targets are briefly described below.

1.2.1 Antibiotics which target cell wall biosynthesis

The Gram-positive cell wall consists of a single 20 – 80 nm thick peptidoglycan layer outside the plasma membrane, while the Gram-negative counterpart consists of a 2 – 7 nm thick peptidoglycan layer surrounded by a 7 – 8 nm thick outer membrane. Major classes of antibiotics which inhibit cell wall biosynthesis are β-lactams and glycopeptides.

β-lactams act by making themselves available for catalytic action of transpeptidases, which are responsible for covalent cross-linking of peptide strands of peptidoglycan. The transpeptidases mistake the β-lactams for substrates, causing the enzyme to be stranded in mid-catalytic cycle (Knox *et al*., 1996).

Glycopeptides prevent the cross-linking of the peptidoglycan units by binding to pentapeptidyl tails terminating in D-Ala_4-D-Ala_5. This process occurs in both polymerized (but yet to be cross-linked) peptidoglycan units and lipid II molecules at the periplasmic face of plasma membrane (Walsh *et al*., 1996).

1.2.2 Antibiotics which target protein biosynthesis

Prokaryotic protein biosynthesis is a complex process that takes place at 50S and 30S ribosomal subunits. The 50S subunit consists of 23S rRNA, 5S rRNA, and ~34 proteins, while the 30S subunit consists of 16S rRNA and ~21 proteins. Major classes of antibiotics which target protein biosynthesis are macrolides, tetracyclines, aminoglycosides, and oxazolidinones.

Macrolides bind to the 23S rRNA in the region of nucleotides 2058 – 2062, blocking the peptide exit tunnel and resulting in premature release of peptidyl-tRNA (Schlunzen *et al*., 2001).

Tetracyclines block the binding of incoming aminoacyl-tRNAs to A-site on the 16S rRNA in the 30S ribosomal subunit. This causes premature release of aminoacyl-tRNA, which terminates the translation process without peptide bond formation (Chopra and Roberts, 2001).

Aminoglycosides bind to a 16S rRNA pocket in the 30S ribosomal subunit, blocking the A-site for aminoacyl-tRNA binding (Carter *et al*., 2000).

Oxazolidinones are the only structurally novel antibiotic to be developed in the last three decades. The exact mechanism of action is unknown but has been proposed that they block the first step in peptide bond formation by binding to P-site in peptidyltransferase centre of the 23S rRNA in the 50S ribosomal subunit (Patel *et al*., 2001).

1.2.3 Antibiotics which target DNA replication and repair

Although not as complex as protein biosynthesis, DNA replication involves ~20 proteins, especially by DNA polymerase III. Fluoroquinolones target DNA replication, causing accumulation of cleaved DNA molecules by preventing their re-ligation by α subunit of DNA gyrase and halt replication by preventing progression of DNA replication forks (Maxwell, 1997).

1.2.4 Antibiotics which target folic acid metabolism

Folic acid is essential for the synthesis of pyrimidines and purines, the precursors of DNA and RNA. Sulfamethoxazole and trimethoprim block essential steps in the folic acid pathway, the former by inhibiting dihydropteroate synthase in a reaction which converts dihydropterin pyrophosphate and *para*-aminobenzoate to dihydropteroate, and the latter by inhibiting dihydrofolate reductase in a reaction which converts dihydrofolate into tetrahydrofolate (Walsh, 2003).

1.2.5 Antibiotics which target DNA transcription

The only antibiotics in clinical use which target transcription are rifamycins such as rifampicin. They inhibit RNA polymerase by binding in the DNA/RNA tunnel of the β subunit, thereby preventing the RNA chain from elongating (Campbell et al., 2001).

1.2.6 Summary of major targets for antibiotics

Major classes of antibiotics and their targets are summarized in Table 1.2.

Table 1.2 Major classes of antibiotics and their targets.

Antibiotic	Target
Cell wall synthesis	
β-Lactams	transpeptidases
Glycopeptides	D-Ala-D-Ala termini of peptidoglycan and of lipid II
Protein synthesis	
Macrolides	23S rRNA
Tetracyclines	A-site on 16S rRNA
Aminoglycosides	A-site on 16S rRNA
Oxazolidinones	P-site on 23S rRNA
DNA replication / repair	
Fluoroquinolones	α subunit, DNA gyrase
Folate coenzyme synthesis	
Sulfamethoxazole	dihydropteroate synthase
Trimethoprim	dihydrofolate reductase
RNA synthesis	
Rifamycins	β subunit, RNA polymerase

1.3 Bacterial resistance to antibiotics

Soon after any antibacterial agent is introduced for clinical use, resistance arises. Selection of resistance is inevitable, as it is a Darwinian consequence of antibiotic usage. Bacteria resist antibiotics using four general mechanisms: drug inactivation, target alteration, prevention of drug influx, and drug efflux (Putman *et al*., 2000).

The pharmaceutical industry has responded with modifications to already existing drugs or with new classes of drugs. For example, penicillinase-stable penicillins, and second generation aminoglycosides were produced to counteract resistant organisms. However, resistance has emerged to the new agents in all cases, thereby repeating the

cycle (Bush, 2004). Usage of a combination of existing drugs is not a complete solution, as evidenced by the emergence of multidrug-resistant pathogens (Livermore, 2004). In order to keep ahead of resistance, development of new antimicrobial agents is vital.

1.4 Antibacterial drug discovery

Despite the emergence of multidrug-resistant bacteria, there has been a serious decline in development of novel antibacterials in the past several decades. In 2000, the oxazolidinones were the first structurally novel antibiotic to be introduced in over thirty years. Most antibiotics introduced since 1968 have been modifications of already existing ones (Powers, 2004). One reason for this decline is difficulty in target identification (Projan and Shlaes, 2004).

Traditionally, antibiotics were developed by screening natural sources for molecules with antimicrobial effects. Most recently, searches for antibacterials have utilized bioinformatics to identify essential and conserved ORFs in bacterial genomes which may be potential targets (Thomson *et al*., 2004). Currently, complete genome sequences of ~ 300 bacterial species are publicly available (http://www.tigr.org/tigr-scripts/CMR2/CMRGenomes.spl), including those of many clinically important pathogens. However, after more than 15 years since the beginning of the genomics era, not a single antimicrobial agent has been developed by this approach (Thomson *et al*., 2004). Other novel methods currently employed for the discovery of new targets are transcriptional profiling and proteomic analysis (Bandow *et al*., 2003).

Besides searching the bacterial genomes, various potential antimicrobial agents are currently under investigation. These include antimicrobial peptides, bacteriophages, and probiotic bacteria (Gillor *et al*., 2005).

1.5 Antimicrobial peptides

1.5.1 Eukaryotic antimicrobial peptides

Gene-encoded, ribosomally synthesized antimicrobial peptides (AMPs) are widely distributed in nature and produced by both eukaryotes and prokaryotes. AMPs are produced in phagocytes and mucosal epithelial cells of mammals, haemolymph of insects, and skin of amphibians, where they provide an early defence against invading microorganisms (Boman, 1995). Some of well-characterized animal AMPs are defensins of humans, cecropins of insects, and magainins of frogs. Animal AMPs display a broad spectrum of activity against both Gram-positive and -negative bacteria. Although various AMP structures exit, those for which the mode of action has been elucidated act by membrane permeabilization (Papagianni, 2003). Examples of eukaryotic AMPs are shown in Table 1.3.

Table 1.3 Examples of eukaryotic AMPs grouped according to structure.

Structure	**Peptide**	**Source**
High content of certain amino acid	Abaecin	Insects
	Bac-5	Sheep
	Prophenin	Pigs
Intramolecular disulfide bridges	α-Defensins	Humans
	Insect defensins	Insects
	Plant defensins	Plants
Amphiphilic α-helical structure	Magainins	Frogs
	Bombinins	Frogs
	Cecropin	Insects

(Modified from Papagianni, 2003)

1.5.2 Prokaryotic antimicrobial peptides

Prokaryotic AMPs, known as bacteriocins, are a large and diverse group of toxins found in both bacteria and archaea. They are narrow spectrum antimicrobials and use a variety of killing mechanisms. Lactic acid bacteria produce the best-known Gram-positive bacteriocins, the lantibiotics, which include nisin, a commercially used food preservative. Bacteriocins produced by Gram-positive bacteria act by membrane depolarization, membrane leakage, and cell wall synthesis inhibition (Pag and Sahl, 2002). Well known bacteriocins from Gram-negative bacteria include colicins, microcins, both of which are produced by *E. coli*, and pyocins produced by *Pseudomonas* spp. Modes of action by colicins include depolarization of cell membrane, degradation of DNA and RNA, and inhibition of cell wall and protein synthesis (Braun *et al*., 1994). Microcins target cell membranes and biosynthesis of DNA, RNA, and proteins (Pons *et al*., 2002; Destoumieux-Garzón *et al*., 2002). Pyocins act by forming pores on cell membranes and degrading DNA and RNA (Duport *et al*., 1995; Parret and De Mot, 2000; Gillor *et al*., 2005). Different groups of bacteriocins and their targets are summarized in Table 1.4.

Table 1.4 Bacteriocin types and their targets.

Bacteriocin	**Class/type**	**Target**
Lactic acid bacteria bacteriocins	IA, II	cell membrane
	IB	cell wall synthesis: lipid II
	III	unknown
Colicins	A, B, E1, Ia, Ib, K, N	cell membrane
	E2, E7, E8, E9	DNA
	E3, E4, E6	protein synthesis: 16S rRNA
	E5, D	protein synthesis: tRNA
	M	cell wall synthesis: lipid carrier
Microcins	Modified	DNA replication: β subunit, DNA gyrase (gyrB)
		Transcription: β' subunit, RNA polymerase
		Protein synthesis
	Non-modified	cell membrane
Pyocins	R, F	cell membrane
	S	DNA
		cell membrane
		protein synthesis: tRNA

1.6 Bacteriophage-mediated antimicrobials

Bacteriophages hold potential for the development of novel antimicrobials. This may be possibly done in three different ways: 1) phage therapy, 2) phage lytic enzymes, and 3) target identification.

1.6.1 Phage therapy

The idea of using bacteriophages, the most abundant organisms on the planet, as antimicrobials is not new. After d'Hérelle coined the term "bacteriophage" in 1916, he successfully treated a patient with severe dysentery using a phage preparation taken orally and set up phage therapy trials across the globe (Stone, 2002). Phage therapy was extensively tested, succeeding against diseases such as dysentery, typhoid, paratyphoid, cholera, pyogenic and urinary-tract infections (Thacker, 2003). However, overall results were variable; due to ineffective trials and the advent of antibiotics, the use of bacteriophages as antimicrobials was abandoned. In hindsight, the ineffective trials were caused by prescription of wrong phages or phages which were no longer viable (Pirisi, 2000).

Although abandoned in the west, phages continued to be used as antibacterial agents in the former Soviet Union and Eastern Europe. For example, during a civil war in the early 1990's, Georgian soldiers carried spray cans with a phage suspension effective against *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Streptococcus pyogenes*, and *Proteus vulgaris* (Stone, 2002). In some Russian villages, civilians relied exclusively on phages as the cheaper option compared to antibiotics (Pirisi, 2000). Although some data have been published, most do not exist in English.

Phage therapy may offer several advantages over antibiotics. Phages are able to mutate in step with evolving bacteria, thereby maintaining their virulence against pathogens which may build resistance. Also, while antibiotics are metabolized by the host, phages multiply as long as host cells are present. In addition, phages are specific in their bacterial targets, thus sparing patients from side effects which cause the destruction of beneficial bacteria and other cells. Normal microbial balance in the body is therefore maintained, reducing risk of secondary infection, often by antibiotic-resistant pathogens (Merril *et al*., 1996). Moreover, owing to the

abundance and diversity of phages in nature, it may be possible to constantly isolate a new variant that maintains its virulence.

Due to the emergence of multidrug-resistant pathogenic bacteria, there has been a recent revival in interest in phage therapy. One advance in phage therapy is against vancomycin-resistant enterococci, for which clinical trials are currently underway (Bradbury, 2004).

1.6.2 Phage lytic enzymes

Instead of phage therapy, there has been a recent interest in purification and characterization of phage lytic enzymes, which can be used to combat target bacteria. Phage enzymes directed against Gram-positive pathogens, including *Streptococcus pneumoniae* and *Bacillus anthracis*, have been isolated and shown to successfully destroy these pathogens both *in vitro* and *in vivo* (Loeffler *et al*., 2001; Schuch *et al*., 2002). For example, lytic enzyme Pal from *S. pneumoniae* phage Dp-1 has been shown to kill fifteen common serotypes of the bacterium, including highly penicillin-resistant strains (Loeffler *et al*., 2001). Lysin PlyG from *B. anthracis* phage γ is host specific, binding to cell wall antigens not recognized by current antibiotics (Schuch *et al*., 2002).

1.6.3 Phage-mediated target identification

Another approach to phage-mediated antimicrobial development is target identification. This has been exemplified by a study by Liu *et al*. (2004), who screened the genomes of *S. aureus* phages for ORFs which were inhibitory to the host bacterium. Targets of these antibacterial ORFs were then biochemically identified, followed by a high-throughput screen to discover small molecules which could

inhibit the host cell by the same mechanism as the phage protein. Using this approach, 31 novel polypeptide families were identified from 26 *S. aureus* phages, several of which were inhibitory to DNA replication and transcription (Liu *et al*., 2004). This approach may be a powerful method to identify effective targets in the host, as phages are believed to have evolved multiple strategies to disable host cells.

Although there is potential in the traditional phage therapy or in the use of their lytic enzymes, the development of novel antibiotics by recruiting phages in the identification of new targets may be the best phage-mediated antimicrobial. Small antimicrobial molecules that are commonly used have been selected to have a sufficient degree of distribution in human tissues and appropriate levels of other pharmacokinetic properties. Therefore, novel antibiotics are expected to better meet these criteria of an effective medicine compared to a phage particle or even a lytic protein, whose complexity is much higher and size much larger (Projan, 2004).

1.7 Phage diversity

A major factor contributing to the potential of phage-mediated target identification is the abundance of these viruses in nature. For instance, > 50 different phages capable of infecting *E. coli* have been included in the Universal Virus Database (http://ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm). In addition, it has been calculated that there are ~10 million bacterial species (Curtis *et al*., 2002). If each of these is a host for ≥ 10 different phages, then the number of phage species may be at least 100 million.

Despite the predicted abundance of phages, the use of phage ORFs in identifying novel targets will be effective only if phages employ diverse mechanisms in inhibiting their hosts. If all phages possess similar genes and use identical

mechanisms to act against their hosts, there are limited prospects of discovering new bactericidal mechanisms.

Recent comparative genomic analyses of phages infecting *Mycobacterium smegmatis* (Hatfull *et al*., 2006; Pedulla *et al*., 2003), *Pseudomonas aeruginosa* (Kwan *et al*., 2006), and *S. aureus* (Kwan *et al*., 2005) imply enormous diversity in their genetic makeup. Only 15% of protein families from 30 sequenced mycobacteriophage genomes have sequence similarity to previously reported proteins (Hatfull *et al*., 2006). This percentage for 18 sequenced *P. aeruginosa* phages and 27 *S. aureus* phages are 35% and 45%, respectively (Kwan *et al*., 2006; Kwan *et al*., 2005). Moreover, all of these phage genomes display little relationship to each other (Hatfull *et al*., 2006; Kwan *et al*., 2006; Kwan *et al*., 2005), suggesting that each individual phage possesses unique genes. Further support for these findings comes from a shotgun sequencing of uncultured viral communities in sea water, which have shown that ~75% of phage sequences are novel (Breitbart *et al*., 2002).

Average number of ORFs from the sequenced *M. smegmatis* phages, *P. aeruginosa* phages, and *S. aureus* phages, are 112, 105, and 80, respectively (Hatfull *et al*., 2006; Kwan *et al*., 2006; Kwan *et al*., 2005). If a conservative assumption is made that all 100 million phage species in the world possess 50 ORFs, of which 50% are unknown, then there are 2.5 billion phage ORFs to be discovered (Rohwer, 2003). Obviously, only a fraction of these ORFs would participate in killing the host; however, phages present an enormous pool of genes from which a large number of novel bactericidal mechanisms may be identified.

1.8 Phage lethal genes

Expression of phage genes may be early, middle, or late, depending on the time at which their transcription begins after initiation of infection. Many lethal genes are

expressed early, as they take part in shutoff of host macromolecular biosynthesis. Knowledge of mechanism by which a phage functions in a host is limited. Similarly, an overwhelming majority of genes believed to be lethal are uncharacterized. Studies also suggest that a single phage may possess several numerous such genes. A small number of them have been studied; when tested individually, a majority of these bactericidal genes have been proven to be dispensable during infection, indicating that a single phage has multiple strategies for killing a host. Therefore, with respect to novel antimicrobial target identification, not only does the potential of phages attribute to the number of species and their genetic heterogeneity, but also to the variety of targets attacked by individual phage species. It should also be noted that killing of hosts may not be the only function of many lethal genes; they may have other benefits for phage replication, while being deleterious to the host. Some of the studied lethal phage genes are summarized below.

1.8.1 Phage T4

Studies on *E. coli* phage T4 began in the 1940s, but of its ~300 ORFs, 127 still have no assigned function (Miller *et al*., 2003). Many of these unknown genes are unclonable, reinforcing the notion that they are lethal to the *E. coli* host, and a high level of their expression occurs early in the infection cycle (Miller *et al*., 2003).

One example of a T4 lethal gene is *alc*, which uniquely recognizes the rapidly elongating form of RNA polymerase complex, leading to termination of transcription (Kutter *et al*., 1994). Another example is Alt protein, which is packaged in the capsid and transported with DNA into the cell upon infection. Alt functions by ADP-ribosylating the α subunit of the host RNA polymerase (Igarashi *et al*., 1991).

T4 possesses at least two other ADP-ribosyltransferase genes, *modA* and *modB*, which are known to be lethal to the host. ModA ADP-ribosylates both α subunits of

the host RNA polymerase (Tiemann *et al*., 1999), whereas ModB acts on the translation machinery of the ribosome by ADP-ribosylating the S1 protein, elongation factor EF-Tu, and the chaperone "trigger factor" (Tiemann *et al*., 1999).

Other T4 gene products inhibiting translation are Lit and Gol proteins. Lit cleaves the host EF-Tu at a region which is central to the Mg-GTP-binding domain (Yu and Snyder, 1994). This process is aided by Gol, which stabilizes the EF-Tu-GDP open complex, making it accessible for cleavage (Bingham *et al*., 2000). Gol is also a part of a T4 head capsid protein, suggesting its dual function (Bingham *et al*., 2000).

Intron- or intein-encoded DNases with highly specific recognition sequences are called homing endonucleases, which help to disseminate the intron/intein into specific regions of the genome. T4 possesses > 15 such genes (Chevalier and Stoddard, 2001). Although purpose served by the introns/inteins is not clear, these endonucleases have been shown to be bactericidal (Sharma *et al*., 1992).

1.8.2 Phage T7

T7 is an *E. coli* phage which has a number of genes implicated in host transcription shutoff. Gp2 binds to promoters of host DNA, thereby abolishing promoter recognition by the host RNA polymerase (Nechaev and Severinov, 1999). Gp0.7 is a serine/threonine protein kinase, which phosphorylates various biosynthetic components: among the transcription machinery includes the β' subunit of RNA polymerase (Zilling *et al*., 1975); RNase III, which is responsible for mRNA processing (Mayer and Schweiger, 1983); RNase E, which is involved in mRNA decay (Marchand *et al*., 2001); among the translation machinery includes IF1, IF2, IF3, elongation factor G, and ribosomal proteins S1 and S6 (Robertson *et al*., 1994). Phosphorylation of the β' subunit of RNA polymerase is lethal (Severinova and

Severinov, 2006). Whether host viability is affected by the phosphorylation of other targets is yet to be discovered.

1.8.3 Phage lambda

Lambda is an extensively studied *E. coli* phage, although most works have focused on non-lethal genes. One of few identified bactericidal gene products is peptidyl-tRNA hydrolase, which cleaves peptidyl-tRNA and aminoacyl-tRNA (Garcia-Villegas *et al*., 1991). In addition, lambda *p* gene causes host death by inhibiting the binding of template DNA and ATP to the DnaA protein (initiator protein), thereby preventing DNA replication to begin (Datta *et al*., 2005). CII gene product is another inhibitor of host DNA replication. Although the exact mechanism is unknown, it seems to interfere with the association of DNA helicase with origin of replication (Kędzierska *et al*., 2003). Another lethal gene product, Kil protein, inhibits cell division by interacting with components of the cell envelope (Sergueev *et al*., 2001).

1.8.4 Phage SPO1

Phage SPO1 of *Bacillus subtilis* possesses a cluster of early genes in a region named "host-takeover module", including ~24 of mostly unknown function but believed to be involved in inhibition of host biosynthesis (Stewart *et al*., 1998). An example of such gene is *e3*, whose expression leads to inhibition of DNA, RNA, and protein synthesis in both *B. subtilis* and *E. coli* (Wei and Stewart, 1993). Products of genes *44*, *50*, and *51* cause inhibition of host transcription, possibly by binding to RNA polymerase (Sampath and Stewart, 2004).

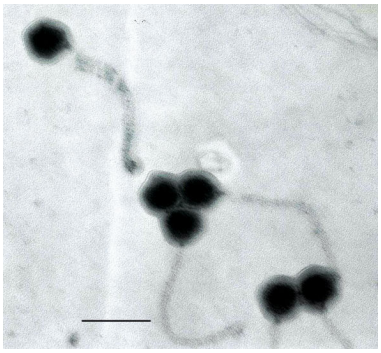
1.8.5 Phages with ssDNA or ssRNA genomes

Where phages with dsDNA genomes encode two to five proteins to carry out host lysis, phages with small, single-stranded nucleic acids usually have only a single gene required for rupturing host cells (Bernhardt *et al*., 2001). Instead of attacking pre-existing peptidoglycan, these phages interfere with cell wall synthesis, thereby causing weakness in the wall, which collapses from osmotic pressure from within. For example, ssDNA phage ϕX174 encodes a single lytic enzyme, E, which blocks MraY, a bacterial membrane protein responsible for transfer of murein precursors to lipid carriers that transport it through the cell membrane (Bernhardt *et al*., 2000). In the ssRNA genome of phage Qβ is a_2 gene, encoding a protein inhibiting MurA, another essential enzyme for cell wall synthesis (Bernhardt *et al*., 2001). Protein A_2 has multiple functions including adsorption to host sex pilus and protection against ribonucleases (Bernhardt *et al*., 2001).

1.9 Phage isolation

For phage-mediated antimicrobial identification, an additional advantage is that they are generally readily isolated from a variety of environments, such as soil, seawater, and sewage. Félix d'Hérelle Reference Center for Bacterial Viruses constantly keeps a global record of new phage descriptions, including many unpublished data. It is evident from their record (Ackermann, 2001) that a majority of isolated phages since 1959 have hosts belonging to the genera *Pseudomonas*, *Rhizobium*, *Escherichia*, *Salmonella*, *Vibrio*, *Lactococcus*, *Staphylococcus*, *Streptococcus*, *Bacillus*, *Clostridium*, *Lactobacillus*, *Listeria*, and *Streptomyces*. Nocardiophages, or phages that infect nocardioforms, are very scarce in their record (Anderson and Bradley, 1961; Brownell *et al*., 1967; Riverin *et al*., 1970; Hiddema *et al*., 1985; Dabbs, 1987; Dabbs, 1998).

In my previous study (Shibayama, 2004), a few nocardiophages were isolated using *Rhodococcus erythropolis* as a host and characterized. Electron micrographs of two of these phages, FA2 and FA3, are shown in Figure 1.1. In the current study, phage FA3 has been renamed Fairland1.



FA2

FA3

Figure 1.1 Electron micrographs of phages FA2 and FA3 (Fairland1).
Bar = 100nm
(Taken from Shibayama, 2004)

1.10 Aim of the project

The aim of this work is to identify genes in novel nocardiophages whose products are inhibitory towards susceptible microorganisms. To achieve this, novel nocardiophages will be isolated from soil and characterized. Phage genomic libraries will be constructed using an *E. coli-Rhodococcus* shuttle vector and screened for genes inhibitory to the host by transformations. For individual clones, minimum necessary DNA will be determined, sequenced, and compared to those in the databases. Where possible, the functions of these genes will be identified by sequence similarity searches.

2. MATERIALS AND METHODS

For all solutions and reagents used in this work refer to Appendix A.

2.1 Bacterial and bacteriophage strains

Bacterial strains used in this work, excluding those used for phage host range study, are listed in Table 2.1. Bacterial strains used for phage host range study are listed in Table 2.2. Phage strains are listed in Table 2.3.

Table 2.1 Bacterial strains used in this work.

Strain	Relevant characteristics	Source / reference
E. coli		
MM294-4	*hsdR17 endA1 gyrA*	Quan *et al*., 1997
MM294-4 [λ]	MM294-4 lysogenized with λcI857S7	Quan *et al*., 1997
GM2929	*dcm-6 dam13::*Tn*9 recF143, hsdR2*	Bachman, B.
R. erythropolis		
SQ1	Highly transformable derivative of ATCC4277	Quan and Dabbs, 1993

Table 2.2 Bacterial strains used for phage host range study.

Strain	Source
Rhodococcus	
R. erythropolis ATCC 4277	Ferreira, N.
R. erythropolis DSM 1069	DSM
R. rhodochrous Ri8	Dabbs, E.
R. opacus HL PM-1	Heiss, G.
R. equi ATCC 14487	Ferreira, N.
Gordonia	
G. australis A554	Ferreira, N.
G. rubropertincta ATCC 25593	Ferreira, N.
G. desulfuricans NB4	van Hamme, J.
Nocardia	
N. mexicana IFO 3927	IFO
Mycobacterium	
M. smegmatis mc^2 155	Durbach, S.
M. parafortuitum 490	Dabbs, E.

ATCC, American Type Culture Collection; DSM, Deutsche Sammlung von Mikroorganismen; IFO, Institute of Fermentation, Osaka.

Table 2.3 Phages used in this work.

Phage	Host organism used for isolation	Source / reference
Fairland1	*R. erythropolis*	Shibayama, 2004
Wits1	*R. erythropolis*	This work
Kanazawa1	*R. equi*	This work
Perouges1	*R. erythropolis*	This work

2.1.1 Media and growth conditions

Liquid cultures of all bacterial strains were grown in LB medium on a shaker or wheel. Plate cultures were grown on the same medium solidified with 1.5% agar.

E. coli MM294-4 and GM2929 were grown overnight at 37°C, and MM294-4 [λ] was grown at 30°C. Growth medium for MM294-4 and MM294-4 [λ] was supplemented with 50 µg/ml nalidixic acid.

Rhodococcus, *Gordonia*, *Nocardia*, and *Mycobacterium* strains were grown for 2 – 4 days at 30°C. Growth medium for *R. erythropolis* SQ1 was supplemented with 50 µg/ml rifampicin.

For short-term storage, all strains were kept on agar plates at 4°C. For long-term storage, cells suspended in 33% glycerol were stored at -70°C.

All phages were propagated using *R. erythropolis* SQ1. Optimal MOIs for phage propagation are shown in Table 3.6. All phage suspensions were stored at 4°C.

2.2 Plasmids

Plasmids used in this work are listed in Table 2.4.

Table 2.4 Plasmids used in this work.

Plasmid	Relevant characteristics	Source / reference
pDA71	*E. coli – Rhodococcus* shuttle vector	Dabbs *et al*., 1995
pUC18/19	*E. coli* cloning vector with lacZ', Amp-R	Fermentas

Plasmid pDA71 was selected in transformations by 200 µg/ml ampicillin for *E. coli* and 40 µg/ml chloramphenicol for *R. erythropolis*. Plasmid pUC18/19 was selected by 200 µg/ml ampicillin for *E. coli*. Plasmids were stored in 6.09 M CsCl or 4 mM Tris-HCl at -20°C.

2.3 Isolation of nocardiophages

2.3.1 Soil assay

One gram of soil was agitated in 10 ml of LB supplemented with 10 mM $CaCl_2$ and $MgCl_2$ in a 100 ml flask on a rotary shaker at ~80 rpm for 8 – 16 hr at 30°C. The soil suspension was centrifuged in a Beckman JA-20 rotor at 15000 rpm for 10 min, after which the supernatant was decanted. 2 ml of this supernatant was added to 50 µl of a stationary phase culture of *R. erythropolis* SQ1, vortexed for 1 sec, and left at 30°C for 30 min. Then 2 ml of sloppy agar was added, mixed by rolling between the palms, and poured on an LA plate supplemented with 100 µg/ml rifampicin, 100 µg/ml kanamycin, 50 µg/ml nystatin, 10 mM $CaCl_2$ and $MgCl_2$. The plate was incubated for ~2 days and checked for plaques.

2.3.2 Production of phage suspensions

Plaques were picked with by stabbing a toothpick through the soft agar and dipping this into 100 µl LB. The suspension was vortexed for 1 sec, left for 5 min, and microfuged for 1 min, after which the phage was in the supernatant.

2.3.3 Single plaque purification

Phage suspension was serially diluted with LB. Order of the dilution depended on the suspected titre of the phage suspension, as it needed to be sufficient for single plaques to appear. For the last dilution, 10 µl of the phage suspension was added to 2 ml of LB and 50 µl of stationary phase *R. erythropolis* SQ1 culture, vortexed for 1 sec, and left at 30°C for 30 min. 2 ml of sloppy agar was added and poured onto an agar plate containing 10 mM $CaCl_2$ and $MgCl_2$. The plate was incubated at 30°C for 2 days. A plaque was picked, and the purification was repeated twice.

2.4 Preliminary phage characterization

2.4.1 Determination of divalent ion requirement

Four LA plates were prepared: 1) LA, 2) LA + 10 mM $CaCl_2$, 3) LA + 10 mM $MgCl_2$, and 4) LA + 10 mM $CaCl_2$ & $MgCl_2$. Each plate was overlaid with 50 µl of *R. erythropolis* SQ1 and dried with the lid open at 47°C for 15 min. Onto each plate, 10 µl of phage suspension was spotted. The plates were incubated at 30°C for 2 days and checked for plaques.

2.4.2 Determination of host range

Phage host range was determined by spotting 10 µl of phage suspension onto an LA plate overlaid with a test strain of bacteria. The plate was incubated at 30°C for 2 days and checked for plaques.

2.4.3 Determination of efficiency of plaquing

Titre of a phage suspension (pfu/ml) was determined on both *R. erythropolis* SQ1 and another strain. Efficiency of plaquing (eop) was calculated by the following formula:

$$\text{eop} = \frac{\text{pfu/ml on another strain}}{\text{pfu/ml on } R.\ erythropolis \text{ SQ1}}$$

2.4.3.1 Determination of plaque forming units (pfu)

A phage suspension was serially diluted 10^{-8} using LB. An LA plate was overlaid with *R. erythropolis* SQ1 and dried with the lid open at 47°C for 15 min. On this plate 10 µl of each phage dilution was spotted. The plate was incubated until plaques were visible. Pfu/ml was calculated by the following formula:

$$\text{pfu/ml} = \frac{\text{number of plaques} \times \text{dilution factor}}{\text{volume}}$$

2.4.3.2 Determination of colony forming units (cfu)

A bacterial culture was serially diluted 10^{-7} using LB. On an LA plate, 10 µl of 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7} dilutions were spread. The plate was incubated until colonies were visible. Cfu/ml was calculated by the following formula:

$$\text{cfu/ml} = \frac{\text{number of colonies} \times \text{dilution factor}}{\text{volume}}$$

2.4.4 Cross checking of phages using lysogens or resistant mutants

On an LA plate overlaid with *R. erythropolis* SQ1, 10 µl of phage suspension was spotted. The plate was incubated at 30°C until bacterial growth was visible inside the clearing region of lysis where the phage suspension had been spotted. These colonies were picked and streaked to single colonies if necessary. Cells from each colony were then grown in LB and tested for immunity to superinfection by the same phage.

2.5 Lysate production

2.5.1 Plate lysate production

A phage suspension was serially diluted 10^{-1}, 10^{-2}, and 10^{-3}. 10 µl of each dilution was added to 2 ml of LB and 50 µl of *R. erythropolis* SQ1 culture, vortexed for 1 sec, left at 30°C for 30 min, and overlaid onto an LA plate containing 10 mM $CaCl_2$ and $MgCl_2$. The plate was incubated at 30°C for 2 days, or until a significant difference was observable between this plate and the cell-only control.

Onto the plate lysate, 2 ml of LB was added. The entire top agar layer, together with all the liquid, was collected with a spatula or pipette and transferred to a centrifuge tube. The tube was vortexed for 1 sec, left for 5 min, and centrifuged in a JA-20 rotor at 15000 rpm for 10 min. The supernatant was collected and titrated.

When a titre of $< 10^9$ pfu/ml was observed, plate lysate production was repeated with different conditions. MOI was altered by changing the amount of added phages or cells until a higher titre was observed.

2.5.2 Small scale liquid lysate production

To each of three 100 ml flasks, 10 ml of LB containing 10 mM $CaCl_2$ and $MgCl_2$ was added. Into the first flask, 200 µl of phage lysate (from plate lysate) was added and mixed by gentle swirling. From this flask, 100 µl of the content was transferred to the second flask and mixed. The same was done from the second flask to the third. Into each flask, 250 µl of a stationary phase *R. erythropolis* SQ1 culture was added. The flasks were left at 30°C for 30 min. They were then agitated on a rotary shaker at 80 rpm for 18 hr. Following the agitation, they were centrifuged in a JA-20 rotor at 15000 rpm for 10 min. The supernatant was collected and titrated. If the titre was $< 10^9$ pfu/ml, other MOIs were tested.

2.5.3 Large scale liquid lysate production

Condition optimal for the production of 10 ml lysate was applied to the production of 500 ml lysate, (5×100ml LB in 1000 ml flasks).

2.6 Phage purification

To the large scale lysate NaCl was added to a final concentration of 1 M and dissolved by slow stirring on a magnetic stirrer. Then polyethylene glycol (PEG 6000) was added to a final concentration of 10% (w/v) and again dissolved in the same manner. The lysate was cooled in ice-water slurry for at least 1 hr and centrifuged in a JA-10 rotor at 9000 rpm for 20 min at 4°C. The supernatant was discarded, and the phage in the pellet was purified in either a step gradient or equilibrium gradient of CsCl.

2.6.1 Phage purification in a step gradient

The phage pellet was dissolved in 3.5 ml of SC buffer containing 50% (w/v) CsCl and transferred to a Beckman Quick-Seal centrifuge tube using a Pasteur pipette. A step gradient was produced by layering 800 µl of CsCl solutions of increasing density (1.45, 1.50, 1.70 g/ml) under one another at the bottom of the tube. The tube was centrifuged in a Beckman VTi 65.2 rotor at 25000 rpm for 1 hr at 4°C. The bluish band of phage was extracted using a needle and a syringe and titrated.

2.6.2 Phage purification in an equilibrium gradient

The precipitated phage was resuspended in 5 ml of SC buffer containing 81.7% (w/v) CsCl. This was loaded into a Beckman Quick-Seal centrifuge tube and centrifuged at 38000 rpm for 24 hr at 4°C. The phage band was extracted and treated as in 2.6.1.

2.7 Transmission electron microscopy

2.7.1 Preparation of Formvar-carbon coated grids

A microscope slide was vertically dipped into 0.35% Formvar solution until about half was immersed, removed, its bottom edge blotted on paper towel, and dried by standing vertically for 2-3 min. Using a razor blade, the Formvar was scrapped off from the edges and corners of the slide. The slide was slowly immersed into a tub of dH_2O at a 45° angle, until the transparent film of Formvar peeled off the slide and floated onto the surface of the water. Using fine forceps, hexagonal 200-mesh copper grids were individually placed onto the Formvar film, with the dull side of the grid facing downward. The film of Formvar with the grids was lifted onto a filter paper

and air-dried. The grids were then coated with a film of carbon by the Electron Microscope Unit at the University of the Witwatersrand.

2.7.2 Negative staining of phages

If a purified phage was to be negatively stained, it was diluted to a titre of ~10^{10} pfu/ml with dH_2O. Phage from a lysate was not diluted. 5 µl of phage suspension was placed on the Formvar-carbon coated side of a copper grid. This was left for 30 min, after which excess fluid was drawn off by gently dabbing with a wedged filter paper. A drop of dH_2O was placed on the grid and immediately removed with a filter paper. Then, 5 µl of 1% uranyl acetate stain was placed on the grid for 20 seconds. Excess stain was removed with a filter paper, and the grid was air-dried for 10 min.

2.7.3 Observation of phage morphology

Negatively stained phage was examined and photographed in a Joel JEM 100-S transmission electron microscope operating at 80 kV. Micrographs were developed at the School of Animal, Plant, and Environmental Sciences at the University of the Witwatersrand by contact-printing the negatives.

2.8 DNA preparations

2.8.1 Phage genomic DNA preparation

2.8.1.1 Dialysis of purified phage in CsCl

Dialysis tubing was prepared according to Sambrook *et al*., 1989. Tubing of ~10 cm length was boiled for 10 min in 800 ml of 2% $NaHCO_3$ and 1 mM EDTA (pH 8.0) and rinsed in dH_2O. The tubing was then boiled for 10 min in 800 ml of 1 mM EDTA (pH 8.0) and rinsed in dH_2O.

150-300 µl of purified phage suspension diluted to 1 ml with dH_2O was pipetted into the dialysis tubing which was then dialyzed for 1 hr against 1000 volumes of a buffer containing 10 mM NaCl, 50 mM Tris-HCl (pH 8.0), and 10 mM $CaCl_2$ while gently stirring with a magnetic stirrer. The dialysis sac was then transferred to a fresh flask of buffer and dialyzed for another 1 hr.

2.8.1.2 Phage genomic DNA extraction

To obtain phage DNA from a plate lysate, DNaseI and RNase were added to a final concentration of 10 µg/ml and incubated at room temperature for 6 hr. To obtain DNA from purified phage, dialyzed suspension was directly subjected to the following steps.

To the phage suspension, EDTA (pH 8.0) was added to a final concentration of 20 mM, proteinase K to a final concentration of 50 µg/ml, and SDS to a final concentration of 0.5% (w/v). The mixture was incubated at 56°C for 1 hr, cooled to room temperature, subjected to three phenol extractions and one chloroform

extraction. A salt-ethanol precipitation was performed on the phage DNA, which was then dissolved in dH_2O.

2.8.2 *E. coli* bulk plasmid preparation

A single colony was inoculated into 100 ml of LB supplemented with 100 µg/ml ampicillin. The culture was grown overnight on a shaker, after which cells were harvested by centrifugation in a JA-10 rotor at 6000 rpm for 10 min, resuspended in 5 ml of solution I and transferred to a JA-20 tube. Then, 10 ml of solution II was added, mixed by gentle inversion, and left for 15 min. Thereafter, 7.5 ml of solution III was added, shaken vigorously, and left in ice-water slurry for 10 min. Cellular debris was removed by centrifugation in a pre-cooled (4°C) JA-20 rotor at 15000 rpm for 10 min. The supernatant was transferred to a fresh JA-20 tube, warmed to room temperature, and then 12 ml of isopropanol was added and left for 10 min. The tube was centrifuged at 15000 rpm for 15 min. The supernatant was decanted, after which the pellet was washed with 2 ml of 96% ethanol and vacuum dried for 20 min. The DNA was resuspended in 4 ml of TE buffer for 2 hr at 30°C. 400 µl of 1% ethidium bromide (EtBr) and 4.1 g of CsCl were added, inverted to mix, and the refractive index was adjusted to between 1.387 and 1.389. This was loaded into a Quick-Seal tube and centrifuged in a VTi 65.2 rotor at 45000 rpm for 16 hr. The plasmid band was extracted under a UV light using a needle and syringe.

EtBr was removed from the DNA solution by adding 1/10 volume of butanol, mixing by inversion, and microfuging for 10 sec. Upper layer containing the EtBr was removed. Butanol extraction was repeated ~3 times to remove all traces of EtBr. DNA in CsCl was stored at -20°C. When required, CsCl was removed from the DNA by ethanol precipitation.

2.8.3 *E. coli* mini plasmid preparation

A single colony was inoculated into 1 ml of LB supplemented with 100 µg/ml ampicillin. After the culture was grown for > 6 hr at 37°C on a shaker, cells were harvested by microfuging for 1 min, the supernatant decanted, and the pellet resuspended in 80 µl of solution I. 160 µl of solution II was mixed in by gentle inversion, left for 15 min, 120 µl of solution III added, shaken vigorously, and left in ice-water slurry for 5 min. Cellular debris was removed by microfuging for 5 min at 4°C. The supernatant was transferred to a fresh microfuge tube and warmed to room temperature. 220 µl of isopropanol was added and after 10 min at room temperature the tube was microfuged for 5 min, the supernatant decanted, then the pellet was washed with 150 µl of 96% ethanol and vacuum desiccated for 20 min. The DNA was dissolved in 150 µl of dH_2O containing freshly boiled RNase (10 µg/ml).

2.8.4 *E. coli* / *Rhodococcus* miniprep of total DNA

For *E. coli*, a single colony was inoculated into 1 ml of LB and grown at 37°C on a shaker overnight. For *R. erythropolis*, a single colony was inoculated into 1 ml of LBG (2% glycine) and grown at 30°C on a shaker for 2-3 days. Cells were harvested by microfuging for 1 min, followed by discarding the supernatant and resuspending the pellet in 500 µl of TE to which lysozyme had been freshly added. The mixture was incubated at 37°C for 1 hr, with inversion every 10 min, after which 100 µl of 10% TE-SDS was added and left for 10 min. 100 µl of 5 M KAc (pH 6.0) was added, shaken vigorously, left in ice-water slurry for 5 min, and microfuged for 5 min at 4°C. The supernatant was subjected to phenol-chloroform extraction, followed by salt and ethanol precipitation. The genomic DNA was resuspended in 150 µl of dH_2O containing freshly boiled RNase (10 µg/ml).

