Molecular analysis of temperate phages in *Salmonella enterica* serovar Typhimurium DT 64 isolated in Australia

Princess Tlou Mmolawa

MSc (P.U. for C.H.E., South Africa)

Infectious Diseases Laboratories
Institute of Medical and Veterinary Science

and

Microbiology and Immunology, Department of Molecular Biosciences
University of Adelaide

Adelaide, South Australia.

A thesis submitted to the University of Adelaide for the degree of

Doctor of Philosophy

January 2001

Table of Contents

Abbreviations

The following abbreviations have been used throughout this thesis and follow the style recommended by the American Society for Microbiology for journals and books.

aa	amino acid
AFLP	Amplified–Fragment Length Polymorphism
Ap^r	ampicillin resistance
ATCC	American type culture collection
ATP	adenosine-5'-triphosphate
att	attachment
bp	base pair
BSA	bovine serum albumin
BYE	Buffered Yeast Extract Broth
C-terminal	Carboxyl-terminal of protein sequence
CA	columbia agar
CDC	Centers for Disease Control and Prevention
CFU	colony forming units
CsCl	caesium chloride gradient
dATP	2'-deoxy-adenosine-5'-triphosphate
dCTP	2'-deoxy-cytidine-5'-triphosphate
DT	definitive phage type
dGTP	2'-deoxy-guanosine-5'-triphosphate
dTTP	2'-deoxy-thymidine-5'-triphosphate
dUTP	2'-deoxy-uridine-5'-triphosphate
EDTA	ethylenediaminetetra acetic acid
FAFLP	fluorescent–amplified fragment length polymorphism
HBA	horse blood agar
HUS	haemolytic uraemic syndrome
IPTG	isopropyl-α-D-thiogalactopyranoside
kb	kilobase
Km^r	kanamycin resistance
kDa	kilodalton
LEE	locus of enterocyte effacement
LPS	lipopolysaccharide
MH	Meuller-Hinton agar
MP	maximum parsimony method
MW	molecular weight
NBT	nitroblue tetrazolium chloride
NCBI	National Center for Biotechnology Information
N-terminal	Amino-terminal of a protein sequence
nt	nucleotide
O/N	over night
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAI	pathogenicity island
PBS	phosphate buffered saline
PCR	Polymerase Chain Reaction
pfu	plaque forming units

PT	phage type
PFGE	Pulsed-Field Gel Electrophoresis
Phage	Bacteriophage
RAPD	Randomly Amplified Polymorphic DNA
RBS	ribosome binding site
REA	Restriction Endonuclease Analysis
RFLP	Restriction Fragment Length Polymorphism
sg	serogroup
SDS	sodium dodecyl sulphate
SPI	*Salmonella* pathogenicity island
SSC	sodium citrate-sodium chloride
TAE	Tris-Acetate-EDTA
TBS	Tris buffered saline
TE	Tris-EDTA
TEM	Transmission electron microscopy
TEMED	N,N,N,N,-tetramethylethylenediamine
Tris	Tris([hydroxymethyl]aminomethane)
UPGMA	unweighted pair group method with arithmetic averages
UV	ultra violet light
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indoyl-α-D-alactopyranoside
X-P	5-bromo-4-chloro-3-indoyl-phosphate toluidine salt

Declaration

I declare that the work described herein contains no material that has been previously submitted for the award of any degree or diploma in any university and to the best of my knowledge and belief contains no material previously published or written by another person, except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Date: 18/01/02 Signature:

Acknowledgements

I would like to acknowledge the support, and thank the following people.
Drs Michael Heuzenroeder, IMVS, Adelaide, and Dr Connor Thomas, University of Adelaide, for their supervision, encouragement and support. I especially wish to thank Dr Heuzenroeder for allowing me to work in his laboratory, for all his guidance during this PhD and for his patience and care to detail while reviewing and editing this thesis.

Dr Jan Lanser, IMVS, Adelaide, for her continual encouragement.

Prof. Horst Schmieger, University of Munich, Germany, for his interest in my work and most of all, for his patience, guidance and assistance since my visit to his laboratory, despite being thousands of kilometres from Australia. I am grateful to him for editing this thesis and assisting me with some of the figures used in this thesis.

Drs Nara Figueroa-Bossi and Lionello Bossi at Center de Génétique Moléculaire, Gif-sur-Yvette Cedex, France, for their fruitful discussions and suggestions during my visit to their laboratory. Prof. Andrew Kropinski for providing his "Online Analysis Tools" website which assisted me in the analysis of my sequence data and for his patience in answering my questions.

Dr Rodney Ratcliff, IMVS, Adelaide, for introducing me to sequencing methods and analyses and for always being there for me. I also wish to thank his wife, Julie and his children, for welcoming me and making Adelaide a home for my family.

Words cannot express my gratitude to staff and students of Microbiology Research and Development Laboratory, IDL, IMVS, Adelaide. Dr Michael Heuzenroeder, Dr Robyn Doyle, Ms Rina Wilmore, Mr Rolf Wise, Mr Alan Goodwin, Dr Ian Ross, Ms Wendy Hart, Ms Carly Tucker, Ms Thiru Vanniasinkam and Ms Ania Gubala, for the friendly working and social atmosphere they created in the laboratory. For their understanding, support, encouragement and friendship. In particular, Rina Wilmore, for working back to back in "our *Salmonella* corner" for the past three years, Wendy Hart, for her magic hand with figures and always being there for me. Dr Robyn Doyle, for introducing me to the mysteries of molecular biology, for partaking in editing my thesis chapters and for being generous with her time. Robyn will always remember the day we saw the two bands on a caesium gradient tube, which was a mystery to us and somehow introduced me to phage genetics.