2.9 DNA manipulations

2.9.1 Phenol-chloroform DNA extraction

For DNA volumes of < 300 µl, TE buffer was added to make it up to this volume. Then, 1/3 volume of TE-saturated phenol was added, mixed by inversion, and microfuged for 5 min at room temperature. The aqueous phase was collected, to which 1/3 volume of chloroform was added, mixed by inversion, and microfuged for 30 sec. The aqueous phase was collected, and the DNA was subjected to salt and ethanol precipitation.

2.9.2 DNA precipitations

2.9.2.1 Salt and ethanol DNA precipitation

DNA was precipitated by adding 1/10 volume of 1 M NaCl and 2 volumes of 96% ethanol. The mixture was mixed by gentle inversion and microfuged for 20 min at 4°C. The supernatant was decanted and the tube was blotted on paper towel, followed by vacuum drying for 20 min. The DNA was resuspended in an appropriate volume of dH_2O or 4 mM Tris-HCl.

2.9.2.2 Ethanol precipitation

DNA was precipitated by adding 2 volumes of 96% ethanol. The mixture was gently inverted and microfuged for 20 min at 4°C. The supernatant was decanted and the tube was blotted on paper towel, followed by vacuum drying for 20 min. The DNA was resuspended in an appropriate volume of dH_2O or 4 mM Tris-HCl.

2.9.3 DNA digestions with restriction endonucleases

Restriction endonucleases were obtained from Fermentas, Roche, and New England Biolabs and used according to their instructions. An appropriate volume of 10× digestion buffer was added to the DNA, tapped to mix, and microfuged for 2 sec. For double digestions, a common buffer in which both enzymes showed suitable activity was chosen. Then, 0.5 µl of restriction endonuclease was added, tapped to mix, and microfuged for 2 sec. The digestion was incubated at the correct temperature for > 3 hr.

If necessary, the endonuclease was inactivated by treatment with either heat or phenol. If the enzyme could be heat-inactivated, the manufacturer's instruction was followed for the correct temperature and incubation time. Otherwise, phenol-chloroform extraction was conducted.

2.9.4 Alkaline phosphatase treatment

Shrimp alkaline phosphatase (Roche) was used to prevent recircularization of vector. An appropriate volume of 10× dephosphorylation buffer and 1 µl of alkaline phosphatase were added and incubated at 37°C for > 1 hr. The enzyme was then heat-inactivated at 65°C for 20 min.

2.9.5 DNA ligation

T4 DNA ligase (Fermentas) was used for ligation reactions. Vector and insert DNA were added, and the volume was made up to 17 µl with dH_2O. 2 µl of 10× ligation buffer was added, tapped to mix, and microfuged for 2 sec. Thereafter, 1 µl of ligase

was added, tapped to mix, microfuged for 2 sec, and incubated at 22°C overnight. The ligase was then inactivated at 65°C for 10 min.

2.9.6 Freeze-squeeze method of DNA extraction from agarose gels

Piece of agarose containing the DNA of interest was placed in a microfuged tube, crushed with a spatula, and frozen at -70°C for 30 min. The agarose was then completely thawed at room temperature, followed by microfuging for 6 min after which the supernatant was collected. The pellet was crushed again, and the process was repeated. The supernatants from the first and second rounds were pooled and subjected to salt and ethanol precipitation, the pellet washed once with ethanol, vacuum dried, and resuspended in 50 µl of dH_2O.

2.9.7 Determination of DNA concentration

In order to measure the concentration of DNA solution, an aliquot was first run on a gel, which was then quantified using UVP LabWorks Image Acquisition and Analysis Software (Ver.4.5) by comparing band intensity between the DNA and the molecular weight marker, for which the concentration had been predetermined by the manufacturers.

2.10 Gel electrophoresis

2.10.1 Agarose gel electrophoresis

Agarose stock solutions were prepared in 0.5× TBE buffer at concentrations of 0.4%, 0.8%, and 1.0% depending on the size of the DNA fragment to be separated. A gel

was prepared by pouring 20 ml of agarose solution containing 1 μg/ml EtBr into a gel tray inserted with a comb, which was then polymerized at 4°C for 30 min. After removing the comb, the tray containing the gel was placed in an electrophoresis unit containing 0.5× TBE buffer with 1 μg/ml EtBr. After DNA samples were mixed with bromophenol blue tracking dye and loaded into the wells, electrophoresis was conducted at 80-120 V at room temperature. The gel was viewed and photographed using UVP BioDoc-It System. Molecular weight markers used were Roche Lambda *Hind*III + *Eco*RI Marker III or Fermentas GeneRuler DNA Ladder Mix.

2.10.2 Low gelling agarose gel electrophoresis

Prior to usage, 20 ml of 1% low gelling agarose in 0.5× TBE buffer was autoclaved. EtBr was added to a final concentration of 1 μg/ml, and the agarose was poured into a tray inserted with a comb and polymerized at 4°C for 30 min. The gel was placed in a pre-cooled electrophoresis unit containing 0.5× TBE with 1 μg/ml EtBr. After DNA samples were mixed with bromophenol blue tracking dye and loaded into the wells, electrophoresis was conducted at 75 V at 4°C. Using a scalpel under a UV light, a band of interest was excised from the gel, from which the DNA was extracted by either phenol-chloroform extraction or freeze-squeeze method. In case of the phenol-chloroform extraction, the piece of agarose carrying the DNA was melted at 60°C for 30 min, followed by three phenol extractions. Between each extraction, 1/15 volume of 1 M Tris-HCl was added. One chloroform extraction was then conducted, followed by precipitation of the DNA.

2.10.3 Pulsed field gel electrophoresis (PFGE)

100 ml of 1% agarose in 0.5× TBE buffer was poured into a PFGE gel tray inserted with a comb and polymerized at 4°C for 30 min. A plug of molecular weight marker

(New England Biolabs PFG Low Range Marker) embedded in agarose was cut with a scalpel to a length of 1 mm, placed inside a well, and the remaining space inside the well was filled with 1% agarose solution in 0.5× TBE buffer, which was then polymerized at room temperature for 5 min. The electrophoresis unit used was BioRad CHEF-DR II Pulsed Field Electrophoresis System. The gel was placed in the unit containing 2000 ml of 0.5× TBE, and DNA samples were mixed with bromophenol blue tracking dye and loaded into the wells. Initial and final switch times were set at 1 and 12 sec, respectively. Electrophoresis was conducted at 6 V/cm at 4°C for 18.5 hr. The buffer circulation pump was set at 750 ml/min and switched on 10 min after the start of the electrophoretic run. Following electrophoresis, the gel was stained in 0.5 µg/ml EtBr solution for 20 min, destained in dH_2O for 3 hr, and viewed.

2.11 Transformations

2.11.1 *E. coli* $CaCl_2$ transformation

20 ml of pre-warmed LB containing 0.5% glucose was inoculated with 200 µl of an overnight culture of *E. coli* and incubated at 37°C with vigorous shaking for 1.75 hr. The culture, while still being shaken, was cooled in ice-water slurry for 5 min and centrifuged in a pre-cooled JA-20 rotor at 10000 rpm for 5 min at 4°C. The supernatant was decanted, and the pellet was resuspended in 10 ml of pre-cooled $CaCl_2$ transformation buffer and placed on ice for 15 min, followed by centrifugation at 10000 rpm for 5 min at 4°C. The supernatant was decanted, and the pellet was resuspended in 1.33 ml of pre-cooled transformation buffer. The cells were left on ice for 2-24 hr.

100 µl of competent cells and 5 µl of plasmid were added together, mixed by bubbling air with a pipette, and left on ice for 15 min. They were heat-shocked at

42°C for 90 sec, after which 0.5 ml of pre-warmed LB was added and incubated at 37°C for 1 hr. The cells were spread on an LA plate containing 200 µg/ml ampicillin and incubated at 37°C overnight.

2.11.2 *Rhodococcus* PEG-mediated transformation

R. erythropolis SQ1 was grown in 5 ml of LBSG (2% glycine) at 30°C for 2 days. Cells from 1 ml of this culture were washed once in 1 ml of B buffer, resuspended in 1 ml of B buffer containing freshly added lysozyme (5 mg/ml), and incubated at 37°C for 1.5 hr with inversion every 10 min. During the incubation, 2 ml of P buffer was prepared. P-PEG buffer was also prepared by UV-sterilizing 0.5 g of polyethylene glycol (PEG 6000) for 10 min and dissolving it in 1 ml of P-buffer by vigorous vortexing.

The cells were microfuged for 10 sec, gently washed in 1 ml of B buffer, and resuspended in 500 µl of P buffer. 25 µl of the protoplast suspension and 5 µl of plasmid were added together, mixed by bubbling air with a pipette, and left for 10 min. An equal volume of P-PEG buffer was added, mixed by bubbling, and spotted onto a pre-cooled regeneration plate. The plate was incubated at 28°C for 12 hr, underlaid with 250 µl of 4 mg/ml chloramphenicol, and further incubated at 28°C for 5 days.

2.12 DNA sequencing and analysis

All sequenced DNA were prepared by the *E. coli* plasmid miniprep method with a phenol-chloroform extraction included prior to isopropanol extraction. Sequencing was conducted by Inqaba Biotec. Sequences were edited and analyzed using the Internet programs listed in Table 2.5.

Table 2.5 Internet addresses used for sequence analysis.

Program	**Address**
BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
FASTA	http://www.ebi.ac.uk/fasta33/
GeneMark	http://opal.biology.gatech.edu/GeneMark/
FramePlot	http://www.nih.go.jp/~jun/research/frameplot/

3. RESULTS

3.1 Phage isolation and preliminary characterization

3.1.1 Isolation of nocardiophages from soil

To isolate nocardiophages, 1 g of soil samples were agitated gently in broth for 16 hr. Based on previous observations (Shibayama, 2004) that longer agitation increased phage yield, it was hypothesized that most phages in soil exist as lysogens, whose induction was expected by the favourable growth conditions in rich medium.

Following agitation and centrifugation, soil supernatants were assayed for nocardiophages that plaqued on *R. erythropolis* SQ1. There were two reasons for using this species. First, there was no health hazard associated with this non-pathogenic species. Secondly, it formed a confluent lawn in soft agar facilitating detection of plaques. Sources of soil samples and their phage yield are shown in Table 3.1. A representative assay plate is shown in Figure 3.1.

Table 3.1 Assayed soil samples and their phage yield.

Soil source	**Plaques/plate**
Fairland, Johannesburg, South Africa	> 350
Wits, Johannesburg, South Africa	223
Emmarentia, Johannesburg, South Africa	115
Rain forest, Mpumalanga, South Africa	57
Augrabies Falls, Northern Cape, South Africa	0
Windhoek, Namibia	22
Kadishi, Namibia	9
Fish River Canyon, Namibia	4
Lüderitz, Namibia	0

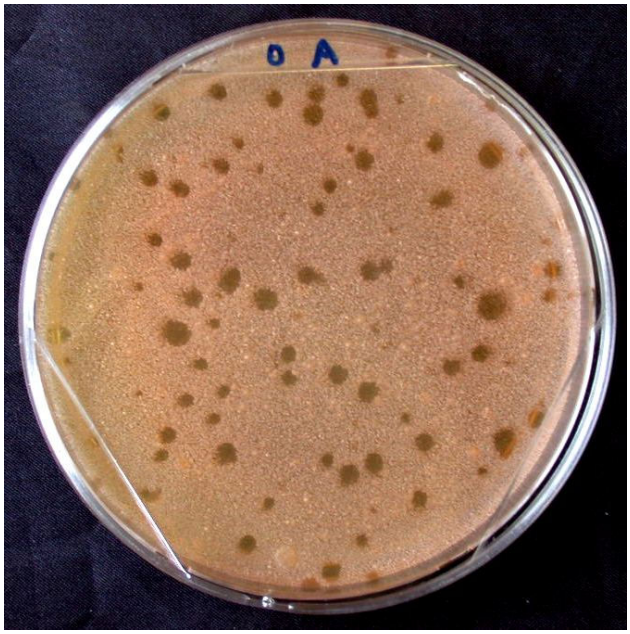


Figure 3.1 Phage assay plate for Emmarentia soil.

Plaques of different sizes and morphologies were present, suggesting presence of different phages. Individual plaques were picked and resuspended in 100 µl of LB. Phages were subjected to three rounds of single plaque purification in order to assure their purity.

3.1.2 Divalent ion requirement

Phage divalent ion requirements were determined by checking for their ability to plaque on *R. erythropolis* SQ1 on agar plates supplemented with no divalent ion, Ca^{++}, or Mg^{++}. Three rounds of this experiment were conducted in order to eliminate the possibility of ion carry-over from the original plate on which the phage was isolated. Divalent ion requirements are shown in Table 3.2. Phages were named by their soil source followed by a number.

Phages Kanazawa1 and Perouges1 were isolated by Dabbs (personal communication), from soil from Japan and France, respectively. Phage Fairland1 was isolated in my previous Honours work (Shibayama, 2004). These three phages were also single plaque purified, and their divalent ion requirements tested.

Table 3.2 Phage divalent ion requirements.

Phage	Divalent ion requirement
Kanazawa1	none
Perouges1	none
Fairland1	none
Fairland2	none
Fairland3	none
Wits1	none
Wits2	none
Emmarentia1	none
Rain Forest1	Ca^{++} or Mg^{++}
Rain Forest2	Ca^{++} or Mg^{++}
Windhoek1	none
Windhoek2	none
Kadishi1	none
Fish River Canyon1	none

3.1.3 Host range

From Table 3.2, four phages were selected for further work – Fairland1, Kanazawa1, Perouges1, and Wits1. They were selected to encompass the diversity in location of their origin, although there is usually no obvious relationship between geographical origin and DNA sequence similarity (Pedulla *et al*., 2003). Phage Wits1 was selected because its relatively large plaques suggested a highly lytic potential. Host ranges of these phages are shown in Table 3.3.

Table 3.3 Host ranges of Fairland1, Wits1, Kanazawa1, and Perouges1.

	Phage			
Host strain	**Fairland1**	**Wits1**	**Kanazawa1**	**Perouges1**
R. erythropolis SQ1	p	p	p	p
R. erythropolis ATCC 4277	p	p	p	p
R. erythropolis DSM 1069	p		p	
R. rhodochrous Ri8	p	p	p	p
R. opacus HL PM-1	p	c	p	
R. equi ATCC 14487	p		p	p
G. australis A554				
G. rubropertincta ATCC 25593				c
G. desulfuricans NB4				
N. mexicana IFO 3927	c		c	
M. smegmatis $mc^2$1155				
M. parafortuitum 490				

p = plaqued; c = cleared at high multiplicity of infection but could not plaque;
where nothing is indicated the phage had no visible effect.

Table 3.3 shows that all four phages could only plaque on *Rhodococcus* species. Fairland1 and Kanazawa1 had identical host range, while those of Wits1 and Perouges1 clearly indicated that these two were different phages from the others.

3.1.4 Efficiency of plaquing

These four phages were propagated *on R. erythropolis* SQ1 and tested for their efficiency of plaquing on the other strains. This was done by taking the ratio of the titre on the other strain to the titre on *R. erythropolis* SQ1 (Table 3.4).

Table 3.4 Efficiency of plaquing of Fairland1, Wits1, Kanazawa1, and Perouges1, propagated on *R. erythropolis* SQ1, on other strains.

Host strain	**Phage**			
	Fairland1	**Wits1**	**Kanazawa1**	**Perouges1**
R. erythropolis SQ1	1.0	1.0	1.0	1.0
R. erythropolis ATCC 4277	1.2	0.7	0.7	4.0
R. erythropolis DSM 1069	0.002		0.001	
R. rhodochrous Ri8	0.8	1.0	2.0	0.2
R. opacus HL PM-1	0.1		0.005	
R. equi ATCC 14487	1.0		0.04	0.01

Where nothing is indicated, the phage could not plaque on the strain.

3.1.5 Cross-checking of phages using lysogens or phage resistant mutants

Lysogen or phage resistant mutant isolation was attempted for each phage by selecting cells inside turbid plaques. The host ranges and efficiencies of plaquing indicated that each phage was different from another. As an additional measure to confirm this, they were cross-checked using phage resistant cells.

No stable lysogens or resistant mutants for Fairland1 and Perouges1 were obtained. whereas cells resistant to Kanazawa1 and Perouges1 were successfully obtained. Table 3.5 shows the plaquing of the four phages on these cells.

Table 3.5 Plaquing of Fairland1, Wits1, Kanazawa1, and Perouges1 on lysogens/resistant mutants of Wits1 and Kanazawa1.

Lysogen / resistant mutant of phage	Phage			
	Fairland1	**Wits1**	**Kanazawa1**	**Perouges1**
Wits1	p		p	p
Kanazawa1	p	p		p

p = plaqued; where nothing is indicated the phage had no visible effect.

These experiments confirmed that all four phages were different.

3.2 Production of lysates and phage purification

To extract a sufficient amount of phage DNA, production of a large volume of a high titre lysate and its purification were necessary. High titre lysates were first produced on plates, which yielded a volume of 3 to 4 ml. Thereafter, the lysates were scaled up in broth. A small scale liquid lysate of 10 ml was first produced and then scaled up to ≥ 500 ml.

3.2.1 Plate lysates

High titre plate lysates were produced by mixing *R. erythropolis* SQ1 and the phage at appropriate concentrations and overlaying them on an agar plate. For each phage, the optimum condition is shown in Table 3.6.

Table 3.6 Conditions used to attain high titre plate lysates.

Phage	Amount of cells (cfu)	Amount of phages (pfu)	MOI	Lysate titre (pfu/ml)
Fairland1	1.5×10^8	2.0×10^6	0.01	4.8×10^{10}
Wits1	1.5×10^8	6.0×10^5	0.004	2.0×10^9
Kanazawa1	1.5×10^7	4.0×10^4	0.003	5.5×10^9
Perouges1	1.5×10^7	4.0×10^5	0.03	1.0×10^9

3.2.2 Liquid lysates

Small scale liquid lysates were produced by mixing *R. erythropolis* SQ1 and the phage at appropriate concentrations in 10 ml broth and gently agitating them on a rotary shaker. Optimum conditions for attaining high titres are shown in Table 3.7.

Table 3.7 Conditions used to attain high titres in 10 ml liquid lysates.

Phage	Amount of cells (cfu)	Amount of phages (pfu)	MOI	Lysate titre (pfu/ml)
Fairland1	7.5×10^8	3.0×10^7	0.04	7.0×10^{10}
Wits1	3.0×10^8	2.0×10^4	0.00007	5.0×10^9
Kanazawa1	3.0×10^8	5.5×10^8	1.8	4.0×10^7
Perouges1	3.0×10^8	1.0×10^8	0.3	4.0×10^7

For Kanazawa1 and Perouges1, maximum titres were 4.0×10^7 pfu/ml so large scale lysate production was not attempted. For Fairland1 and Wits1, conditions used to produce the small scale lysates were applied to produce large scale lysates of 500 ml; more host cells and phages were added, but at an identical MOI. Conditions used for the production of large scale lysates are shown in Table 3.8.

Table 3.8 Conditions used to attain high titres in 500 ml liquid lysates for Fairland1 and Wits1.

Phage	Amount of cells added (cfu)	Amount of phages added (pfu)	MOI	Lysate titre (pfu/ml)
Fairland1	3.75×10^{10}	1.5×10^{9}	0.04	7.0×10^{10}
Wits1	1.5×10^{10}	1.0×10^{6}	0.00007	9.0×10^{9}

3.2.3 Phage purification

Phages from large scale lysates were precipitated by adding NaCl and PEG, resuspended in a smaller volume, and purified by CsCl gradient centrifugation. Phage purification may be done in either a step gradient or an equilibrium gradient of CsCl. When using a step gradient, the gradient was manually loaded into a tube prior to centrifugation, forcing the phages to migrate to a position corresponding to their buoyant density. When using an equilibrium gradient, the gradient was formed by the centrifugal force.

Fairland1 was purified in a step gradient with densities of 1.70, 1.50, 1.45 and 1.15 g/ml (Sambrook *et al*., 1989). Following centrifugation, a bluish mass was visible at the bottom of the tube, indicating that the density of phage Fairland1 is ≥ 1.70 g/ml. Wits1 was purified in an equilibrium gradient with an initial density of 1.50 g/ml. A bluish band of this phage was visible near the middle of the tube, indicating that its density is ~1.50 g/ml. Post-purification titres and phage recovery are shown in Table 3.9.

Table 3.9 Titres and volumes before and after purification and phage recovery.

Phage	Pre-purification		Post-purification		Recovery (%)
	Titre (pfu/ml)	Volume (ml)	Titre (pfu/ml)	Volume (ml)	
Fairland1	7.0×10^{10}	500	4.8×10^{13}	0.5	68.6
Wits1	9.0×10^{9}	500	6.0×10^{11}	0.5	6.7

3.3 Phage morphology

Phage morphology was observed by viewing negatively stained phages under the transmission electron microscope. Electron micrographs of Fairland1 are shown in Figure 1.1, as they were obtained in my previous work (Shibayama, 2004). Head diameter, tail diameter, and tail length of the four phages were estimated from the micrographs. These are shown in Table 3.10.

3.3.1 Wits1 morphology

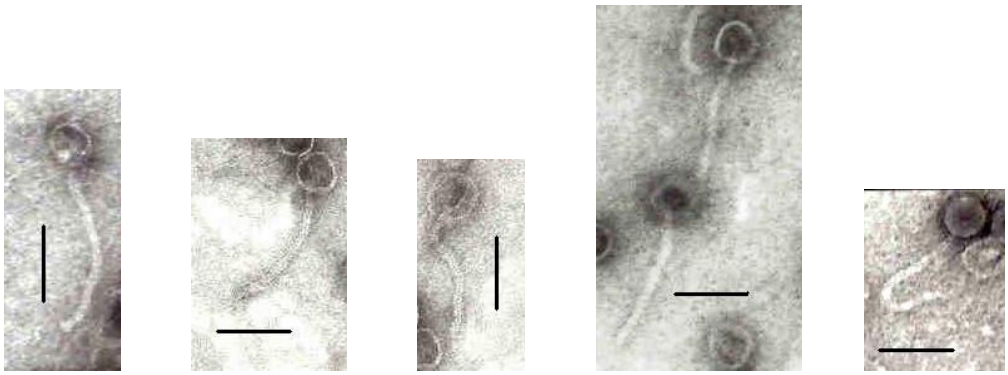
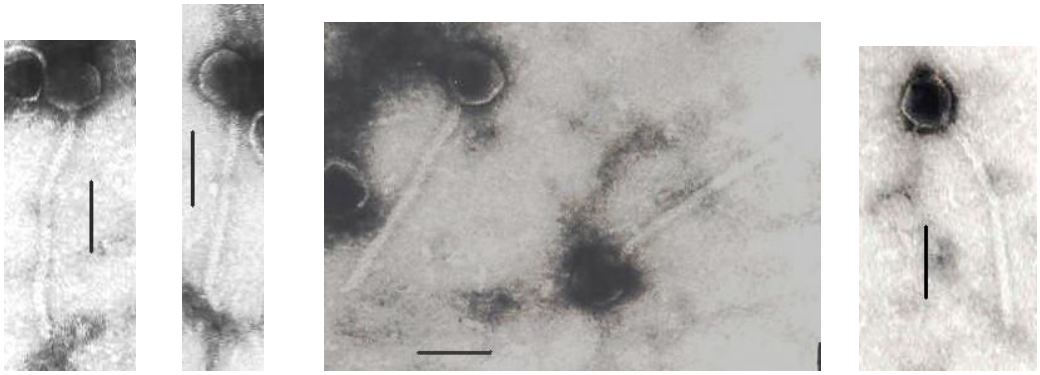


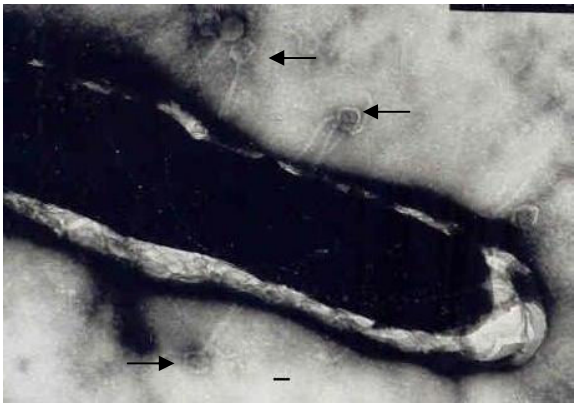
Figure 3.2 Electron micrographs of Wits1.
Bar = 100 nm. All micrographs were taken at ×100,000 magnification

According to morphology, Wits1 was assigned to the family *Siphoviridae*, characterized by their long non-contractile tails.

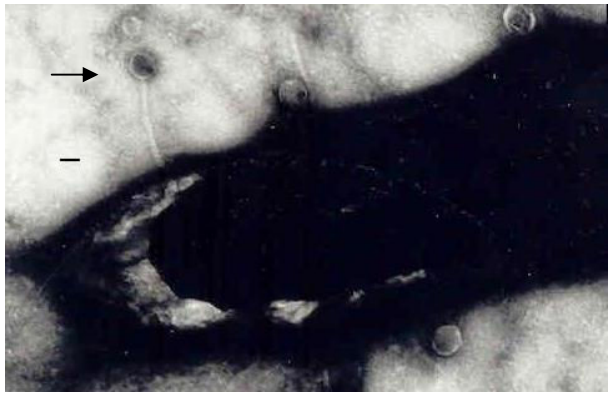
3.3.2 Kanazawa1 morphology



A **B** **C** **D**



E



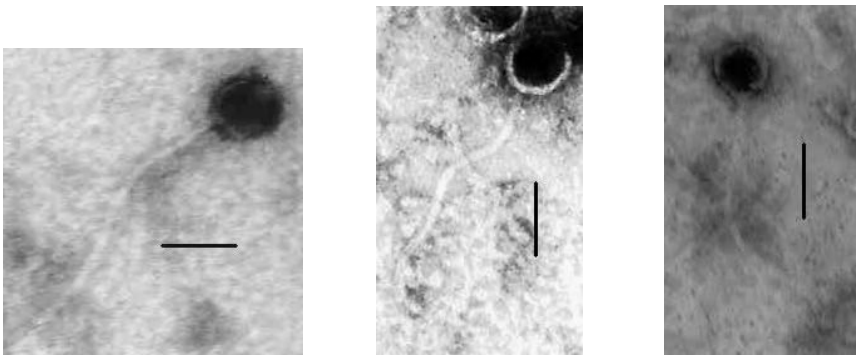
F

Figure 3.3 Electron micrographs of Kanazawa1.

Bar = 100 nm. Micrographs A – D were taken at ×100,000 magnification. Micrographs E and F were taken at ×40,000 and ×50,000 magnifications, respectively. In E and F, arrows point to phages which are adsorbed to the host cell.

Kanazawa1 was also assigned to the family *Siphoviridae*.

3.3.3 Perouges1 morphology



A **B** **C**

Figure 3.4 Electron micrographs of Perouges1.

Bar = 100 nm. All micrographs were taken at ×100,000 magnification.

Perouges1 was also assigned to the family *Siphoviridae*.

3.3.4 Summary of phage morphologies

Table 3.10 Summary of phage morphologies.

Phage	Family	Head diameter (nm)	Tail diameter (nm)	Tail length (nm)
Fairland1	*Siphoviridae*	60	14	280
Wits1		60	10	250
Kanazawa1		80	20	330
Perouges1		85	15	380

3.4 Phage genome size

Fairland1 and Wits1 genomic DNA were extracted from purified phage particles. Kanazawa1 and Perouges1 DNA were extracted from plate lysates. Gels from pulsed field electrophoresis of these DNAs are shown in Figure 3.5. Phage genome sizes are shown in Table 3.11.

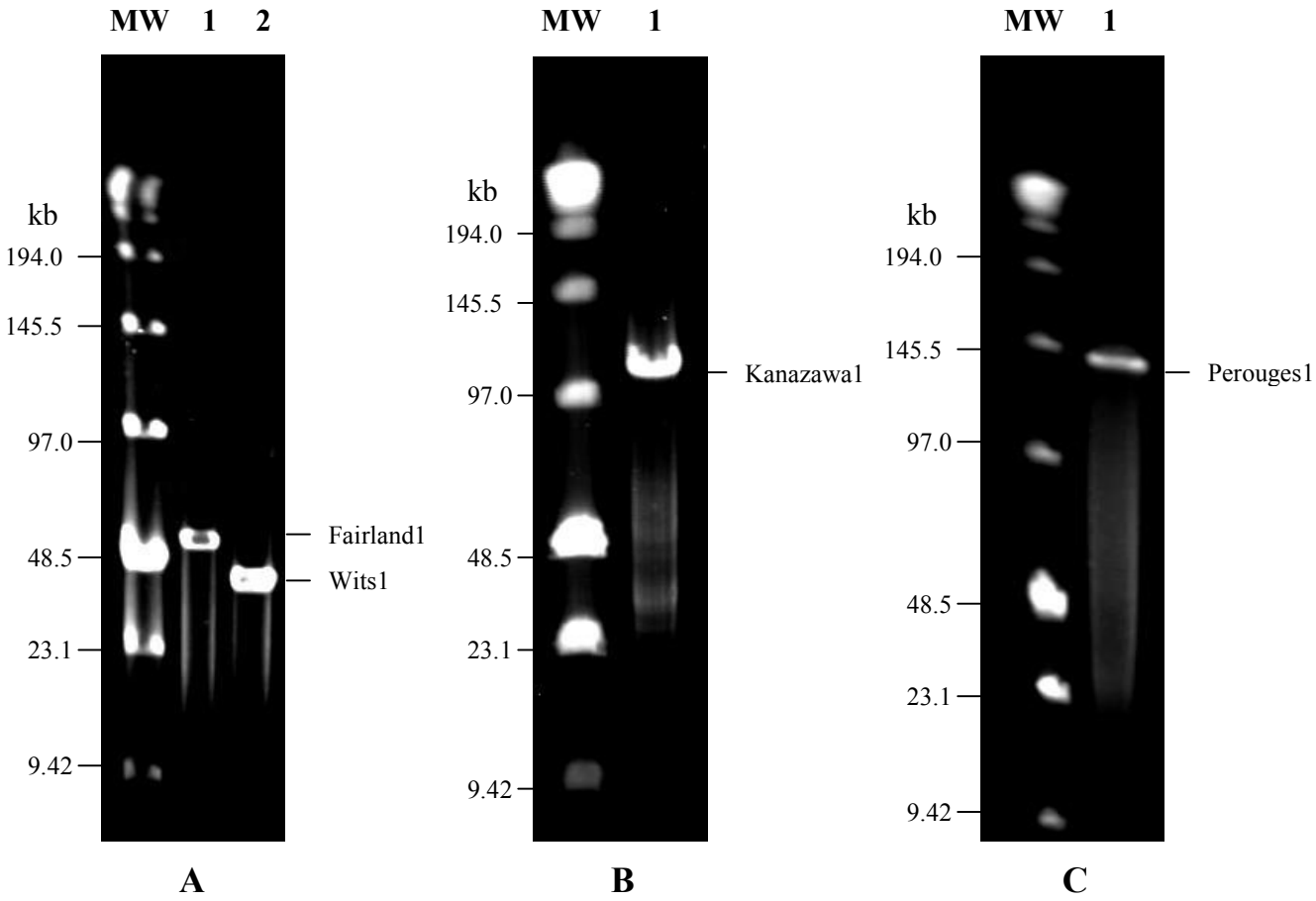


Figure 3.5 Pulsed field gel electrophoresis of phage DNA on 1% agarose gels.

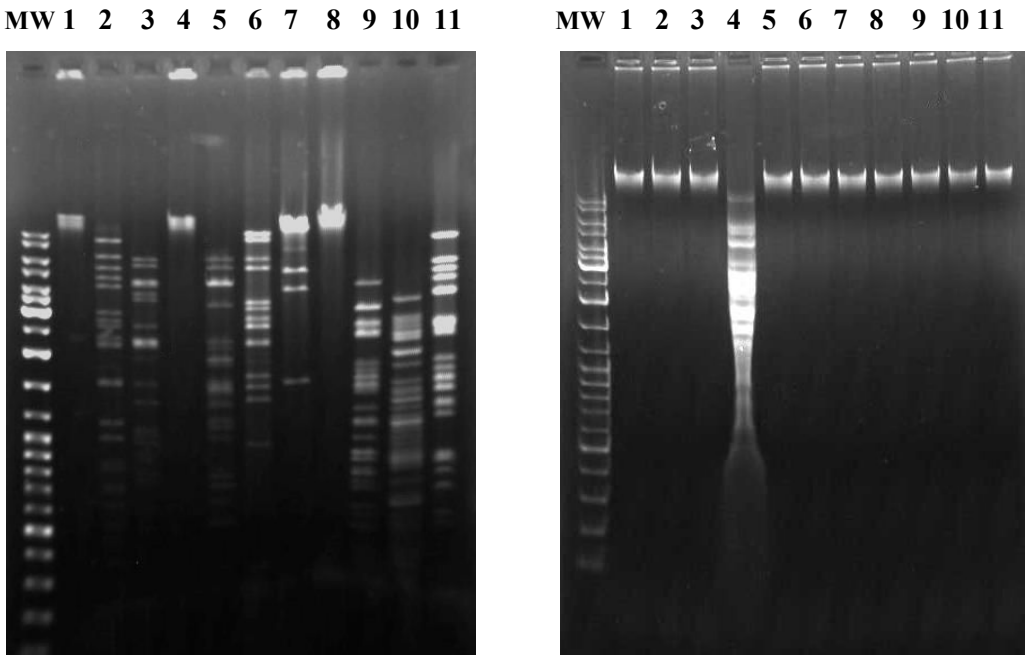
Table 3.11 Phage genome sizes.

	Fairland1	**Wits1**	**Kanazawa1**	**Perouges1**
Genome size (kb)	55	39	106	140

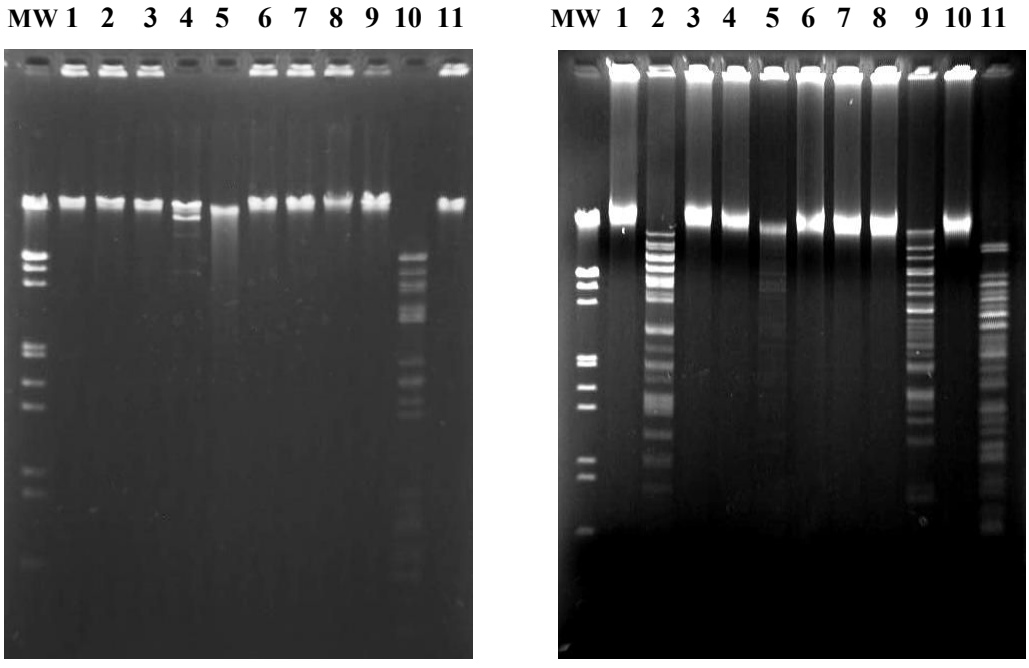
3.5 Restriction digestion of phage genomic DNA

3.5.1 Endonucleases screening for library construction

To select restriction endonucleases for phage genomic library construction, their DNA were subjected to digestion by those with unique recognition sites in the suicide (*Eco*RI) gene of the *E.coli – Rhodococcus* shuttle vector pDA71. These enzymes were *Hind*III, *Bgl*II, *Pst*I, and *Sfu*I. Various isocaudamers of these endonucleases were also tested. These digestions are shown in Figure 3.6. As a control, *E. coli* GM2929 and *R. erythropolis* SQ1 genomic DNA were subjected to digestion with the same endonucleases. All were able to digest the DNA from both sources (data not shown).



A. Fairland1 **B. Kanazawa1**



C. Wits1 **D. Perouges1**

Figure 3.6 Digestions of phage genomic DNA on 0.8% agarose gels.

Lane number corresponds to the same enzyme in all gels. Lane 1, undigested DNA; 2, *Hind*III; 3, *Bgl*II; 4, *Bam*HI; 5, *Bcl*I; 6, *Pst*I; 7, *Nsi*I; 8, *Bmy*I; 9, *Sfu*I; 10, *Acc*I; 11, *Cla*I.

3.5.2 Further digestion of Wits1 DNA

Many endonucleases used above did not digest the DNA from Wits1, Kanazawa1, and Perouges1. Wits1 DNA was therefore subjected to digestions by other enzymes (Huang *et al*., 1982) to observe whether a pattern in recognition sequences existed for those which did not cut the DNA (Table 3.12).

Table 3.12 Digestion of Wits1 DNA.

Enzyme	
*Alu*I	+
*Cfo*I	+
*Dde*I	
*Hae*III	+
*Hinf*I	+
*Hpa*I	+
*Sau*3A	+
*Dpn*I	
*Sau*96I	+
*Taq*I	+
*Bam*HI	+
*Bcl*I	

Enzyme	
*Bgl*II	
*Xho*II	+
*Pvu*I	+
*Bst*EII	+
*Bst*NI	+
*Eco*RI	
*Hind*III	
*Hpa*I	
*Xba*I	
*Sac*I	
*Cfr*42I	

+ = digested; where nothing is indicated the enzyme did not digest the DNA.

3.6 Phage genomic library construction

Construction of genomic library was attempted for all four phages. All libraries were constructed in *E. coli* MM294-4 using pDA71. Twenty clones from each library were screened for their insert sizes.

3.6.1 Fairland1 library

Two libraries – one with *Bgl*II and another with *PstI* – were constructed for Fairland1.

3.6.1.1 Fairland1 *Bgl*II library

Plasmid mini preparations were done on randomly chosen twenty clones, which are shown in Figure 3.7.

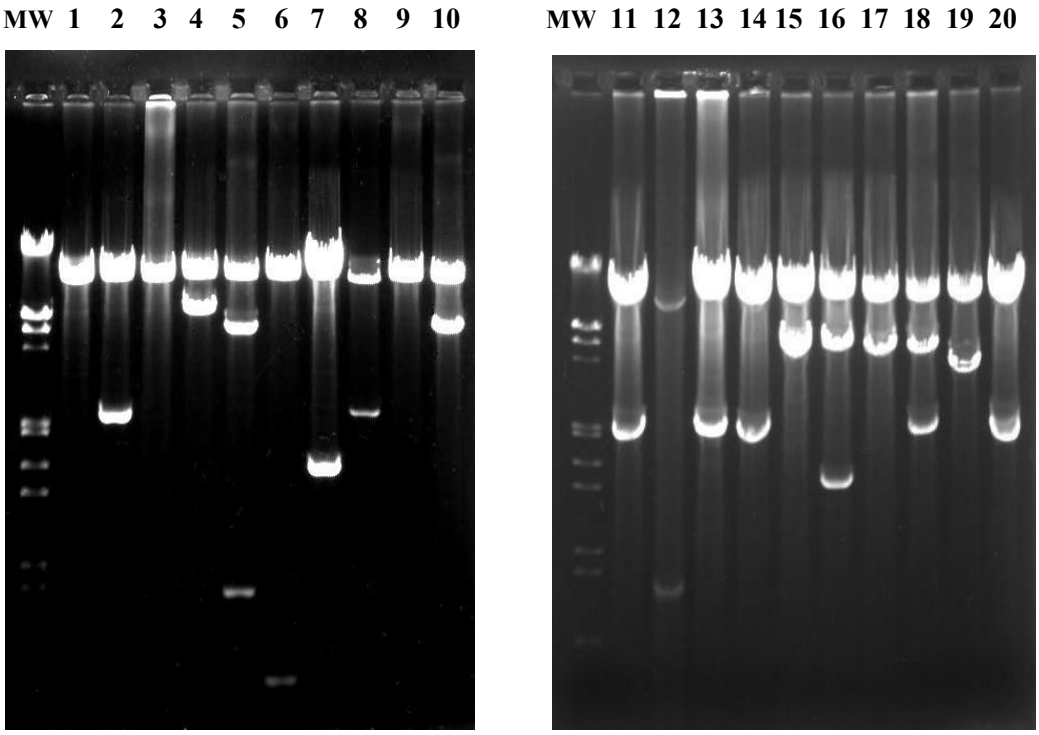


Figure 3.7 Twenty clones from Fairland1 *Bgl*II library, on 0.8% agarose gels. Inserts were released by *Bgl*II digestion. The upper band was linearized pDA71, and the lower bands were inserts.

To confirm heterogeneity of these twenty clones, they were also subjected to a double digestion by *Bgl*II and *Pst*I (data not shown). This revealed that clones [11 and 20] and [13 and 14] may have identical inserts (clone numbers correspond to the lanes in the gel in which they were run).

Average insert size from this library was ~2.5 kb. As shown below, number of transformants necessary to have 99% probability of any Fairland1 DNA sequence present in the library was calculated to be 99, and ~140 transformants were obtained.

$$N = \frac{\ln (1 - p)}{\ln (1 - a/b)}$$

$$= \frac{\ln (1 - 0.99)}{\ln (1 - 2.5/55)}$$

= 99 clones

where N = number of clones required; p = probability of any given DNA being present in the library; a = average insert size; b = genome size

3.6.1.2 Fairland1 *Pst*I library

Fairland1 genomic DNA was completely digested with *Pst*I for another library construction. Plasmid mini preparations were done on randomly chosen twenty clones, which are shown in Figure 3.8.

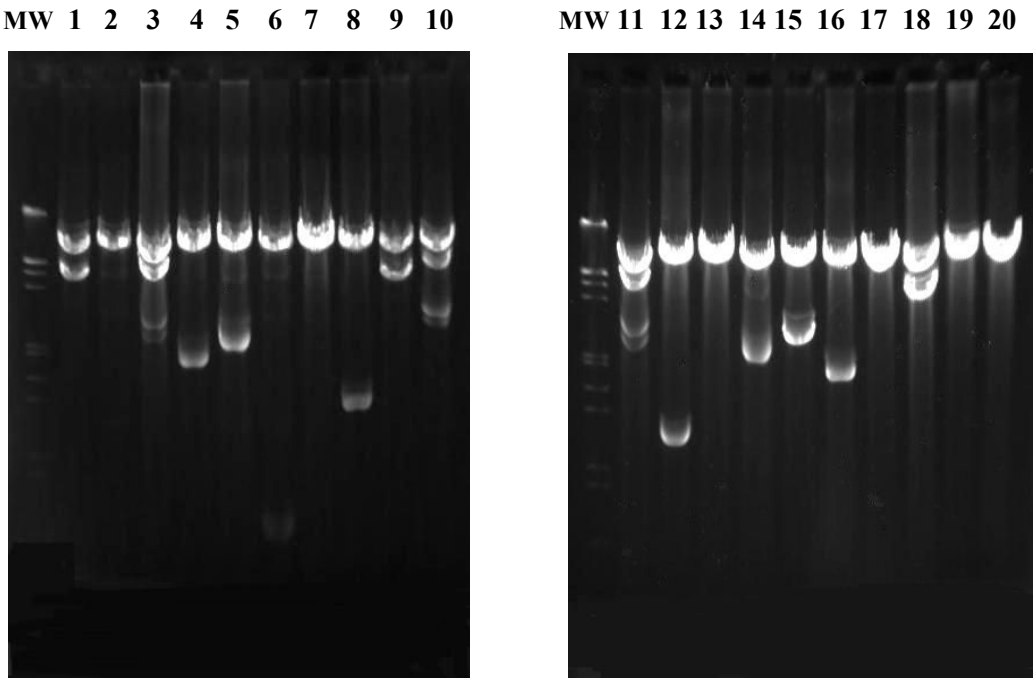


Figure 3.8 Twenty clones from Fairland1 *Pst*I library, on 0.8% agarose gels. Inserts were released by *Pst*I digestion.

To confirm heterogeneity of these twenty clones, they were also subjected to a double digestion by *Pst*I and *Bgl*II (data not shown). This revealed that clones [1 and 9] and [3, 10, and 11] may have identical inserts.

Average insert size from this library was ~2.7 kb. Number of transformants necessary to have 99% probability of any Fairland1 DNA sequence present in the library was calculated to be 92, and ~95 transformants were obtained.

3.6.2 Wits1 library

Wits1 library construction was attempted by *Xho*II (isocaudamer of *Bgl*II) partial digestion of its genome to generate fragments of > 2 kb (Figure 3.9A). Plasmid minipreps were done on ten randomly chosen clones from the library. However, these clones mostly possessed inserts of approximately 100 bp, as shown in Figure 3.9B. DNA partial digestion and library construction were repeated twice, with the same outcome, so these experiments were not continued.

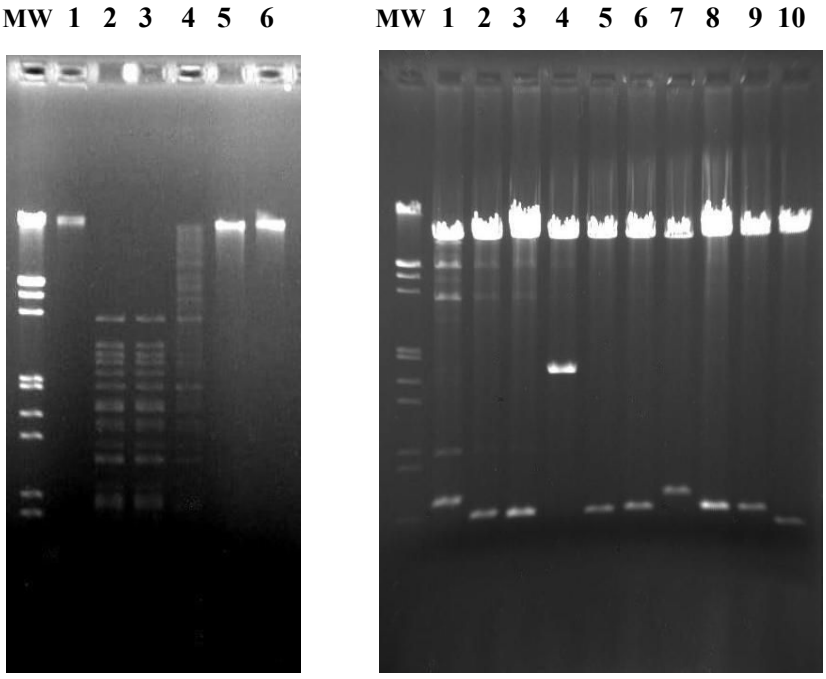


Figure 3.9 Wits1 genomic library construction with *Xho*II, on 0.8% agarose.

A. Partial digestion of Wits1 genomic DNA with *Xho*II. DNA digested with 1/16 diluted enzyme (lane 4) was used for the library construction.

Lane 1, undigested; 2, digestion with undiluted *Xho*II; 3, 1/4 dilution; 4, 1/16 dilution; 5, 1/64 dilution; 6, 1/256 dilution.

B. Ten clones from Wits1 *Xho*II library.

Inserts were released by double digestion with *Hind*III and *Pst*I. All inserts included a part of the vector between the *Hind*III and *Pst*I sites, which was ~420 bp. Therefore, the actual inserts were approximately 100 bp.

3.6.3 Perouges1 libraries

Two Perouges1 libraires were constructed, one by complete digestion with HindIII and another with partial digestion with same enzyme. The partial digestion was done because the library from the complete digestion did not contain many antimicrobial clones (reported below, in Table 3.15).

3.6.3.1 Perouges1 *Hind*III (complete digest) library

From Perouges1 *Hind*III (complete digest) library, plasmid mini preparations were done on twenty randomly chosen clones (Figure 3.10).

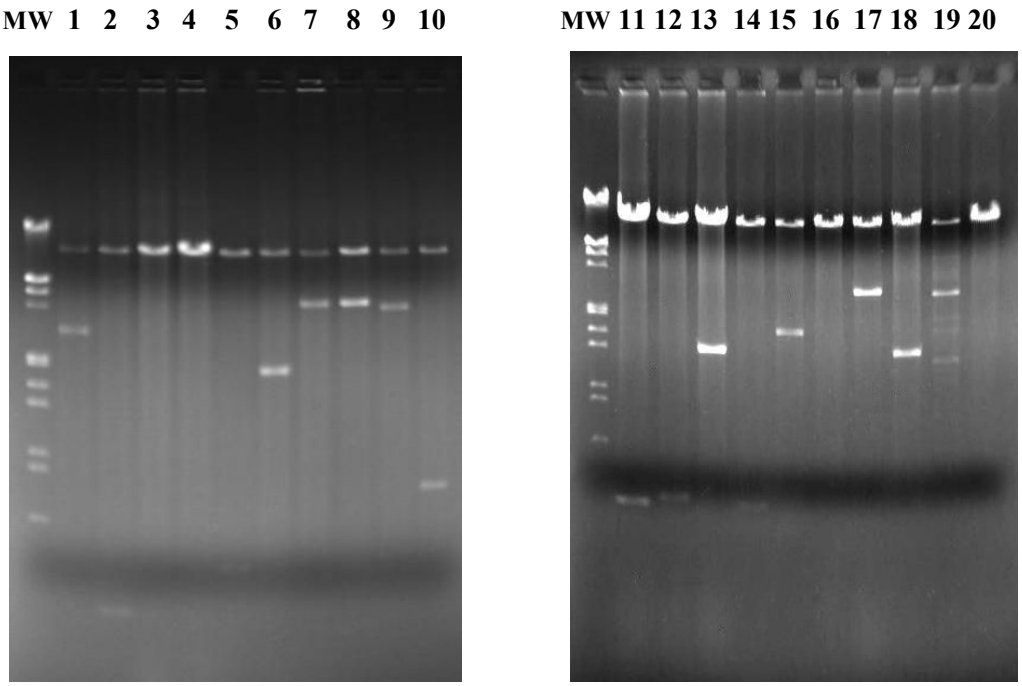


Figure 3.10 Twenty clones from Perouges1 *Hind*III (complete digest) library, on 0.8% agarose gels.
Inserts were released by *Hind*III digestion.

To confirm heterogeneity of these twenty clones, they were also subjected to a double digestion by *Pst*I and *Cla*I (data not shown). This revealed that clones 7 and 8 apparently had identical inserts.

Average insert size from this library was ~1.3 kb. Number of transformants necessary to have 99% probability of any given Perouges1 DNA sequence present in the library was 494, and ~1145 transformants were obtained.

3.6.3.2 Perouges1 *Hind*III (partial digest) library

From Perouges1 *Hind*III (partial digest) library, plasmid mini preparations were done on randomly chosen twenty clones, which are shown in Figure 3.11.

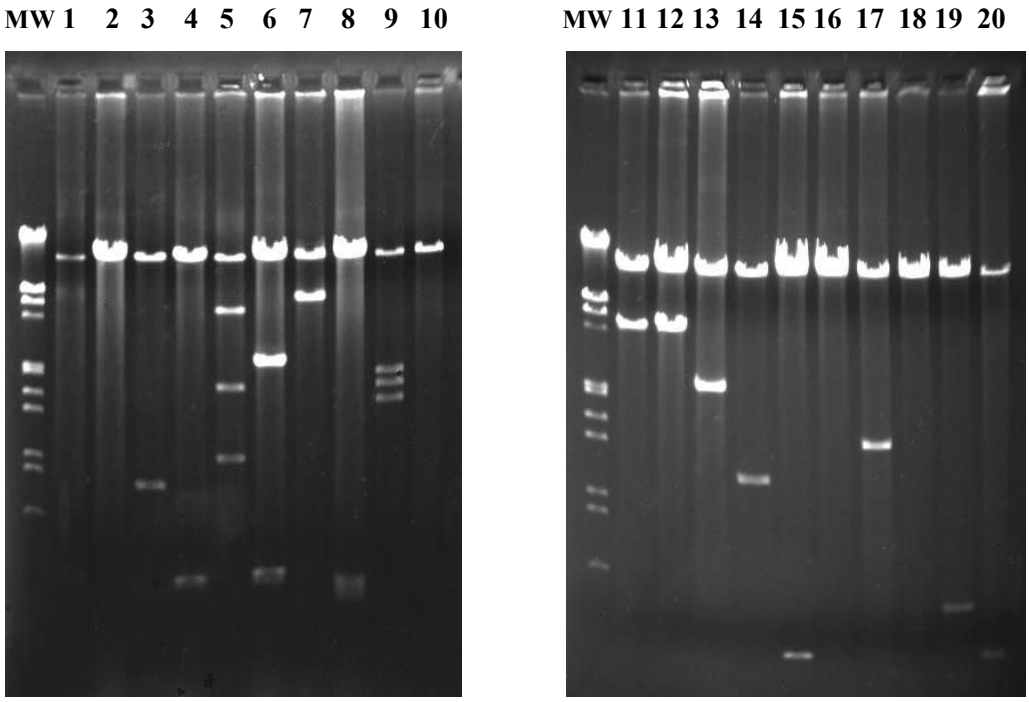


Figure 3.11 Twenty clones from Perouges1 *Hind*III (partial digest) library, on 0.8% agarose gels.
Inserts were released by *Hind*III digestion.

Average insert size from this library was ~1.6 kb. Number of transformants necessary to have 99% probability of any Perouges1 DNA sequence present in the library was 400, and ~2750 transformants were obtained.

4.6.4 Phage Kanazawa1 library

Phage Kanazawa1 library construction was attempted with *Bam*HI, since this was the only one enzyme of the ten tested that digested its DNA. However, despite several attempts, cloning of its DNA was unsuccessful.

3.6.5 Summary of phage library construction

Restriction endonucleases used for the construction of phage libraries and their outcome are shown in Table 3.13.

Table 3.13 Summary of phage library construction.

Phage	Endonuclease	Successful?
Fairland1	*Bgl*II	yes
	*Pst*I	yes
Wits1	*Xho*II	no
Kanazawa1	*Bam*HI	no
Perouges1	*Hind*III	yes

3.6.6 Further analysis on phage Wits1 genomic DNA

Upon attempting to construct Wits1 genomic library, there was a clone with an insert of ~4.5 kb. This construct was re-cloned in *E.coli* GM2929 and subjected to digestion by the eight endonucleases which could not digest the phage DNA (Figure 3.6C). None of them were able to digest the insert of the clone (data not shown), thereby strengthening the possibility of lack of restriction sites.

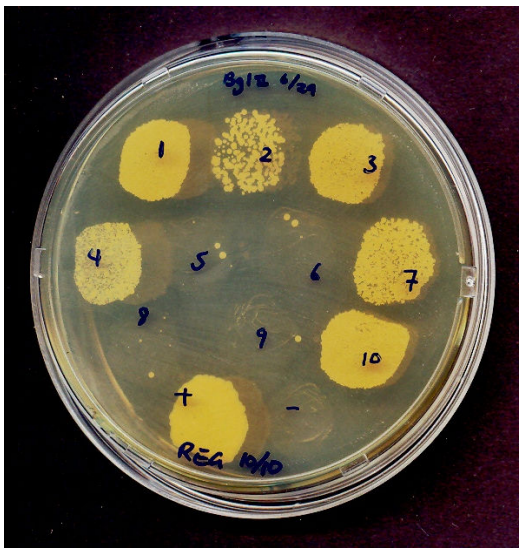
3.7 Screening for clones with antimicrobial activity

Clones from Fairland1 and Perouges1 libraries were screened for antimicrobial activity by transforming *R. erythropolis* SQ1 with each clone. A clone was identified to be inhibitory to the host if no or few transformants grew up. Vector pDA71* (suicide gene inactivated) was always included as a vector-only control. From each library, the twenty clones previously screened for insert size were used for this experiment. Ten clones were screened per agar plate.

Screening of clones from Fairland1 *Bgl*II and *Pst*I libraries is shown in Figures 3.12 and 3.13, respectively. Number of transformants/μg of DNA is shown in Table 3.14.

Screening of clones from Perouges1 *Hind*III libraries is shown in Figures 3.14 (complete digest) and 3.15 (partial digest). Number of transformants/μg of DNA is shown in Table 3.15.

A

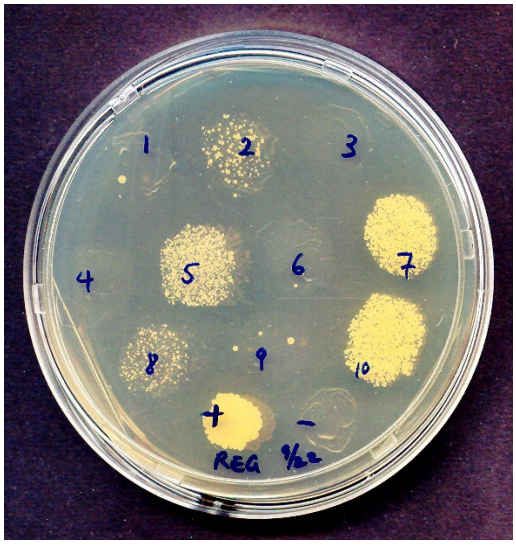


B

Figure 3.12 Screening of Fairland1 *Bgl*II library for clones inhibitory to *R. erythropolis* SQ1.

On plate **A**, spots 1 – 10 correspond to clones 1 – 10 from the library. On plate **B**, spots 1 – 10 correspond to clones 11 – 20. On both plates, "+" was the vector-only control, while "–" was the negative control with no DNA.

A



B

Figure 3.13 Screening of Fairland1 *Pst*I library for clones inhibitory to *R. erythropolis* SQ1.

On plate **A**, spots 1 – 10 correspond to clones 1 – 10 from the library. On plate **B**, spots 1 – 10 correspond to clones 11 – 20. On both plates, "+" was the vector-only control, while "–" was the negative control with no DNA.

Table 3.14 Transformants/µg of DNA for clones from Fairland1 libraries.

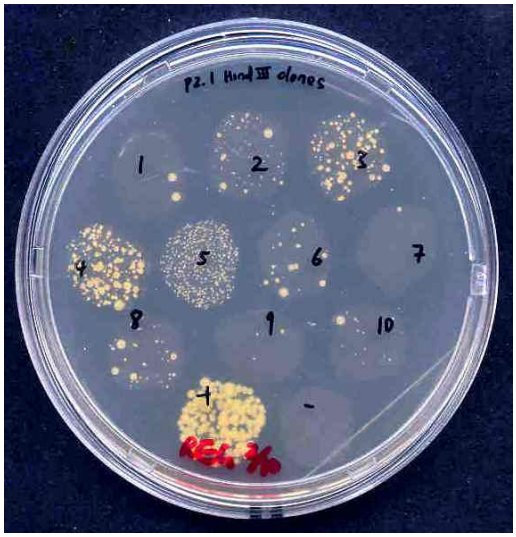
***Bgl*II library**

Clone #	# transformants ($\times 10^4$)/µg of DNA
1	0.56
2	0.27
3	0.74
4	0.06
5	0.56
6	0.46
7	0.02
8	< 0.0017
9	0.37
10	< 0.0017
11	0.83
12	0.37
13	1.01
14	0.93
15	0.006
16	0.007
17	0.85
18	0.007
19	< 0.0017
20	1.1
Vector only	2.22

***Pst*I library**

Clone #	# transformants ($\times 10^4$)/µg of DNA
1	0.03
2	0.22
3	0.63
4	0.004
5	0.12
6	0.24
7	0.22
8	0.006
9	0.009
10	0.007
11	0.007
12	0.28
13	0.006
14	< 0.0017
15	1.02
16	0.006
17	1.07
18	0.46
19	0.006
20	1.04
Vector only	2.59

Clones with transformation efficiency of < 0.01×10^4/µg of DNA are highlighted.

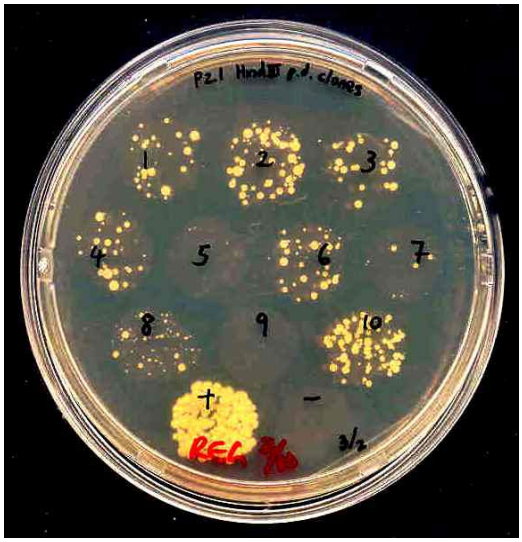


A

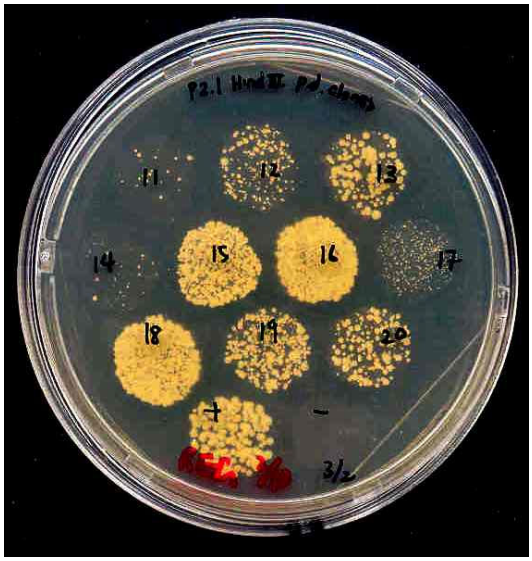
B

Figure 3.14 Screening of Perouges1 *Hind*III (complete digest) library for clones inhibitory to *R. erythropolis* SQ1.

Spots 1 – 20 correspond to clones 1 – 20. On both plates, "+" was the vector-only control, while "–" was the negative control with no DNA.



A



B

Figure 3.15 Screening of Perouges1 *Hind*III (partial digest) library for clones inhibitory to *R. erythropolis* SQ1.

Spots 1 – 20 correspond to clones 1 – 20. On both plates, "+" was the vector-only control, while "–" was the negative control with no DNA.

Table 3.15 Transformants/µg of DNA for clones from Perouges1 libraries.

***Hind*III (complete digest) library**

Clone #	# transformants ($\times 10^4$)/µg of DNA
1	0.004
2	0.08
3	0.15
4	0.19
5	0.46
6	0.03
7	0.002
8	0.07
9	0.006
10	0.02
11	0.96
12	1.2
13	0.19
14	0.28
15	0.28
16	1.85
17	0.37
18	1.3
19	0.19
20	0.99
Vector only	1.29

***Hind*III (partial digest) library**

Clone #	# transformants ($\times 10^4$)/µg of DNA
1	0.05
2	0.12
3	0.04
4	0.05
5	< 0.0017
6	0.07
7	0.009
8	0.07
9	< 0.0017
10	0.13
11	0.03
12	0.33
13	0.13
14	0.1
15	0.88
16	0.72
17	0.39
18	0.59
19	0.24
20	0.21
Vector only	1.11

Clones with transformation efficiency of $< 0.01 \times 10^4$/µg of DNA are highlighted.

3.8 Analysis of inhibitory clones

Clones with transformation efficiency of $< 0.01\times10^4/\mu g$ of DNA in *R. erythropolis* SQ1 were selected for analysis. Restriction maps of their inserts were constructed. Restriction endonucleases *Hind*III, *Bgl*II, *Pst*I, *Sfu*I, and their isocaudamers were screened, as the purpose was to later subclone them using pDA71. Those which cut the insert once or twice were used for restriction map construction by single and double digestions of the purified inserts. Orientation of inserts with respect to the vector was then identified by digestions with certain endonucleases and checking for the sizes of released fragments.

Purified inserts were digested with the appropriate enzymes and subcloned in *E. coli* MM294-4 using pDA71, and subclones screened for retention of antimicrobial activity in *R. erythropolis* SQ1. For transformations, pDA71* was always included as a vector-only control. Once the minimum inhibitory DNA was identified, this was cloned into vector pUC18 or pUC19. Sequencing was done by Inqaba Biotec.

Sequences were analyzed using BLASTn for similar nucleotide sequences, while BLASTx was used to search for similar amino acid sequences. An E value of 0.009 was used as the cutoff for similarity searches. This number is more lenient than those used by other phage studies; for example, Hatfull *et al*. (2006) and Kwan *et al*. (2006) used a cutoff E value of 10^{-4}.

ORFs were predicted on DNA sequences using FramePlot and GeneMark. Prediction of genes by FramePlot is based on an analysis between overall GC content of a prokaryotic DNA sequence and base composition at each position in potential codons (Bibb *et al*., 1984). Prediction of ORFs by GeneMark uses a heuristic method of Markov model (Besemer and Borodovsky, 1999) – it is the program of choice by many authors for predicting phage ORFs. Only ORFs predicted by both programs were included in this report. When a DNA sequence initially had a significant match

using BLAST, each predicted ORF was then translated and a similarity search was conducted using BLASTp in order to identify the specific ORF corresponding to the alignment. GC content of sequenced DNA was also determined by both programs.

In sections below, the analysis as mentioned above is reported for every selected inhibitory clone.

3.8.1 Analysis of inhibitory clones from Fairland1 libraries

From Fairland1 *Bgl*II library, clones 7, 8, 10, and 19 were selected for analysis. From the *Pst*I library, clones 8, 14, and 16 were selected.

3.8.1.1 Clone 7 of Fairland1 *Bgl*II library

3.8.1.1.1 DNA sequence

Restriction map construction and subcloning of clone 7 from Fairland1 *Bgl*II library were not conducted due to its insert size of ~1600 bp. The insert was cloned into the *Bam*HI site of pUC19 and partially sequenced. Front and back regions of the insert were sequenced, as shown in Figures 3.16 and 3.17, respectively.

ggatctgtgt ttgtgcaagg caagtatggg accgttgaaa agatgctcgt caccgagtac
acgagaactg ccgatgcaaa cggcgtcagc ggattcccaa ctctagttag atgggacaat
cctaatgaca aaatctgatg gtttgactga cttgggtcta agtggccttc tcgctccaag
tcctaagtgg aaccatatca agccctggaa tcgacacggc ctgattctac tcacagccgg
cttcgcttat attctaatcg gcatcgcttt catgctttct gaagagaccc ctactcgaag
agagaatcta caccacatgc tttcgctcat gccattcgat tggtggtgta tcggctttat
catagtcggt gttttatcgg ttctctcttc acgatggcct tcgaaaccac gaaccttggg

ttatggcgca cttaccggat ggactgctct ttggtctggc ttctatattc tcggtggttt
gtcacatccc ttgaattggg gttacgttgc tacaggcttg ggttgggctat gctaggcttt
ctttggtggg cagttagcgg actaatctgt cctccatctg aaaaggggag tcatgggagc
agtgcttccg catatcgcca ctgtgctcga gcccttatag aacgtggctg cgatctcgac
ccaacgtc

Figure 3.16 DNA sequence of the front region of the *Bgl*II 1600 bp insert.
The *Bam*HI – *Bgl*II hybrid site is shown in blue.

gaacacggag acgatcatac gcccagaacc ggatgatcga cgaactgaat gaaagagtcg
gataaactcc gtctcgttgt gagggatcaa cgtcgccgac tggagacata cgagggtaac
tcccattaag ggtatatttg tgtccgcagg aaaaacacgg cctataatga acccctatga
aaggaaccca catgtttatc gaaaagactg agtccaacaa gaccaaagta accgaagcca
tcaacgcttt gctcgatgag atgtctggac atcctgccga ttcttctgaa tacgccaaga
tggtggaaca gctgaccaag ctctatgctc tgaaggaagt tgactcaaaa gtcgactccc
ccagacgtgt gagcgcggat acgctggcca tcgtcggcgg caatattctc ggaatcgtca
tgatcgtcgg acatgaacga gcgagcgtgt tgacctcgaa gctctgacct tgctgaccaa
gctacggtag actgtcgaac ccacacagga acaaaagttg ggagggcttg taagacacga
atttacaag ccttctcaat ttttccgcg gaggaatttt gggattctaa attttgtctc
gcataggaaa cacgggttat aatgagaccc cacctatgaa aggaattacc atgcctgaga
tcacttaccg acagtggcgc ggagaactca cccagattca gctcaagata cgatgcattc
gacgaagctt tctacgcctt cccattcacg aattctgtaa aaccgacgag acagtgataa
ccgtataaca **agatcc**

Figure 3.17 DNA sequence of the back region of the *Bgl*II 1600 bp insert.
The *Bam*HI – *Bgl*II hybrid site is shown in blue.

3.8.1.1.2 DNA sequence analysis

3.8.1.1.2a BLASTn result

Sequences of both the front and back regions had no significant nucleotide match.

4.8.1.1.2b BLASTx result

Translated sequences of both the front and back regions had no significant protein match.

4.8.1.1.2c ORF prediction

On each of the front and back regions of the insert, one ORF was predicted, as shown in Figures 3.18 and 3.19, respectively. GC content was 50.2% for the front and 48.6% for the back region.

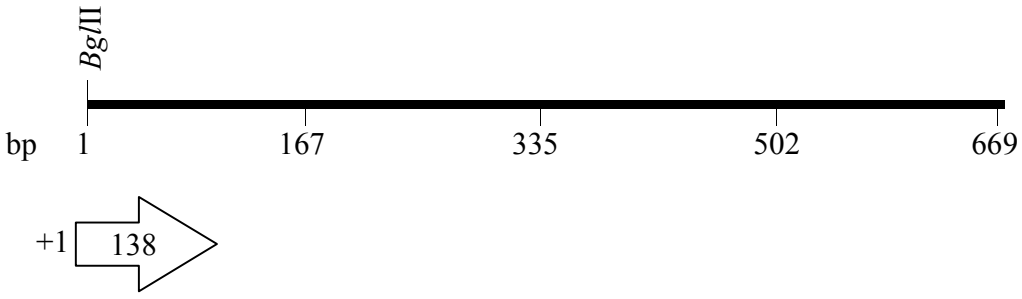


Figure 3.18 Predicted ORF on the front region of the *Bgl*II 1600 bp insert.
The numbers inside and at the beginning of the arrow indicate the length and the frame of the ORF, respectively.

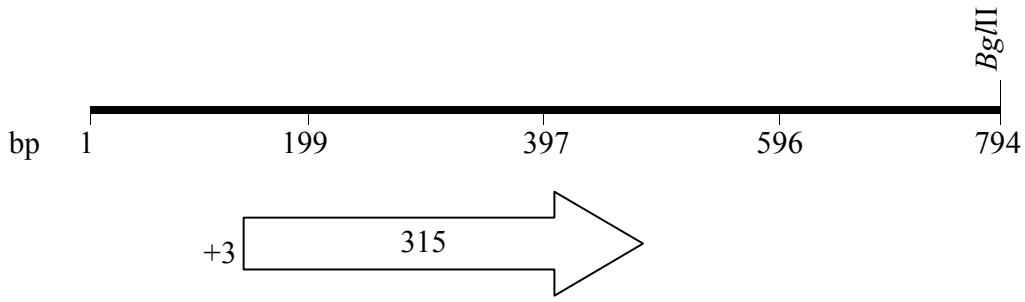


Figure 3.19 Predicted ORF on the back region of the *Bgl*II 1600 bp insert.
The numbers inside and at the beginning of the arrow indicate the length and the frame of the ORF, respectively.

3.8.1.2 Clone 8 of Fairland1 *Bgl*II library

3.8.1.2.1 Restriction map

Single and double digestions for restriction map construction are shown in Figure 3.20. The restriction map is shown in Figure 3.21.

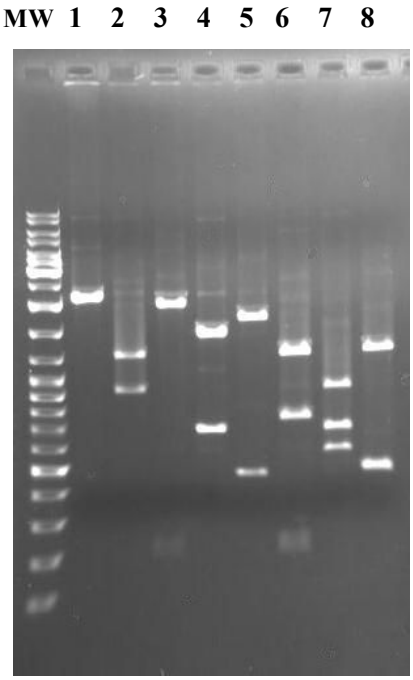


Figure 3.20 Single and double digestions of purified insert from clone 8 of Fairland1 *Bgl*II library, on 0.8% gel.

Lane 1, undigested; 2, *Pst*I; 3, *Eco*RI; 4, ScaI; 5, *Sal*I; 6, *Pst*I + *Eco*RI; 7, *Pst*I + ScaI; 8, *Pst*I + *Sal*I.

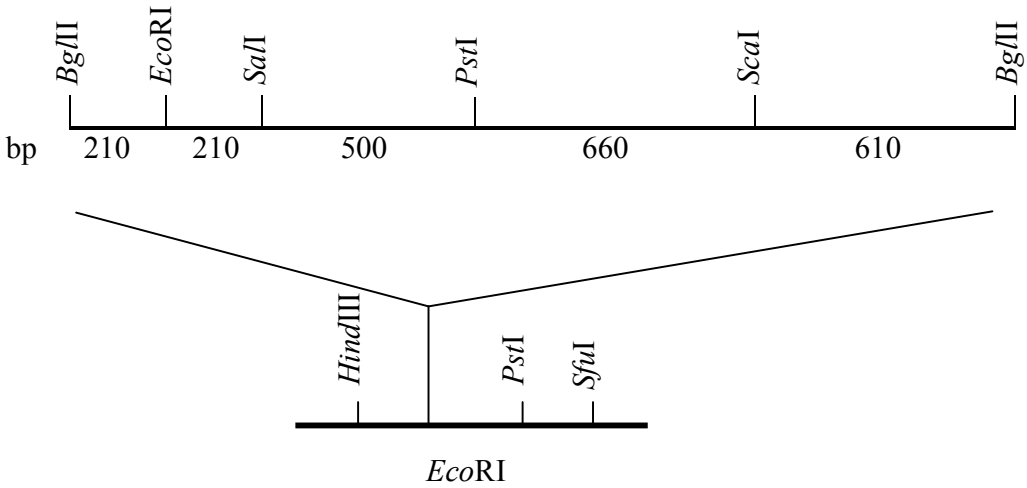


Figure 3.21 Restriction map of the insert from clone 8 of Fairland1 *Bgl*II library.

The orientation of the insert with respect to the suicide gene of pDA71 was identified by digesting the clone with *Pst*I, which released a fragment of ~1400 bp (gel not shown).

3.8.1.2.2 Screening of subclones

DNA fragments generated by digestion of the insert with *Pst*I were subcloned into pDA71, as shown in Figure 3.22.

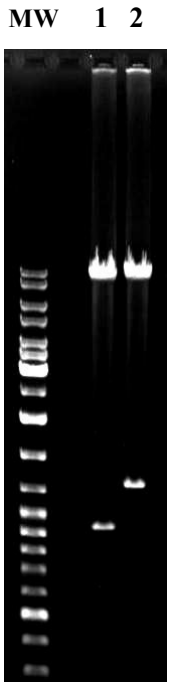


Figure 3.22 Subcloned DNA fragments from clone 8 of Fairland1 *Bgl*II library, on 0.8% agarose gel.
Inserts were released by digestions with respective endonucleases as indicated below. Lane 1, *Bgl*II – *Pst*I 920 bp fragment; 2, *Pst*I – *Bgl*II 1270 bp fragment

Retention of antimicrobial activity in these subclones was tested in *R. erythropolis* SQ1. Number of transformants/µg of DNA for each subclone is shown in Table 3.16.

Table 3.16 Transformants/µg of DNA for each subclone from clone 8 of Fairland1 *Bgl*II library.

Restriction sites at the ends of subclone insert	**Subclone insert size (bp)**	**# transformants ($\times 10^4$)/µg of DNA**
*Bgl*II – *Pst*I	920	0.004
*Pst*I – *Bgl*II	1270	0.26
Vector only	NA	1.11

Subclone with transformation efficiency of $< 0.01 \times 10^4$/µg of DNA is highlighted.

3.8.1.2.3 DNA sequence

The *Bgl*II - *Pst*I 920 bp fragment was cloned into the *Bam*HI and *Pst*I sites of pUC19 and partially sequenced from the side containing the *Bam*HI – *Bgl*II hybrid site. This sequence is shown in Figure 3.23.

ggatctctac aagaagtcgg ttcggatgag tttatctgtc aggccgctcg agtctcgacc
aaggcgctga tgctttcggc actggtgagt ccgaaggtct gttaaacttc ttgatggagg
gacggcatgg gtcacccttc gaacacggct tgatgagctt ccgaatcacg gcgcctatat
ttgtgtggcg tgaattcatg cgtcatcgca tcggttggtc gtacaacgag cagtccggac
gttacatgga gatgcttccg atcttctacg tccccaacat cgagcgtcct ctcgtccagg
tcggtaaggc cggcgcatat tcgttcgtgc ctggcgacgc tgctcagtac gatcagatgt
tcaagtcgat gagtatctcg tacgagactg tgtgggctga gtacgtcaac caactcaagg
ccggcgtggc caaggaggtt gctcgcatgg tgttgtcgac gggcatatat tcgtcggcgt
acgtgacatg caaccctcgt tcgatgatga gcttcttgtc tctcagaacg aaagacgaga
cttcgatgtt cccgtcatat cctcagtggg agatcaatca ggtcgccgac aagatggaag

cagtcttcaa ggagctctac cctctcaccc atgccacatt ccaggccaac ggccgagtat
ctccgtaaag gaacccaaga tgaatctcat tccagtaaca gaaaaggtcg atgtcaagac
cgcctgccat ttcaacggat cacccgtgcc

Figure 3.23 Partial DNA sequence of the *Bgl*II - *Pst*I 920 bp fragment.
The *Bam*HI – *Bgl*II hybrid site, *Eco*RI, and *Sal*I sites are shown in blue, green, and red, respectively.

3.8.1.2.4 DNA sequence analysis

3.8.1.2.4a BLASTn result

This DNA sequence had no significant nucleotide match.

3.8.1.2.4b BLASTx result

Translated protein sequence of the *Bgl*II – *Pst*I fragment showed significant alignment with a putative thymidylate synthase complementing protein from *Streptomyces coelicolor*, with an E value of 2e-59, along with the same protein from various organisms. Included were two mycobacteriophage proteins gp90 and gp91. BLASTx search results are shown in Table 3.17 and Figure 3.24. Features on this protein are shown in Appendix D.

Table 3.17 BLASTx search result: proteins whose sequences aligned to the translated *Bgl*II - *Pst*I fragment.

Sequences producing significant alignments:	Score (Bits)	E Value
gi|3413826|emb|CAA20294.1| conserved hypothetical protein SC9...	231	2e-59
gi|29829059|ref|NP_823693.1| thymidylate synthase [Streptomyc...	226	7e-58
gi|88931859|ref|ZP_01137552.1| Thymidylate synthase complemen...	218	2e-55
gi|28410912|emb|CAD67297.1| conserved hypothetical protein [T...	201	2e-50
gi|28493583|ref|NP_787744.1| thymidylate synthase [Tropheryma...	201	2e-50
gi|109393301|ref|YP_656098.1| gp90 [Mycobacteriophage Catera]...	186	7e-46
gi|29566599|ref|NP_818164.1| gp91 [Mycobacterium phage Bxz1] ...	186	1e-45
gi|14916788|sp|Q9WYT0|THYX_THEMA Thymidylate synthase thyX (T...	111	3e-23
gi|33357640|pdb|1O2B|D Chain D, Crystal Structure Of Thymidyl...	111	3e-23
gi|20150963|pdb|1KQ4|D Chain D, Crystal Structure Of Thy1-Com...	107	7e-22
gi|94986126|ref|YP_605490.1| Thymidylate synthase (FAD) [Dein...	100	2e-20

Only results with an E value ≤ 1e-20 are shown.

```
>  gi|3413826|emb|CAA20294.1|    conserved hypothetical protein SC9A10.07
[Streptomyces coelicolor
A3(2)]
 gi|21224089|ref|NP_629868.1|    thymidylate synthase [Streptomyces coelicolor
A3(2)]
 gi|24638242|sp|O86840|THYX_STRCO  Thymidylate synthase thyX (TS) (TSase)
Length=246

 Score =  231 bits (590),  Expect = 2e-59
 Identities = 115/212 (54%), Positives = 144/212 (67%), Gaps = 2/212 (0%)
 Frame = +3

Query  30   VYLSGRSSLDQGADAFGTGESEGLLNFLMEGRHGSPFEHGLMSFRITAPIFVWREFMRHR  209
            V  +G  SLD+         S+GL+N+LM  RHGSPFEH  M+F ++APIFV+REFMRHR
Sbjct  37   VSTAGEQSLDELKK--DPERSKGLINYLMRDRHGSPFEHNSMTFFVSAPIFVFREFMRHR  94

Query  210  IGWSYNEQSGRYMEMLPIFYVPNIERPLVQVGKAGAYSFVPGDAAQYDQMFKSMSISYET  389
            +GWSYNE+SGRY E+ P+FY P+  R LVQ G+ G Y FV G   Q++ +  +M  SY
Sbjct  95   VGWSYNEESGRYRELQPVFYAPDASRKLVQQGRPGKYVFVEGTPEQHELVGSAMEDSYRQ  154
```

```
Query  390  VWAEYVNQLKAGVAKEVARMVLSTGIYSSAYVTCNPRSMMSFLSLRTKDETSMFPSYPQW  569
             +A Y   L AGVA+EVAR VL  G+YSS Y TCN RS+M FL LRT+ E +  PS+PQ
Sbjct  155  AYATYQQMLAAGVAREVARAVLPVGLYSSMYATCNARSLMHFLGLRTQHELAKVPSFPQR  214

Query  570  EINQVADKMEAVFKELYPLTHATFQANGRVSP  665
             EI    +KMEA +  L PLTHA F ANGRV+P
Sbjct  215  EIEMAGEKMEAEWARLMPLTHAAFNANGRVAP  246
```

Figure 3.24 BLASTx search result: alignment between the translated *Bgl*II - *Pst*I fragment and the thymidylate synthase complementing protein from *Streptomyces coelicolor*.