Mr Arthur Mangos and the Molecular Pathology Sequencing Laboratory, for their patience whilst sequencing over 160 kb phage DNA.

Professor Hans–Wolfgang Ackermann, University of Laval, Canada, for confirming the TEM results, Ms Dianne Davos and the *Salmonella* Reference Laboratory staff for providing bacterial isolates and for assisting with phage typing. Profs. Bruce Stocker and Kenneth Sanderson at the *Salmonella* Genetic Stock Center in Calgary, Canada, for providing Q1 and ATCC 14028s strains, Professor Horst Schmieger for providing bacterial isolates used mostly in transduction and immunity experiments.

The Australian Agency for International Development (AusAID) for awarding me a scholarship and the Rural Industrial Research and Development Corporation (Chicken Meat), Australia, for their generous financial support.

The University of North West, Mafikeng, South Africa, for awarding me a four year study leave. Professor Dale Beighle, Head of Animal Health Department, as well as his wife, Ann, for their support, prayers, and encouragement throughout my study period.

My friend, Lyn Thompson, for her love, support, encouragement and understanding during my PhD study period. For taking care of my two sons for three weeks, whilst I attended a conference in Canada and visited several phage laboratories in Europe. I lack words to express my gratitude to her, she has been a shoulder to lean on.

My African friends in Australia. In particular, Dr Joseph Masika and his wife Dr Lillian Mwanri, Mr and Mrs Daniel and Patience Dzimano and Mr and Mrs Ben and Ada Daka, for their love, support and understanding, for caring for my children whilst busy with endless experiments and thesis writing. Their friendship is highly valued. My friends and family in South Africa, who were always there for me.

My family; my husband Patrick, for allowing me to undertake this study, for being most understanding, caring and supportive over the past four years whilst being separated from us, living alone in South Africa. For his endless love and endurance. My children, Onkabetse and Oratile, for somehow sharing these hard times with me. Many times when they needed motherly love and care, I was not there for them. I thank them for their endurance, love and support. Most of all, for being the most wonderful and understanding children, who over the years became independent and mature. I also thank my extended family members (sisters, brothers, nephews and nieces) in South Africa, who never ceased to support me.

Summary

Salmonella enterica serovar Typhimurium is the most common *Salmonella* serovar isolated from humans and accounts for 40–70% of human salmonellosis in Australia. The most common phage types include 9, 64 and 135. Phage typing is the traditional phenotypic method used for subtyping *Salmonella,* but it can be of limited use. Crude phage lysate induced from *S.* Typhimurium DT 64 (Definitive Phage Type) was capable of mediating phage type conversion where *S.* Typhimurium DT 9 was converted to DT 64 whereas DT 135 was converted to DT 16. Amplified–fragment length polymorphism revealed differences between the parental isolates and the DT convertants. Phage type conversion raises the question of the stability of the bacterial phage types in natural settings and the possibility of its occurrence during an outbreak scenario.

Induction of lysogenic phages from *S.* Typhimurium DT 64 with mitomycin C followed by caesium chloride gradient purification, resulted in separation of two populations of phage particles. DNA extracted from these particles and digested with *Sma*I, exhibited two distinct patterns of banding. Transmission Electron Microscopy showed that both phage particles belonged to the C1 morphotype of the family *Podoviridae*. However, sequence analysis of phage ST64B showed many putative tail genes in contrast to one tail gene found in ST64T. These CsCl gradient purified two phages were designated ST64T and ST64B (*Salmonella* Typhimurium DT 64 Top or Bottom) to reflect their respective banding density in the CsCl gradient. ST64T is capable of mediating both generalised transduction and bacteriophage type conversion.

The complete sequence of the double–stranded DNA genomes of both *S.* Typhimurium ST64T and ST64B bacteriophages were determined. Although the two temperate phages are carried by *S.* Typhimurium DT 64 isolate, they share very little

sequence similarity. The 40,679 bp genomic sequence of ST64T has an overall GC content of 47.5% and was reminiscent of lambdoid phages, in particular, P22. The ST64T putative proteins which exhibited a high degree of sequence similarity to P22 proteins (>90%) included the antigen conversion, integrase, excisionase, Abc1, Abc2, early antitermination (gp24), NinD, NinH, NinZ, packaging (gp3 and gp2), head (with the exception of gp26, gp7, gp20 and gp 16) and tail proteins. The putative immunity genes were highly related to those of *S.* Typhimurium phage L whereas the lysis genes were similar to those of *S.* Typhimurium PS3.

The 40,149 bp genomic sequence of ST64B has an overall GC content of 51.3% and was distinct from both P22 and ST64T. The genome architecture was similar to that of the lambdoid phages, in particular, coliphage λ. Whereas most tail genes showed sequence similarity with a non-lambdoid phage Mu tail genes, most head genes shared similarity with the lambdoid phages HK97 and HK022. ST64B contains putative genes that are related by sequence similarities to several other dsDNA phages infecting diverse bacterial hosts, including *Escherichia, Salmonella, Shigella, Pseudomonas, Haemophillus, Mesorhizobium, Rhodobacter, Agrobacterium, Caulobacter, Xylella, Pyrococcus, Streptomyces* and *Xenorhabdus.*

Both ST64T and ST64B like most phages, are genomic mosaics most probably as a result of random recombinational exchange of gene modules or cassettes. For example, the putative lysis gene cassette in ST64T is highly similar to that of bacteriophage, PS3 whilst the putative lysis genes found in ST64B were highly similar to putative proteins in prophage Fels–1. These data suggest that all dsDNA phages share common ancestry and exchange genes from a common genetic pool.

Publications

Tribe, G. I., H. Tsimogiannis, P. Mmolawa, D. Davos. 2001. An outbreak of *Salmonella* Typhimurium phage type 29 linked to a noodle restaurant in South Australia. Outbreak Report. Communicable Disease Intelligence. **25** (2).

Mmolawa, P. T., R. Wilmore, C. J. Thomas, M. W. Heuzenroeder. 2001. Temperate phages in *Salmonella enterica* serovar Typhimurium: Implications for epidemiology. Int J Med Microbiol. **291** (8). In Press.

Manuscripts in preparation

Mmolawa, P. T., C. J. Thomas and M. W. Heuzenroeder. 2002. Genomic sequence of *Salmonella enterica* serovar Typhimurium bacteriophage ST64T.

Mmolawa, P. T., C. J. Thomas and M. W. Heuzenroeder. 2002. Genetic characterisation of *Salmonella enterica* serovar Typhimurium bacteriophage ST64B.

Mmolawa, P. T., C. J. Thomas and M. W. Heuzenroeder. 2002. Comparison of the *Salmonella* Typhimurium phage type 64 phages ST64T and ST64B: Implications for evolution.

Chapter 1

Introduction

1.1 Salmonellae

The genus *Salmonella* was first described in 1880 by Eberth and cultured in 1884 by Gaffky (Burrows, 1959). Members of this genus belong to the family of Enterobacteriaceae (Darwin and Miller, 1999, Hook, 1990). They are Gram negative, non-spore forming rods, usually 0.7-1.5 x 2.0-5.0 μm in size (Cruickshank, 1975). Most *Salmonellae* are motile with peritrichous flagella, with the exception of *Salmonella enterica* serovar Gallinarum and *Salmonella enterica* serovar Pullorum (Hook, 1990). *Salmonellae* produce circular, smooth, raised convex colonies, which are generally 2-4 mm in diameter.

Salmonella bacteria are facultative anaerobes that grow on simple media. Faecal specimens with high concentration of other microorganisms are grown on selective or differential media such as deoxycholate agar or bismuth sulphate agar in order to inhibit non-pathogenic bacteria of the normal flora (Hook, 1990). Selenite or tetrathionate broths are used as enrichment media for faecal specimens (Cruickshank, 1975, Hook, 1990). *Salmonella* reduce nitrates to nitrites, usually produce gas from glucose and hydrogen sulphide (H_2S) from triple-sugar iron (TSI). They are urease negative, indole negative and utilise citrate as a sole carbon source (Cruickshank, 1975, Le Minor, 1984).

1.2 Nomenclature

The concept of species in the genus *Salmonella* has historically evolved through four overlapping phases, with emphasis on clinical evidence, antigenic specificities,

biochemical properties and deoxyribonucleic acid (DNA) relatedness (Le Minor and Popoff, 1987). In 1972, only three distinct species were recognised by the Enteric Reference Laboratories, at the Centres for Disease Control and Prevention (CDC), namely, *S. typhi*, *S. cholera-suis* and *S. enteritidis*, where *S. enteritidis* had many serotypes such as Typhimurium, Hadar, Agona and Enteritidis (Ewing, 1972). This nomenclature was not adopted by most National *Salmonella* reference centres in other countries leading to confusion in the United States (Farmer III, 1985). With the advent of DNA studies, two species are currently recognised in the genus *Salmonella.* Species one, *Salmonella enterica,* is composed of six subspecies: *S. enterica* subsp. *enterica* (subspecies I), *S. enterica* subsp. *salamae* (subspecies II), *S. enterica* subsp. *arizonae* (subspecies IIIa), *S. enterica* subsp. *diarizonae* (subspecies IIIb), *S.* enterica subsp. *houtenae* (subspecies IV), *S.* enterica subsp. *indica* (subspecies VI) (Le Minor, 1987, Le Minor, 1988, Popoff and Le Minor, 1997). The second species, *Salmonella bongori* was previously designated subspecies V (Popoff and Le Minor, 1997). The World Health Organisation (WHO) collaborative Centre for Reference and Research on *Salmonella* designates serovars (serotypes) belonging to *S. enterica* subspecies I with a name related to the geographical location where the serovar was first isolated (Popoff and Le Minor, 1997). Some workers including those at the National *Salmonella* Reference Laboratory at CDC essentially use this nomenclature but use the term, serotype instead of serovar. Currently, there are more than 2,400 serovars (Brenner, *et al.*, 2000). Most of these (including *S.* Typhimurium), belong to subspecies I and are found in O groups A, B, C_1, C_2, D, E_1, E_2 E_3 and E_4 (Hook, 1990).