3.8.1.2.4c ORF prediction

On this DNA sequence, two ORFs were predicted, as shown in Figure 3.25. One of these ORFs aligned with the thymidylate synthase complementing protein from *Streptomyces coelicolor*. GC content of the DNA fragment was 53.9%.

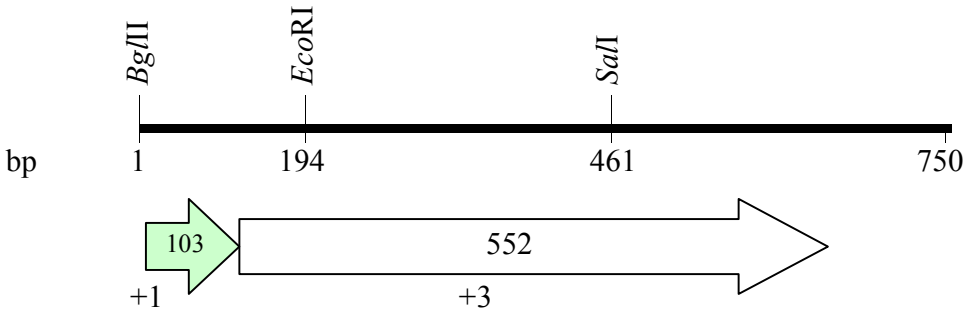


Figure 3.25 Predicted ORFs on the sequenced region of the *Bgl*II – *Pst*I fragment. The numbers inside and at the bottom of each arrow indicate the length and the frame of the ORF, respectively. ORF aligning with the thymidylate synthase complementing protein from *Streptomyces coelicolor* is shown in light green.

3.8.1.3 Clone 10 of Fairland1 *Bgl*II library

3.8.1.3.1 Restriction map

Single and double digestions for restriction map construction are shown in Figure 3.26. The restriction map is shown in Figure 3.27.

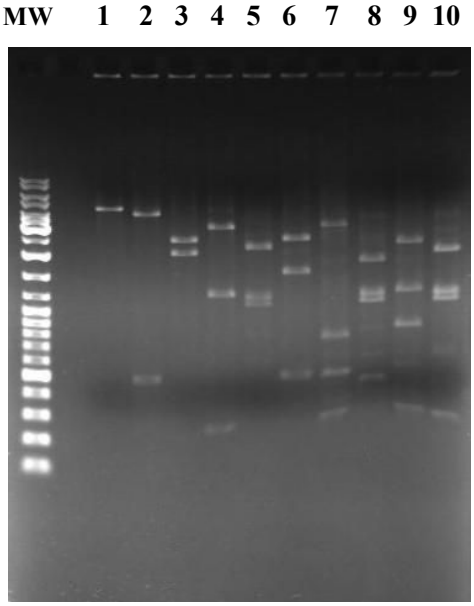


Figure 3.26 Single and double digestions of purified insert from clone 10 of Fairland1 *Bgl*II library, on 0.8% gel.
Lane 1, undigested; 2, *Hind*III; 3, *Pst*I; 4, *Sfu*I; 5, *Acc*I; 6, *Hind*III + *Pst*I; 7, *Hind*III + *Sfu*I; 8, *Hind*III + *Acc*I; 9, *Pst*I + *Sfu*I; 10, *Pst*I + *Acc*I.

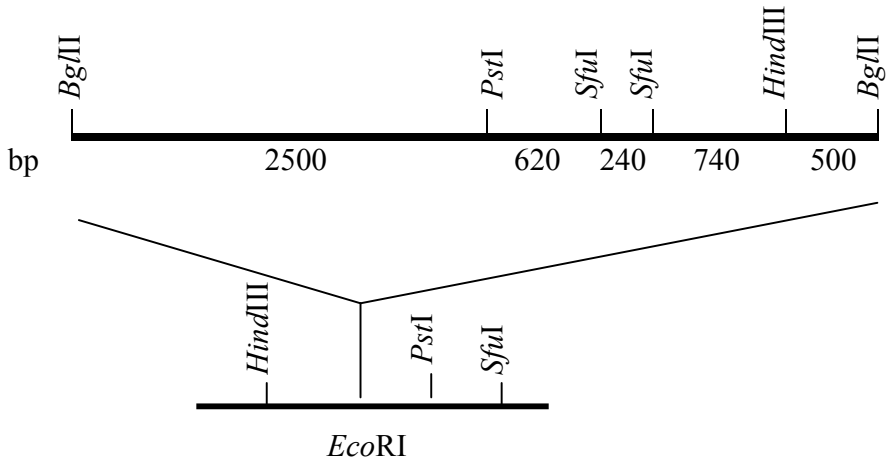


Figure 3.27 Restriction map of the insert from clone 10 of Fairland1 *Bgl*II library. The orientation of the insert with respect to the suicide gene of pDA71 was identified by digesting the clone with *Hind*III, which released a fragment of ~4500 bp (gel not shown).

3.8.1.3.2 Screening of subclones

DNA fragments generated by single digestions of the purified insert with *Pst*I, *Hind*III, *Sfu*I, and double digestion with *Pst*I and *Hind*III, were subcloned into pDA71, as shown in Figure 3.28.

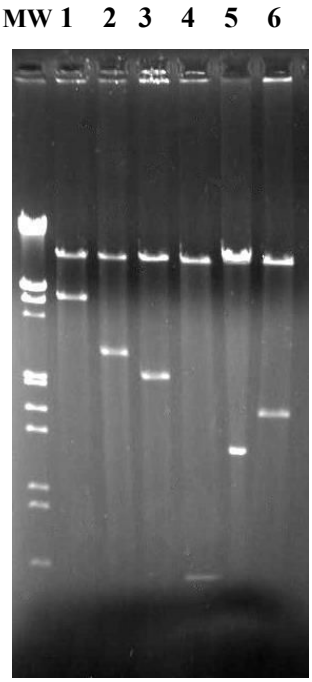


Figure 3.28 Subcloned DNA fragments from clone 10 of Fairland1 *Bgl*II library, on 0.8% agarose gel.

Inserts were released by digestions with respective endonucleases as indicated below. Lane 1, original clone 10, insert released by *Bgl*II; 2, *Bgl*II – *Pst*I 2500 bp fragment; 3, *Pst*I – *Bgl*II 2100 bp fragment; 4, *Hind*III - *Bgl*II 500 bp fragment; 5, *Sfu*I – *Bgl*II 1240 bp fragment; 6, *Pst*I – *Hind*III 1600 bp fragment.

Retention of antimicrobial activity in each subclone was tested in *R. erythropolis* SQ1. Transformants/µg of DNA for each subclone is shown in Table 3.18.

Table 3.18 Transformants/µg of DNA for each subclone from clone 10 of Fairland1 *Bgl*II library.

Restriction sites at the ends of subclone insert	**Subclone insert size (bp)**	**# transformants ($\times 10^4$)/µg of DNA**
*Bgl*II – *Pst*I	2500	0.007
*Pst*I – *Bgl*II	2100	< 0.0017
*Hind*III – *Bgl*II	500	0.08
*Sfu*I – *Bgl*II	1240	0.17
*Pst*I – *Hind*III	1600	0.002
Vector only	NA	1.3

Subclones with transformation efficiency of < 0.01×10^4/µg of DNA are highlighted.

Both the *Bgl*II – *Pst*I 2500 bp and the *Pst*I – *Bgl*II 2100 bp fragments, which together made up the entire insert from the original clone, were inhibitory. This suggested the presence of at least two different inhibitory determinants on the original insert, with at least one in each of these two fragments. From the data in Table 3.18, the inhibitory region was narrowed down as shown in Figure 3.29.

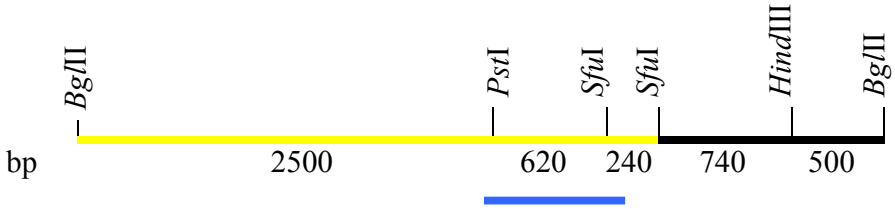


Figure 3.29 Region containing inhibitory determinants, as suggested by the screening of subclones.

Inhibitory region is shown in yellow. Region which was sequenced is indicated with the blue line.

3.8.1.3.3 DNA sequence

The region between the *Pst*I site and the second (from left) *Sfu*I site, as indicated by the blue line in Figure 3.29, was selected for sequencing. The *Pst*I – *Hind*III 1600 bp fragment was cloned into pUC19 and sequenced from the side containing the *Pst*I site. This sequence is shown in Figure 3.30.

ctgcagccgg gaatatagag cgtccaggtc aggttgatct tctcgttggt gtccatcggg
acgagaatct caggctcttc gttggtcagg acctcacgaa cgctgtgctc gtgttcgcga
cgctcttcct cttctgcaat cagctttgca gccttgaacg aggcctggcc ggccaagatg
gttgccgtaa cagcacctgt aacagccaca gctgtcagga tggcggggga ctttgaaatg
gtgaaccgtt cgaggttctt tgccagatcg gaaatcgaca ttgtttactc cagtttagtc
ttgaatgatg ggtgtcgtgc taagtcaagc gtcgtaatat ggcatttgac gttgttcggc
ttcttctgcc tcgaggagct ctgcggcttt ccggtattca gcaaagaagc gtgtggacac
aacccatgtg gctgcgagga tgacggtgaa gatgacggat acagccaaga tcccaagacc
caagaggagg tagaggaata tggtccatac gaactcgaga acgctcatgc tgctacgctt
tcttccggga tttcattgta gaaccgatcg atcattgcaa agaactcctc agcggtctgt
tccttcgaac gtcgatggcg cggtgcggga cggggagggc tggtgtactt ggggttagtc
ccatttacct tggcgcgacg acgatcagtg acgagaacgg ttgatgtcac caatgccaac
ggaagtccga tgacctgc

Figure 3.30 DNA sequence between the *Pst*I and the second *Sfu*I sites, in the *Pst*I – *Hind*III 1600 bp fragment.
The *Pst*I site and the first *Sfu*I site are shown in blue and green, respectively.

3.8.1.3.4 DNA sequence analysis

3.8.1.3.4a BLASTn result

The *Pst*I – *Sfu*I DNA sequence had no significant nucleotide match.

3.8.1.3.4b BLASTx result

Translated sequence of the *Pst*I – *Sfu*I DNA fragment had no significant protein match.

3.8.1.3.4c ORF prediction

Two ORFs were predicted on this sequence, as shown in Figure 3.31. GC content was 53.8%.

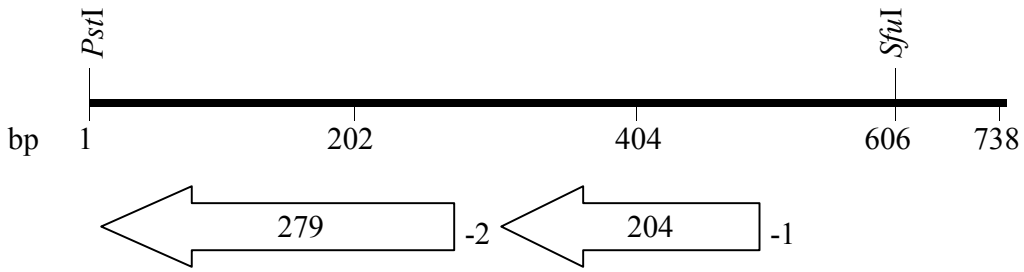


Figure 3.31 Predicted ORFs on the *Pst*I –*Sfu*I fragment.
The numbers inside and at the beginning of each arrow indicate the length and the frame of the ORF, respectively.

3.8.1.4 Clone 19 of Fairland1 *Bgl*II library

3.8.1.4.1 Restriction map

Single and double digestions for the construction of a restriction map are shown in Figure 3.32. The restriction map is shown in Figure 3.33.

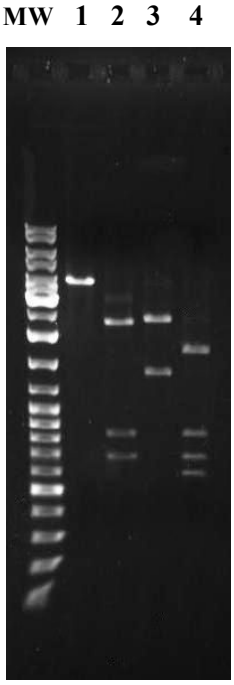


Figure 3.32 Single and double digestions of purified insert from clone 19 of Fairland1 *Bgl*II library, on 0.8% gel.
Lane 1, undigested; 2, *Hind*III; 3, *Cla*I; 4, *Hind*III + *Cla*I.

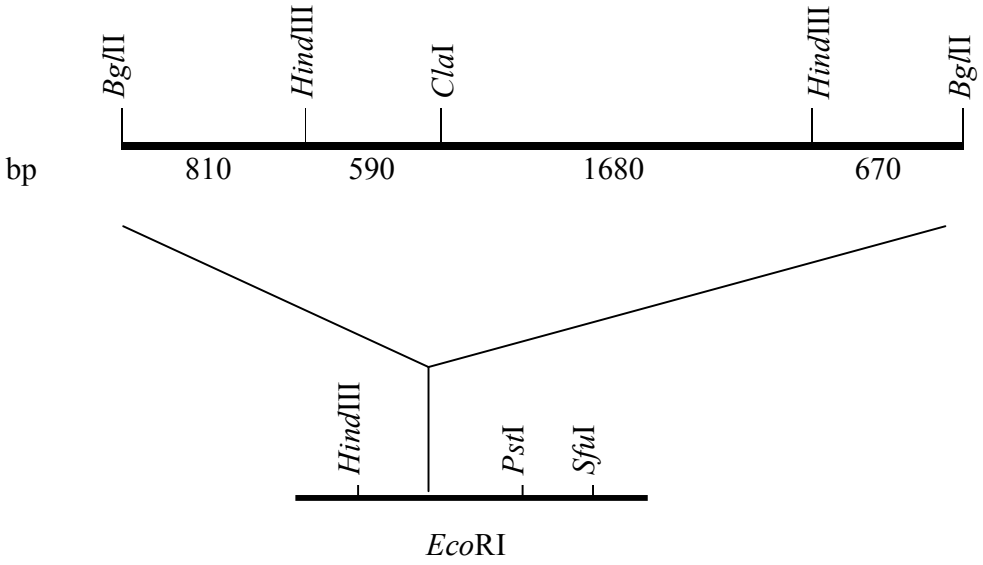


Figure 3.33 Restriction map of the insert from clone 19 of Fairland1 *Bgl*II library. Orientation of the insert with respect to the suicide gene of pDA71 was identified by digesting the clone with *Hind*III, which released fragments of ~2300 bp and ~1000 bp (gel not shown).

3.8.1.4.2 Screening of subclones

DNA fragments generated by digestion of the purified insert with *Hind*III were subcloned into pDA71, as shown in Figure 3.34.

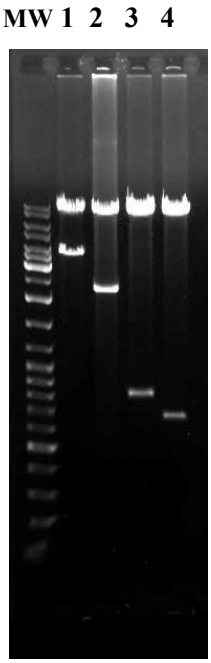


Figure 3.34 Subcloned DNA fragments from clone 19 of Fairland1 *Bgl*II library, on 0.8% agarose gel.

Inserts were released by digestions with respective endonucleases as indicated below. Lane 1, original clone 19, insert released by *Bgl*II digestion; 2, *Hind*III 2270 bp fragment; 3, *Bgl*II - *Hind*III 810 bp fragment; 4, *Hind*III – *Bgl*II 670 bp fragment.

Retention of antimicrobial activity in these subclones was tested in *R. erythropolis* SQ1. Transformants/µg of DNA for each subclone is shown in Table 3.19.

Table 3.19 Transformants/µg of DNA for each subclone from clone 19 of Fairland1 *Bgl*II library.

Restriction sites at the ends of subclone insert	**Subclone insert size (bp)**	**# transformants ($\times 10^4$)/µg of DNA**
*Hind*III	2270	0.004
*Bgl*II - *Hind*III	810	0.11
*Hind*III – *Bgl*II	670	0.23
Vector only	NA	1.07

Subclone with transformation efficiency of $< 0.01 \times 10^4$/µg of DNA is highlighted.

3.8.1.4.3 DNA sequence

The *Hind*III 2270 bp fragment was cloned into pUC18 and partially sequenced. Sequences of the front and back regions are shown in Figures 3.35 and 3.36, respectively.

aagcttttca gccaactcga caggatcggc aaactcaccg gtcttgttaa gctggctgta
atatagccgc gccttggcga cttcttcttc gatctgaccc tcgtactgag cgatcatttg
cttcttggtc aagatgtggg cgattgctgc acccacaccg agtgacacca ctgaggatgc
cactacgatg ccagtgtttg tcatcagtcc tccagaagtt cgagcacgtt gccgtcgacg
ttgaagtcaa gccacaccga gcggctgtct ccgttgacaa agcgagttcc ctctgccaag
ttgttgaaga cgccgaagtc gatgaagttg tctccgttac cgttcagaac ccatccgacg
atctggccag ccttggtacg atccagtccg agcatgtcgt gaacatcgtt caagaagaca
acgcccttgg cacgaagcaa gtcgttggca tagttctgct gcgacgagat gaagaactgg
ttgtactcag ttgccttgtt ccagttcggg ttccgctcat cgaagagaac cttgtacggc
gatccaccac gagccgtcgg aacagcgacc ttgaccttct tgcccgtctc atcgattgca
tcgacctgct cgaccggctg ccaaataaga gattccttat cagccccgat cgattcgatg

acgcgtccgc ggtactcacg gaaggacttc tcgacgccgg cgtaagtcc

Figure 3.35 DNA sequence of the front region of the *Hind*III 2270 bp fragment. The *Hind*III and the *Cla*I sites are shown in blue and green, respectively.

gcggttgttg tagctgacgt aacccgagcc agttcctgga cgactgcttg cgctggatcg
agtacgactc ctcgcctcac cgaacagcat cttctcgaca ccctgagaaa cagcatcagc
catcatgtcc ttggcagccg gcaagagaac gtccgtcgca acgtagacca agacgccctg
agcatctccg gcaccaaata tctctgcgaa gcgcttaccg agcggcttct tacgacgtgt
aactgcaccg acggtgattc gttcgatctt cttctcaggc gctgccttct cggcgacggg
tgccggctgc ttaacggtct tgctgttgct ggggtagttc tccattttag ttcctcaact
ttggtaggat aaaacgtaga taccgtgtta gggtatctag tttcgagatg ttatgattgt
gcttgttcag gattcttgaa ggatttgatc agtgcgaccg cttcatctat cttggtatcg
gtgtagttct tcgatgcgtc ggccaccatc gctccgatga caactgatcc tgcgtacacc
gttacgtgtt caacgggtgt ttcaggttcg acattgtttc gaatgatggt ggagacgatc
ttggtggttc cgattccaac aacagtggaa acgacgttct tgagtagatc cagcttgttc
atggcaggtc ctttcatagg ggtctcatta tagggtgtgt ttttactgcg tttgcgtggg
cacatgttct actgtttcac gcgacatatc agctcgaagg gctcagtgat taagctt

Figure 3.36 DNA sequence of the back region of the *Hind*III 2270 bp fragment. The *Hind*III site is shown in blue.

3.8.1.4.4 DNA sequence analysis

3.8.1.4.4a BLASTn result

DNA sequences of the forward and back regions of the *Hind*III 2270 bp fragment had no nucleotide match.

3.8.1.4.4b BLASTx result

Translated sequences of both the forward and back regions of the *Hind*III 2270 bp fragment had no protein match.

3.8.1.4.4c ORF prediction

Two ORFs were predicted on the front region of the *Hind*III 2270 bp fragment, and one ORF was predicted on the back region, as shown in Figures 3.37 and 3.38, respectively. GC content of the front and back regions were 54.9% and 51.4%, respectively.

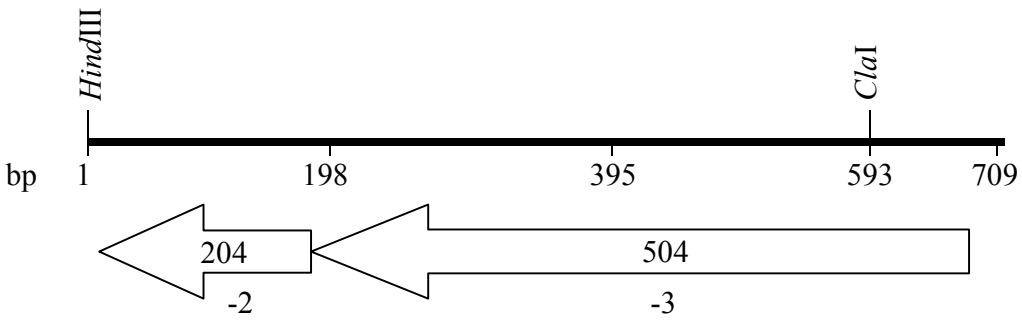


Figure 3.37 Predicted ORFs on the front region of the *Hind*III 2270 bp fragment. The numbers inside and at the bottom of each arrow indicate the length and the frame of the ORF, respectively.

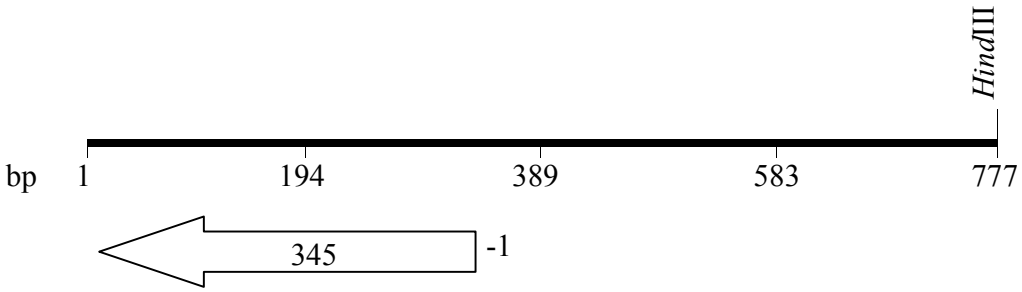


Figure 3.38 Predicted ORF on the back region of the *Hind*III 2270 bp fragment. The numbers inside and at the beginning of the arrow indicate the length and the frame of the ORF, respectively.

3.8.1.5 Clone 8 of Fairland1 *Pst*I library

3.8.1.5.1 DNA sequence

Subcloning of clone 8 from Fairland1 *Pst*I library was not conducted due to its size of ~1600 bp, and it was cloned directly into pUC19. Sequences of the front and back regions are shown in Figures 3.39 and 3.40, respectively.

ctgcaggtga taccgtaata catggtccat gagctgtctt cctgtgtgga aaatgttata
ccactaacga tgaaggcagc caagcgtatc gtaaacactc cagcaagcgg ctcgaagcca
tcgctagaat tggaacttag ccgatgacag cgaccaagtt atattaccga caatcgtcat
tacaactcaa agaaaggggg cggtaaacat ggaacatgat ttcagcggtt acgccacaaa
agccaatatc agatgctctg atggccgaaa ttttcttggc ttgacgcctt caagcatatg
gatggcgaac gagttcctct tgtgtggcag catggacaca atgaacccgg taacgttctc
gggcatatga atctcgaagc tcgagatggt gatgcctatg gctacgcctt cttcaacgat
actccgactg gtcagagtac gaagttgatg gtcaagcacg gcgatattac tgctctctca
atctacgcca acaagctggt tgagcgagat aaggcagtca tgcacggtca gatccgtgaa
gtcagcctag tcctctcggg tgcgaacccc ggtgccttga tcgacaacgt caacttgccc

atggtgatgg atctattacg ttggccgatg aagaggcagt tatttacacc ggagctccac
tcgaactcaa ccaacccgaa gagagtacta tgtctcacag caatcagtcc ggtgaccttt
agtctgttct cgactccctg gacgagacac agcgtggagc agtcgacttc ctccttgagc
aggctcttgc tcatcg

Figure 3.39 DNA sequence of the front region of the insert from clone 8 of Fairland1 *Pst*I library.

The *Pst*I site is shown in blue.

ctgagcgagg acgaagaagt tctccggaag attcatccgc caagcactct gacctcgccg
gaaaaccctt tggaaggaac caactccagt cgcacaacgc cttcgaatcc aacactctcg
ccgccaccga cgccggcaag tcggctacgc tcagccacgc tgacactgct tccatcttcg
aaggcgcaaa gcgtctcggt tccatgaagg aatccctcga ggactacgct ctgtcgcacg
gcatccagga catcgatatc ctgttcccgg acgccaaggc aatctcgtcc acccctgagt
tcgagaagcg tcgcaccgag tgggtcgccg aggtcatgaa cggcacccgc aaggacccgt
tctcgcgcat caagagcctg acggccaacc tggccctcga ggaagctcgg gcgaagggtt
acatcaaggg caacctgaag aaggaagagt tcttccgggt gtcgaagcga accaccaccc
ctcagaccat ctacaagaag cagaagctgg accgggatga catcctggac atcaccgact
ttggatgtcg tggctttggc tcaagggtga aaatgcgcct catgctcgag gaggagattg
ctcgcgcagt cctgctcggc gatggccgtt cggttggcga cgacgacaag atctccgagg
atcacgttcg tccgatcgct acggatgctg acctgtacgt cacgtacgtc aacgtcaacc
tgggtgatgc caattctcgg ccgaagattc ccagtgcaga gttctgtaaa accgacgagg
ccagtgatag ccgtatactg cag

Figure 3.40 DNA sequence of the back region of the insert from clone 8 of Fairland1 *Pst*I library.

The *Pst*I site is shown in blue.

3.8.1.5.2 DNA sequence analysis

3.8.1.5.2a BLASTn result

Both the front and back regions of the *Pst*I 1600 bp insert had no nucleotide match.

3.8.1.5.2b BLASTx result

A translated sequence of the front region aligned to a putative phage head maturation peptidase U35 from *Mycobacterium* sp. MCS and KMS, with an E value of 5e-4. BLASTx search result for the front region is shown in Table 3.20 and Figure 3.41. Features of this protein are shown in Appendix D.

A translated sequence of the back region aligned to a putative phage capsid protein from *Streptococcus agalactiae*, with an E value of 4e-4. In a different reading frame, the sequence also aligned to a putative dihydropteroate synthase from *Nocardioides* sp. JS614, with an E value of 0.003. BLASTx search result for the back region is shown in Table 3.21 and Figure 3.42. Features of these proteins are shown in Appendix D.

Table 3.20 BLASTx search result: proteins whose sequences aligned to the translated front region of the *Pst*I 1600 bp insert.

Sequences producing significant alignments:	Score (Bits)	E Value
gi|108800480|ref|YP_640677.1| peptidase U35, phage prohead HK...	47.8	5e-04
gi|92916876|ref|ZP_01285493.1| Peptidase U35, phage prohead H...	47.8	5e-04

```
>  gi|108800480|ref|YP_640677.1|    peptidase U35, phage prohead HK97
[Mycobacterium sp. MCS]
 gi|108770899|gb|ABG09621.1|    peptidase U35, phage prohead HK97 [Mycobacterium
sp. MCS]
Length=526

 Score = 47.8 bits (112),  Expect = 5e-04
 Identities = 29/98 (29%), Positives = 48/98 (48%), Gaps = 2/98 (2%)
 Frame = +1

Query  304  GERVPLVWQHGHNEPGNVLGHMN--LEARDGDAYGYAFFNDTPTGQSTKLMVKHGDITAL  477
            G  VPL W+HGH++P  ++G +   +E   G         DT  G +    VK   I  L
Sbjct  49   GGVVPLFWEHGHDDPRAIVGEVTAAVETTRGLEIVGKLDTDTERGAAAYRAVKGRRIRGL  108

Query  478  SIYANKLVERDKAVMHGQIREVSLVLSGANPGALIDNV  591
            S+       R  +++   + E+SLV+  +N  AL++V
Sbjct  109  SVGMRPTQRRGASIIAADLCEISLVMRPSNSRALVESV  146


>  gi|92916876|ref|ZP_01285493.1|  Peptidase U35, phage prohead HK97
[Mycobacterium sp. KMS]
 gi|92438995|gb|EAS96845.1|  Peptidase U35, phage prohead HK97 [Mycobacterium sp.
KMS]
Length=526

 Score = 47.8 bits (112),  Expect = 5e-04
 Identities = 29/98 (29%), Positives = 48/98 (48%), Gaps = 2/98 (2%)
 Frame = +1

Query  304  GERVPLVWQHGHNEPGNVLGHMN--LEARDGDAYGYAFFNDTPTGQSTKLMVKHGDITAL  477
            G  VPL W+HGH++P  ++G +   +E   G         DT  G +    VK   I  L
Sbjct  49   GGVVPLFWEHGHDDPRAIVGEVTAAVETTRGLEIVGKLDTDTERGAAAYRAVKGRRIRGL  108

Query  478  SIYANKLVERDKAVMHGQIREVSLVLSGANPGALIDNV  591
            S+       R  +++   + E+SLV+  +N  AL++V
Sbjct  109  SVGMRPTQRRGASIIAADLCEISLVMRPSNSRALVESV  146
```

Figure 3.41 BLASTx search result: alignment between the translated front region of the *Pst*I 1600 bp insert and the phage head maturation peptidase U35 from *Mycobacterium* sp. MCS and KMS.

Table 3.21 BLASTx search result: proteins whose sequences aligned to the translated back region of the *Pst*I 1600 bp insert.

Sequences producing significant alignments:	Score (Bits)	E Value
gi|22534892|gb|AAN00715.1| conserved domain protein [Streptoc...	48.1	4e-04
gi|71369843|ref|ZP_00660252.1| Dihydropteroate synthase [Noca...	45.1	0.003

>gi|22534892|gb|AAN00715.1| conserved domain protein [Streptococcus agalactiae 2603V/R]
gi|22537991|ref|NP_688842.1| hypothetical protein SAG1852 [Streptococcus agalactiae 2603V/R]
Length=420

Score = 48.1 bits (113), Expect = 4e-04
Identities = 41/168 (24%), Positives = 75/168 (44%), Gaps = 9/168 (5%)
Frame = +3

```
Query  201  SMKESLEDYALSHGIQDIDILFPDAKAISSTPEFEKRRTEWVAEVMNGTRKDPFSRIKSL  380
            S +++ E    +  G+ ++   + P+    I+     F          ++N   KDP   ++
Sbjct  120  SARKAWEANLVEKGVTNLTKILPEPVLIAIQDAFTNYNG-----ILNHVSKDPRYAVR--  172

Query  381  TANLALEEARAKGYIKGNLKKEEFFRVSKRTTTPQTIYKKQKLDRDDILDITDFGCRGFG  560
               L   + ++AKG+  G  KK+E F     T    T+Y K    +  D+    T      +
Sbjct  173  -VALQTQVSQAKGHKAGKTKKDEDFTFLDFTINSATVYIKYAFEYSDLKKDTTGAYFNYV  231

Query  561  SRVKMRLMLEEEIARAVLLGDGRSVGDDDKISEDHVRPIATDADLYVT  704
              +   +  +   I RAV++GDG+S   +DKI+E       T+A+L+ T
Sbjct  232  MKELAQGFIRT-IERAVVIGDGKSNSAEDKITEIKSIAEETEANLFET  278
```

>gi|71369843|ref|ZP_00660252.1| Dihydropteroate synthase [Nocardioides sp. JS614]
gi|71154437|gb|EAO04955.1| Dihydropteroate synthase [Nocardioides sp. JS614]
Length=466

Score = 45.1 bits (105), Expect = 0.003
Identities = 31/87 (35%), Positives = 38/87 (43%), Gaps = 16/87 (18%)
Frame = +2

```
Query  188  KASRFHEGIPRGLRSVARHPGHRYPVPGRQGNLVHP*VREASHRVGRRGHERHPQGPVLA  367
            +A++  + + RG R   RHP H  P  GR              R GR  H R P  PV A
Sbjct  100  RAAQRADRVRRGARRPRRHPAH--PRAGRA-------------RPGRPRHPR-PPAPVAA  143
```

```
Query  368  HQEPDGQPGPRGSSGEGLHQGQPEEGR  448
            + DG P PRG  G G H  +   GR
Sbjct  144  RPQRDGHPDPRGRPGPGRHPHRGRAGR  170
```

Figure 3.42 BLASTx search result: alignment between the translated back region of the *Pst*I 1600 bp fragment and the phage capsid protein from *Streptococcus agalactiae*, and the dihydropteroate synthase from *Nocardioides* sp. JS614.

3.8.1.5.2c ORF prediction

One ORF was predicted on each of the front and back regions of the *Pst*I 1600 bp insert, as shown in Figures 3.43 and 3.44, respectively. The ORF predicted on the front segment corresponded to the phage head maturation peptidase from *Mycobacterium* sp. MCS and KMS. The ORF predicted on the back segment did not correspond to either of the aligned proteins from the BLASTx search. GC content of the front and back regions were 50.4% and 58.3%, respectively.

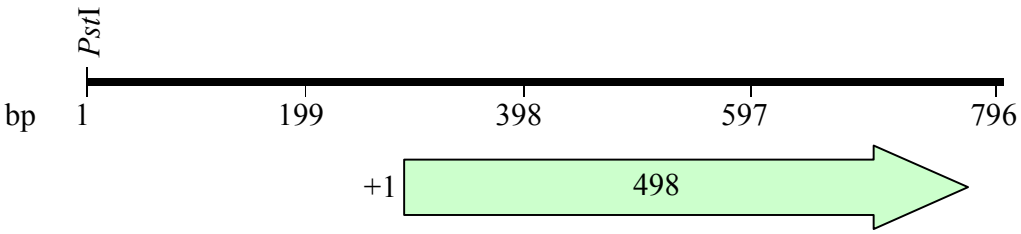


Figure 3.43 Predicted ORF on the front segment of the *Pst*I 1600 bp insert.
The numbers inside and at the beginning of the arrow indicate the length and the frame of the ORF, respectively. This ORF corresponded to the phage head maturation peptidase from *Mycobacterium* sp. MCS and KMS.

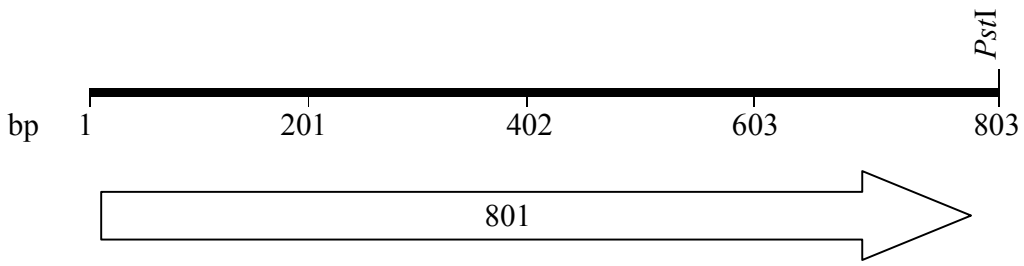


Figure 3.44 Predicted ORF on the back segment of the *Pst*I 1600bp insert.
The number inside the arrow indicates the length of the ORF. This ORF corresponded to neither of the aligned proteins from the BLASTx search.

3.8.1.6 Clone 14 of Fairland1 *Pst*I library

3.8.1.6.1 Restriction map

Single and double digestions for the construction of a restriction map are shown in Figure 3.45. The restriction map is shown in Figure 3.46.

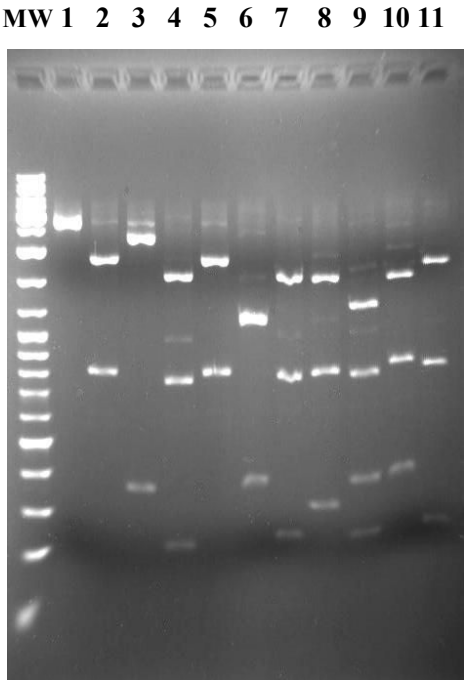


Figure 3.45 Single and double digestions of purified insert from clone 14 of Fairland1 *Pst*I library, on 0.8% gel.

Lane 1, undigested; 2, *Bgl*II; 3, *Hind*III; 4, *Acc*I; 5, *Sal*I; 6, *Bgl*II + *Hind*III; 7, *Bgl*II + *Acc*I; 8, *Bgl*II + *Sal*I; 9, *Hind*III + *Acc*I; 10, *Hind*III + *Sal*I; 11, *Acc*I + *Sal*I.

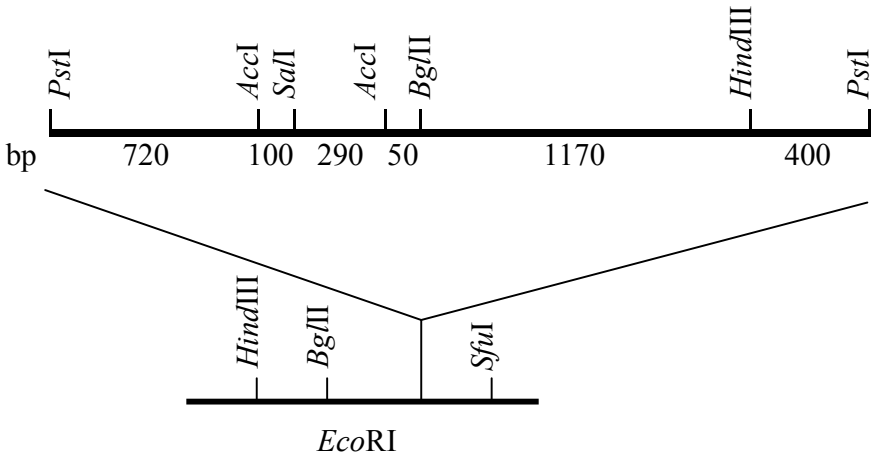


Figure 3.46 Restriction map of the insert from clone 14 of Fairland1 *Pst*I library.

Orientation of the insert with respect to the suicide gene of pDA71 was identified by digesting the clone with *Bgl*II and *Hind*III, individually, which released fragments of ~1300 bp and ~2900 bp, respectively (gel not shown).

3.8.1.6.2 Screening of subclones

DNA fragments generated by single digestions of the purified insert with *Bgl*II and *Pst*I, and double digestion with both enzymes, were subcloned into pDA71, as shown in Figure 3.47.

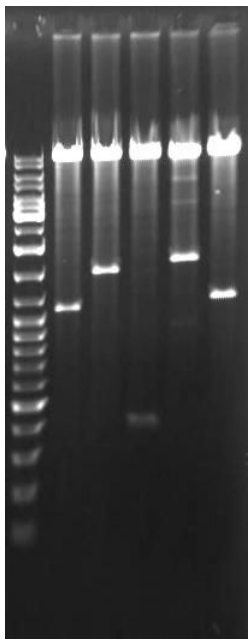


Figure 3.47 Subcloned DNA fragments from clone 14 of Fairland1 *Pst*I library, on 0.8% agarose gel.
Inserts were released by digestions with respective endonucleases as indicated below. Lane 1, *Pst*I – *Bgl*II 1160 bp fragment; 2, *Bgl*II – *Pst*I 1570 bp fragment; 3, *Hind*III – *Pst*I 400 bp fragment; 4, *Pst*I – *Hind*III 2330 bp fragment; 5, *Bgl*II – *Hind*III 1170 bp fragment.

Retention of antimicrobial activity in these subclones was checked in *R. erythropolis* SQ1. Transformants/µg of DNA for each subclone is shown in Table 3.22.

Table 3.22 Transformants/µg of DNA for each subclone from clone 14 of Fairland1 *Pst*I library.

Restriction sites at the ends of subclone insert	**Subclone insert size (bp)**	**# transformants ($\times 10^4$)/µg of DNA**
*Pst*I – *Bgl*II	1160	0.9
*Bgl*II – *Pst*I	1570	< 0.0017
*Hind*III – *Pst*I	400	< 0.0017
*Pst*I – *Hind*III	2330	0.05
*Bgl*II – *Hind*III	1170	0.24
Vector only	NA	1.13

Subclones with transformation efficiency of < 0.01×10^4/µg of DNA are highlighted.

The *Hind*III – *Pst*I 400 bp fragment is part of the *Bgl*II – *Pst*I 1570 bp fragment. Thus the smaller fragment was cloned into pUC19 and sequenced.