The proposals for nomenclature changes in *Salmonella* have been previously summarised (Ewing, 1986, Farmer III, 1999, Le Minor and Popoff, 1987). The current system of nomenclature (Le Minor, 1987) described above will be used throughout this thesis ie. the previously designated *Salmonella typhimurium, Salmonella enteritidis* etc.,

will be referred to as *Salmonella enterica* serovar Typhimurium or *S.* Typhimurium, *Salmonella enterica* serovar Enteritidis or *S.* Enteritidis.

1.3 Host range of *Salmonella*

The primary reservoir of *Salmonella* is the intestinal tract of animals (Hook, 1990). The most common serovars which cause infections in humans and food animals belong to subspecies I (Le Minor, 1988). Approximately, 99% of *Salmonella* clinical isolates from humans belong to subspecies I (Le Minor and Popoff, 1988). In contrast, serovars belonging to subspecies II, IIIa, IIIb, IV, VI and *S. bongori* are common in cold blooded animals (Le Minor and Popoff, 1988, Bopp *et al.*, 1999). *Salmonella* strains have been isolated from nearly all animal species, including domestic animals (Bennet and Hook, 1959) and human illness is often linked to foods of animal origin (Bopp *et al.*, 1999).

Some *Salmonella* serovars are host specific. For example, *S.* Typhi, the causative agent of typhoid, only infects humans (Cohen and Tauxe, 1986, Hook, 1990). However, research has shown that mice and chimpanzees can be experimentally infected with *S.* Typhi, but this does not occur in natural settings (Rubin and Weistein, 1977). Broad-host range serovars such as *S.* Typhimurium and *S.* Enteritidis have been isolated from a variety of animal hosts including cows, pigs, cats, dogs, sheep, and poultry. Animals are usually the reservoir of infection for man (Bennet and Hook, 1959). For example, the human pathogens *S.* Heidelberg and *S.* Litchfield have primary avian and reptilian reservoirs, whereas chickens and other birds are the major reservoir of *S.* Gallinarum, *S.* Pullorum and *Salmonella enterica* subspecies II Sofia (Cohen and Tauxe, 1986). Although different serovars have different habitats, it should be emphasised that almost all *Salmonella* serovars can be isolated from both man and animals (Hook, 1990).

1.4 Epidemiology

Salmonellosis in man remains one of the three most important meat associated infections (Silliker, 1980), the other two are due to *Campylobacter* infections and *Staphylococcus* intoxication. *Salmonella* infection is still classified globally as a major health hazard by the WHO. Infection with nonTyphoid *Salmonella* serovars represents one of the leading causes of diarrhoea in developed countries (Mirold, *et al.*, 1999). In the 1980s, the reported annual incidence of human salmonellosis was 17.4 cases per 100,000 population in the United States of America, 23.4 in England and Wales, 40.8 in Canada, 77.1 in West Germany and 27 in Australia (D'Aoust, 1989).

The incidence of salmonellosis in the USA has steadily increased since World War II accounting for more than 11% of all instances of gastroenteritis (Cohen and Tauxe, 1986, Rubin, 1977). The most common serovar isolated was *S.* Typhimurium with an increase in the incidence of *S.* Enteritidis, *S.* Newport, *S.* Heidelberg and *S.* Infantis (Cohen and Tauxe, 1986, Todd, 1990). Furthermore, the incidence of *S.* Enteritidis has increased from 6% in 1972 to 25% in 1996, making *S.* Enteritidis the predominating serovar (Angulo, 1998) in the USA. In Canada, *S.* Typhimurium ranked first among human and non-human infections (Khakhria, 1991) with *S.* Heidelberg, *S.* Enteritidis, *S.* Hadar and *S.* Thompson being amongst the five most commonly isolated serovars (Lior and Khakhria, 1985, Rowe, 1984-1987).

The incidence of salmonellosis has also increased dramatically in European countries. In England, the number of cases increased six fold between 1982 and 1988 (Anonymous, 1990). In France, the incidence increased by seven fold between 1986 and 1990 (Hubert, *et al.*, 1991). In Germany, more than 100,000 cases of salmonellosis were recorded in 1990 with the yearly overall loss estimated to be DM600 million (Schulte, 1994). *S.* Typhimurium has predominated mostly in Sweden and Denmark from the 1970s (Fonden, *et al.*, 1976, Hakansson, *et al.*, 1976) and is still regarded as the most prevalent

serovar in many countries (D'Aoust, 1989). The increase of *S.* Typhimurium PT 12 in these countries was associated with an increase in the consumption of pork (Wegener, *et al.*, 1994). The majority of human infections with *S.* Enteritidis in Europe, are due to PT 4 (Binkin, 1993, Cowden, *et al.*, 1989b, Perales and Audicana, 1988, Stanley, *et al.*, 1992, Ward, *et al.*, 1987) whereas in Canada and USA, they are mostly caused by PT 8 (Hickman-Brenner, *et al.*, 1991, Lior, 1989, Poppe, *et al.*, 1991). *S.* Enteritidis outbreaks are mainly associated with consumption of eggs or food containing eggs and consumption of broiler chickens due to transovarian transmission of PT 4 by infected hens (Cowden, *et al.*, 1989a, Lister, 1988, Perales and Audicana, 1988).

The incidence of salmonellosis has also been documented in Southern Africa (Mokgatla, *et al.*, 1998, Simango and Mbewe, 2000, Staedler and Nesbit, 1990, Suzuki, *et al.*, 1995), Central Africa (Gendrel, 1994) and West Africa (Oboegbulem, 1990).

In Australia, the incidence of non-typhoidal salmonellosis is on the increase. As is the case for developed countries, *S.* Typhi is rarely isolated in Australia. A few cases of *S.* Typhi infection are usually acquired overseas or from contact with carriers from overseas (Australian *Salmonella* Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide, South Australia, Annual Reports, 1995-1999). In 1985 and 1986, *S.* Enteritidis ranked eighth as the most common serovar isolated from humans (Murray, 1987) whereas in the 1990s, it was the second most common isolate from humans. Since 1960, *S.* Typhimurium has been the most common serovar isolated from man in Australia (Australian *Salmonella* Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide, Annual Reports, 1960 – 1999).

Until 1970, there were no reports of the isolation of *S.* II serovar Sofia in Australia (Harrington, *et al.*, 1991). *S.* II Sofia was first isolated from chickens in 1980. Since then, *S.* II serovar Sofia remains the most common *Salmonella* isolate from chickens, followed by *S.* Typhimurium. (Murray, 1987; Australian *Salmonella* Reference Laboratory, Institute

of Medical and Veterinary Sciences, Adelaide, Annual Reports, 1960 – 1999). However, *S.* II Sofia rarely causes disease in humans. The high incidence of this serovar (*S.* II Sofia) in chickens and the very low incidence in humans, correlate with the known low virulence of subspecies II *Salmonella* serovars for humans.

Concurrent outbreaks of *S.* Typhimurium have occurred in South Australia on a number of occasions. Between December 1996 and January 1997, a number of cases of *S.* Typhimurium were reported. Random Amplified Polymorphic DNA (RAPD) analysis of the isolates indicated that there were three outbreaks involved. Phage typing confirmed that the isolates belonged to PTs 64, 135 and 44 (Andrews, *et al.*, 1997). PT 135 was associated with the consumption of bread rolls with meat filling, however, investigators failed to identify any association between four cases of PT 44. Recently, another outbreak of *S.* Typhimurium PT 29 was linked with the consumption of prepared food at a noodle restaurant in South Australia (Tribe, *et al.*, 2001).

Although many serovars of subspecies I have been identified world wide, the majority are rarely isolated from humans. For organism tracing during an outbreak of salmonellosis, subdivision of the most common serovars and other serovars of clinical importance is necessary (Threlfall and Frost, 1990). Epidemiological studies of *S.* Typhimurium are complicated by the fact that this organism has multiple indigenous sources. Additionally, contamination of foodstuffs and raw materials from overseas, utensils used for food preparation, or a food handler who is excreting this organism, may all contribute to the high incidence of *S.* Typhimurium (Callow, 1959). Because of the ubiquitous nature of the *Salmonella* organisms, the use of a powerful epidemiological tool is necessary to discriminate amongst the strains involved in an outbreak.

1.5 Criteria for evaluating typing methods

Typing systems are used to define specific characteristics of the object under study (van Belkum, *et al.*, 2001). Factors such as typeability, reproducibility, discriminating power, ease of interpretation and ease of performance are considered in the evaluation of typing methods (Maslow, *et al.*, 1993a). Evaluating a particular method according to these criteria requires information that is often unavailable. Evaluation of a method involves analysis of adequate numbers of sporadic and epidemic isolates and directly comparing the results with those of previously studied approaches (Maslow, *et al.*, 1993a). The ideal typing method should be inexpensive, rapid and technically simple. Typing systems may be categorised as phenotypic techniques (those that detect characteristics expressed by the microorganisms) or genotypic techniques (those involving DNA-based analyses of chromosomal or extra-chromosomal genetic elements) (Maslow, *et al.*, 1993a).

1.5.1 Genotypic typing methods

Genotypic methods such as plasmid profile analysis, pulsed-field gel electrophoresis (PFGE), ribotyping, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and genome sequencing, offer many advantages over traditional phenotypic methods. The discriminatory power of DNA–based methods is usually greater than that of phenotypic methods (Farber, 1996, van Belkum *et al.*, 2001). Plasmid profile analysis, ribotyping and RAPD are reviewed in (Farber, 1996, Hunter, 1990, Maslow, *et al.*, 1993a).

1.5.1.1 Restriction endonuclease analysis (REA) of chromosomal DNA

Chromosomal DNA restriction analysis involves isolation of DNA from the organism, digestion of DNA with a frequently cutting restriction enzyme and separation of fragments by electrophoresis (Farber, 1996). Fragments are then stained in ethidium bromide (a fluorescent dye which intercalates between the bases of nucleic acids) and visualised under ultra violet (UV) light (Farber, 1996). Restriction fragment length polymorphisms (RFLP) can then be identified. The major disadvantage of this method is the complexity of the banding patterns, since numerous bands may be obtained by using enzymes with relatively frequent restriction sites (4- or 6-base pair cutters) against whole chromosomal DNA (Maslow, *et al.*, 1993a, Threlfall and Frost, 1990). Furthermore, restriction fragments derived from plasmids may result in undue emphasis placed on isolates that differ only in their plasmid content. Different strains of the same species will have different REA patterns because of variation in their DNA sequences that result in RFLPs (Maslow, *et al.*, 1993a). Conversely, the main advantage of chromosomal (genomic) fingerprinting is that it can be used to differentiate strains that do not carry plasmids. Using genomic fingerprinting, small differences could be observed among strains of *S.* Typhi from Peru and Chile after digestion with *Pst*I restriction enzyme (Maher, *et al.*, 1986).