3.8.1.6.3 DNA sequence

DNA sequence of the *Hind*III – *Pst*I 400 bp fragment is shown in Figure 3.48.

aagcttgcga atgaggatgg tgtgcgcaag acgaagtccg ttgctcttct tgtagctcag
gcctacttgg caccacccaa gaatcacctc tacaattcta tgatccattt ggatggggat
cgtagtaacc tcaatgctcg caacttgatg tggcgtccgc gttggtacgc gattcgctat
caccagatgt tcaatctcga tcctatcgat atctcggtca tgatcgaaga cacaggcgag
gtctttggaa ccatcaggga cctgtgtgta aagtacggac tcgaacaaca cctcacctac

atggatcttc tgaatgggaa caagctcttt cactatggat accgcatcag acgagtggat tcctaaatag ggtattaacc **ctgcag**

Figure 3.48 DNA sequence of the *Hind*III – *Pst*I 400 bp fragment.
*Hind*III and *Pst*I sites are shown in blue and green, respectively.

3.8.1.6.4 DNA sequence analysis

3.8.1.6.4a BLASTn result

The *Hind*III – *Pst*I 400 bp DNA sequence had no nucleotide match.

3.8.1.6.4b BLASTx result

A translated sequence of the *Hind*III – *Pst*I 400 bp fragment showed alignment to a putative HNH endonuclease from *Lactobacillus plantarum* bacteriophage LP65, with an E value of 0.006. BLASTx search results are shown in Table 3.23 and Figure 3.49. Features of this protein are shown in Appendix D.

Table 3.23 BLASTx search result: protein whose sequence aligned to the translated *Hind*III – *Pst*I fragment.

Sequences producing significant alignments:	Score (Bits)	E Value
gi|56693191|ref|YP_164778.1| orf143 [Lactobacillus plantarum ...	42.4	0.006

```
>  gi|56693191|ref|YP_164778.1|    orf143 [Lactobacillus plantarum
bacteriophage LP65]
 gi|54633692|gb|AAV35963.1|    orf143 [Lactobacillus plantarum bacteriophage
LP65]
Length=354

 Score = 42.4 bits (98),  Expect = 0.006
 Identities = 24/59 (40%), Positives = 31/59 (52%), Gaps = 2/59 (3%)
 Frame = +1

Query  16   DGVRKTKSVALLVAQAYLAPPKNHLYNSMIHLDGDRSNLNARNLMWRPRWYAIRYHQMF  192
            DG   TK V  LVAQ ++  P N     + H DGD+SN N  NL W    Y  +Y ++F
Sbjct  63   DGKYVTKKVHRLVAQTFIPNPNN--MPQVNHKDGDKSNNNVSNLEWCTNSYNDKYKKVF  119
```

Figure 3.49 BLASTx search result: alignment between the translated *Hind*III – *Pst*I fragment and the HNH endonuclease from *Lactobacillus plantarum* bacteriophage LP65.

4.8.1.6.4c ORF prediction

On the *Hind*III – *Pst*I fragment one ORF was predicted, as shown in Figure 3.50, and it corresponded to the HNH endonuclease from *Lactobacillus plantarum* bacteriophage LP65. GC content was 48.6%.

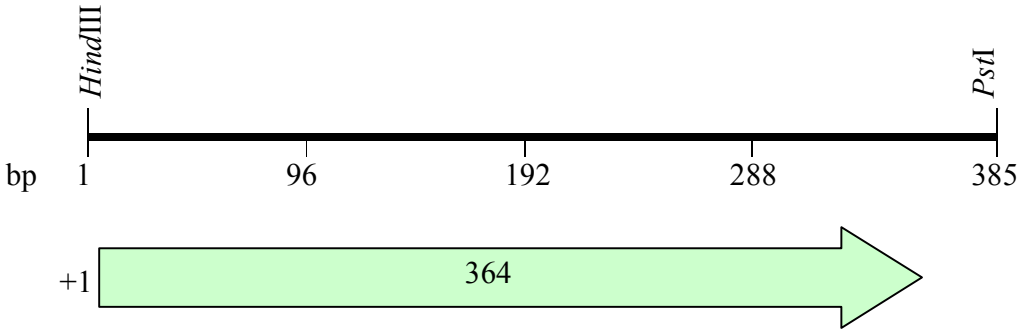


Figure 3.50 Predicted ORF on the *Hind*III – *Pst*I fragment.

The numbers inside and at the beginning of the arrow indicate the length and the reading frame of the ORF, respectively. This ORF corresponded to the HNH endonuclease of *Lactobacillus plantarum* phage LP65.

3.8.1.7 Clone 16 of Fairland1 *Pst*I library

3.8.1.7.1 DNA sequence

Subcloning of the insert from clone 16 of Fairland1 *Pst*I library was not conducted due to its size of ~2100 bp, and it was directly cloned into pUC18. Sequences of the front and back regions are shown in Figures 3.51 and 3.52, respectively.

ctgcaggtgc ataccgttat acatgctcac tagcctgttt cctgtgtgca aatgttatgc
cgcatattct cgttggtgtc catcgggacg agaatctcag gctcttcgtt ggtcaggacc
tcacgaacgc tgtgctcgtg ttcgcgacgc tcttcctctt ctgcaatcag ctttgcagcc
ttgaacgagg cctggccggc caagatggtt gccgtaacag cacctgtaac agccacagct
gtcaggatgg cgggggaatt tggaaatggt gaaccgttcg aggttctttg ccagatcgga
aatcgacatt gtttactcca gtttagtctt gaatgatggg tgtcgtgcta agtcaagcgt
cgtaatatgg catttgacgt tgttcggctt cttctgcctc gaggagctct gcggctttcc
ggtattcagc aaagaagcgt gtggacacaa cccatgtggc tgcgaggatg acggtgaaga
tgacggatac agccaagatc ccaagaccca agaggaggta gaggaatatg gtccatacga
actcgagaac gctcatgctg ctacgctttc ttccgggatt tcattgtaga accgatcgat
cattgcaaag aactcctcag cggtctgttc cttgctacgt cgatggcgcg gtgcgggacg
gggagggctg gtgtacttgg ggttagtccc atttaccttg gcgcgacgac gatcagtgac
gagaacggtt gatgtcacca atgccaacgg aagtccgatg accatgaatg c

Figure 3.51 DNA sequence of the front region of the insert from clone 16 of Fairland1 *Pst*I library.
The *Pst*I site is shown in blue.

tgatcgttcc agattgatct ttcatgatga gcctttcaaa tatggtgggg gagaggatgt
aaaagacaaa gcttaaatac cgtgctaggg tatttagctt ttgggccttt acttgggttt
tgtagccgga caattttcgg tgcattcatg tacagcgcta ccaggaaggt aacggtgaag
caatggtgca atgatggtgg tcatggtaca tcctctcata gggggtctca ttataggatg
tgtttttct gcgagacaaa aatgaaaggg cgagtactgt taagtagctc atctccttac
cacggttgtg gtgacaacat tgaagttgtc tgagcagttt tcctctcctt tcattatagg
gatgtgtttt ttctgcgact ctaatctcgc tcgggatctt gcttgtacaa ggcgcccttg
gtctcggatg atcgtctcaa gagcaatggc gtcatactcg ctcattttcg acaccgttct
tgatcagaat atcggcgtag tagacaagca tagcttcgag ggtgtctgct tccttttttc
gatcgtcgct gatcgtcaaa gatcttgaac agaagtgccg tctcgagacc tacgacaaat
atgagcacca gatcgaacgt gtccttgcgt aaccatctca tgtgactcct taagaaaact
gagaaaccgt gttagggtct ctcaggtcgg atcgactgtg cgaatgcggg ttgcatcttt
tgcgaactcg gtgctgctgc attcccagtt cacgaggttg taaaagcgac gagaccatga
tatccgtcat aacactgcag

Figure 3.52 DNA sequence of the back region of the insert from clone 16 of Fairland1 *Pst*I library.
The *Pst*I site is shown in blue.

3.8.1.7.2 DNA sequence analysis

3.8.1.7.2a BLASTn result

DNA sequences of the front and back regions of the *Pst*I 2100 bp insert had no nucleotide match.

3.8.1.7.2b BLASTx result

Translated sequences of the front and back regions of the *Pst*I 2100 bp insert had no protein match.

3.8.1.7.2c ORF prediction

Two ORFs in the front, and one ORF in the back of the *Pst*I 2100 bp fragment were predicted, as shown in Figures 3.53 and 3.54, respectively. GC contents for the front and back regions were 52.9% and 46.3%, respectively.

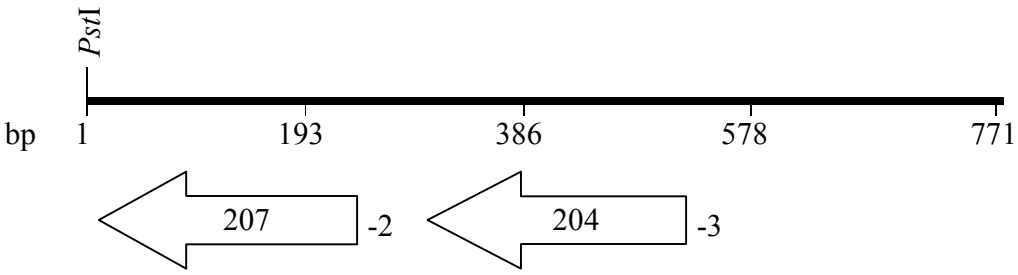


Figure 3.53 Predicted ORFs on the front region of the *Pst*I 2100bp insert.
The numbers inside and at the beginning of each arrow indicate the length and the frame of the ORF, respectively.

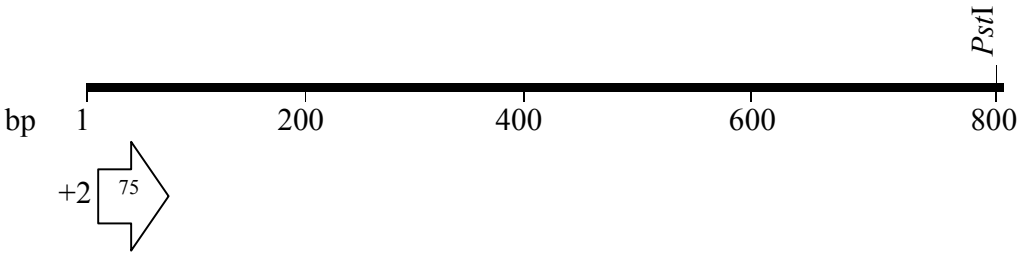


Figure 3.54 Predicted ORF on the back region of the *Pst*I 2100 bp insert.

The numbers inside and at the beginning of the arrow indicate the length and the frame of the ORF, respectively.

3.8.2 Analysis of clones from Perouges1 libraries

Clone 1 from Perouges1 *Hind*III (complete digest) library, and clone 7 from *Hind*III (partial digest) library were selected for analysis.

3.8.2.1 Clone 1 of Perouges1 *Hind*III (complete digest) library

3.8.2.1.1 Restriction map

Single and double digestions for the construction of a restriction map are shown in Figure 3.55. The restriction map is shown in Figure 3.56

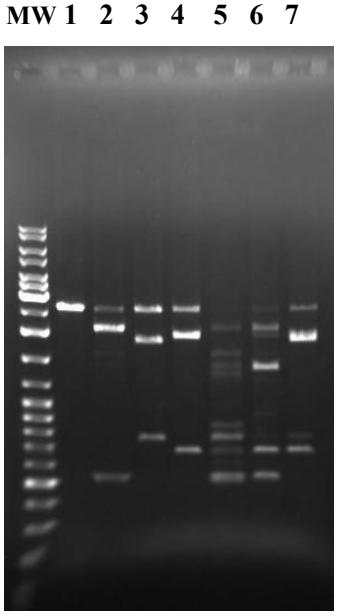


Figure 3.55 Single and double digestions of purified insert from clone 1 of Perouges1 *Hind*III (complete digest) library, on 0.8% gel.

Lane 1, undigested; 2, *Nsi*I; 3, *Sfu*I; 4, *Cla*I; 5, *Nsi*I + *Sfu*I; 6, *Nsi*I + *Cla*I; 7, *Sfu*I + *Cla*I.

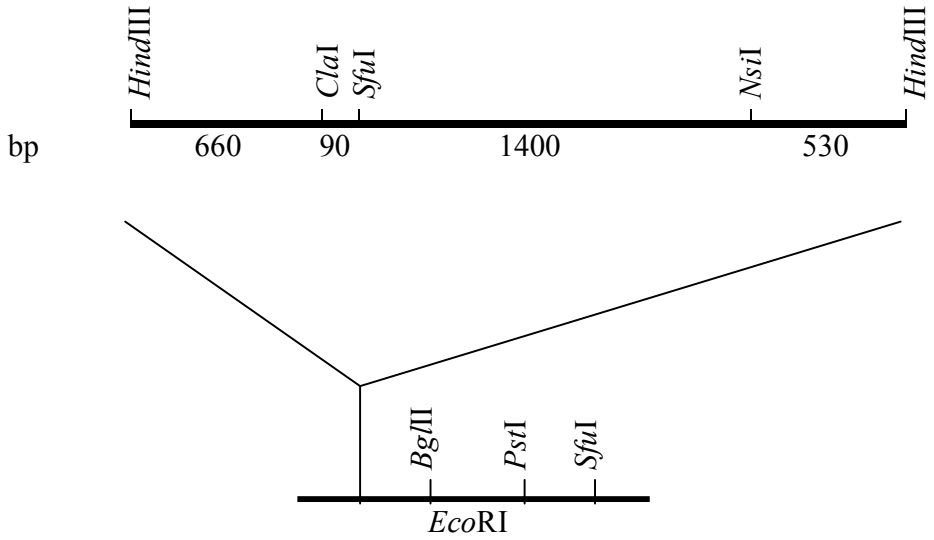


Figure 3.56 Restriction map of the insert from clone 1 of Perouges1 *Hind*III (complete digest) library.
Orientation of the insert with respect to the suicide gene of pDA71 was identified by digesting the clone with *Sfu*I, which released a fragment of ~2500 bp (gel not shown).

3.8.2.1.2 Screening of subclones

DNA fragments generated by single digestions of the purified insert with *Sfu*I and *Nsi*I, and double digestion with both enzymes, were subcloned into pDA71, as shown in Figure 3.57.

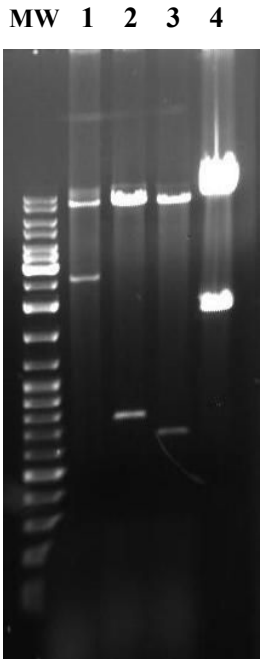


Figure 3.57 Subcloned DNA fragments from clone 1 of Perouges1 *Hind*III (complete digest) library, on 0.8% agarose gel.

As the purified insert was digested with *Nsi*I and ligated to the *Pst*I site of pDA71, resulting subclone inserts were released by different endonucleases in flanking restriction sites, as indicated below, thereby increasing their sizes.

Lane 1, original clone 14, insert released by *Hind*III; 2, *Hind*III – *Sfu*I 750 bp fragment; 3, *Nsi*I - *Hind*III 530 bp fragment, insert released by *Sfu*I + *Hind*III; 4, *Sfu*I – *Nsi*I 1400 bp fragment, insert released by *Sfu*I + *Hind*III.

Retention of antimicrobial activity in these subclones was tested in *R. erythropolis* SQ1. Transformants/µg of DNA for each subclone is shown in Table 3.24.

Table 3.24 Transformants/µg of DNA for each subclone from clone 1 of Perouges1 *Hind*III library.

Restriction sites at the ends of subclone insert	**Subclone insert size (bp)**	**# transformants ($\times 10^4$)/µg of DNA**
*Hind*III – *Sfu*I	750	1.9
*Nsi*I - *Hind*III	530	0.56
*Sfu*I – *Nsi*I	1400	< 0.0017
Vector only	NA	1.56

Subclone with transformation efficiency of < 0.01×10^4/µg of DNA is highlighted.

4.8.2.1.3 DNA sequence

The *Sfu*I – *Nsi*I 1400 bp fragment was cloned into the *Acc*I and *Pst*I site of pUC18. The sequence is shown in Figure 3.58.

ctgcatatca cttatctgac tcaaagtatg agtatgagcg gtgggagcaa aggtagtagg
cttatcggta atatcagccc atagatgagt atgactggta ttagctttac cagctaatga
ggcggttagt ccagtaacat cagcttctac gtgtgagtga gcactcggag cgaatgtact
aggaattcct gtcaaagtat tgtaggcatg agtatgggca ctaggagtaa atgtagcagg
aatgctggta atagcactcc aatctacgcc aacatttaca gcaccgagac catcaggact
cacaccattt actgtcttaa tgtactcggg atgagtatga gaagtacttg acttaccact
tagtgcagtt gcagtagcat tactgatagg cttattcaaa tcggaggtat tatctacgtt
acccaatccg atatcagcct tggtgctagg aatggtaggc ttggcagtaa ggtcgttgta
tgatccgctg gtagcgacag catgcaacga agtaatgttc accttaagat taagtgcagt
ttgtgtagca cttgatacag gcttattaac gtctgaggta ttgtctacgt tgcctagtcc
tacggcaccc ttggttaatc cggcaggctg gacactagta tcagcctttg atagcgtggt
acgtacgtca ggatgcaact tagctgatgt aatcttatcg ttaccaatga caatagcgtt
ggcattaccg gctaggtcac cagcgataag aatcttacca aaaacacttt gataggcgtg

gtacgtacgt caggatgcaa cttagctgat gtaatcttat cgttaccaat gacaatagcg
ttggcatttc cggctaggtc accagcggat aagaatctta ccctttttact tcgggtggta
gcatcaacaa cgccaccctc actaccgtcc ccggactggt agcgcaaaag gtgtttgtgg
taccatcatc gtaatgtacg gtaagctacg agtccctgtg tggtaatctc aacaatgccc
gcgtccagtg tcacccttgg tgattaatga aaccaggcac ctcggcggca ccaaaggaag
tcattcaagt caataggggc cggtaccagc attaaacagg tacactagca acacgctcgc
cccagctagg aacaacccgg tagggtccag ccgttgttag gagttatgcc aggctgatta
gtaatgatag cttcaaatga gaacgtaccg tctggttcaa tttctacacg ctgaggatag
atgcttgaca gatcaaggat caacccggta ctgtctacac gtatgctctt agcactagga
taaaccatca cgaaaccgtg cttaggacgt gacttgtact gctgatctat ttgatcaaaa
tacaggtcag cgtactggcc tgtgacgatt cgggatgcaa ttcctgctgg ataactcata
catcaatttt aacacaaaac aaaggaacga gattaaatgg aactcagtgg atacgac**ttc**
gac

Figure 3.58 DNA sequence of the *Sfu*I – *Nsi*I 1400 bp fragment.
The *Pst*I – *Nsi*I hybrid site and the AccI – *Sfu*I hybrid site are shown in blue and green, respectively.

3.8.2.1.4 DNA sequence analysis

3.8.2.1.4a BLASTn result

The *Sfu*I – *Nsi*I fragment had no nucleotide match.

3.8.2.1.4b BLASTx result

Translated sequence of the *Sfu*I – *Nsi*I fragment had several protein matches. It aligned to a putative head decoration protein of prophage MuMc02 from *Roseobacter* sp. MED193, with an E value of 8e-7. Another alignment was found to a putative

phage-related tail protein from *Xanthomonas axonopodis*, with an E value of 4e-5. It also aligned to gp4 from mycobacteriophage Bethlehem and gp133 from mycobacteriophage Catera, with E values of 3e-4 and 4e-4, respectively. These two proteins were also hypothetical. Lastly, an alignment was found to a putative cell wall surface anchor family protein from *Streptococcus agalactiae*, with an E value of 5e-4. All of alignments were in the same reading frame. BLASTx search results are shown in Table 3.25 and Figure 3.59. Features on these proteins are shown in Appendix D.

Table 3.25 BLASTx search result: proteins whose sequences aligned to the translated *Sfu*I - *Nsi*I fragment.

Sequences producing significant alignments:	Score (Bits)	E Value
gi|86139877|ref|ZP_01058443.1| prophage MuMc02, head decorati...	58.5	8e-07
gi|21107209|gb|AAM35948.1| phage-related tail protein [Xantho...	52.8	4e-05
gi|40769419|gb|AAR89725.1| gp4 [Mycobacteriophage Bethlehem]	50.1	3e-04
gi|109393343|ref|YP_656141.1| gp133 [Mycobacteriophage Catera...	49.7	4e-04
gi|22534491|gb|AAN00330.1| cell wall surface anchor family pr...	49.3	5e-04

> gi|86139877|ref|ZP_01058443.1| prophage MuMc02, head decoration protein, putative [Roseobacter sp. MED193]
gi|85823506|gb|EAQ43715.1| prophage MuMc02, head decoration protein, putative [Roseobacter sp. MED193]
Length=419

Score = 58.5 bits (140), Expect = 8e-07
Identities = 57/189 (30%), Positives = 82/189 (43%), Gaps = 24/189 (12%)
Frame = -1

```
Query  627  PAGLTKGAVGLGNVDNTSDVNKPVSSATQTALNLKVNITSLHAVATSGSYNDLTAKPTIP  448
            PA L  G V    VD+T  + K              N TS+ A+A SG +   T   TI
Sbjct  18   PAALKSGEVAHNEVDDTLYIGKGDDGGG--------NATSIVAIAGSGGFVAKTGTQTIA  69
```

```
Query  447  STKAD--IGLGNVDNTSDLNKPISNATATALSGKSSTSHTHPEYIKTVNGV--SPDGLGA  280
              K    +   + D  +  +    +   T L GK++TSH+H   I  V G+  + DG  A
Sbjct  70   GKKTFSLVPTASQDAAAGSDLVRKSQLDTLLGGKANTSHSHA--IADVTGLQGALDGKAA  127

Query  279  V---NVGVDWSAITSIPATFTPSAHTHAYNTLTGI-------PSTFAPSAHSHVEADVTG  130
             V   +   + S  TS   T   +A   A +T  G+        + FA +AH H  +DV+G
Sbjct  128  VSHDHTAAEISDSTSAGRTLLKAADVAAQHTALGLGTAALMSSTAFAAAAHGHAISDVSG  187

Query  129  LTASLAGKA  103
             L  +L GKA
Sbjct  188  LQTALNGKA  196


 Score = 42.0 bits (97),  Expect = 0.075
 Identities = 20/36 (55%), Positives = 25/36 (69%), Gaps = 0/36 (0%)
 Frame = -1

Query  171  APSAHSHVEADVTGLTASLAGKANTSHTHLWADITD  64
             A ++HSH  ADVTGL  +L GKA  SH H  A+I+D
Sbjct  104  ANTSHSHAIADVTGLQGALDGKAAVSHDHTAAEISD  139


 Score = 40.8 bits (94),  Expect = 0.17
 Identities = 39/154 (25%), Positives = 69/154 (44%), Gaps = 15/154 (9%)
 Frame = -1

Query  552  SATQTALNLK----VNITSLHAVATSGSYNDLTAKPTIPSTKADIGLGNVDNTSDLNKPI  385
             +A  TAL L     ++ T+  A A   + +D++   T  + KA +   +   T
Sbjct  154  AAQHTALGLGTAALMSSTAFAAAAHGHAISDVSGLQTALNGKAPLASPSFTGTP------  207

Query  384  SNATATALSGKSSTSHTHPEYIKTVNGVSPDGLGAVNVGVDWSAITSIPATFTPSAHTHA  205
               A  TA  G ++T      ++++   ++  G G + +   + +           A   A
Sbjct  208  --AAPTAAGGTNTTQIATTAFVQSA--IASFGAGDM-LKATYDSDNDGKVDAAELADAVA  262

Query  204  YNTLTGIPSTFAPSAHSHVEADVTGLTASLAGKA  103
             +  +TG P+TF PSAH+H  + VTGL ++L  KA
Sbjct  263  WTGVTGKPATFPPSAHNHPISQVTGLQSALDAKA  296


>  gi|21107209|gb|AAM35948.1|    phage-related tail protein [Xanthomonas
axonopodis pv. citri str. 306]
 gi|21241830|ref|NP_641412.1|    phage-related tail protein [Xanthomonas
axonopodis pv. citri str. 306]
Length=686

 Score = 52.8 bits (125),  Expect = 4e-05
 Identities = 37/93 (39%), Positives = 49/93 (52%), Gaps = 10/93 (10%)
 Frame = -1

Query  600  GLGNVDNTSDVNKPVSSATQTALNLKVNITSLHAVATSGSYN---DLTAKPTI--PS---  445
             G  ++DNT  V      +    L +  +I+ L  +AT  S +    +T KP    PS
Sbjct  291  GWDHIDNTEQVTSVAGKSGAVQL-VPSDISGLGTLATRDSVDFSSHVTNKPASYPPSGHM  349

Query  444  -TKADIGLGNVDNTSDLNKPISNATATALSGKS  349
```

```
             TK D+GL NVDNTSD NKP+SNA   AL  K+
Sbjct  350  HTKGDVGLSNVDNTSDANKPVSNAQQAALDTKA  382


 Score = 50.1 bits (118),  Expect = 3e-04
 Identities = 28/54 (51%), Positives = 34/54 (62%), Gaps = 2/54 (3%)
 Frame = -1

Query  669  DVRTTLSKADTSVQPAGL--TKGAVGLGNVDNTSDVNKPVSSATQTALNLKVNI  514
            D  + ++    S  P+G   TKG VGL NVDNTSD NKPVS+A Q AL+ K  I
Sbjct  331  DFSSHVTNKPASYPPSGHMHTKGDVGLSNVDNTSDANKPVSNAQQAALDTKAPI  384


>  gi|40769419|gb|AAR89725.1|  gp4 [Mycobacteriophage Bethlehem]
Length=356

 Score = 50.1 bits (118),  Expect = 3e-04
 Identities = 27/79 (34%), Positives = 38/79 (48%), Gaps = 26/79 (32%)
 Frame = -1

Query  264  DWSAITSIPATFTPSAHTH-------------------------AYNTLTGIPSTFAPS  163
            DW A++++P  F PSAH H                          A++ +TG P+TF PS
Sbjct  113  DWDALSNVPVEFPPSAHDHVAADVTDLDAAIADYLADNPPESSSVAWDDVTGKPTTFTPS  172

Query  162  AHSHVEADVTGLTASLAGK  106
             H+H  A+VTGL  +L  K
Sbjct  173  THTHSIANVTGLQTALDEK  191


 Score = 38.1 bits (87),  Expect = 1.1
 Identities = 21/56 (37%), Positives = 28/56 (50%), Gaps = 1/56 (1%)
 Frame = -1

Query  267  VDWSAITSIPATFTPSAHTHAYNTLTGIPSTFAPSA-HSHVEADVTGLTASLAGKA  103
            V W  +T  P TFTPS HTH+   +TG+ +          V+A V   TA+L   A
Sbjct  157  VAWDDVTGKPTTFTPSTHTHSIANVTGLQTALDEKLDEDAVDARVAVGTAALVDSA  212


>  gi|109393343|ref|YP_656141.1|    gp133 [Mycobacteriophage Catera]
 gi|91981165|gb|ABE67880.1|    gp133 [Mycobacteriophage Catera]
 gi|29566641|ref|NP_818207.1|    gp134 [Mycobacterium phage Bxz1]
 gi|29425366|gb|AAN16790.1|    gp134 [Mycobacterium phage Bxz1]
Length=489

 Score = 49.7 bits (117),  Expect = 4e-04
 Identities = 23/36 (63%), Positives = 29/36 (80%), Gaps = 0/36 (0%)
 Frame = -1

Query  456  TIPSTKADIGLGNVDNTSDLNKPISNATATALSGKS  349
            T+  TK+D+GL NVDNTSD  KPIS+ATA AL+ K+
Sbjct  186  TVTLTKSDVGLNNVDNTSDAAKPISSATALALAAKA  221
```

```
 Score = 48.1 bits (113),  Expect = 0.001
 Identities = 60/216 (27%), Positives = 84/216 (38%), Gaps = 27/216 (12%)
 Frame = -1

Query  690  TSAKLHPDVRTTLSKADTSVQ----PAGLTKGAVGLGNVDNTSDVNKPVSSATQTALNLK  523
            T+ K + D R   +    SV        LTK  VGL NVDNTSD  KP+SSAT  AL  K
Sbjct  161  TTNKYYTDARAAAAAPVQSVAGRTGTVTLTKSDVGLNNVDNTSDAAKPISSATALALAAK  220

Query  522  VNITSLHA------VATSGSYNDLTAKPTIPST-----KADIG-LGNVDNTSDLNKPISN  379
             ++ S         +     Y DL++ P           AD G L   D+ ++  + I +
Sbjct  221  ADLVSGRVPPEQLPIPPVAEYADLSSFPAPGEAGTLYIAADTGNLYRWDSVAENFELIVS  280

Query  378  ATATALS-------GKSSTSHTHPEYIKTVNGVSPDGLGAVNVGVDWSAITSIPATFTPS  220
            AT    +       G ++   T+      V  +  D  G V  G D      +  T TP+
Sbjct  281  ATGGGAASTDELPEGGTNLYFTNARVTARVEAMFGDAAGTVTQGND----PRLSNTRTPT  336

Query  219  AHTHAYNTLTGIPSTFAPSAHSHVEADVTGLTASLA  112
             +T +   L  +  T    A   V  D       LA
Sbjct  337  DNTVSTAKLVDLAVTTDKIADEAVTLDKLAEDVQLA  372


>  gi|22534491|gb|AAN00330.1|    cell wall surface anchor family protein
[Streptococcus agalactiae 2603V/R]
 gi|22537606|ref|NP_688457.1|    cell wall surface anchor family protein
[Streptococcus agalactiae 2603V/R]
Length=970

 Score = 49.3 bits (116),  Expect = 5e-04
 Identities = 51/241 (21%), Positives = 100/241 (41%), Gaps = 10/241 (4%)
 Frame = -1

Query  789  RTYHAYQSVFGKILIAGDLAGNANAIVIGNDKITSAKLHPDVRTTLS---KADTSVQPAG  619
            +TY +  S +G   I+GD   NA   ++     TSA        ++S    A TS   +
Sbjct  615  QTYGSKYS-YGHTNISGDSDANAEIKLLSESASTSASTSASTSASMSASTSASTSASMSA  673

Query  618  LTKGAVGL---GNVDNTSDVNKPVSSATQTALNLKVNITSLHAVATSGSYNDLTAKPTIP  448
             T  +       ++  ++  +   S++  T+ +   ++++  + +TS S +  T+  T
Sbjct  674  STSASTSASTSASMSASTSASTSASTSASTSASMSASTSASTSASTSASTSA  733

Query  447  STKADIGLGNVDNTS---DLNKPISNATATALSGKSSTSHTHPEYIKTVNGVSPDGLGAV  277
            ST A +      +TS     +   S + +T+ S  +STS +    +      S     +
Sbjct  734  STSASMSASTSASTSASTSASTSASMSASTSASTSASTSASMSASTSASTSASTSA  793

Query  276  NVGVDWSAITSIPATFTPSAHTHAYNTLTGIPSTFAPSAHSHVEADVTGLTASLAGKANT  97
            +     SA TS   + + SA T A  + +   ST A ++ S   +     +AS++   +
Sbjct  794  STSASMSASTSASTSASTSASTSASMSASTSASTSASTSASMSASTSASTSASMSASTSA  853

Query  96   S  94
            S
Sbjct  854  S  854
```

Figure 3.59 BLASTx search result: alignment between the translated *Sfu*I – *Nsi*I fragment and the various proteins.

3.8.2.1.4.3 ORF prediction

Two ORFs were predicted on the *Sfu*I – *Nsi*I fragment, as shown in Figure 3.60. One of them corresponded to the various proteins to which the alignments were found. GC content was 46.2%.

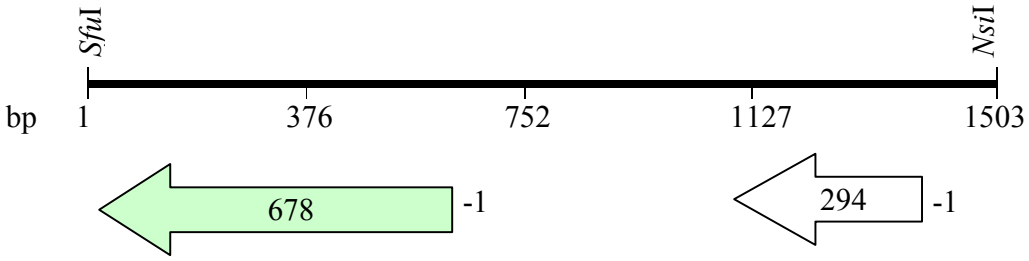


Figure 3.60 Predicted ORFs on the *Sfu*I – *Nsi*I fragment.
The numbers inside and at the beginning of each arrow indicate the length and the frame of the ORF, respectively. The ORF corresponding to the various proteins to which the alignments were found, is shown in light green.

3.8.2.2 Clone 7 of Perouges1 *Hind*III (partial digest) library

3.8.2.2.1 Restriction map

Single and double digestions for the construction of a restriction map are shown in Figure 3.61. The restriction map is shown in Figure 3.62.

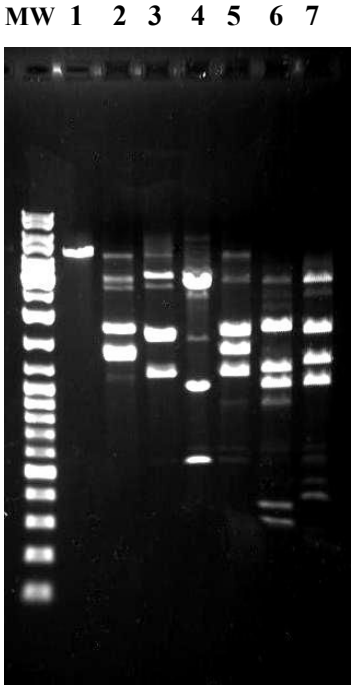


Figure 3.61 Single and double digestions of purified insert from clone 7 of Perouges1 *Hind*III (partial digest) library, on 0.8% gel.

Lane 1, undigested; 2, *Nsi*I; 3, *Sfu*I; 4, *Acc*I; 5, *Nsi*I + *Sfu*I; 6, *Nsi*I + *Acc*I; 7, *Sfu*I + *Acc*I.

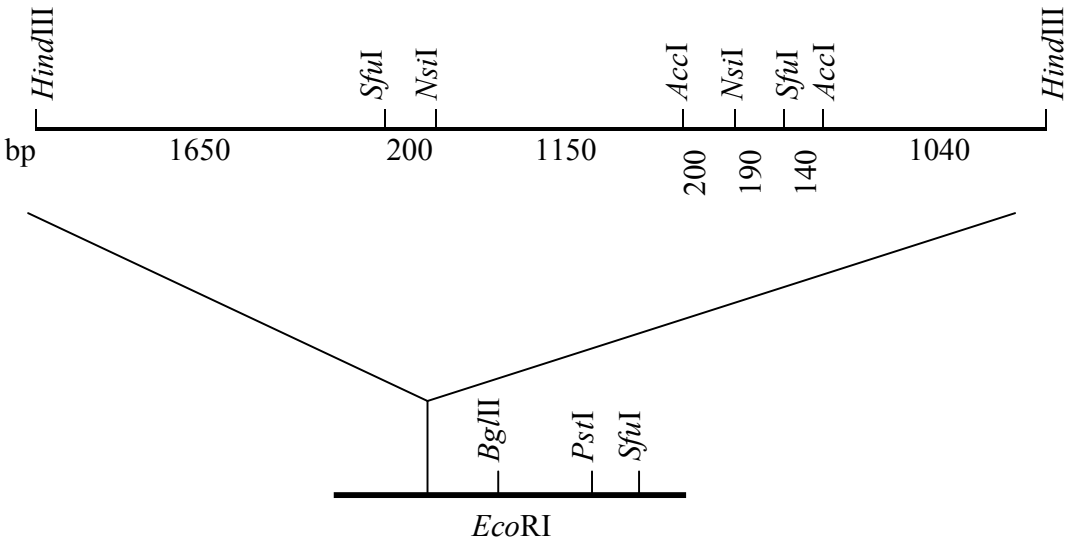


Figure 3.62 Restriction map of clone 7 from Perouges1 *Hind*III (partial digest) library.

The orientation of the insert with respect to the suicide gene of pDA71 was identified by digesting the clone with *Sfu*I, which released two fragments of ~1700 bp (gel not shown).

3.8.2.2.2 Screening of subclones

DNA fragments generated by single digestions of the purified insert with *Sfu*I and *Nsi*I were subcloned into pDA71, as shown in Figure 3.63.

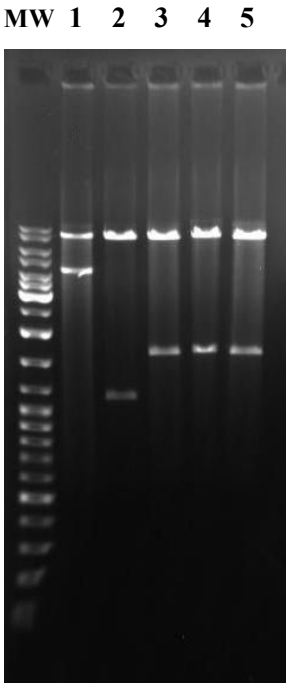


Figure 3.63 Subcloned DNA fragments from clone 7 of Perouges1 *Hind*III (partial digest) library, on 0.8% agarose gel.

As the purified insert was digested with *Nsi*I and ligated to the *Pst*I site of pDA71, resulting subclone inserts were released by different endonucleases in flanking restriction sites, as indicated below, thereby increasing their sizes.

Lane 1, original clone 7, insert released by *Hind*III; 2, *Hind*III – *Sfu*I 1320 bp fragment; 3, *Sfu*I - *Hind*III 1650 bp fragment; 4, *Nsi*I 1350 bp fragment, insert

released by *Bgl*II + *Sfu*I;5, *Hind*III – *Nsi*I 1510 bp fragment, insert released by *Hind*III + *Sfu*I.

Retention of antimicrobial activity in these subclones was tested in *R. erythropolis* SQ1. Transformants/µg of DNA for each subclone is shown in Table 3.26.

Table 3.26 Transformants/µg of DNA for each subclone from clone 7 of Perouges1 *Hind*III library.

Restriction sites at the ends of subclone insert	**Subclone insert size (bp)**	**# transformants ($\times 10^4$)/µg of DNA**
*Hind*III - *Sfu*I	1320	0.03
*Sfu*I - *Hind*III	1650	0.29
*Nsi*I	1350	0.06
*Hind*III - *Nsi*I	1510	0.15
Vector only	NA	0.93

Subclone with the most retention of antimicrobial activity is highlighted.

The inhibitory activity was retained the most in the *Hind*III – *Sfu*I 1320 bp fragment, although its transformation efficiency was > 0.01×10^4/µg of DNA. This fragment was cloned into the *Hind*III and *Acc*I sites of pUC18 and sequenced.

3.8.2.2.3 DNA sequence

Sequence of the entire *Hind*III - *Sfu*I 1320 bp fragment is shown in Figure 3.64.

aagctttaag catgtcgctt ttatgcgagg tagctgtacc gaagatgcag aactctatcg
agaggactgg ggactatacc taggagccac agagcaccag gccccggac gcgacggatt
ccaggcacgc gcattggata cagaagagct tgaaggcgca gactgggcac aggagcagca
tatcgaaaac ccacgtctag gatggggaaa gcgtgaaaac ttctacatca ctgtagccgt
agaatcattt cgatatcggc acaggagaaa aggtagttgt accagcaatc ctatctcgtc
caatcggaca cgaattcgta gggaaagatc attcagttca acgatgagtc taatggtgag
ggaatcactg gcaagactta ctggctaact agaacaggtc agaagaccga aactaagtgg
cttatcgagg aagagaaaga tccagacaag cagatcgtat tcggagatat cgctccgtac
gatttgaagg agacagcagt tatccacgta ccttatgaca agcagaagga attctacatg
cgtaaggcag atatcgaggg tgcacagaag gccgcagatg cccgtgcagc taagagaaac
ggtggagaga ccaacaacag ctcatctaac ggagactcac agggtggtaa caagcccgca
gcatccggtg gcaagcgtta tagctggtaa tccagtttga tcggggaggc tctacaatac
gtggagcctc cctaaactaa tactagacac gaccaaggtc gctaaagagg agaagaatga
ctaaattatg ggatggtaaa gagttcgttg aacttcataa ccacacttat tacagcccac
tggacggatt gagttcgcca gaggtatact ttattcgagc acatgagcta ggactcacta
agctcgccat tacagaccac ggcacattgg ccgggtatcg ccactttcag agggcagcca
aagaagcagg cattgatcct attcttggcg tagaagcata cattagtcag accaaccgct
ttgaccgtcg taccaaggca aatcgtcagg acggcaccga ggtctacaat cacattatcc
tgttggcaaa gaatgatgcc ggggtgcaca gccttaacgt cttgtcagag attggctgga
ccgaaggatt ctacaagaag ccacgtatcg atttcgac

Figure 3.64 DNA sequence of the *Hind*III – *Sfu*I fragment.

The *Hind*III, *Acc*I, and *Acc*I – *Sfu*I hybrid sites are shown in blue, green, and red, respectively.

3.8.2.2.4 DNA sequence analysis

3.8.2.2.4a BLASTn result

The *Hind*III – *Sfu*I fragment had no significant match.