1.5.1.2 Southern blot analysis of chromosomal DNA

This technique involves the transfer of restriction fragments, after separation by agarose gel electrophoresis, onto a nitrocellulose or nylon membrane. Labelled DNA probes are then used for the detection of restriction fragment(s) containing sequences homologous to the probe. Restriction fragment length polymorphism (RFLP) reflect variations in both the number of loci homologous to the probe and the location of

restriction sites flanking those loci (Southern, 1975). The results are highly reproducible. For strain differentiation, the most effective probes are those that detect multiple bands simultaneously (Maslow, *et al.*, 1993a).

Tompkins and coworkers (1986) used ^{32}P-labelled *S.* Enteritidis genomic DNA fragments as probes to identify RFLPs in strains of *S.* Typhimurium, *S.* Dublin and *S.* Enteritidis. In their study, six fingerprints were identified in both *S.* Typhimurium and *S.* Dublin. In contrast, the majority of strains of *S.* Enteritidis had a homogeneous fingerprint, indicating little evolutionary divergence in this serovar.

1.5.1.3 Pulsed-field gel electrophoresis (PFGE)

Pulsed-field gel electrophoresis (PFGE) is regarded as a very discriminating and reproducible typing method that is becoming more common (Farber, 1996, Maslow, *et al.*, 1993a). PFGE is a form of macro–restriction analysis, which relies on the use of restriction enzymes, that recognise 6– to 8–base sequences and only infrequently cut genomic DNA. This allows a small number of well resolved, large fragments (~10–800 kb) to be viewed (Arbeit, *et al.*, 1990, Farber, 1996, Maslow, *et al.*, 1993a, Tenover, *et al.*, 1995). PFGE involves embedding bacteria in agarose plugs, lysing the bacteria *in situ* and digesting chromosomal DNA with restriction enzymes (Finney, 1993, Maslow, *et al.*, 1993b). Agarose plugs containing the digested chromosomal DNA are inserted into the wells of an agarose gel which is electrophoresed horizontally in a tank connected to an apparatus that switches the direction of current according to a determined pattern (Tenover, *et al.*, 1995). The restriction fragments are then resolved into a pattern of discrete bands that can then be viewed after staining with ethidium bromide. The relatedness of isolates is assessed by comparison of the DNA restriction patterns using software such as GelCompar II (Garaizar, *et al.*, 2000, Gerner-Smidt, *et al.*, 1998, Liebana, *et al.*, 2001). PFGE has

been successfully applied to a variety of Gram negative and Gram positive bacteria for epidemiological investigations. The method is laborious and takes almost six days from isolation of a pure colony until results are available. However, a rapid method that yields results in less than three days has been reported (Matushek, *et al.*, 1996).

PFGE has been employed in the typing of *Salmonella*. Thong *et al.* (1995) evaluated 20 different restriction enzymes for their usefulness in PFGE. *Xba*I, *Spe*I and *Avr*II were found to be the most useful restriction enzymes, as reported previously for *S.* Typhi (Thong, *et al.*, 1994b). Using these restriction enzymes, Thong *et al.* (1995) showed an identical PFGE pattern in all the outbreak-related isolates indicating that a single clone may have been involved in the outbreak (Thong, *et al.*, 1995b). Analysis of a different group of isolates (poultry and human) using the same restriction enzymes, demonstrated a common pattern in both isolates, thus indicating that infection in humans may have originated from poultry products (Thong, *et al.*, 1995b). A different study indicated that 21 out of 33 *S.* Enteritidis PTs formed one cluster by PFGE (Olsen, *et al.*, 1994). In this study, more than one PFGE pattern was observed among strains belonging to PTs 1, 6, 7 and 14b. Thus PFGE, when performed with a set of suitable restriction enzymes, proved to be the method of choice in subdividing isolates of the same PT in *S.* Enteritidis (Weide-Botjes, *et al.*, 1998). However, studies on *S.* Typhimurium DT 009 showed one predominant PFGE pattern when *Xba*I, *Spe*I and *Bln*I restriction enzymes were used individually to digest chromosomal DNA. However, when isolates of the same *Xba*I pattern were further subdivided using *Spe*I and *Bln*I, three different *Spe*I and four different *Bln*I patterns were displayed (Frech, *et al.*, 1998). Although better discrimination of strains is achieved by increasing the number of restriction enzymes for DNA digestion, the limitation of running electrophoresis in separate gels for each enzyme makes this approach impractical for routine use (Liebana, *et al.*, 2001). *Xba*I seems to be the most

discriminative enzyme for PFGE of *Salmonella* serovars (Messenger, *et al.*, 1999, Ridley, *et al.*, 1998, Thong, *et al.*, 1994a, Thong, *et al.*, 1998, Weide-Botjes, *et al.*, 1998).

PFGE was also used to differentiate outbreak and non–outbreak strains of *S.* Bovismorbificans and *S.* Stanley in epidemics in Finland (Puohiniemi, *et al.*, 1997). In each separate outbreak, the PFGE pattern of *Xba*I digested outbreak strains (40 strains of *S.* Bovismorbificans and 28 of *S.* Stanley) were indistinguishable from each other, but were clearly different from those of the non–outbreak strains (26 of *S.* Bovismorbificans and 40 of *S.* Stanley) (Puohiniemi, *et al.*, 1997). In this study, *S.* Bovismorbificans infection in humans was traced to the consumption of a sprout salad. Furthermore, PFGE supported findings that *S.* Stanley outbreak isolates originated from contaminated alfalfa seeds (Puohiniemi, *et al.*, 1997). In another study, Liebana and coworkers used PFGE to subdivide 54 PT 4 isolates into three types designated X1, X2 and X4 (Liebana, *et al.*, 2001). Most isolates (89%) showed pattern X1 whereas 9.3% showed pattern X2. Pattern X4 was only found in a single PT 4 isolate, indicating that PFGE could not clearly distinguish amongst the PT 4 isolates. Furthermore, PFGE could not distinguish amongst PTs 1, 4, 6, 7, 21, 24, 29, 35 and 36 since an identical PFGE pattern was observed in all these PTs (Liebana, *et al.*, 2001).

Although many authors recommend PFGE as the most sensitive and discriminatory molecular method for subdividing isolates within a serovar, other studies have shown that it has limitations (Letellier, *et al.*, 1999, Liebana, *et al.*, 2001, Thong, *et al.*, 1995a). However, when combined with other typing methods, excellent discrimination can be achieved.

1.5.1.4 Amplified– fragment length polymorphism (AFLP)

AFLP is one of the newest and promising typing methods (Blears, *et al.*, 2000, Janssen, *et al.*, 1996, Vos, *et al.*, 1995). It is based on the selective amplification of genomic restriction fragments (Huys, *et al.*, 1996, Janssen, *et al.*, 1996, Vos, *et al.*, 1995). Adapters are ligated to genomic restriction fragments, followed by a PCR–based amplification using adapter specific primers (Vaneechoutte, 1996). Oligonucleotide adapters are designed in such a way that the initial restriction site is not restored after ligation. This allows simultaneous restriction and ligation, while re-ligated fragments are cleaved again (Savelkoul, *et al.*, 1999). Adapter–specific primers with an extension of one to three nucleotides which run into the unknown chromosomal restriction fragments, are used for PCR amplification under high stringent conditions (Savelkoul, *et al.*, 1999). The ultimate result obtained usually has a highly informative pattern of 40 to 200 bands (Savelkoul, *et al.*, 1999). For small bacterial and fungal DNA genomes, it has been shown that it is sufficient to use a single PCR amplification with one and two selective nucleotides, respectively, on both primers (Duim, *et al.*, 1997, Janssen, *et al.*, 1996, Janssen and Dijkshoorn, 1996, Mueller, *et al.*, 1996). A study on *Salmonella* using the primer combination M00 and E11 (two selective nucleotides) resulted in well defined fingerprints containing 50 scorable bands (Aarts, *et al.*, 1998).

An advantage of the AFLP technique is that it can be used to generate fingerprints from DNA of any origin and complexity without prior sequence knowledge (Huys, *et al.*, 1996, Janssen, *et al.*, 1996, Lin, *et al.*, 1996, Vos, *et al.*, 1995). AFLP seems to be a powerful fingerprinting technique which can be used in identification, epidemiology and taxonomy (Folkertsma, *et al.*, 1996, Huys, *et al.*, 1996, Janssen, *et al.*, 1996, Lin, *et al.*, 1996). Aarts *et al.* (1998) were able to classify strains of *Salmonella* into serovars and strain–specific identification of *Salmonella* serovars was also possible. Lindstedt and

coworkers compared PFGE with fluorescent–AFLP (FAFLP) of *Salmonella* enterica serovars (Lindstedt, *et al.*, 2000). These researchers separated the dye–labelled fragments by capillary electrophoresis which is one of the most suitable method for fragment analysis and is often used to type small alterations at microsatellite loci (Kleparnik, *et al.*, 1998, Le, *et al.*, 1998). The addition of an internal standard as used by Lindstedt, *et al.* (2000), enabled the sizing of FAFLP (with a resolution of ±1 bp) and direct comparison of fingerprints. Using *Eco*RI and *Mse*I restriction enzymes, both the FAFLP and PFGE methods generated about the same number of distinct fingerprint profiles. Although these two methods have the same discriminatory power, PFGE is a very laborious method. By contrast, the use of fluorescent dyes totally removes the use of radioactive isotopes in AFLP and there is no need for a gel fixation step and development of a film. Furthermore, there is only one PCR amplification step and a 5-hour restriction/ligation reaction step. In addition, FAFLP is run on automated sequencing apparatus that gives a pattern with less variability (as a result of gel conditions) than PFGE. All these steps make FAFLP a rapid fingerprinting method (Lindstedt, *et al.*, 2000), that is much simpler than PFGE and generate results with much less gel to gel variation.

1.5.1.5 Gene sequencing

It is technically possible to compare multiple isolates by sequencing the same locus from each isolate. Introduction of improved sequencing methods such as PCR sequencing with dye terminator chemistry, has made sequencing the method of choice for some organisms which are difficult to identify by traditional methods (Amann, *et al.*, 1988, Ludwig and Schleifer, 1994, Sallen, *et al.*, 1996). Many genotypic schemes utilise variation in the rRNA genes, by targeting conserved regions with universal primers in order to amplify regions of sequence variability (Collins, 1986, Ludwig and Schleifer,

1994, Sallen, *et al.*, 1996). PCR based sequence analysis of bacterial rRNA has successfully been used to detect bacteria identified as an unusual Gram positive actinomycete in the tissues of patients with Whipple's disease (Relman, 1992). However, for epidemiological purposes, identification of the most suitable loci for sequence analysis remains a critical issue (Maslow, *et al.*, 1993a).