3.8.2.2.4b BLASTx result

A translated sequence of the *Hind*III – *Sfu*I fragment showed significant alignment to the alpha subunit of DNA polymerase III from various organisms. The highest E value of 2e-24 was observed with the protein from *Aquifex aeolicus*. BLASTx results are shown in Table 3.27 and Figure 3.65. Features of this protein are shown in Appendix D.

Table 3.27 BLASTx result: proteins whose sequences aligned to the translated *Hind*III – *Sfu*I fragment.

Sequences producing significant alignments:	Score (Bits)	E Value
gi|6014991|sp|O67125|DPO3A_AQUAE DNA polymerase III subunit a...	116	2e-24
gi|106886892|ref|ZP_01354219.1| DNA polymerase III, alpha sub...	109	2e-22
gi|73660858|emb|CAI83465.1| DNA polymerase III, alpha subunit...	108	6e-22
gi|88932976|ref|ZP_01138658.1| DNA polymerase III, alpha subu...	108	6e-22
gi|3676693|gb|AAC62072.1| DNA polymerase; DnaE [Treponema dentic	107	8e-22

Only the top five hits are shown.

```
>  gi|6014991|sp|O67125|DPO3A_AQUAE  DNA polymerase III subunit alpha
 gi|2983510|gb|AAC07087.1|    DNA polymerase III alpha subunit [Aquifex aeolicus
VF5]
 gi|15606309|ref|NP_213688.1|    DNA polymerase III alpha subunit [Aquifex
aeolicus VF5]
Length=1161

 Score =  116 bits (291),  Expect = 2e-24
 Identities = 59/127 (46%), Positives = 83/127 (65%), Gaps = 1/127 (0%)
 Frame = +3

Query  798   KEFVELHNHTYYSPLDGLSSPEEYFIRAHELGLTKLAITDHGTLAGYRHFQRAAKEAGID  977
             K+FV LH HT +S LDG    +E   +A E G   + ++DHG L G   F +A K  GI
Sbjct  3     KDFVHLHLHTQFSLLDGAIKIDELVKKAKEYGYKAVGMSDHGNLFGSYKFYKALKAEGIK  62

Query  978   PILGVEAYISQTNRFDRRTKANRQDGTEVYN-HIILLAKNDAGVHSLNVLSEIGWTEGFY  1154
```

```
             PI+G+EAY +  +RFDR+TK +  + T+ YN H+IL+AK+D G+ +L  LS + + EGFY
Sbjct  63    PIIGMEAYFTTGSRFDRKTKTSEDNITDKYNHHLILIAKDDKGLKNLMKLSTLAYKEGFY  122

Query  1155  KKPRIDF  1175
              KPRID+
Sbjct  123   YKPRIDY  129


>  gi|106886892|ref|ZP_01354219.1|  DNA polymerase III, alpha subunit
[Clostridium phytofermentans ISDg]
 gi|106765595|gb|EAT22347.1|  DNA polymerase III, alpha subunit [Clostridium
phytofermentans ISDg]
Length=1186

 Score =  109 bits (273),  Expect = 2e-22
 Identities = 56/123 (45%), Positives = 76/123 (61%), Gaps = 4/123 (3%)
 Frame = +3

Query  804   FVELHNHTYYSPLDGLSSPEEYFIRAHELGLTKLAITDHGTLAGYRHFQRAAKEAGIDPI  983
             F  LH HT +S LDG    +E   +A +LG+  LAITDHG + G   F RA    GI PI
Sbjct  3     FTHLHVHTEFSLLDGSGKIKEMAAQAKKLGMDALAITDHGVMYGVIDFYRACLAEGIKPI  62

Query  984   LGVEAYISQTNRFDRRTKANRQDGTEVYNHIILLAKNDAGVHSLNVLSEIGWTEGFYKKP  1163
             +G E Y++  +RFD+   A+     E Y+H++LLA+ND G H+L  +   G+TEGFY KP
Sbjct  63    IGCEVYVAPNSRFDKEGSASE----ERYHHLVLLAENDIGYHNLIKIVSRGFTEGFYYKP  118

Query  1164  RID  1172
             R+D
Sbjct  119   RVD  121


>  gi|73660858|emb|CAI83465.1|    DNA polymerase III, alpha subunit
[Dehalococcoides sp. CBDB1]
 gi|73749142|ref|YP_308381.1|    DNA polymerase III, alpha subunit
[Dehalococcoides sp. CBDB1]
Length=1170

 Score =  108 bits (269),  Expect = 6e-22
 Identities = 60/123 (48%), Positives = 77/123 (62%), Gaps = 6/123 (4%)
 Frame = +3

Query  804   FVELHNHTYYSPLDGLSSPEEYFIRAHELGLTKLAITDHGTLAGYRHFQRAAKEAGIDPI  983
             F  LH HT +S LDG+    +  +RA ELG+  LAITDHGT+ G   F + AKEAGI PI
Sbjct  2     FCHLHVHTEFSLLDGMCRISQLVVRAKELGMEALAITDHGTMHGIPRFYKEAKEAGIKPI  61

Query  984   LGVEAYISQTNRFDRRTKANRQDGTEVYNHIILLAKNDAGVHSLNVLSEIGWTEGFYKKP  1163
             LG E Y++  +R   +T A++      Y H++LLAKN+ G H+L  LS     EGFY KP
Sbjct  62    LGSEFYLAYGDR-TAKTAADKN-----YYHLVLLAKNNKGYHNLIQLSSRSHLEGFYYKP  115

Query  1164  RID  1172
             RID
Sbjct  116   RID  118
```

```
>  gi|88932976|ref|ZP_01138658.1|  DNA polymerase III, alpha subunit
[Dehalococcoides sp. BAV1]
 gi|88914829|gb|EAR34210.1|  DNA polymerase III, alpha subunit [Dehalococcoides
sp. BAV1]
Length=1170

 Score =  108 bits (269),  Expect = 6e-22
 Identities = 60/123 (48%), Positives = 77/123 (62%), Gaps = 6/123 (4%)
 Frame = +3

Query  804   FVELHNHTYYSPLDGLSSPEEYFIRAHELGLTKLAITDHGTLAGYRHFQRAAKEAGIDPI  983
             F  LH HT +S LDG+    +  +RA ELG+  LAITDHGT+ G   F + AKEAGI PI
Sbjct  2     FCHLHVHTEFSLLDGMCRISQLVVRAKELGMEALAITDHGTMHGIPRFYKEAKEAGIKPI  61

Query  984   LGVEAYISQTNRFDRRTKANRQDGTEVYNHIILLAKNDAGVHSLNVLSEIGWTEGFYKKP  1163
             LG E Y++  +R   +T A++      Y H++LLAKN+ G H+L  LS     EGFY KP
Sbjct  62    LGSEFYLAYGDR-TAKTAADKN-----YYHLVLLAKNNKGYHNLIQLSSRSHLEGFYYKP  115

Query  1164  RID  1172
             RID
Sbjct  116   RID  118


>  gi|3676693|gb|AAC62072.1|  DNA polymerase; DnaE [Treponema denticola]
Length=361

 Score =  107 bits (268),  Expect = 8e-22
 Identities = 56/124 (45%), Positives = 78/124 (62%), Gaps = 4/124 (3%)
 Frame = +3

Query  801   EFVELHNHTYYSPLDGLSSPEEYFIRAHELGLTKLAITDHGTLAGYRHFQRAAKEAGIDP  980
             +FV LH H+ +S LDG +S +    +A +LG T LA+TDHG + G  +F  A KE  I P
Sbjct  4     DFVHLHVHSDFSLLDGAASVKALAAKAEKLGQTALALTDHGNMFGAINFYNACKEKNIKP  63

Query  981   ILGVEAYISQTNRFDRRTKANRQDGTEVYNHIILLAKNDAGVHSLNVLSEIGWTEGFYKK  1160
             I+G E Y++  +RF+++       G + Y H+ILLAKN+ G  +L VL    G+TEG Y K
Sbjct  64    IIGCEFYVAPESRFEKK----ETPGGKKYYHLILLAKNETGYRNLMVLCSKGYTEGMYYK  119

Query  1161  PRID  1172
             PRID
Sbjct  120   PRID  123
```

Figure 3.65 BLASTx search result: alignment between the translated *Hind*III - *Sfu*I fragment and the alpha subunit of DNA polymerase III from various organisms.

3.8.2.2.4c ORF prediction

On the *Hind*III – *Sfu*I fragment one ORF was predicted, which corresponded to the alpha subunit of DNA polymerase III, as shown in Figure 3.66. GC content was 49.5%.

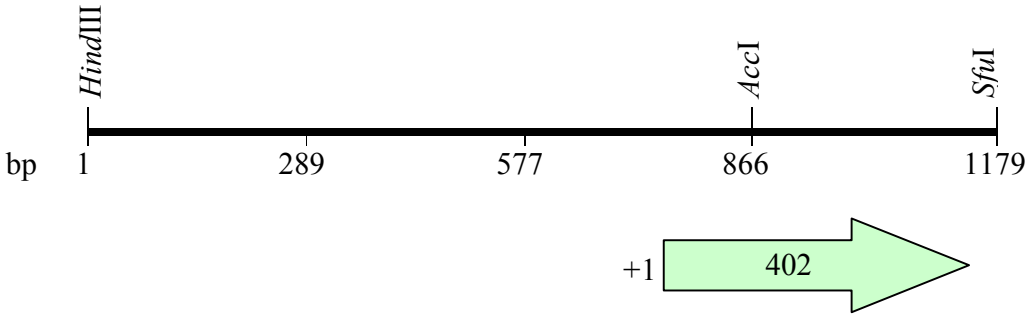


Figure 3.66 Predicted ORF on the *Hind*III – *Sfu*I fragment.
The numbers inside and at the beginning of the arrow indicate the length and the frame of the ORF.

3.8.3 Summary of DNA sequence analyses

Average length of ORF from the sequenced regions of Fairland1 and Perouges1 DNA was 342.6 bp. Average GC content of Fairland1 and Perouges1 DNA was 51.8% and 47.9%, respectively. A summary of all DNA sequence analyses is presented in Table 3.28.

Table 3.28 Summary of sequence analyses of phage DNA inhibitory to *Rhodococcus erythropolis* SQ1.

Phage	Library	Clone #	BLASTx alignment	E value	GenBank Accession#	# ORFs	# ORFs with no match
Fairland1	*Bgl*II	7	none		DQ981382 DQ981383	2	2
		8	thymidylate synthase complementing protein, *Streptomyces coelicolor*	2e-59	DQ981384	2	1
		10	none		DQ981385	2	2
		19	none		DQ981386 DQ981387	3	3
	*Pst*I	8 front	phage head maturation peptidase, *Mycobacterium* sp. MCS and KMS	5e-04	DQ981388	1	0
		8 back	phage capsid protein, *Streptococcus agalactiae*	4e-04	DQ981389	1	1
			dihydropteroate synthase, *Nocardioides* sp. JS614	0.003			
		14	HNH endonuclease, *Lactobacillus plantarum* phage LP65	0.006	DQ981390	1	0
		16	none		DQ981391 DQ981392	3	3
Perouges1	*Hind*III	1	head decoration protein, prophage MuMc02, *Roseobacter* sp. MED193	8e-07	DQ981393	2	1
			phage-related tail protein, *Xanthomonas axonopodis*	4e-05			
			cell wall surface anchor family protein, *Streptococcus agalactiae*	5e-04			
		7	α subunit, DNA polymerase III, *Aquifex aeolicus*	2e-24	DQ981394	1	0
					TOTAL	18	13

3.8.4 Modification and absence of restriction sites on phage DNA

As was shown in Figure 3.6D, Perouges1 genomic DNA was only susceptible to digestions by *Hind*III, *Bcl*I, *Sfu*I, and *Cla*I. However, DNA from clones 1 and 7 of its library in *E. coli* were digested by *Nsi*I and *Acc*I, which were used for the construction of their restriction maps (except for *Acc*I for clone 1, as it cut the insert many times). Cloned DNA turning sensitive to these endonucleases suggests that Perouges1 DNA had an unusual base or a modification. The recognition sequence of *Nsi*I is ATGCAT, while those of *Acc*I are GTATAC and GTCGAC. No obvious pattern was present in these sequences. Endonucleases, such as *Bgl*II, *Bam*HI, *Pst*I, and *Bmy*I, did not digest these inserts from the clones, whose total length was ~7.2 kb, so absence of restriction sites was probably the reason in these cases.

*Bam*HI and *Bmy*I digested neither Fairland1 genomic DNA (Figure 3.6A) nor the clones from its library in *E. coli*, which in total were ~13 kb, so absence of their restriction sites was definitely the reason.

4. DISCUSSION

4.1 Characteristics of novel nocardiophages

Although nocardioform infections are rare, treatment is challenging, as the pathogens are manifested in a variety of diseases. Number of reported cases is increasing mainly due to rise in number of immunocompromised HIV/AIDS patients. Both phylogenetically and pathogenically, they are closely related to *M. tuberculosis*, one of the leading causes of death worldwide. Nocardioforms themselves are also now considered as important human pathogens associated with a high mortality rate. Combination antibiotic therapy may no longer be a solution, as reported cases of multidrug resistant strains are increasing. However, novel antibacterial drug discovery faces challenges at a time when they are desperately needed.

Abundance of bacteriophages in nature and their genetic heterogeneity suggest a huge number of undiscovered lethal genes targeting essential host metabolism, which can be studied to identify novel bactericidal mechanisms. Nocardiophage isolation was therefore undertaken, with the aim of identifying novel genes inhibitory to the host.

Nocardiophages were readily isolated from various soil samples using *R. erythropolis* as the host. Most of these phages had no divalent ion requirement. All of the four selected phages – Fairland1, Wits1, Kanazawa1, and Perouges1- could infect other species of *Rhodococcus*, including the pathogenic *R. equi* (except for Wits1), confirming that studies using non-pathogenic species can be applied to closely related pathogenic ones.

Although there is strong pressure to incorporate DNA or RNA sequence information into phage classification, morphology is still the principal basis. As shown in Figure 4.1, 21 phage morphologies have been recorded (Ackermann, 2001). Characteristics and family name associated with each morphotype is summarized in Table 4.1.

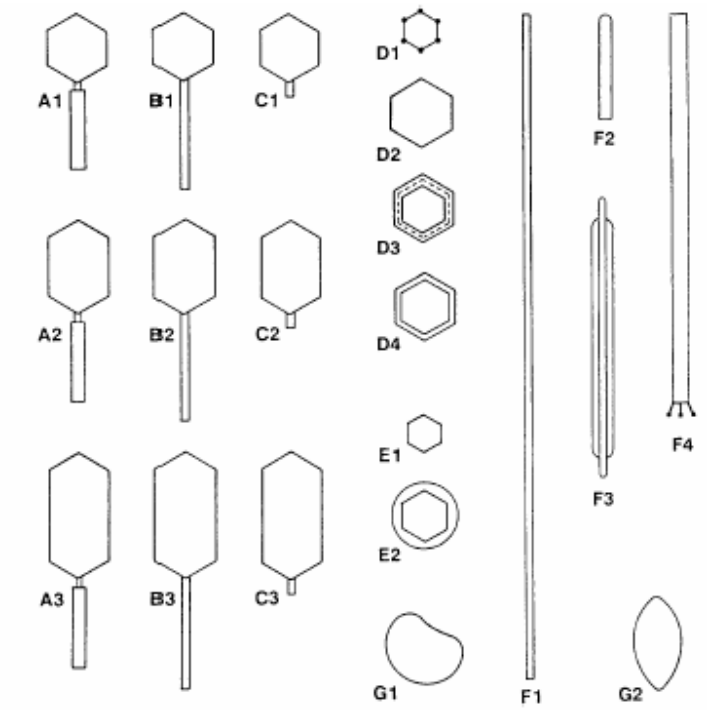


Figure 4.1 Phage morphotypes.
(Taken from Ackermann, 2001)

Table 4.1 Basic properties associated with phage morphotypes in Figure 4.1.

Morphotype	Shape	Nucleic acid	Family	Description
A1 to A3	Tailed	Linear dsDNA	*Myoviridae*	Tail contractile
B1 to B3			*Siphoviridae*	Tail long, noncontractile
C1 to C3			*Podoviridae*	Tail short
D1	Polyhedral	Circular ssDNA	*Microviridae*	Conspicuous capsomers
D3		Circular dsDNA, superhelical	*Corticoviridae*	Complex capsid, lipid
D4		Linear dsDNA	*Tectiviridae*	Lipid vesicle, pseudotail
E1		Linear ssRNA	*Leviviridae*	No envelope, no lipids
E2		Linear segmented dsRNA	*Cystoviridae*	Envelope, lipids
F1	Filamentous	Circular ssDNA	*Inoviridae*	Long filaments
F2				Short rods
F3		Linear dsDNA	*Lipothrixviridae*	Envelope, lipids
F4		Linear dsDNA	*Rudiviridae*	TMV-like
G1	Pleomorphic	Circular dsDNA, superhelical	*Plasmaviridae*	Envelope, lipids, no capsid
G2			*Fuselloviridae*	

Morphotype D2 was excluded as it is awaiting classification; TMV = Tobacco Mosaic Virus.

(Modified from Ackermann, 2001).

All four phages in this study were assigned to the family *Siphoviridae* in the order *Caudovirales*. This was not surprising, as the majority of recorded phages belong to this family. In fact, all *Rhodococcus* phages included in the records of the Félix

d'Hérelle Reference Center for Bacterial Viruses belong to *Siphoviridae* (Ackermann, 2001).

4.2 Phage genome size and GC content

Phage genome sizes ranged from 39 kb (Wits1) to 140 kb (Perouges1). In comparison to mycobacteriophages (Hatfull *et al*., 2006), Wits1 genome was smaller than the smallest sequenced mycobacteriophage genome (42 kb of phage Halo), while that of Perouges1 was smaller than the largest sequenced mycobacteriophage genome (156 kb of phage Bxz1). Average ORF length of 342.6 bp from Fairland1 and Perouges1 was significantly shorter than that from the mycobacteriophages, which is 600 bp (Hatfull *et al*., 2006). However, the ORFs in this work may be biased towards shorter sizes as each sequenced region was short.

Average GC contents of 51.8% from Fairland1 and 47.9% from Perouges1, although these may not be accurate representations of their genomes, are significantly lower than what is known of their host, which is > 60% (Geize *et al*., 2002; Stecker *et al*., 2003). GC contents of phages for *E. coli* (Miller *et al*., 2003), *M. smegmatis* (Hatfull et al., 2006), *S. pneumoniae* (Kwan *et al*., 2005), and *S. aureus* (Kwan *et al*., 2005) are very similar to those of their hosts. On the other hand, GC content of *P. aeruginosa* phages is significantly lower than that of the host (Kwan *et al*., 2006). Discrepancy in GC content between phage and host genomes may be due to 1) horizontal gene transfer from other hosts or phages, 2) recent invasion of phages into *R. erythropolis* from other hosts, or 3) presence of selective pressures which have maintained such GC content (Kwan *et al*., 2006).

4.3 Phage anti-restriction

Bacteria have evolved protective mechanisms against attacking phages at almost all stages of phage life cycle. One of these is the restriction and modification system, which attacks foreign DNA inside the cells while protecting their own. This function is important enough to have apparently evolved independently several times, as exhibited by the presence of different classes of restriction modification systems. In order to survive restriction, phages have evolved anti-restriction mechanisms including 1) blocking of restriction enzyme, 2) phage-encoded modification of DNA, 3) stimulation of host modification functions, 4) possession of unusual bases, 5) co-injection of protective proteins, 6) destruction of endonuclease co-factors, and 7) loss of restriction sites (Krüger and Bickle, 1983). Of these mechanisms, ones that are functional *in vitro* are the modification of DNA, presence of unusual bases, and absence of restriction sites. These have been mostly studied in *E. coli* phages and *B. subtilis* phages.

Modified bases refer to those whose modifications take place after the synthesis of normal DNA. *B. subtilis* phages SPβ and ϕ3T possess a methyltransferase gene that methylates the central cytosine in the sequence GGCC, which is the recognition sequence of endonuclease *Bsu*RI of *B. subtilis* (Cregg *et al*., 1980). *E. coli* phage Mu produces the Mom protein, which acetimidates ~15% of its adenine residues, thereby producing N^6-(1-acetamido)adenine, making its DNA resistant to type I and III enzymes and partially resistant to type II enzymes (Hattman, 1980). *E. coli* phages T2 and T4 encode a methyltransferase that methylates a fraction of adenine residues to produce 6-methylaminopurine (Gold *et al*., 1966). *Shigella sonnei* phage, DDVI, methylates the 7 position of guanine in about a quarter of its DNA, producing 7-methylguanine (Nikolskaya *et al*., 1979).

Unusual bases are those which are not adenine, cytosine, guanine, or thymine, and are produced at the level of nucleotide metabolism. Various unusual bases in phages

have been discovered, summarized in Table 4.2. Their structures are illustrated in Figure 4.2.

Table 4.2 Unusual bases in phage DNA.

Phage	**Host**	**Base change**	**% change**
T4	*E. coli*	5-hydroxymethylcytosine (glucosylated) replacing cytosine	100%
SPO1	*B. subtilis*	5-hydroxymethyluracil replacing thymine	100%
PBS1	*B. subtilis*	Uracil replacing thymine	100%
SP15	*B. subtilis*	5-dihydroxypentyluracil replacing thymine	41%
SP10	*B. subtilis*	α-glutamylthymine replacing thymine	15 – 20%
XP12	*Xanthomonas oryzae*	5-methylcytosine replacing cytosine	100%
S2L	*Synechococcus elongatus*	2-aminoadenine replacing adenine	100%
ϕW14	*Pseudomonas acidovorans*	α-putrescinylthymine replacing thymine	50%

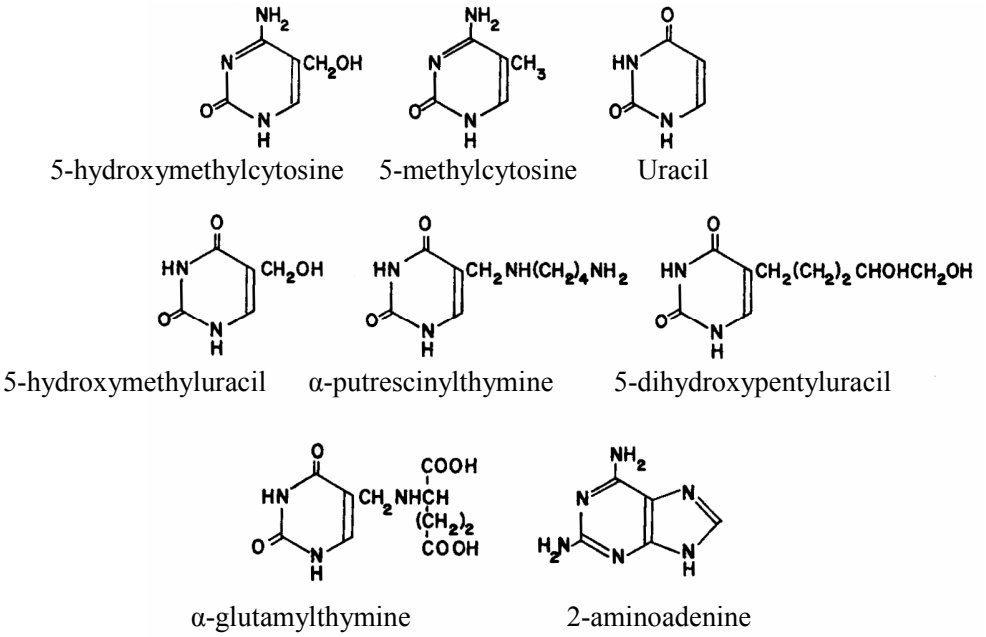


Figure 4.2 Structures of unusual bases in phage DNA (Modified from Warren, 1980).

Loss of endonuclease restriction sites has been shown in many phages, by comparing the actual number of recognition sites to its statistically expected number. For example, the recognition sequence of *B. subtilis* endonuclease *Bsu*RI, GGCC, does not occur once in *B. subtilis* phage ϕ1, although 400 such sites can be predicted statistically in its genome (Kawamura *et al*., 1981). Phage SPO1, which possesses the unusual base hydroxymethyluracil, also lacks the same recognition sequence (Reeve *et al*., 1980), showing multiple strategies used to counter restriction.

None of the phage genomic DNA in this study was digested by all restriction endonucleases. *Bam*HI and *Bmy*I were unable to digest Fairland1 DNA, while most were unable to digest Wits1, Kanazawa1, and Perouges1 DNA, confirming the presence of anti-restriction mechanisms. Absence of restriction sites for *Bam*HI and *Bmy*I was observed in Fairland1 DNA, perhaps a result of selection pressure imposed by restriction enzymes in their natural hosts. Known *Rhodococcus* restriction endonucleases are listed in Table 4.3. Five enzymes, *Rhs*I, *Rlu*4I, *Rsp*LKII, *Rlu*1I, and *Rlu*3I, are isoschizomers of *Bam*HI, whose recognition sequence is GGATCC.

Therefore it is highly credible that Fairland1 lost this palindromic sequence due to the presence of such restriction enzymes in its hosts.

Table 4.3 *Rhodococcus* restriction endonucleases.

Enzyme	**Recognition sequence**	**Organism**
*Rca*I	TCATGA	*R. capsulatum*
*Rer*ORFAP	Unknown	*R. erythropolis* BD2
*Rer*PR4ORFAP	Unknown	*R. erythropolis* PR4
*Rlu*1I	GATC	*R. luteus* RFL1
*Rlu*3I	GGNNCC	*R. luteus* RFL3
*Rlu*4I	GGATCC	*R. luteus* RFL4
*Rrh*I	GTCGAC	*R. rhodochrous*
*Rrh*II	Unknown	*R. rhodochrous*
*Rrh*4273I	GTCGAC	*R. rhodochrous*
*Rro*I		*R. rhodochrous*
*Rhe*I		*Rhodococcus* species
*Rhp*I		*R.* species
*Rhp*II	Unknown	*R.* species
*Rhs*I	GGATCC	*R.* species
*Rsp*XI	TCATGA	*R.* species
*Rsp*LKI	GCATGC	*R.* species LK2
*Rsp*LKII	GGATCC	*R.* species LK2
*Rhc*I	TCATGA	*R.* species SE1991

(Modified from Roberts *et al*., 2005).

Presence of a modification or an unusual base was detected in Perouges1 DNA, in addition to potential loss of recognition sites. In DNA of phages infecting the closely related *Mycobacterium*, modified or unusual bases have also been found (Coene *et al*., 1993). Detection is usually made when a discrepancy exists between values of

%GC from thermal transition temperature and buoyant density, as modified or unusual bases only influence the latter (Warren, 1980).

Wits1 DNA was subjected to digestion by additional enzymes to observe any similarity to known unusual bases (Huang *et al*., 1982), as shown in Table 4.4.

Table 4.4 Digestion of phage DNA containing unusual bases and Wits1 DNA.

Enzyme	Recognition Sequence	Phage					
		XP12	T4	SP15	SPO1	PBS1	Wits1
*Alu*I	AGCT	-	-	-	s	s	+
*Cfo*I	GCGC	-	-	-	+	+	+
*Hha*I	GCGC	-	-	-	+	+	
*Dde*I	CTNAG	-	-	s	s	+	-
*Hae*III	GGCC	-	-	-	+	s	+
*Hinf*I	GANTC	s	-	-	s	s	+
*Hpa*II	CCGG	-	-	-	+	+	
*Msp*I	CCGG	-	-	-	+	+	
*Mbo*I	GATC	s	-	-	+	+	
*Sau*3AI	GATC	-	-	-	s	+	+
*Sau*96I	GGNCC	-	-	-	+	s	+
*Taq*I	TCGA	+	s	s	+	+	+
*Tha*I	CGCG	-	-	s	-	s	
*Ava*II	GG(A/T)CC	-	-	-	s	+	
*Bam*HI	GGATCC	-	-	-	-	+	+
*Bcl*I	TGATCA	-	-	-	s	+	-
*Bgl*I	CCCNNNNNGGC	-	-	-	s	s	
*Bgl*II	AGATCT	-	-	-	s	-	-
*Bst*EII	GGTNACC	s	-	-	s	s	+
*Bst*NI	CC(A/T)GG	s	-	s	+	s	+
*Eco*RII	CC(A/T)GG	-	-	-	s	s	
*Eco*RI	GAATTC	-	-	-	s	s	-
*Hae*II	(A/G)GCGC(T/C)	-	-	-	s	s	
*Hinc*II	GT(T/C)(A/G)AC	-	-	-	s	s	
*Hind*III	AAGCTT	-	-	-	s	s	-
*Hpa*I	CTTAAC	-	-	-	s	-	-
*Sst*I	GAGCTC	-	-	-	-	s	
*Sst*II	CCGCGG	-	-	-	s	s	
*Xba*I	TCTAGA	-	-	-	s	s	-
*Xor*II	CGATCG	-	-	-	-	-	

+ = digested; - = did not digest; s = digested more slowly than normal DNA; where nothing is indicated the digestion was not done; when a highlighted base is

methylated the enzyme no longer recognizes the sequence. Unusual bases possessed by each phage are listed in Table 4.2.
(Huang et al., 1982).

The digestion pattern of Wits1 DNA does not match those of the known unusual bases. However, potential presence of multiple anti-restriction mechanisms in phages makes it difficult to establish a link between unusual or modified bases and digestion patterns.

Presence of anti-restriction mechanisms hindered phage library construction, which has also been experienced with some well known phages. For example, it was not possible to clone T4 DNA (containing glucosylated hydroxymethylcytosine) until mutant strain defective in the synthesis of the unusual base was constructed (Snyder *et al*., 1976). For phage Kanazawa1, whose library construction was unsuccessful, it is not certain which step in cloning was affected. However, there have been reports of T4 DNA ligase (used in this work) not being able to ligate DNA containing certain modified or unusual bases (Cao *et al*., 1983). With phage Wits1, a great majority of cloned DNA fragments were very small (~100 bp). No conclusions concerning their nature were made of Wits1 DNA and Kanazawa1 DNA, despite the obvious presence of anti-restriction mechanisms.

4.4 Non-annotated phage genes

Thirteen phage DNA fragments were sequenced in this study. In these a total of 18 ORFs were predicted, 13 of which had no nucleotide or amino acid sequence match in the databases. A relaxed cutoff E value of 0.009 was used in this study because even such alignment may possibly indicate the potential function of those genes. In other similar studies (Hatfull *et al*., 2006; Kwan *et al*., 2006; Kwan *et al*., 2005), a more stringent cutoff E value of 10^{-4} has been used. If this value were to be used, 16

out of the 18 ORFs (89%) would not be annotated for structure or function, which illustrates the amount of unknown phage genes.

4.5 Frequency of lethal genes

For phage Fairland1, 30% of 20 randomly chosen clones from the *Bgl*II library and 45% from the *Pst*I library were significantly inhibitory to *R. erythropolis* SQ1, which emphasizes the frequency of potential lethal genes in a single phage genome. For phage Perouges1, 15% of clones from the *Hind*III libraries were inhibitory. When screening the libraries, many inhibitory clones gave rise to a few transformants (where the transformation efficiencies were > 0 but < 0.01×10^4/μg of DNA). These were probably mutants in either the antimicrobial gene of plasmid or in a gene of the target in the host. This could be determined by extracting their plasmids and using them to transform *R. erythropolis*; high transformation efficiency would imply plasmid mutants, as the non-mutated plasmids should be inhibitory. In parallel experiments, the original transformants could be cured of their plasmids and re-transformed with the inhibitory plasmids extracted prior to curing; high transformation efficiency here would mean a mutation in the host chromosome.

4.6 Inhibitory genes

In *Rhodococcus*, vector pDA71 has a low copy number of ~5/cell (Dabbs, personal communication). It was unlikely that genes were expressed off the lambda promoter at the beginning of the cloning site of the vector. Thus all genes should have been expressed off respective phage promoters, which usually allow high levels of expression that is necessary during infection.

The genes for which potential functions were found could be grouped into three categories: 1) essential components of DNA metabolism, 2) protein with functions

related to phage structure, and 3) other proteins. Most were hypothetical genes and lacked experimental evidence, but their potential mechanisms in cellular inhibition are speculated upon below.

4.6.1 Essential components of DNA metabolism

Inhibitory genes which are essential components of DNA metabolism were dihydropteroate synthase, thymidylate synthase complementing protein, and α subunit of DNA polymerase III.

4.6.1.1 Dihydropteroate synthase

An inhibitory DNA from phage Fairland1 was most similar to dihydropteroate synthase gene from *Nocardioides* sp. JS614. The same DNA, in a different ORF, was also similar to a phage capsid protein from *Streptococcus agalactiae*. It is uncertain whether the dihydropteroate synthase or the capsid protein, or both, were involved in the inhibitory effect. However, each of these proteins will be discussed as a potential antimicrobial agent.

As shown in Figure 4.3, dihydropteroate synthase catalyses the condensation of 6-hydroxymethyl-7,8-dihydropterin pyrophosphate (DHPPP) with *para*-aminobenzoic acid (pABA) to form dihydropteroate and pyrophosphate, a key step in folate biosynthesis pathway (Vinnicombe and Derrick, 1999). This pathway is unique and essential to bacterial cells, which require folate derivatives as cofactors in the synthesis of nucleic acids. It is no coincidence that dihydropteroate synthase is the target for sulphonamide class of antibiotics, which are analogues of pABA and act as substrates for the enzyme. Sulphonamides therefore deplete cells of folate cofactors.

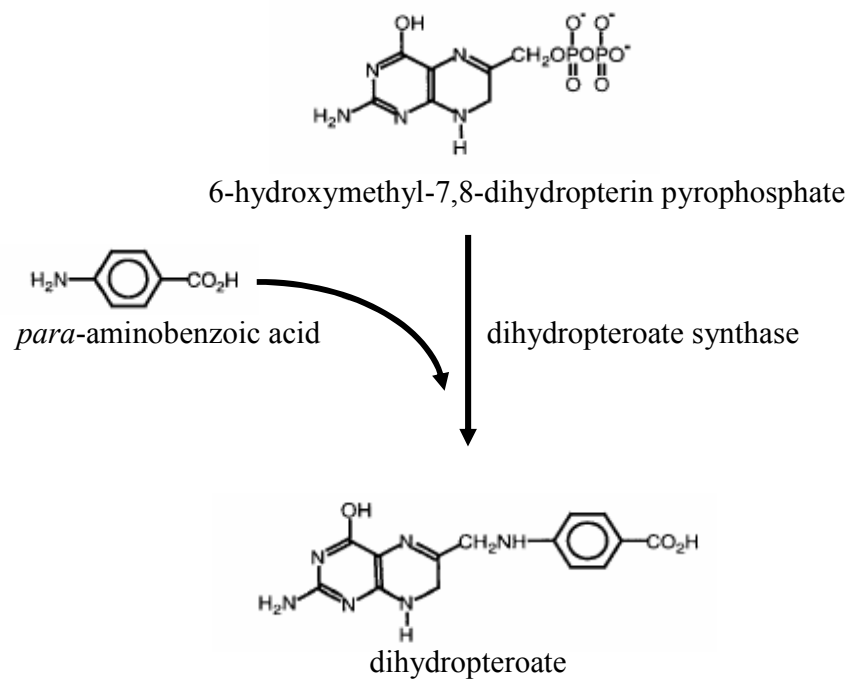


Figure 4.3 Catalytic action of dihydropteroate synthase.

It is possible that expression of Fairland1 dihydropteroate synthase is disrupting the balance of enzymes in the folate biosynthesis pathway, leading to nucleotide depletion. In some organisms, the enzyme is expressed together with hydroxyl-methylpterin pyrophosphate and dihydroneopterin aldolase, two preceding enzymes in the pathway, on the same polypeptide (Brooks *et al*., 1994), suggesting a tight regulation of their concentration.

4.6.1.2 Thymidylate synthase complementing protein

Another inhibitory DNA from Fairland1 was most similar to *Streptomyces coelicolor* thymidylate synthase complementing protein, which is distinct from the classical thymidylate synthase. Thymidylate synthase, encoded by *thyA*, catalyses the conversion of deoxyuridine monophosphate (dUMP) to thymidine monophosphate

(dTMP), which requires the replacement of the 5-hydrogen on the dUMP pyrimidine ring with a methyl group of a cofactor 5,10-methylenetetrahydrofolate (MTF) (Carreras and Santi, 1995). This catalysis is only a few reactions ahead of where dihydropteroate synthase is used, and ThyA is also an essential enzyme in the production of dTMP. Deprivation of thymine in both prokaryotic and eukaryotic cells leads to an activation of genetic suicide mechanism in which DNA cleavage accumulates (Ahmad *et al*., 1998), leading to what has been known as "thymineless cell death" (Barner and Cohen, 1954).

Recent genomic and proteomic analyses have revealed that some organisms possess thymidylate synthase complementing protein (Thy1 or ThyX) instead of thymidylate synthase (Myllykallio *et al*., 2002). Although its function is to produce dTMP, the thymidylate synthase complementing protein shares no sequence or structural homology to thymidylate synthase, and the details of its catalytic mechanism are yet to be discovered. Thy1 is currently being investigated as a potential antimicrobial drug target, as its distribution is mostly within bacteria and archaea. Simultaneous occurrence of Thy1 and ThyA is observed only in a small number of organisms, including *M. tuberculosis*, *M. leprae* (Mathews *et al*., 2003), and *Rhodococcus* sp. RHA1 (McLeod *et al*., 2006), although the coordination, if any, between the two enzymes is unknown.

Many phages are known to possess *thyA*, as it plays a role in the increased rate dNTP synthesis during infection. There has been no report of ThyA or Thy1 directly being deleterious to the host, although mutations in these essential genes may lead to thymineless cell death. Therefore, one possibility is that Fairland1 *thy1* expression in *R. erythropolis* is interfering with dTMP production, leading to cell death. Possession of both *thyA* and *thy1* by *R.* sp. RHA1 makes it credible for *R. erythropolis* to also have both enzymes. Should there be a tight regulation in the activity of these enzymes, the expression of *thy1* from pDA71 may be disrupting their coordination. In fact, even when a functional ThyA is present in a cell, expression of an inactive mutant protein from a plasmid has been shown to be lethal (West *et al*.,

2004). This occurs as the mutant ThyA protein folds with the subunits of the functional host ThyA, forming an inactive dimer (West *et al*., 2004).

During phage T4 infection of *E. coli* cells, T4 thymidylate synthase is known to interact directly with host nucleoside diphosphate (NDP) kinase, which is a multifunctional enzyme involved in dNTP/NTP synthesis, genetic and metabolic regulation, DNA repair, and signal transduction (Shen *et al*., 2004). Therefore, it is also possible that the expression of phage Fairland1 *thy1* in *R. erythropolis* may be depriving host NDP kinase or other similar proteins which are essential in other cellular processes, thereby killing the cell.

4.6.1.3 α subunit, DNA polymerase III

Another inhibitory gene involved in DNA metabolism was most similar to the α subunit of DNA polymerase III (PolIII) from *Aquifex aeolicus*. PolIII is a ten-subunit, 18-polypeptide holoenzyme responsible for most of DNA replication in bacteria and occur in many phage genomes. *Aquifex* PolIII has been shown to be similar to that of *E. coli* (Bruck *et al*., 2002). The α subunit, encoded by *dnaE*, is the catalytic polymerase lacking an exonuclease (Maki and Kornberg, 1985). The proofreading exonuclease activity belongs to the ε subunit, which forms a 1:1 complex with the α subunit (O'Donnell and Studwell 1990). Recent sequencing of bacterial genomes has revealed the presence of more than one type of subunit responsible for the catalytic polymerase activity in many species. Some organisms possess two types of *dnaE*, while others have *polC* in addition to *dnaE* (Le Chatelier *et al*., 2004). In *Rhodococcus* sp. RHA1, *dnaE1* and *E2* have been found (McLeod *et al*., 2006).

Besdies PolIII, there are alternative DNA polymerases in bacterial cells such as *E. coli* PolII, PolIV, and PolV (Tang *et al*., 1999). Interference of DNA replication by alternative polymerases is sometimes mutagenic or even lethal (Viguera *et al*., 2003),

so their activities are tightly regulated in the cell. One of the ways in which these polymerases are coordinated is by their intracellular concentration (Becherel and Fuchs, 2001); in an *E. coli* cell usually there are only 10 to 20 copies of PolIII (Witkin and Roegner-Maniscalco, 1992). This regulation extends to the subunit level (Delmas and Matic, 2006); therefore it is highly likely that this balance is disrupted by the expression of phage Perouges1 α subunit in *R. erythropolis*.

Another factor for the coordination of alternative polymerases is their binding to the β clamp, which clamps onto DNA and tethers the polymerase for high processivity (Stukenberg *et al*., 1991). All polymerases compete for binding to the same β clamp. It has been shown that sub-optimal interactions between the different PolIII subunits lead to interference of DNA synthesis by PolII and PolIV, which take over the binding to available β clamps (Delmas and Matic, 2006). This may be occurring when the phage Perouges1 α subunit is expressed in the *Rhodococcus* cell. When PolII and PolIV out-compete PolIII for β clamp binding and take part in chromosomal replication, translesion DNA synthesis could occur, which is the potentially mutagenic or lethal replication over DNA lesions that usually act as blocks to continued replication (Delmas and Matic, 2006). In addition, the α subunit is crucial in activation of translesion synthesis (Sutton and Walker, 2001).

Another possible inhibitory mechanism by the α subunit is sequestration of the ε subunit, thereby arresting replication. This is implicated as a regulatory mechanism when PolIII encounters damaged DNA, whose synthesis should be halted (Sutton and Walker, 2001). On the other hand, accumulation of the α subunit in the cell may lead to polymerase activity without the proofreading exonuclease activity by the ε subunit, causing frameshift mutations (Le Chatelier *et al*., 2004) and eventually cell death.

4.6.2 Proteins involved in phage structure assembly

Inhibitory proteins with similarity to functions related to phage structure were head maturation peptidase from *Mycobacterium* sp. MCS and KMS, capsid protein from *Streptococcus agalactiae*, head decoration protein from *Roseobacter* sp. MED193, and tail protein from *Xanthomonas axonopodis*.

4.6.2.1 Phage capsid and head maturation peptidase

The head maturation peptidase from *Mycobacterium* sp. MCS and KMS belong to the U35 family of peptidases, represented by that from *E. coli* phage HK97. This gene was found on an inhibitory Fairland1 DNA fragment together with a capsid protein. For phage HK97, a pro-head is formed from 415 copies of a coat protein which are in an uncleaved form (Duda *et al*., 1995). The U35 peptidase removes an N-terminal portion from these coat proteins, which then allows them to cross-react, leading to head expansion and maturation (Duda *et al*., 1995). The catalytic mechanisms of the U35 peptidase are undiscovered, although the coat protein is processed at the lysyl bond (Rawlings *et al*., 2002). There is no report of this peptidase alone being inhibitory to a host cell.

The capsid protein that was present on the same DNA fragment belongs to family 05065, which is a group of hypothetical capsid proteins. Although information on this family is scarce, it is reasonable to assume that this capsid protein, with help from the U35 peptidase, forms a mature head of phage Fairland1. There have been many reports on viral coat proteins associating with host translation factors. For example, phage T4 head protein has the Gol peptide, which binds to and activates the proteolysis of host elongation factor EF-Tu, leading to cell death (Bingham *et al*., 2000). Similar mechanisms have been observed with eukaryotic viruses including HIV-1 and herpesvirus 1 (Cimarelli and Luban, 1999; Kawaguchi *et al*., 1997). Thus it has been speculated that association of viral capsids with host translation factors is

a normal process in viral maturation (Bingham *et al*., 2000). In addition, the U35 peptidase is homologous to the herpesvirus protease used in coat maturation (Cheng *et al*., 2004). These observations strengthen the possibility that the mature capsid of phage Fairland1 associates and inhibits a *Rhodococcus* elongation factor.

4.6.2.2 Head decoration protein and tail protein

Sequence of an inhibitory DNA fragment from phage Perouges1 aligned with the head decoration protein from *Roseobacter* sp. MED193. This fragment also aligned with a phage-related tail protein from *Xanthomonas axonopodis* and cell wall anchor family protein from *Streptococcus agalactiae*, all in the same ORF. The differences in E values between these proteins were insignificant, suggesting a similarity between these proteins.

Head decoration proteins are semiessential or nonessential proteins that stabilize the coat protein by binding to its outer side (Gilcrease *et al*., 2005). There is no report of such proteins being bactericidal. On the other hand, phage tail-like structures are well-known features of many bacteriocins produced by both Gram-positive and negative bacteria. Phage tail-like bacteriocins generally belong to two families: contractile (R-type) and non-contractile (F-type). Genetic analyses have shown the similarity of these structures to actual phage tails (Nakayama *et al*., 2000). They are believed to be variations of defective phages whose tails have been evolutionarily specialized as bacteriocins (Nakayama *et al*., 2000). Treatment of cells with SOS response inducing agents, such as mitomycin C and UV, which causes phage induction, usually leads to the production of these bacteriocins, often together with detached, empty head particles. Phage tail-like bacteriocins being encoded by viable phages have also been reported (Pond *et al*., 1989). Although the protein from *X. axonopodis* is a hypothetical one, other related species have been shown to produce functional bacteriocins (Pham *et al*., 2004). Both types of phage tail-like bacteriocins kill sensitive cells by depolarizing the plasma membrane (Gillor *et al*., 2005), and are

being investigated as potential therapeutic agents, as they lack nucleic acid and are specific in their target organisms.

Cell wall surface anchor family protein from *Streptococcus agalactiae* is the M protein which is covalently anchored in the cell wall (Bergmann and Hammerschmidt, 2006). This protein is recognized as the major virulence determinant of group A *Streptococcus*, as it is vital in escaping phagocytosis (Fischetti, 1989), although its mode of action has not been determined. In addition, the M protein is responsible for enhanced attachment to epithelial cells, which is also necessary for its virulence (Ellen and Gibbons, 1972). Therefore, it is possible that membrane binding, similar to the phage-tail like bacteriocins, is involved, causing cellular inhibition. Various membrane-associated proteins have been shown in phage T4, including baseplate proteins of its tail. During infection, T4 baseplate is assembled on the host cell membrane and remains attached until lysis (Simon, 1969). Other examples include the holin gene, responsible for membrane disruption during lysis, and the *ndd* gene, which is believed to be responsible for binding bacterial DNA to the membrane (Miller *et al*., 2003). Phage ϕX174 encodes protein E, which is associated with host membrane and halts cell wall synthesis by blocking the activity of MraY, a membrane protein that transfers murein precursors to lipid carriers (Bernhardt *et al*., 2000).

4.6.3 Other proteins

Inhibitory genes which were neither essential components of DNA metabolism nor proteins involved in phage structure assembly were the cell wall surface anchor family protein from *Streptococcus agalactiae* (discussed above) and HNH endonuclease from *Lactobacillus plantarum* phage LP65.

4.6.3.1 HNH endonuclease

HNH endonuclease is a member of the family of rare-cutting enzymes called homing endonucleases encoded by introns and inteins. They initiate homing, or transport, of these mobile elements into intronless/inteinless sites of DNA. Homing endonucleases are present in all three biological kingdoms and differ from restriction endonucleases in many features (Belfort and Roberts, 1997). So-called because of the three invariant amino acids that span the core motif, HNH endonucleases, along with introns and inteins, occur in many phages. Translocation of these mobile elements is site-specific; most phage introns and inteins are found within genes of DNA metabolism such as DNA polymerase (Derbyshire and Belfort, 1998) and ribonucleotide reductase (Lazarevic, 2001), suggesting selective cleavage at these sites by these nucleases. Therefore one possibility is that the endonuclease from phage Fairland1 is cleaving important genes in *R. erythropolis*.

Besides their role in homing, HNH endonucleases are responsible for the cytotoxic activity of colicins E2, E7, E8, and E9 (James *et al*., 2002). After binding to cell surface receptors, they are translocated to the cytoplasm where they degrade chromosomal DNA non-specifically (James *et al*., 2002). These nucleases are thus reported to have both specific and non-specific activity. Whichever the case, their nuclease activity should be responsible for the inhibition of *Rhodococcus* host.

4.7 Conclusion

Inhibitory genes for which putative functions could be assigned may be directly interfering with nucleotide metabolism, DNA synthesis, protein synthesis, and integrity of the plasma membrane and DNA. Indirectly interfered processes may include genetic and metabolic regulation, signal transduction and DNA repair. Many of these are well-appreciated targets of antibiotics; however, further studies are required to identify the specific inhibitory mechanisms and target molecules. That

most E values were rather low may make the above speculations out-of-context; however most genes had convincing connection to inhibitory mechanisms.

It would be possible to screen phage genomes for lethal genes very rapidly, by excluding some steps in this work. A plate lysate provides enough DNA for construction of a phage library, so large scale lysate production, purification, and electron microscopy can be sacrificed if the sole purpose is to identify inhibitory DNA. One obstacle, however, may be the difficulty in cloning some phage DNA. To succeed, mechanical shearing followed by chemical modification of the sheared ends may be necessary.

Host inhibition by a phage *in vivo* cannot be simulated *in vitro* by expressing a limited number of genes in a plasmid in the host. Regulation of phage gene expression by operons and temporally modified RNA polymerases for recognition of specific promoters may interfere with identification of some lethal genes. However, obvious inhibitory mechanisms suggested by a few genes in this work validate the effectiveness of the screening experiment. Thus, the large number of inhibitory genes with no assigned function confirms that phages provide an untapped pool of novel bactericidal mechanisms.

Exact molecular targets of these genes can be identified by both biochemical and genetic methods. A biochemical approach would involve affinity chromatography and mass spectrometry. A genetic approach would involve construction of a genomic library of a mutant resistant to the inhibitory gene, followed by screening for the resistant gene. This gene should be a mutant in the target molecule, which may then be identified by its sequence.

5. APPENDICES

APPENDIX A - MEDIA AND SOLUTIONS

1. MEDIA

LB

3 g tryptone
1.5 g sodium chloride
1.5 g yeast extract
300 ml distilled water

LA

3 g tryptone
1.5 g sodium chloride
1.5 g yeast extract
4.5 g agar
300 ml distilled water

Sloppy agar

1 g tryptone
0.5 g sodium chloride
0.5 g yeast extract
0.75 g agar
100 ml distilled water

LBG

1 g tryptone

0.5 g sodium chloride

0.5 g yeast extract

2 g glycine

100 ml distilled water

LBSG

1 g tryptone

0.5 g sodium chloride

0.5 g yeast extract

10.3 g sucrose

2 g glycine

100 ml distilled water

***Rhodococcus* regeneration medium**

3 g tryptone

1.5 g yeast extract

0.9 g sodium chloride

30.9 g sucrose

1 g glucose

1 g magnesium chloride hexahydrate

250 ml distilled water

Solution was microwaved to dissolve sucrose. Then, 5.5 g agar was added, autoclaved, and cooled to 60°C, after which the following were added:

10 ml 0.25 M TES pH 7.2

6 ml 1 M calcium chloride

3 ml 0.5% potassium dihydrogen orthophosphate

1.5 ml 10 mg/ml nystatin

0.9 ml 10 mg/ml rifampicin

Plates of constant volume of 25 ml were poured on a horizontal surface.

2. SOLUTIONS FOR PHAGE PURIFICATION

SC buffer

0.58 g sodium chloride
0.2 g calcium chloride hexahydrate
5 ml 1 M Tris-hydrochloride pH 8.0
0.5 ml 2% gelatin
95 ml distilled water

2% Gelatin

0.02 g gelatin
1 ml distilled water

Caesium chloride solutions for gradient centrifugation

(Density values are relative to that of water at 20°C)

1.45 g/ml

3.53 g caesium chloride
5 ml SC buffer

1.50 g/ml

4.09 g caesium chloride
5 ml SC buffer

1.70 g/ml

6.33 g caesium chloride
5 ml SC buffer

3. SOLUTIONS FOR TRANSMISSION ELECTRON MICROSCOPY

0.35% Formvar

0.35 g Formvar
100 ml chloroform

1% uranyl acetate

0.05 g uranyl acetate
5 ml distilled water

4. SOLUTIONS FOR DNA PREPARATION

Solution I

0.9 g glucose
2.5 ml 1 M Tris-hydrochloride pH 8.0
2 ml 0.5 M EDTA
95.5 ml distilled water

Solution II

0.8 g sodium hydroxide
1 g SDS
100 ml distilled water

Solution III

60 ml 5 M potassium acetate
11.5 ml glacial acetic acid
28.5 ml distilled water

1 M Tris-hydrochloride pH 8.0

12.1 g Tris base, dissolved in distilled water
pH adjusted to 8.0 with hydrochloric acid
Distilled water added to 100 ml

0.5 M EDTA pH 8.0

18.61 g EDTA, dissolved in distilled water
pH adjusted to 8.0 with sodium hydroxide
Distilled water added to 100 ml

TE

1 ml 1M Tris-hydrochloride pH 8.0
2 ml 0.5 M EDTA pH 8.0
100 ml distilled water

10% SDS in water

1 g SDS
10 ml distilled water

10% SDS in TE

1 g SDS
10 ml TE buffer

5 M potassium acetate pH 6.0

4.9 g potassium acetate
pH adjusted to 6.0 with acetic acid
Distilled water added to 10 ml

TE-saturated phenol

10 g phenol

10 ml TE

1 M sodium chloride

5.8 g sodium chloride

100 ml distilled water

1% Ribonuclease

0.01 g ribonuclease

1 ml distilled water

Solution was boiled for 10 min before use.

Dialysis buffer

0.58 g sodium chloride

2.2 g calcium chloride hexahydrate

50 ml 1 M Tris-hydrochloride pH 8.0

950 ml distilled water

5. SOLUTIONS FOR ELECTROPHORESIS

5× TBE

54 g Tris base

27.5 g boric acid

20 ml 0.5 M EDTA pH 8.0

Distilled water added to 1000 ml

Agarose gel

0.8 g (for 0.4%), 1.6 g (for 0.8%), or 2.4 g (for 1.2%) agarose
20 ml 5× TBE
180 ml distilled water

Running buffer

20 ml 5× TBE
180 ml distilled water
20 µl 1% EtBr

1% EtBr

0.1 g EtBr
10 ml distilled water

Tracking dye

0.01 g bromophenol blue
0.01 g xylene cyanol
10 ml 30% glycerol in TE

6. SOLUTIONS FOR TRANSFORMATION

Calcium chloride transformation buffer

4.38 g calcium chloride hexahydrate
2 ml 1 M Tris-hydrochloride pH 8.0
198 ml distilled water

Basal (B) buffer

10.3 g sucrose

0.025 g potassium sulfate

0.202 g magnesium chloride hexahydrate

10 ml 0.25 M TES pH 7.2

Protoplast (P) buffer

2 ml B buffer

20 µl 0.5% potassium dihydrogen orthophosphate

50 µl 1 M calcium chloride

P-PEG buffer

0.5 g PEG 6000

1 ml P buffer

PEG sterilized by UV irradiation for 10 min and vortexed vigorously to dissolve.

20% glucose

4 g glucose

20 ml distilled water

0.25 M TES pH 7.2

17.2 g TES, dissolved in distilled water

pH adjusted to 7.2 with sodium hydroxide

Distilled water added to 300 ml

1 M calcium chloride

21.9 g calcium chloride hexahydrate

100 ml distilled water

0.5% potassium dihydrogen orthophosphate

0.5 g potassium dihydrogen orthophosphate

100 ml distilled water

7. ANTIMICROBIAL AGENTS

Table 5.1 Antimicrobial agents.

Agent	**Stock concentration (mg/ml)**	**Solvent**
Ampicillin	100	70% ethanol, 30% distilled water
Chloramphenicol	4	ethanol
Kanamycin	100	distilled water
Nalidixic acid	10	70% ethanol, 30% distilled water
Nystatin	10	methanol
Rifampicin	10	methanol
Streptomycin	20	distilled water

APPENDIX B - RESTRICTION MAPS OF PLASMIDS

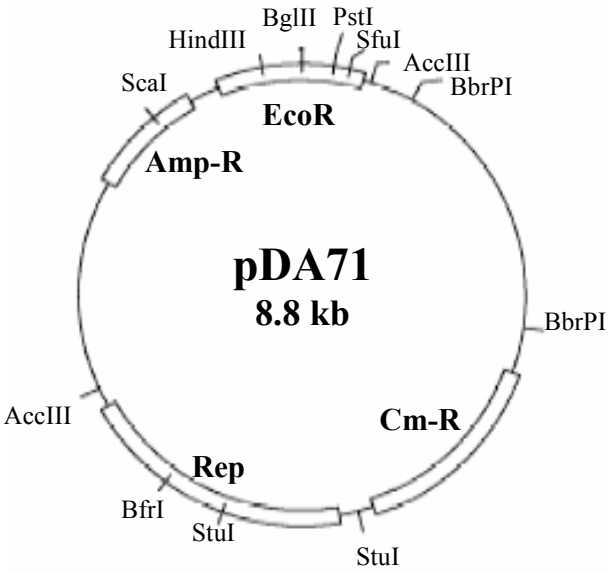


Figure 5.1 Restriction map of pDA71.

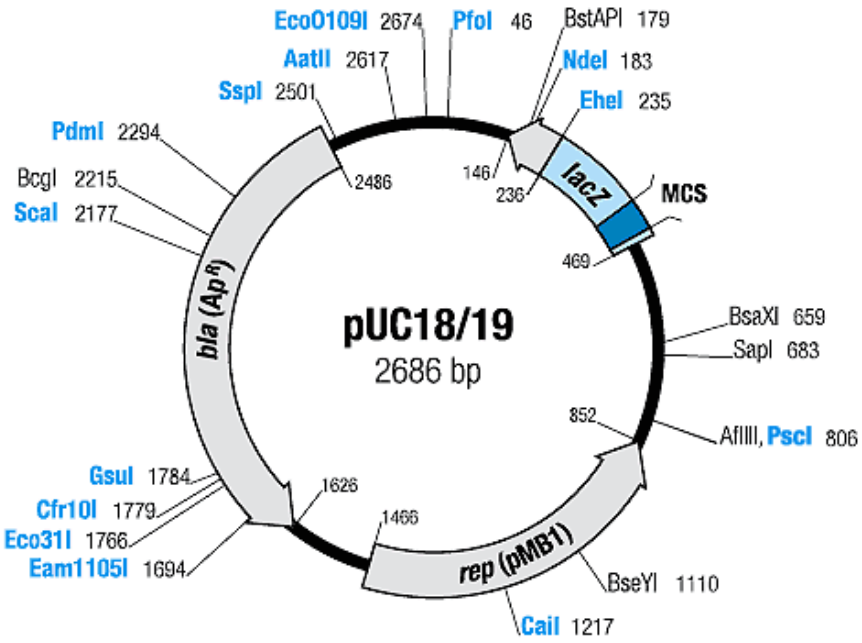
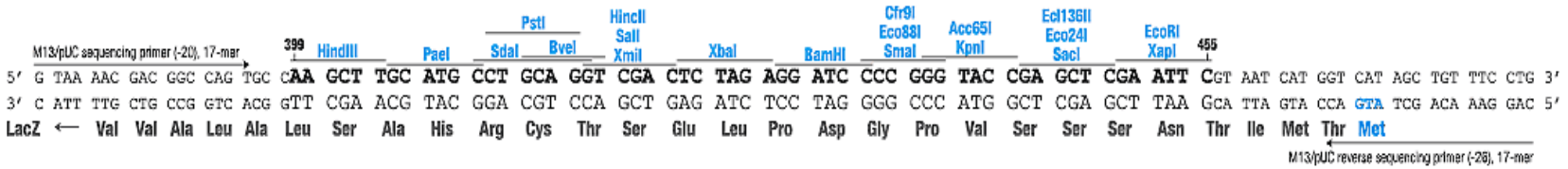
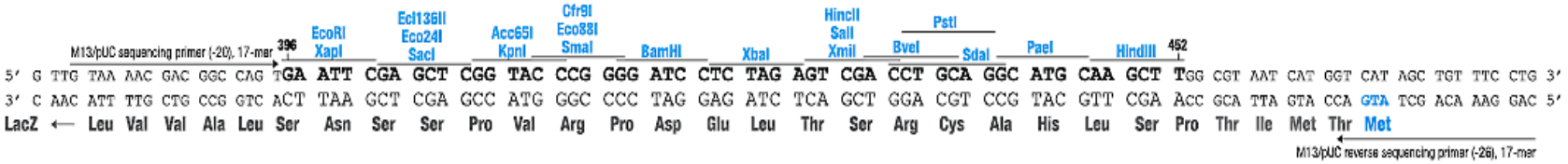


Figure 5.2 Restriction map of pUC18/19.



pUC18



pUC19

Figure 5.3 Polylinker of pUC18/19.

APPENDIX C – MOLECULAR WEIGHT MARKERS

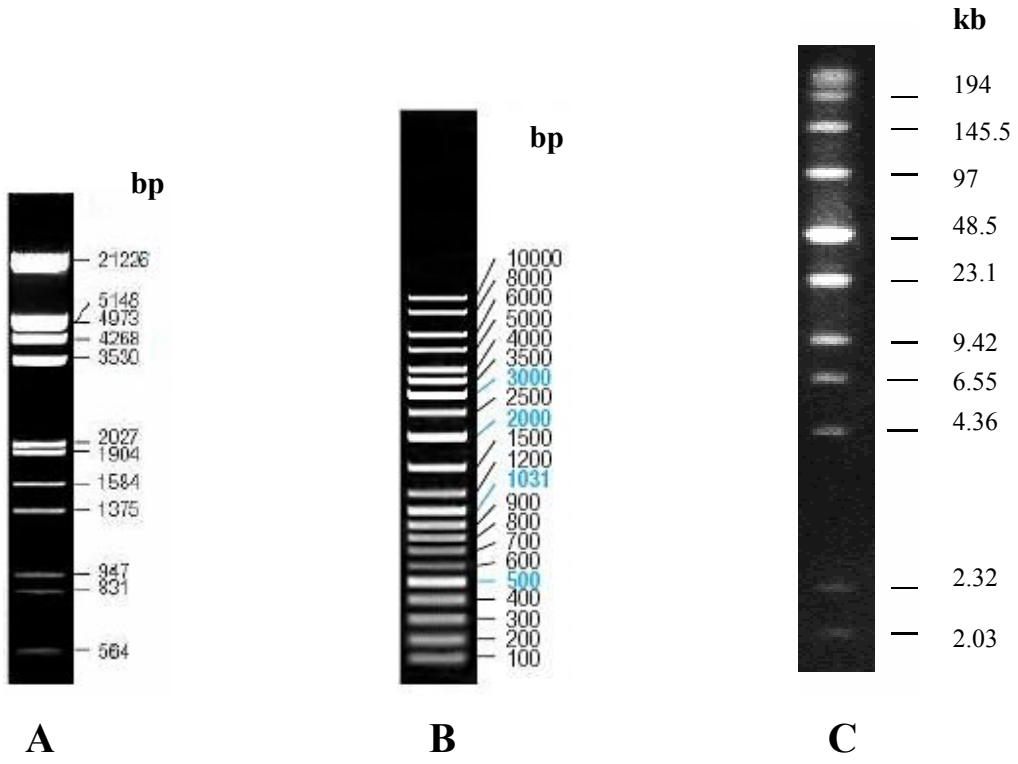


Figure 5.4 Molecular weight markers.

A. Roche Lambda *Eco*RI+*Hind*III Marker III

B. Fermentas GeneRuler DNA Ladder Mix

C. New England Biolabs PFG Low Range Marker

APPENDIX D – FEATURES OF PROTEINS SIMILAR TO INHIBITORY PHAGE PRODUCTS

Table 5.2 Features of thymidylate synthase complementing protein from *Streptomyces coelicolor*.

```
LOCUS       CAA20294                 246 aa            linear   BCT 16-APR-2005
DEFINITION  conserved hypothetical protein SC9A10.07 [Streptomyces coelicolor
            A3(2)].
ACCESSION   CAA20294
VERSION     CAA20294.1  GI:3413826
DBSOURCE    embl locus SCO939124, accession AL939124.1
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2)
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
            Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
            Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
            Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
            Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S.,
            Rabbinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
            Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
            Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G.,
            Parkhill,J. and Hopwood,D.A.
  TITLE     Complete genome sequence of the model actinomycete Streptomyces
            coelicolor A3(2)
  JOURNAL   Nature 417 (6885), 141-147 (2002)
   PUBMED   12000953
REFERENCE   2  (residues 1 to 246)
  AUTHORS   Bentley,S.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
            sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
FEATURES             Location/Qualifiers
     source          1..246
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
     Protein         1..246
                     /product="conserved hypothetical protein SC9A10.07"
     Region          6..229
                     /region_name="Thy1"
                     /note="Thymidylate synthase complementing protein;
                     pfam02511"
                     /db_xref="CDD:42487"
     CDS             1..246
                     /gene="SCO5743"
                     /coded_by="AL939124.1:284793..285533"
```

```
                     /note="SC9A10.07, unknown, len: 246 aa; similar to e.g.
                     Dictyostelium discoideum THY1_DICDI THY1 protein (260 aa),
                     fasta scores; opt: 308 z-score: 443.0 E(): 2.1e-17, 35.4%
                     identity in 198 aa overlap"
                     /transl_table=11
                     /db_xref="GOA:O86840"
                     /db_xref="InterPro:IPR003669"
                     /db_xref="UniProtKB/Swiss-Prot:O86840"
ORIGIN
        1 mtdipaddpk ielrsditve lvksaatdsd vlfaarvsta geqsldelkk dperskglin
       61 ylmrdrhgsp fehnsmtffv sapifvfref mrhrvgwsyn eesgryrelq pvfyapdasr
      121 klvqqgrpgk yvfvegtpeq helvgsamed syrqayatyq qmlaagvare varavlpvgl
      181 yssmyatcna rslmhflglr tqhelakvps fpqreiemag ekmeaewarl mplthaafna
      241 ngrvap
```

Table. 5.3 Features of phage head maturation peptidase from *Mycobacterium* sp. MCS.

```
LOCUS       YP_640677                526 aa            linear   BCT 13-JUN-2006
DEFINITION  peptidase U35, phage prohead HK97 [Mycobacterium sp. MCS].
ACCESSION   YP_640677
VERSION     YP_640677.1  GI:108800480
DBSOURCE    REFSEQ: accession NC_008146.1
KEYWORDS    .
SOURCE      Mycobacterium sp. MCS
  ORGANISM  Mycobacterium sp. MCS
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE   1  (residues 1 to 526)
  AUTHORS   Copeland,A., Lucas,S., Lapidus,A., Barry,K., Detter,J.C., Glavina
            del Rio,T., Hammon,N., Israni,S., Dalin,E., Tice,H., Pitluck,S.,
            Martinez,M., Schmutz,J., Larimer,F., Land,M., Hauser,L.,
            Kyrpides,N., Kim,E., Miller,C.D., Hughes,J.E., Anderson,A.J.,
            Sims,R.C. and Richardson,P.
  CONSRTM   US DOE Joint Genome Institute
  TITLE     Complete sequence of chromosome of Mycobacterium sp. MCS
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 526)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2006) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 526)
  AUTHORS   Copeland,A., Lucas,S., Lapidus,A., Barry,K., Detter,J.C., Glavina
            del Rio,T., Hammon,N., Israni,S., Dalin,E., Tice,H., Pitluck,S.,
            Martinez,M., Schmutz,J., Larimer,F., Land,M., Hauser,L.,
            Kyrpides,N., Kim,E., Miller,C.D., Hughes,J.E., Anderson,A.J.,
            Sims,R.C. and Richardson,P.
  CONSRTM   US DOE Joint Genome Institute
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2006) US DOE Joint Genome Institute, 2800
            Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from ABG09621.
            Method: conceptual translation.
```

```
FEATURES             Location/Qualifiers
     source          1..526
                     /organism="Mycobacterium sp. MCS"
                     /strain="MCS"
                     /db_xref="taxon:164756"
     Protein         1..526
                     /product="peptidase U35, phage prohead HK97"
                     /calculated_mol_wt=55053
     Region          3..150
                     /region_name="COG3740"
                     /note="Phage head maturation protease [General function
                     prediction only]; COG3740"
                     /db_xref="CDD:33535"
     CDS             1..526
                     /locus_tag="Mmcs_3514"
                     /coded_by="complement(NC_008146.1:3740139..3741719)"
                     /note="PFAM: peptidase U35, phage prohead HK97
                     KEGG: nse:NSE_0554 phage prohead protease, HK97 family"
                     /transl_table=11
                     /db_xref="InterPro:IPR006433"
                     /db_xref="GeneID:4112346"
ORIGIN
        1 mhtravelte vrtdddagtf tglaagydnv dthgtvlqrg afasslaggg vvplfwehgh
       61 ddpraivgev taavettrgl eivgkldtdt ergaaayrav kgrrirglsv gmrptqrrga
      121 siiaadlcei slvmrpsnsr alvesvrsad dalqtraasa vatfetiakd ttmtepitte
      181 rrdelvaetr glvaaaqgrt ltaeevatie tntetirrhd eqaletrnda qaaniaralg
      241 qaidtrsggr qspfmlsadn vttletarkr fenitvletr aalattdmgt areygpnglq
      301 aprslwrsag ipttapdgys gvvpqftlpg gavlvgegvd hqefdgvnpd avtigragaw
      361 stltsealls tsitevsaah ariiarnvdr atvakiedas pdtmsidqal vtvaaecacd
      421 vsdlwivgap aavaalvgna tftpanggda esyasrygga avypttsata dtltvfhpqs
      481 frafasplss gvfvdpksgk qdfgqwmfyg lgqalvgaai tvdttp
```

Table 5.4 Features of phage capsid protein from *Streptococcus agalactiae*.

```
LOCUS       AAN00715                420 aa            linear   BCT 23-SEP-2002
DEFINITION  conserved domain protein [Streptococcus agalactiae 2603V/R].
ACCESSION   AAN00715
VERSION     AAN00715.1  GI:22534892
DBSOURCE    accession AE014275.1
KEYWORDS    .
SOURCE      Streptococcus agalactiae 2603V/R
  ORGANISM  Streptococcus agalactiae 2603V/R
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (residues 1 to 420)
  AUTHORS   Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
            Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
            Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
            Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
            Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
            Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
            Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
            Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,
```

```
            Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
  TITLE     Complete genome sequence and comparative genomic analysis of an
            emerging human pathogen, serotype V Streptococcus agalactiae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
   PUBMED   12200547
REFERENCE   2  (residues 1 to 420)
  AUTHORS   Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
            Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
            Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
            Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
            Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
            Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
            Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
            Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,
            Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..420
                     /organism="Streptococcus agalactiae 2603V/R"
                     /strain="2603V/R"
                     /serotype="V"
                     /db_xref="taxon:208435"
     Protein         1..420
                     /product="conserved domain protein"
     Region          <254..387
                     /region_name="Phage_capsid"
                     /note="Phage capsid family; pfam05065"
                     /db_xref="CDD:44968"
     CDS             1..420
                     /gene="SAG1852"
                     /coded_by="complement(AE014275.1:18904..20166)"
                     /note="identified by Glimmer2; putative"
                     /transl_table=11
ORIGIN
        1 mkkslielle arqkatdela evklkkatie akmksstied ddleqlktda eslvsqatai
       61 ketiagldsd ieeteeelsk aakiikekqk gntpmdylkt kaaaldfvri lmdnegsans
      121 arkaweanlv ekgvtnltki lpepvliaiq daftnyngil nhvskdprya vrvalqtqvs
      181 qakghkagkt kkdedftfld ftinsatvyi kyafeysdlk kdttgayfny vmkelaqgfi
      241 rtieravvig dgksnsaedk iteiksiaee teanlfetqe invagvfdna vletlvagid
      301 kmvpnttpil vtskaiarkl klvkdaegry idpqpfapia tngnviagfq vyiydwmdga
      361 tnpiiafadq aykmigdeaa adrfddydvt vnrrhielas vmggrlaqyk savkftnpag
```

Table 5.5 Features of dihydropteroate synthase from *Nocardioides* sp. JS614.

```
LOCUS       ZP_00660252              466 aa            linear   BCT 28-JUL-2005
DEFINITION  Dihydropteroate synthase [Nocardioides sp. JS614].
ACCESSION   ZP_00660252
VERSION     ZP_00660252.1  GI:71369843
DBSOURCE    REFSEQ: accession NZ_AAJB01000146.1
KEYWORDS    .
SOURCE      Nocardioides sp. JS614
  ORGANISM  Nocardioides sp. JS614
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Propionibacterineae; Nocardioidaceae; Nocardioides.
REFERENCE   1  (residues 1 to 466)
  AUTHORS   Copeland,A., Lucas,S., Lapidus,A., Barry,K., Detter,J.C.,
            Glavina,T., Hammon,N., Israni,S., Pitluck,S. and Richardson,P.
  CONSRTM   US DOE Joint Genome Institute (JGI-PGF)
  TITLE     Sequencing of the draft genome and assembly of Nocardioides sp.
            JS614
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 466)
  AUTHORS   Larimer,F. and Land,M.
  CONSRTM   US DOE Joint Genome Institute (JGI-ORNL)
  TITLE     Annotation of the draft genome assembly of Nocardioides sp. JS614
  JOURNAL   Unpublished
REFERENCE   3  (residues 1 to 466)
  AUTHORS   Copeland,A., Lucas,S., Lapidus,A., Barry,K., Detter,J.C.,
            Glavina,T., Hammon,N., Israni,S., Pitluck,S. and Richardson,P.
  CONSRTM   US DOE Joint Genome Institute (JGI-PGF)
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-2005) US DOE Joint Genome Institute, 2800
            Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     PREDICTED REFSEQ: The mRNA record is supported by experimental
            evidence; however, the coding sequence is predicted. The reference
            sequence was derived from EAO04955.
            URL -- http://www.jgi.doe.gov/
            Source DNA and bacteria available from James Gossett
            (jmg18@cornell.edu)
            Whole genome sequencing and draft assembly at JGI-PGF
            Annotation by JGI-ORNL
            Contact: Paul Richardson (microbes@cuba.jgi-psf.org).
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..466
                     /organism="Nocardioides sp. JS614"
                     /strain="JS614"
                     /db_xref="taxon:196162"
     Protein         1..466
                     /product="Dihydropteroate synthase"
                     /EC_number="2.5.1.15"
                     /calculated_mol_wt=49552
     Region          194..450
                     /region_name="DHPS"
                     /note="DHPS subgroup of Pterin binding enzymes; cd00739"
                     /db_xref="CDD:29545"
     Region          194..450
                     /region_name="DHPS"
                     /note="DHPS subgroup of Pterin binding enzymes; cd00739"
```

```
                     /db_xref="CDD:29545"
     CDS             1..466
                     /locus_tag="NocaDRAFT_4752"
                     /coded_by="NZ_AAJB01000146.1:4589..5989"
                     /inference="non-experimental evidence, no additional
                     details recorded"
                     /transl_table=11
                     /db_xref="InterPro:IPR000489"
                     /db_xref="InterPro:IPR006390"
ORIGIN
        1 megavrdhrr rarvhgrrpa rlghplrarp arpaahpprg ragqgllraa dalrrrrgpv
       61 rgrpghragr aaadpavvpr argpapaarg rvggttrpvr aaqradrvrr garrprrhpa
      121 hpragrarpg rprhprppap vaarpqrdgh pdprgrpgpg rhphrgragr ratagggram
      181 tlklgrhtfp dtatlmmaiv nrtpdsfydk gatwaedkaf ervalvageg aeivdiggik
      241 aapgveisea eekarvvdfv arvrasypdl visvdtwrae vgsavcaaga dlindawgga
      301 dpeladaaae agaaivctht ggvtprtrpy rmeyddvvaa aiadtvayae raiaagvdpa
      361 svvidpahdf gkntfhslev trrlgemvat gypvlvslsn kdfvgesldl pvgerltgtl
      421 aatavcalag ariyrvhevv etrqtvdmvd tiagrrpprl airglq
```

Table 5.6 Features of HNH endonuclease from *Lactobacillus plantarum* phage LP65.

```
LOCUS       YP_164778                354 aa            linear   PHG 07-APR-2006
DEFINITION  orf143 [Lactobacillus plantarum bacteriophage LP65].
ACCESSION   YP_164778
VERSION     YP_164778.1  GI:56693191
DBSOURCE    REFSEQ: accession NC_006565.1
KEYWORDS    .
SOURCE      Lactobacillus plantarum bacteriophage LP65
  ORGANISM  Lactobacillus plantarum bacteriophage LP65
            Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
            SPO1-like viruses; unclassified SPO1-like viruses.
REFERENCE   1  (residues 1 to 354)
  AUTHORS   Chibani-Chennoufi,S., Dillmann,M.L., Marvin-Guy,L.,Rami-Shojaei,S.
            and Br.
  TITLE     Lactobacillus plantarum Bacteriophage LP65: a New Member of the
            SPO1-Like Genus of the Family Myoviridae
  JOURNAL   J. Bacteriol. 186 (21), 7069-7083 (2004)
   PUBMED   15489418
REFERENCE   2  (residues 1 to 354)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-2004) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
REFERENCE   3  (residues 1 to 354)
  AUTHORS   Chibani-Chennoufi,S., Dillmann,M.-L., Marvin-Guy,L.,
            Rami-Shojaei,S. and Brussow,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2004) Nutrition & Health, Nestle Research Centre,
            Lausanne 26, Vers-chez-les-Blanc CH-1000, Switzerland
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AAV35963.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..354
```

```
                     /organism="Lactobacillus plantarum bacteriophage LP65"
                     /virion
                     /db_xref="taxon:298338"
     Protein         1..354
                     /product="orf143"
                     /calculated_mol_wt=39890
     Region          71..114
                     /region_name="HNH"
                     /note="HNH endonuclease; pfam01844"
                     /db_xref="CDD:41868"
     Region          200..247
                     /region_name="IENR1"
                     /note="Intron encoded nuclease repeat motif; Repeat of
                     unknown function, but possibly DNA-binding via
                     helix-turn-helix motif (Ponting, unpublished); smart00497"
                     /db_xref="CDD:47801"
     Region          272..321
                     /region_name="IENR1"
                     /note="Intron encoded nuclease repeat motif; Repeat of
                     unknown function, but possibly DNA-binding via
                     helix-turn-helix motif (Ponting, unpublished); smart00497"
                     /db_xref="CDD:47801"
     CDS             1..354
                     /locus_tag="orf143"
                     /coded_by="NC_006565.1:121037..122101"
                     /note="similar to SPO1 gene 112"
                     /transl_table=11
                     /db_xref="GeneID:3197340"
ORIGIN
        1 mkkpvelwkt fpgsagieis sfgnvrsldg vkcsegyehy isgnyytnsr issgyrgvci
       61 pidgkyvtkk vhrlvaqtfi pnpnnmpqvn hkdgdksnnn vsnlewctns yndkykkvfg
      121 kslerpvyai nlstlevswf psqieagrsl gvnvghvnsv vkgkmkklsg ywfvedngdn
      181 fnldkdrlre irvraqfryg vyainldtqe afhfksqqea gqslgvdsss insvlkgrlk
      241 qaggyyfved tggdfkldkd klreikdemr fkggvyaidl etleafyfkt qrvagqelgi
      301 dnsninkvlk gkqktakgyw fisdnekive ivrekfgdtv aneiamliak rqai
```

Table 5.7 Features of head decoration protein of prophage MuMc02 from *Roseobacter* sp. MED193.

```
LOCUS       ZP_01058443              419 aa            linear   BCT 30-JAN-2006
DEFINITION  prophage MuMc02, head decoration protein, putative [Roseobacter sp.
            MED193].
ACCESSION   ZP_01058443
VERSION     ZP_01058443.1  GI:86139877
DBSOURCE    REFSEQ: accession NZ_AANB01000014.1
KEYWORDS    .
SOURCE      Roseobacter sp. MED193
  ORGANISM  Roseobacter sp. MED193
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
            Rhodobacteraceae; Roseobacter.
REFERENCE   1  (residues 1 to 419)
  AUTHORS   Pinhassi,J., Pedros-Alio,C., Ferriera,S., Johnson,J., Kravitz,S.,
            Halpern,A., Remington,K., Beeson,K., Tran,B., Rogers,Y.-H.,
            Friedman,R. and Venter,J.C.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (26-JAN-2006) J Craig Venter Institute, 9704 Medical
            Center Drive, Rockville, MD 20850, USA
COMMENT     PREDICTED REFSEQ: The mRNA record is supported by experimental
            evidence; however, the coding sequence is predicted. The reference
            sequence was derived from EAQ43715.
            Annotation was added by the NCBI Prokaryotic Genomes Automatic
            Annotation Pipeline Group.  Information about the Pipeline can be
            found here:
            http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html. Please be
            aware that the annotation is done automatically with little or no
            manual curation.
            Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..419
                     /organism="Roseobacter sp. MED193"
                     /strain="MED193"
                     /db_xref="taxon:314262"
     Protein         1..419
                     /product="prophage MuMc02, head decoration protein,
                     putative"
                     /calculated_mol_wt=40699
     CDS             1..419
                     /locus_tag="MED193_12628"
                     /coded_by="NZ_AANB01000014.1:95778..97037"
                     /note="COG5295 Autotransporter adhesin"
                     /transl_table=11
ORIGIN
        1 mantiqlkrr vsgnagapaa lksgevahne vddtlyigkg ddgggnatsi vaiagsggfv
       61 aktgtqtiag kktfslvpta sqdaaagsdl vrksqldtll ggkantshsh aiadvtglqg
      121 aldgkaavsh dhtaaeisds tsagrtllka advaaqhtal glgtaalmss tafaaaahgh
      181 aisdvsglqt alngkaplas psftgtpaap taaggtnttq iattafvqsa iasfgagdml
      241 katydsdndg kvdaaelada vawtgvtgkp atfppsahnh pisqvtglqs aldakapkvs
      301 ptftgtptap taaggtsntq iattafvsaa iaalidaapg amdtlnelaa algddpdfat
      361 tvtnglagkl ektsnlsdlt naatarsnlg lgsmatqaan nvaitggsis gialdggtf
```