In conclusion, integration of taxonomic, evolutionary and phylogenetic studies, population genetics and epidemiology can essentially be achieved by means of full genome sequencing (van Belkum, *et al.*, 2001). For example, once full genome sequences are available for multiple isolates of a single bacterial species, all genetic variables can be catalogued and the nature of mutations identified may elucidate the relatedness between these isolates. Furthermore, if multiple isolates from multiple species have been fully sequenced, their data may define the relatedness, or lack thereof, between microbial species and genera. At the time of writing this thesis (2001), more than 30 bacteriophage genome sequences from different bacterial hosts were available. These genome sequences (although inclusion of additional full bacteriophage genome sequences is still required) illustrate the type of taxonomic, evolutionary and population genetics information that can be obtained from detailed experimental genetic identification (van Belkum, *et al.*, 2001). It has recently been suggested that even though it is more difficult to sequence whole prokaryotic genomes compared with phage genomes, it is necessary to have full genome sequences for meaningful results in genetic typing (van Belkum, *et al.*, 2001).

1.5.2 Phenotypic typing methods

In the past, the relatedness of bacterial isolates was determined by testing for one or several phenotypic markers, using methods such as serotyping, phage typing, biotyping, antibiotic sensitivity testing and bacteriocin typing (Farber, 1996). Typing of *Salmonella*

is done by serotype analysis of O (lipopolysaccharide) and H (flagella) antigens according to the scheme developed by Kauffmann and White (Kauffmann, 1954). The Kauffmann and White scheme is based on the immunologically distinct variations in somatic O antigen of the cell wall and flagellar H antigens which provide each *Salmonella* serovar with its own unique antigenic combination (Kauffmann, 1954). *Salmonella* strains identified by this scheme have traditionally been given their own species designation (with the new nomenclature these are now called serovars) for example, *S.* Typhimurium (Farmer III, 1984). A limitation of this system is that a small number of *S. enterica* serovars is responsible for a majority of infections (Harvey, *et al.*, 1993). For organism tracing during an outbreak of salmonellosis, the need for a method to subdivide isolates within a serovar is indispensable. Traditionally, this has been achieved by phage typing (Callow, 1959; Hickman-Brenner *et al.*, 1991, Rodrigue *et al.*, 1992). Other phenotypic methods (mentioned above) normally used to subdivide isolates within a serovar, are reviewed in (Maslow, *et al.*, 1993a, Threlfall and Frost, 1990). Table 1.1 shows the general characteristics of different bacterial typing methods.

1.5.2.1 Phage typing

The underlying principle of phage typing is the host specificity of bacteriophages (Threlfall and Frost, 1990). Studies have shown that by varying the host strain on which the phage was propagated, different adaptations on the host strain occur which form the basis of phage typing (Anderson, 1959). The first phage typing scheme was developed in 1938 for the differentiation of *S.* Typhi isolates (cited in Callow, 1959). *S.* Typhi isolates were grouped on the basis of their susceptibility to lysis by Vi phages (cited in Callow, 1959). Eleven phage types were initially identified and by 1986, about 106 Vi PTs were internationally recognised (Edelman and Levine, 1986). Nevertheless, this phage typing

scheme of *S.* Typhi depends on adaptations of a single phage, namely Vi-phage II (Threlfall and Frost, 1990). By contrast, phage typing schemes for other *Salmonella* serovars are based on patterns of lysis produced by distinct phages isolated from a variety of sources (Threlfall and Frost, 1990). Phage typing has been used to subdivide isolates within serovars Typhi (Anderson and Williams, 1956), Typhimurium (Callow, 1959, Anderson, 1977), Enteritidis (Ward, *et al.*, 1987), Virchow (Chambers, *et al.*, 1987), Hadar (De Sa, *et al.*, 1980) and Heidelberg (Harvey, *et al.*, 1993). Additional phage typing schemes are being developed for other *Salmonella*, for example, the phage typing scheme of *S. enterica* serovar Bovismorbificans.

Phage typing is based on the assumption that strains which are epidemiologically related will exhibit the same PT while unrelated strains will have different PTs. For this to be valid, the phage typing system must have a high discriminatory power in the bacterial population where the system is applied and the PTs must demonstrate high stability within the population (Hunter, 1990; Olsen *et al.*, 1993). Although phage typing is essential for the subdivision of *Salmonella* serovars, the method can prove inadequate for serovars in which a small number of phage types predominate. For example, from 1981 to 1986, most isolates of *S.* Enteritidis in Britain and Wales belonged to PTs 4 and 8 (Ward, *et al.*, 1987). Thus methods which could differentiate within these two PTs are epidemiologically valuable.

Phage typing can also be subjective in nature and relies on the experience of the operator and uniformity between laboratories (Harvey, *et al.*, 1993). Although phage typing is still used widely for epidemiological purposes, several studies, which will be detailed below, have shown that some PTs are able to spontaneously convert from one type to another. Thus the instability of PTs could potentially have serious implications for