Table 5.8 Features of phage-related tail protein from *Xanthomonas axonopodis*.

```
LOCUS       AAM35948                686 aa            linear   BCT 29-MAY-2002
DEFINITION  phage-related tail protein [Xanthomonas axonopodis pv. citri str.
            306].
ACCESSION   AAM35948
VERSION     AAM35948.1  GI:21107209
DBSOURCE    accession AE011736.1
KEYWORDS    .
SOURCE      Xanthomonas axonopodis pv. citri str. 306
  ORGANISM  Xanthomonas axonopodis pv. citri str. 306
            Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
            Xanthomonadaceae; Xanthomonas.
REFERENCE   1  (residues 1 to 686)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida,N.F.
            Jr., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A.,
            Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P.,
            Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H.,
```

```
            Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T.,
            Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A.,
            Katsuyama,A.M., Kishi,L.T., Leite,R.P. Jr., Lemos,E.G.M.,
            Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N.,
            Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M.,
            Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K.,
            Oliveira,M.C., Oliveira,V.R., Pereira,H.A. Jr., Rossi,A.,
            Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A.,
            Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M.,
            Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
   PUBMED   12024217
REFERENCE   2  (residues 1 to 686)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida,N.F.
            Jr., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A.,
            Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P.,
            Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H.,
            Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T.,
            Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A.,
            Katsuyama,A.M., Kishi,L.T., Leite,R.P. Jr., Lemos,E.G.M.,
            Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N.,
            Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M.,
            Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K.,
            Oliveira,M.C., Oliveira,V.R., Pereira,H.A. Jr., Rossi,A.,
            Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A.,
            Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M.,
            Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
COMMENT     Method: conceptual translation supplied by author.
FEATURES             Location/Qualifiers
     source          1..686
                     /organism="Xanthomonas axonopodis pv. citri str. 306"
                     /strain="306"
                     /db_xref="taxon:190486"
                     /note="pathovar: citri"
     Protein         1..686
                     /product="phage-related tail protein"
     CDS             1..686
                     /gene="stf"
                     /coded_by="complement(AE011736.1:3489..5549)"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /transl_table=11
ORIGIN
        1 mdsskdsdne elmaetpgyv snaklaeris alvdrwntre gqmialltqp qgavtvtdgl
       61 gkdhvlpsfl qlskdvselv deltgavsga sefanmarly aegaqasand aqtsaedaaq
      121 qladatdqae asaasatasa asasasessn qsaglhdsaa sasaadadad aiqaanskaq
      181 avasasaaas seaqaasyah qaedwsdhaq swanapagte iepgafsakh wseqakaslt
      241 gtlvyrggwd agtgsfpsgg ntgafykvtt ggsiggrqyn pgdqivhngs gwdhidnteq
      301 vtsvagksga vqlvpsdisg lgtlatrdsv dfsshvtnkp asyppsghmh tkgdvglsnv
      361 dntsdankpv snaqqaaldt kapidkpsft ggvtvrnssl rlsgwggian dgvvyfgasd
```

```
      421 swlfkqgdaf rfqiegkfna sldssgtvwt sgnfnpaskv dmggqprfag vfvganndqy
      481 fyaedanqsv vlrfgygtsf kyakwngvtg aftaphliag gsgngayvqi gddvrlvdig
      541 rshttaiqst angnigflqf gtgptigwdg ttlqaagqel khagntgyyg vsgrqhsdwn
      601 tcvtqgiwma vgatnspdns dwwlghvtvh ngdwvqqevi nftsnppkqf rrqrqggswg
      661 acqqcgmviv sttdpggadg vlwiqp
```

Table 5.9 Features of cell wall surface anchor family protein from *Streptococcus agalactiae*.

```
LOCUS       AAN00330                970 aa            linear   BCT 23-SEP-2002
DEFINITION  cell wall surface anchor family protein [Streptococcus agalactiae
            2603V/R].
ACCESSION   AAN00330
VERSION     AAN00330.1  GI:22534491
DBSOURCE    accession AE014259.1
KEYWORDS    .
SOURCE      Streptococcus agalactiae 2603V/R
  ORGANISM  Streptococcus agalactiae 2603V/R
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (residues 1 to 970)
  AUTHORS   Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
            Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
            Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
            Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
            Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
            Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
            Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
            Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,
            Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
  TITLE     Complete genome sequence and comparative genomic analysis of an
            emerging human pathogen, serotype V Streptococcus agalactiae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
   PUBMED   12200547
REFERENCE   2  (residues 1 to 970)
  AUTHORS   Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
            Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
            Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
            Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
            Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
            Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
            Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
            Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,
            Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..970
                     /organism="Streptococcus agalactiae 2603V/R"
                     /strain="2603V/R"
                     /serotype="V"
                     /db_xref="taxon:208435"
     Protein         1..970
```

```
                     /product="cell wall surface anchor family protein"
     CDS             1..970
                     /gene="SAG1462"
                     /coded_by="complement(AE014259.1:196..3108)"
                     /note="identified by match to PFAM protein family HMM
                     PF00746"
                     /transl_table=11
ORIGIN
        1 msqktfgkql tvvdtksrvk mhkseknwvr tvmshfnlfk aikgratvea dvciqdveke
       61 drlssgnlty lkgilaagal vggasltsrv yadetpvvqe qsssvptlae qtevtvkttt
      121 vqnhqdgtvs kniidsnsvs msesaststs esvsmsmsgs tltsvsesvs tsaltsases
      181 istsasesvs kstsisevsn iletqasltd kgresfsanq ivtesslvtd agknasvssl
      241 ieitkpksel qtskmsnesl itpeksqvmi asdktgnesl tptirlksvi qprsmnlmtl
      301 ssemdliple evsdtemlgk dvsselqkvn ialkdntlse pgtvkldsse nlvlnfafsi
      361 asvnegdvft vklsdnldtq gigtilkvqd imdetgqlla tgsysplthn itytwtryas
      421 tlnnikarvn mpvwpdqrii skttsdkqcf tatlnnqvas ieervqynsp svtehtnvkt
      481 nvrsrimkld derqtetyit qinpegkemy fasglgnlyt iigsdgtsgs pvnllnaevk
      541 ilktnsknlt dsmdqnydsp efedvtsqys ytndgskiti dwktnsisst tsyvvlvkip
      601 kqsgvlystv sdinqtygsk ysyghtnisg dsdanaeikl lsesastsas tsastsasms
      661 astsastsas msastsasts astsasmsas tsastsasts astsastsas msastsasts
      721 astsastsas tsastsasms astsastsas tsastsasms astsastsas tsastsasms
      781 astsastsas tsastsasms astsastsas tsastsasms astsastsas tsasmsasts
      841 astsasmsas tsastsasms astsastsas msastsasts asmsastsas msattsasts
      901 vstsastsas tsastsssss vtsnsskekv ysalpstgdq dysvtatalg lglmtgatll
      961 grkkskkdkd
```

Table 5.10 Features of α subunit of DNA polymerase III from *Aquifex aeolicus*.

```
LOCUS       O67125                  1161 aa            linear   BCT 13-JUN-2006
DEFINITION  DNA polymerase III subunit alpha.
ACCESSION   O67125
VERSION     O67125  GI:6014991
DBSOURCE    swissprot: locus DPO3A_AQUAE, accession O67125;
            class: standard.
            created: Dec 15, 1998.
            sequence updated: Aug 1, 1998.
            annotation updated: Jun 13, 2006.
            xrefs: AE000657.1, AAC07087.1, B70387
            xrefs (non-sequence databases): GenomeReviews:AE000657_GR,
            BioCyc:AAEO63363:AQ_1008-MONOMER, InterPro:IPR011708,
            InterPro:IPR012340, InterPro:IPR004365, InterPro:IPR004013,
            InterPro:IPR003141, InterPro:IPR004805, Pfam:PF07733, Pfam:PF02811,
            Pfam:PF01336, SMART:SM00481, TIGRFAMs:TIGR00594
KEYWORDS    Complete proteome; DNA replication; DNA-directed DNA polymerase;
            Nucleotidyltransferase; Transferase.
SOURCE      Aquifex aeolicus
  ORGANISM  Aquifex aeolicus
            Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
REFERENCE   1  (residues 1 to 1161)
  AUTHORS   Deckert,G., Warren,P.V., Gaasterland,T., Young,W.G., Lenox,A.L.,
            Graham,D.E., Overbeek,R., Snead,M.A., Keller,M., Aujay,M.,
            Huber,R., Feldman,R.A., Short,J.M., Olsen,G.J. and Swanson,R.V.
  TITLE     The complete genome of the hyperthermophilic bacterium Aquifex
            aeolicus
```

```
  JOURNAL   Nature 392 (6674), 353-358 (1998)
   PUBMED   9537320
  REMARK    NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
            STRAIN=VF5
COMMENT     On Jan 20, 2006 this sequence version replaced gi:7434831.
            [FUNCTION] DNA polymerase III is a complex, multichain enzyme
            responsible for most of the replicative synthesis in bacteria. This
            DNA polymerase also exhibits 3' to 5' exonuclease activity. The
            alpha chain is the DNA polymerase (By similarity).
            [CATALYTIC ACTIVITY] Deoxynucleoside triphosphate + DNA(n) =
            diphosphate + DNA(n+1).
            [SUBUNIT] DNA polymerase III contains a core (composed of alpha,
            epsilon and theta chains) that associates with a tau subunit. This
            core dimerizes to form the PolIII' complex. PolIII' associates with
            the gamma complex (composed of gamma, delta, delta', psi and chi
            chains) and with the beta chain to form the complete DNA polymerase
            III complex (By similarity).
            [SUBCELLULAR LOCATION] Cytoplasm (By similarity).
            [SIMILARITY] Belongs to the DNA polymerase type-C family. DnaE
            subfamily.
FEATURES             Location/Qualifiers
     source          1..1161
                     /organism="Aquifex aeolicus"
                     /db_xref="taxon:63363"
     gene            1..1161
                     /gene="dnaE"
                     /locus_tag="aq_1008"
     Protein         1..1161
                     /gene="dnaE"
                     /locus_tag="aq_1008"
                     /product="DNA polymerase III subunit alpha"
                     /EC_number="2.7.7.7"
     Region          1..1161
                     /gene="dnaE"
                     /locus_tag="aq_1008"
                     /region_name="Mature chain"
                     /experiment="experimental evidence, no additional details
                     recorded"
                     /note="DNA polymerase III subunit alpha.
                     /FTId=PRO_0000103308."
     Region          2..1138
                     /gene="dnaE"
                     /locus_tag="aq_1008"
                     /region_name="DnaE"
                     /note="DNA polymerase III, alpha subunit [DNA replication,
                     recombination, and repair]; COG0587"
                     /db_xref="CDD:30932"
     Region          6..70
                     /gene="dnaE"
                     /locus_tag="aq_1008"
                     /region_name="POLIIIAc"
                     /note="DNA polymerase alpha chain like domain; DNA
                     polymerase alpha chain like domain, incl; smart00481"
                     /db_xref="CDD:47787"
     Region          115..222
                     /gene="dnaE"
                     /locus_tag="aq_1008"
```

```
                     /region_name="PHP_C"
                     /note="PHP domain C-terminal region; pfam02811"
                     /db_xref="CDD:42765"
     Region          1000..1075
                     /gene="dnaE"
                     /locus_tag="aq_1008"
                     /region_name="tRNA_anti"
                     /note="OB-fold nucleic acid binding domain; pfam01336"
                     /db_xref="CDD:41388"
ORIGIN
        1 mskdfvhlhl htqfslldga ikidelvkka keygykavgm sdhgnlfgsy kfykalkaeg
       61 ikpiigmeay fttgsrfdrk tktsednitd kynhhlilia kddkglknlm klstlaykeg
      121 fyykpridye llekygegli altaclkgvp tyyasinevk kaeewvkkfk difgddlyle
      181 lqannipeqe vanrnlieia kkydvkliat qdahylnped ryahtvlmal qmkktihels
      241 sgnfkcsned lhfappeymw kkfegkfegw ekallntlev mektadsfei fenstyllpk
      301 ydvppdktle eylrelaykg lrqriergqa kdtkeywerl eyelevinkm gfagyflivq
      361 dfinwakknd ipvgpgrgsa ggslvayaig itdvdpikhg flferflnpe rvsmpdidvd
      421 fcqdnrekvi eyvrnkyghd nvaqiitynv mkakqtlrdv aramglpyst adklaklipq
      481 gdvqgtwlsl eemyktpvee llqkygehrt diednvkkfr qiceespeik qlvetalkle
      541 gltrhtslha agvviapkpl selvplyydk egevatqydm vqleelgllk mdflglktlt
      601 elklmkelik erhgvdinfl elplddpkvy kllqegkttg vfqlesrgmk ellkklkpds
      661 fddivavlal yrpgplksgl vdtyikrkhg kepveypfpe lepvlketyg vivyqeqvmk
      721 msqilsgftp geadtlrkai gkkkadlmaq mkdkfiqgav ergypeekir klwediekfa
      781 sysfnkshsv aygyisywta yvkahypaef favklttekn dnkflnlikd aklfgfeilp
      841 pdinksdvgf tiegenrirf glarikgvge etakiivear kkykqfkgla dfinktknrk
      901 inkkvvealv kagafdftkk krkellakva nsekalmatq nslfgapkee veeldplkle
      961 kevlgfyisg hpldnyekll knrytpiedl eewdkeseav ltgvitelkv kktkngdyma
     1021 vfnlvdktgl iecvvfpgvy eeakelieed rvvvvkgfld edletenvkf vvkevfspee
     1081 fakemrntly iflkreqaln gvaeklkgii ennrtedgyn lvltvdlgdy fvdlalpqdm
     1141 klkadrkvve eieklgvkvi i
```

6. REFERENCES

Ackermann, H.-W., 2001. Frequency of morphological phage descriptions in the year 2000. Archives of Virology **146:** 843 – 857.

Ahmad, S.I., Kirk, S.H., and Eisenstark, A., 1998. Thymine metabolism and thymineless death in prokaryotes and eukaryotes. Annual Review of Microbiology **52:** 591 – 625.

Anderson, D.L., and Bradley, S.G., 1961. Susceptibility of *nocardiae* and mycobacteria to actinophage. Antimicrobial Agents and Chemotherapy **3:** 894 – 903.

Bandow, J.E., Brotz, H., Leichert, L.I., Labischinski, H., and Hecker, M., 2003. Proteomics approach to understanding antibiotic action. Antimicrobial Agents and Chemotherapy **47:** 948 – 955.

Barner, H.D., and Cohen, S.S., 1954. The induction of thymine synthesis by T2 infection of a thymine requiring mutant of *Escherichia coli*. Journal of Bacteriology **68:** 80 – 88.

Becherel, O.J., and Fuchs, R.P., 2001. Mechanism of DNA polymerase III-mediated frameshift mutagenesis. Proceedings of the National Academy of Sciences of USA **98:** 8566 – 8571.

Belfort, M., and Roberts, R.J., 1997. Homing endonucleases: keeping the house in order. Nucleic Acids Research **25:** 3379 – 3388.

Bergmann, S., and Hammerschmidt, S., 2006. Versatility of pneumococcal surface proteins. Microbiology **152:** 295 – 303.

Bernhardt, T.G., Roof, W.D., and Young, R., 2000. Genetic evidence that the bacteriophage ϕX174 lysis protein inhibits cell wall synthesis. Proceedings of the National Academy of Sciences of USA **97:** 4297 – 4302.

Bernhardt, T.G., Wang, I.N., Struck, D.K., and Young, R., 2001. A protein antibiotic in the phage Qβ virion: diversity in lysis targets. Science **292:** 2263-2264.

Besemer, J., and Borodovsky, M., 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Research **27:** 3911 – 3920.

Bibb, M.J., Findlay, P.R., and Johnson, M.W., 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene **30:** 157 – 166.

Bingham, R., Ekunwe, S.I., Falk, S., Snyder, L., and Kleanthous, C., 2000. The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. Journal of Biological Chemistry **275:** 23219 – 23226.

Boman, H.G., 1995. Peptide antibiotics and their role in innate immunity. Annual Review of Immunology **13:** 61 – 92.

Bradbury, J., 2004. My enemy's enemy is my friend. The Lancet **363:** 624-625.

Braun, V., Pilsl, H., and Gross, P., 1994. Colicins: structures, modes of action, transfer through membranes, and evolution. Archives of Microbiology **161:** 199 – 206.

Breitbart M., Salamon, P., Andersen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., and Rohwer, F., 2002. Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of the Sciences of USA **99:** 14250 – 14255.

Brooks, D.R., Wang, P., Read, M., Watkins, W.M., Sims, P.F., and Hyde, J.E., 1994. Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: dihydropteroate synthase gene in lines of the human malaria parasite, Plasmodium falciparum, with differing resistance to sulfadoxine. European Journal of Biochemistry **224:** 397 – 405.

Brown-Elliot, B.A., Brown, J.M., Conville, P.S., and Wallace, R.J.Jr, 2006. Clinical and laboratory features of the *Nocardia* spp. based on current molecular taxonomy. Clinical Microbiology Reviews **19:** 259 – 282.

Brownell, G., Adams, J., and Bradley, S., 1967. Growth and characterization of nocardiophages for *Nocardia canicruria* and *Nocardia erythropolis* mating types. Journal of General Microbiology **47:** 247 – 256.

Bruck, I., Yuzhakov, A., Yurieva, O., Jeruzalmi, D., Skangalis, M., Kuriyan, J., and O'Donnel, M., 2002. Analysis of a mulitcomponent thermostable DNA polymerase III replicase from an extreme thermophile. Journal of Biological Chemistry **277:** 17334 – 17348.

Bush, K., 2004. Antibacterial drug discovery in the 21st century. Clinical Microbiology and Infection **10:** S10 -S17.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A., 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell **104:** 901 – 912.

Cao, X.-M, Huang, L.-H., Farnet, C.M., and Ehrlich, M., 1983. Ligation of highly modified bacteriophage DNA. Biochimica et Biophysica Acta **741:** 237 – 243.

Carreras, C.W., and Santi, D.V., 1995. The catalytic mechanism and structure of thymidylate synthase. Annual Review of Biochemistry **64:** 721 – 762.

Carter, A.P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T., Ramakrishnan, V., 2000. Functional insights from the structure of the 30S ribosomal subunit and its interaction with antibiotics. Nature **407:** 340 – 348.

Cheng, H., Shen, N., Pei, J., and Grishin N.V., 2004. Double-stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. Protein Science **13:** 2260 – 2269.

Chevalier, B.S., and Stoddard, B.L., 2001. Homing endonucleases: structural and functional insight into the catalysis of intron/intein mobility. Nucleic Acids Research **29:** 3757 – 3774.

Chopra, I., and Roberts, M., 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiology and Molecular Biology Reviews **65:** 232 – 260.

Cimarelli, A., and Luban, J., 1999. Translation elongation factor 1-alpha interacts specifically with the human immunodeficiency virus type 1 gag polyprotein. Journal of Virology **73:** 5388 – 5401.

Coene, M., De Kesel, M., Hong, N.T.T., Gheysen, A., Jezierska-Anczukow, A., and Cocito, C., 1993. Comparative analysis of the genomes of mycobacteriophages infecting saprophytic and pathogenic mycobacteria. Archives of Virology **133:** 39 – 49.

Corbett, E.L., Watt, C.J., Walker, N., Maher, D., Williams, B.G., Raviglione, M.C., Dye, C., 2003. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Archives of Internal Medicine **163:** 1009 – 1021.

Corti, M.E., and Fioti, M.F.V., 2003. Nocardiosis: a review. International Journal of Infectious Diseases **7:** 243 – 250.

Cregg, J.M., Nguyen, A.H., and Ito, J., 1980. DNA modification induced during infection of *Bacillus subtilis* by phage ϕ3T. Gene **12:** 17 – 24.

Curtis, T.P., Sloan, W.T., and Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of USA **99:** 10494 – 10499.

Dabbs, E.R., 1987. A generalized transducing bacteriophage for *Rhodococcus erythropolis*. Molecular and General Genetics **206:** 116 – 120.

Dabbs, E.R., 1998. Cloning of genes that have environmental and clinical importance from rhodococci and related bacteria. Antonie van Leeuwenhoek **74:** 155 – 168.

Dabbs, E.R., Gowan, B., Quan, S., and Andersen, S.J., 1995. Development of improved *Rhodococcus* plasmid vectors and their use in cloning genes of potential commercial and medical importance. Biotekhnologiia **7-8:** 129 – 135.

Datta, I., Sau, S., Sil, A.K., Mandal, N.C., 2005. The bacteriophage λ DNA replication protein P inhibits the *ori*C DNA- and ATP-binding functions of the DNA replication initiator protein DnaA of *Escherichia coli*. Journal of Biochemistry and Molecular Biology **38:** 97 – 103.

Delmas, S., and Matic, I., 2006. Interplay between replication and recombination in *Escherichia coli*: impact of the alternative DNA polymerases. Proceedings of the National Academy of Sciences of USA **103:** 4564 – 4569.

Derbyshire, V., and Belfort, M., 1998. Lightning strikes twice: intron-intein coincidence. Proceedings of the National Academy of Sciences of USA **95:** 1356 – 1357.

Destoumieux-Garzón, D., Peduzzi, J., and Rebuffat, S., 2002. Focus on modified microcins: structural features and mechanisms of action. Biochimie **84:** 511 – 519.

Duda, R.L., Martincic, K., and Hendrix, R.W., 1995. Genetic basis of bacteriophage HK97 prohead assembly. Journal of Molecular Biology **247:** 636 – 647.

Duport, C., Baysse, C., and Michel-Briand, Y., 1995. Molecular characterization of pyocin S3, a novel S-type pyocin from *Pseudomonas aeruginosa*. Journal of Biological Chemistry **270:** 8920 – 8927.

Ellen, R.P., and Gibbons, R.J., 1972. M-protein associated adherence of *Streptococcus pyogenes* to epithelial surfaces: prerequisite for virulence. Infections and Immunology **5:** 826 – 830.

Fischetti, V.A., 1989. Streptococcal M protein: molecular design and biological behaviour. Clinical Microbiology Reviews **2:** 285 – 314.

Garcia-Villegas, M.R., De La Vega, F.M., Galindo, J.M., Segura, M., Buckingham, R.H., and Guarneros, G., 1991. Peptidyl-tRNA hydrolase is involved in λ inhibition of host protein synthesis. EMBO Journal **10:** 3549 – 3555.

Geize, R., Gerda, I.H., and Dijkhuizen, L., 2002. Molecular and functional characterization of the kstD2 gene of *Rhodococcus erythropolis* SQ1 encoding a second 3-ketosteroid Δ'-dehydrogenase isoenzyme. Microbiology **148:** 3285 – 3292.

Gilcrease, E.B., Winn-Stapley, D.A., Hewitt, F.C., Joss, L., and Casjens, S.R., 2005. Nucleotide sequence of the head assembly gene cluster of bacteriophage L and decoration protein characterization. Journal of Bacteriology **187:** 2050 – 2057.

Gillor, O., Nigro, L.M., and Riley, M.A., 2005. Genetically engineered bacteriocins and their potential as the next generation of antimicrobials. Current Pharmaceutical Design **11:** 1067 – 1075.

Gold, M., Gefter, M., Hausmann, R., and Hurwitz, J., 1966. Methylation of DNA. Journal of General Physiology **49:** 5 – 28.

Hatfull, G.F., Pedulla, M.L., Jacobs-Sera, D., Cichon, P.M., Foley, A., Ford, M.E., Gonda, R.M., Houtz, J.M., Hryckowian, A.J., Kelchner, V.A., Namburi, S., Pajcini, K.V., Popovich, M.G., Schleicher, D.T., Simanek, B.Z., Smith, A.L., Zdanowicz, G.M., Kumar, V., Peebles, C.L., Jacobs Jr.,

W.R., Lawrence, J.G., and Hendrix, R.W., 2006. Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS Genetics **2:** 1 – 13.

Hattman, S., 1980. Specificity of the bacteriophage MU mom^{+}-controlled DNA modification. Journal of Virology **34:** 277 – 279.

Hiddema, R., Curran, M.D., Ferreira, N.P., Coetzee, J.N., and Lecatsas, G., 1985. Characterization of phages derived from strains of *Rhodococcus australis* and *R. equi*. Intervirology **23:** 109 – 111.

Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley, J.T. *et al*., 1994. Group 22, nocardioform actinomycetes, pp. 625 – 650. In: Holt J.G. (Ed.), 9th ed., Bergey's Manual of Determinative Bacteriology, Williams & Wilkins, Baltimore, MD.

Hsueh, P.R., Hung, C.C., Teng, L.J., Yu, M.C., Chen, Y.C., Wang, H.K., and Luth, K.T., 1998. Report of invasive *Rhodococcus equi* infections in Taiwan, with an emphasis on the emergence of multidrug-resistant strains. Clinical Infectious Diseases **27:** 370-375.

Huang, L.-H., Farnet, C.M., Ehrlich, K.C., and Ehrlich, M., 1982. Digestion of highly modified bacteriophage DNA by restriction endonucleases. Nucleic Acids Research **10:** 1579 – 1591.

Igarashi, K., Fujita, N., and Ishihama, A., 1991. Identification of a subunit assembly domain in the alpha subunit of *Escherichia coli* RNA polymerase. Journal of Molecular Biology **218:** 1 – 6.

James, R., Penfold, C.N., Moore, G.R., and Keanthous, C., 2002. Killing of *E. coli* cells by E group nuclease colicins. Biochimie **84:** 381 – 389.

Kawaguchi, Y., Bruni, R., and Roizman, B., 1997. Interaction of herpes simplex virus 1 alpha regulatory protein ICP0 with elongation factor 1delta: ICP0 affects translational machinery. Journal of Virology **71:** 1019 – 1024.

Kawamura, F., Mizukami, T., Shimotsu, H., Anzai, H., Takahashi, H., and Saito, H., 1981. Unusually infrequent cleavage with several endonucleases and physical map construction of *Bacillus subtilis* bacteriophage ϕ1 DNA. Journal of Virology **37:** 1099 – 1102.

Kedlaya, I., Ing, M.B., and Wong, S.S., 2001. *Rhodococcus equi* infections in immunocompetent hosts: case report and review. Clinical Infectious Diseases **32:** e39-46.

Kędzierska, B., Glinkowska, M., Iwanichi, A., Obuchowski, M., Sojka, P., Thomas, M.S., Węgrzyn, G., 2003. Toxicity of the bacteriophage λ *cII* gene product to *Escherichia coli* arises from inhibition of host cell DNA replication. Virology **313:** 622 – 628.

Knox, J.R., Moews, P.C., and Frere, J.M., 1996. Molecular evolution of bacterial beta-lactam resistance. Chemical Biology **3:** 937 – 947.

Krüger and Bickle, 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiological Reviews **47:** 345 – 360.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J., 2005. The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. Proceedings of the National Academy of Sciences of USA **102:** 5174 – 5179.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J., 2006. Comparative genome analysis of 18 *Pseudomonas aeruginosa* bacteriophages. Journal of Bacteriology **188:** 1184 – 1187.

Kutter, E., Guttman, B., and Carlson, K., 1994. The transition from host to phage metabolism after T4 infection, p. 343 – 346. In Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserling, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., and Miller, S. (ed.), Molecular biology of bacteriophage T4. American Society for Microbiology, Washington, D.C.

Lazarevic, V., 2001. Ribonucleotide reductase genes of *Bacillus* prophages: a refuge to introns and intein coding sequences. Nucleic Acids Research **29:** 3212 – 3218.

Le Chatelier, E., Bécherel, O.J., d'Alençon, Canceill, D., Ehrlich, S.D., Fuchs, R.P.P., and Jannière, L., 2004. Involvement of DnaE, the second replicative DNA polymerase from *Bacillus subtilis*, in DNA mutagenesis. Journal of Biological Chemistry **279:** 1757 – 1767.

Lederman, E.R., and Crum, N.F., 2004. A case series and focused review of nocardiosis: clinical and microbiological aspects. Medicine **83:** 300 – 313.

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferreti, V., Ha, N., Kwan, T., McCarty, J., Srikumar, R., Williams, D., Wu, J., Gros, P., Pelletier, J., and DuBow, M., 2004. Antimicrobial drug discovery through bacteriophage genomics. Nature Biotechnology **22:** 185-191.

Livermore, D.M., 2004. The need for new antibiotics. Clinical Microbiology and Infection **10:** S1 – S9.

Loeffler, J.M., Nelson, D., and Fischetti, V.A., 2001. Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science **294:** 2170-2172.

Maki, H., and Kornberg, A., 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli*. Purification of the alpha subunit, devoid of nuclease activities. Journal of Biological Chemistry **260:** 12987 – 12992.

Marchand, I., Nicholson, A.W., and Dreyfus, M., 2001. Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. Molecular Microbiology **42:** 767 – 776.

Mathews, I.I., Deacon, A.M., Canaves, J.M., McMullan, D., Lesley, S.A., Agarwalla, S., and Kuhn, P., 2003. Functional analysis of substrate and cofactor complex structures of a thymidylate synthase-complementing protein. Structure **11:** 677 – 690.

Maxwell, A., 1997. DNA gyrase as a drug target. Trends in Microbiology **5:** 102 – 109.

Mayer, J.E., and Schweiger, 1983. RNase III is positively regulated by T7 protein kinase. Journal of Biological Chemistry **258:** 5340 – 5343.

McNeil, M.M., and Brown, J.M, 1994. The medically important aerobic actinomycetes: epidemiology and microbiology. Clinical Microbiology Reviews **7:** 357 – 417.

Merril, C.R., Biswas, B., Carlton, R., Jensen, N.C., Creed, G.J., Zullo, S., and Adhya, S., 1996. Long-circulating bacteriophage as antimicrobial agents. Proceedings of the National Academy of Sciences of USA **93:** 3188-3192.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Rüger, W., 2003. Bacteriophage T4 genome. Microbiology and Molecular Biology Reviews **67:** 86 – 156.

Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P., and Liebl, U., 2002. An alternative flavin-dependent mechanism for thymidylate synthesis **297:** 105 – 107.

Nakayama, K., Takashima, K., Ishihara, H., Shinomiya, T., Kageyama, M., Kanaya, S., Ohnishi, M., Murata, T., Mori, H., and Hayashi, T., 2000. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Molecular Microbiology **38:** 213 – 231.

Nechaev, S., and Severinov, K., 1999. Inhibition of *E. coli* RNA polymerase by bacteriophage T7 gene 2 protein. Journal of Molecular Biology **289:** 815 – 826.

Nikolskaya, I.I., Tediashvili, M.G., Vasielieva, M.B., Chanishvili, T.G., and Debov, S.S., 1979. Specificity and functions of guanine methylase of *Shigella sonnei* DDVI phage. Biochimica et Biophysica Acta **561:** 232 – 239.

Nordmann, P., Rouveix E., Guenounou, M., and Nicolas, M.H., 1992. Pulmonary abscess due to a rifampin and fluoroquinolone resistant *Rhodococcus equi* strain in a HIV infected patient. European Journal of Clinical Microbiology and Infectious Diseases **11:** 557-558.

O'Donnell, M., and Studwell, P.S., 1990. Total reconstitution of DNA polymerase III holoenzyme reveals dual accessory protein clamps. Journal of Biological Chemistry **265:** 1171 – 1178.

Pag, U., and Sahl, G., 2002. Multiple activities in lantibiotics – models for the design of novel antibiotics. Current pharmaceutical design **8:** 815 – 833.

Papagianni, M., 2003. Ribosomally synthesized peptides with antimicrobial properties: biosynthesis, structure, function, and applications **21:** 465 – 499.

Parret, A., and D Mot, R., 2000. Novel bacteriocins with predicted tRNase and pore-forming activities in *Pseudomonas aeruginosa* PAO1. Molecular Microbiology **35:** 472 – 475.

Patel, U., Yan, Y.P., Hobbs, F.W., Kaczmarczyk, J., Slee, A.M., Pompliano, D.L., Kurilla, M.G., and Bobkova, E.V., 2001. Oxazolidinones mechanism of action: inhibition of the first peptide bond formation. Journal of Biological Chemistry **276:** 37199 – 37205.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J.A., Pannunzio, N.R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardoarov, S., Kriakov, J., Lawrence, J.G., Jacobs Jr., W.R., Hendrix, R.W., and Hatfull, G.F., 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113:** 171-182.

Pham, H.T., Riu, K.Z., Jang, K.M., Cho, S.K., and Cho, M., 2004. Bactericidal activity of glycinecin A, a bacteriocin derived from *Xanthomonas campestris* pv. glycines, on phytopathogenic *Xanthomonas campestris* pv. vesicatoria cells. Applied and Environmental Microbiology **70:** 4486 – 4490.

Pirisi, A., 2000. Phage therapy-advantage over antibiotics? The Lancet **356:** 1418.

Pond, F.R., Gibson, I., Lalucat, J., and Quackenbush, R.L., 1989. R-body-producing bacteria. Microbiological Reviews **53:** 25 – 67.

Pons, A.-M., Lanneluc, I., Cottenceau, G., and Sable, S., 2002. New developments in non-post translationally modified microcins. Biochimie **84:** 531 – 537.

Powers, J.H., 2004. Antimicrobial drug development – the past, the present, and the future. Clinical Microbiology and Infection **10:** S23 – S31.

Projan, S.J., 2004. Phage-inspired antibiotics? Nature Biotechnology **22:** 167 – 168.

Projan, S.J., and Shlaes, D.M., 2004. Antibacterial drug discovery: is it all downhill from here? Clinical Microbiology and Infection **10:** S18 – S22.

Putman, M., Van Veen, H.W., and Konings, W.N., 2000. Molecular properties of bacterial multidrug transporters. Microbiology and Molecular Biology Reviews **64:** 672 – 693.

Quan, S., and Dabbs, E.R., 1993. Nocardioform arsenic resistance plasmid characterization and improved *Rhodococcus* cloning vectors. *Plasmid* **29:** 74 – 79.

Quan, S., Venter, H., Dabbs, E.R., 1997. Ribosylative inactivation of rifampin by *Mycobacterium smegmatis* is a principal contributor to its low susceptibility to this antibiotic. Antimicrobial Agents and Chemotherapy **41:** 2456 – 2460.

Rawlings, N.D., O'Brien, E., and Barrett, A.J., 2002. MEROPS: the protease database. Nucleic Acids Research **30:** 343 – 346.

Reeve, J.N., Amann, E., Tailor, R., Günthert, U., Scholz, K., and Trautner, T.A., 1980. Unusual behaviour of SPO1 DNA with respect to restriction and modification enzymes recognizing the sequence 5'-G-G-C-C. Molecular and General Genetics **178:** 229 – 231.

Riverin, M., Beaudoin, J., and Vezina, C., 1970. Characterization of a nocardiophage for *Nocardia restrictus*, Journal of General Virology **6:** 395 – 407.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D., 2005. REBASE restriction enzymes and methyltranferases. Nucleic Acids Research **33:** D230 – D232.

Robertson, E.S., Aggison, L.A., and Nicholson, A.W., 1994. Phosphorylation of elongation factor G and ribosomal protein S6 in bacteriophage T7-infected *Escherichia coli*. Molecular Microbiology **11:** 1045 – 1057.

Rohwer, F., 2003. Global phage diversity. Cell **113:** 141.

Salinas-Carmona, M.C., 2000. *Nocardia brasiliensis*: from microbe to human and experimental infections. Microbes and infection **2:** 1373 – 1381.

Sambrook, J., Fritsch, E.F., and Maniatis, T., 1989. Bacteriophage λ growth, purification, and DNA extraction, pp. 2.60 – 2.81, E.39. In: Molecular cloning, a laboratory manual, 2nd ed., vol.3. Cold Spring Harbour Laboratory Press, New York.

Sampath, A., and Stewart, C.R., 2004. Roles of genes 44, 50, and 51 in regulating gene expression and host takeover during infection of *Bacillus subtilis* by bacteriophage SPO1. Journal of Bacteriology **186:** 1785 – 1792.

Schlunzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilij, A., Albrecht, R., Yonath, A., and Franceschi, R., 2001. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. Nature **413:** 814 – 821.

Schuch, R., Nelson, D., and Fischetti, V.A., 2002. A bacteriolytic agent that detects and kills *Bacillus anthracis*. Nature **418:** 825-826.

Sergueev, K., Yu, D., Austin, S., and Court, D., 2001. Cell toxicity caused by products of the *pL* operon of bacteriophage lambda. Gene **272:** 227 – 235.

Severinova, E., and Severinov, K., 2006. Localization of the *Escherichia coli* RNA polymerase β' subunit residue phosphorylated by bacteriophage T7 kinase gp0.7. Journal of Bacteriology **188:** 3470 – 3476.

Sharma, M.S., Ellis, R.L., and Hinton, D.M., 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. Proceedings of the National Academy of Sciences of USA **89:** 6658 – 6662.

Shen, R., Olcott, M.C., Kim, J.H., Rajagopal, I., and Mathews, C.K., 2004. *Escherichia coli* nucleoside diphosphate kinase interactions with T4 phage proteins of deoxyribonucleotide synthesis and possible regulatory functions. Journal of Biological Chemistry **279:** 32225 – 32232.

Shibayama, Y., 2004. Isolation and characterization of novel bacteriophages as potential biocontrol agents against *Rhodococcus equi*. BSc Honours research report. University of the Witwatersrand, Johannesburg.

Simon, L.D., 1969. The infection of *Escherichia coli* by T2 and T4 bacteriophages as seen in the electron microscope: membrane-associated intracellular bacteriophages. Virology **38:** 285 – 296.

Snyder, L., Gold, L., and Kutter, E., 1976. A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. Proceedings of the National Academy of Sciences of USA **73:** 3098 – 3102.

Stecker, C., Johann, A., Herzberg, C., Averhoff, B., and Gottschalk, 2003. Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. Journal of Bacteriology **185:** 5269 – 5274.

Stewart, C.R., Gaslightwala, I., Hinata, K., Krolikowski, K., Needleman, D.S., Peng, A.S.-Y., Peterman, M.A., Tobias, A., and Wei, P., 1998. Genes and regulatory sites of the "host-takeover module" in the terminal redundancy of *Bacillus subtilis* bacteriophage SPO1. Virology **246:** 329 – 340.

Stone, R., 2002. Stalin's forgotten cure. Science **298:** 728-732.

Stukenberg, P.T., Studwell-Vaughan, P.S., O'Donnell, M., 1991. Mechanism of the sliding beta-clamp of DNA polymerase holoenzyme. Journal of Biological Chemistry **266:** 11328 – 11334.

Sutton, M.D., and Walker, G.C., 2001. Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. Proceedings of the National Academy of Sciences of USA **98:** 8342 – 8349.

Takasugi, J.E., and Godwin, J.D., 1991. Lung abscess caused by *Rhodococcus equi*. Journal of Thoracic Imaging **6:** 72 – 74.

Tang, M., Shen, X., Frank, E.G., O'Donnell, M., Woodgate, R., and Goodman, M.F., 1999. UmuD'$_2$C is an error-prone DNA polymerase, *Escherichia coli* polV. Proceedings of the National Academy of Sciences of USA **96:** 8919 – 8924.

Thacker, P.D., 2003. Set a microbe to kill a microbe. Journal of the American Medical Association **290:** 3183.

Thomson, C.J., Power, E., Ruebsamen-Waigmann, H., and Labischinski, H., 2004. Antibacterial research and development in the 21st century – an industry perspective of the challenges. Current Opinion in Microbiology **7:** 445 – 450.

Tiemann, B., Depping, R., and Ruger, W., 1999. Overexpression, purification, and partial characterization of ADP-ribosyltransferases modA and modB of bacteriophage T4. Gene Expression **8:** 187 – 196.

Viguera, E., Petranovic, M., Zahradka, D., Germain, K., Ehrlich, D.S., and Michel, B., 2003. Lethality of bypass polymerases in *Escherichia coli* cells with a defective clamp loader complex of DNA polymerase III. Molecular Microbiology **50:** 193 – 204.

Vinnicombe, H.G., and Derrick, J.P., 1999. Dihydropteroate synthase from *Streptococcus pneumoniae*: characterization of substrate binding order and sulfonamide inhibition. Biochemical and Biophysical Research Communications **258:** 752 – 757.

Walsh, C.T., 2003. Other targets of antibacterial drugs, pp. 79 – 83. In: Antibiotics: actions, origins, resistance. ASM Press, Washington, DC.

Walsh, C.T., Fisher, S.L., Park, I.S., Prahalad, M., and Wu, Z., 1996. Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. Chemical Biology **3:** 21 – 28.

Warren, R.A.J., 1980. Modified bases in bacteriophage DNAs. Annual Review of Microbiology **34:** 137 – 58.

Wei, P., and Stewart, C.R., 1993. A cytotoxic early gene of *Bacillus subtilis* bacteriophage SPO1. Journal of Bacteriology **175:** 7887 – 7900.

West, D.K., Porter, D.C., Saxl, R.L., and Maley, F., 2004. A trojan horse approach for silencing thymidylate synthase. Biochemistry **43:** 9177 – 9184.

Witkin, E.M., and Roegner-Maniscalco, V., 1992. Overproduction of DnaE protein (α subunit of DNA polymerase III) restores viability in a conditionally inviable *Escherichia coli* strain deficient in DNA polymerase I. Journal of Bacteriology **174:** 4166 – 4168.

Yazawa, K., Mikami, Y., Maeda, A., Akao, M., Morisaki N., and Iwasaki, S., 1993. Inactivation of rifampin by *Nocardia brasiliensis*. Antimicrobial Agents and Chemotherapy **37:** 1313 – 1317.

Yu, Y.-T.N., and Snyder, L., 1994. Translational elongation factor Tu cleaved by a phage-exclusion system. Proceedings of the National Academy of Sciences of USA **91:** 802 – 806.

Zilling, W., Fujiki, H., Blum, W., Janekovic, D., Schweiger, M., Rahmsdorf, H.J., Ponta, H., and Hirsch-Kauffmann, M., 1975. *In vivo* and *in vitro* phosphorylation of DNA-dependent RNA polymerase of *Escherichia coli* by bacteriophage-T7-induced protein kinase. Proceedings of the National Academy of Sciences of USA **7:** 2506 – 2510.

References from the World Wide Web

Complete genome sequence of *Rhodococcus* sp. RHA1. Direct submission to the National Center for Biotechnology Information. Accession NC_008268.1.
McLeod, M.P., Warren, R.L., Araki, N., Hsiao, W.W.L., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A.L., Wang, D., Dosanjh, M., Petrescu, A., Morin, R.D., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Smailus, D., Siddiqui, A.S., Marra, M.A., Davies, J.E., Mohn, W.W., and Eltis, L.D., 2006.
http://www.ncbi.nlm.nih.gov/

The TIGR database of bacterial genomes:
http://www.tigr.org/tigr-scripts/CMR2/CMRGenomes.spl

The Universal Virus Database:
http://ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm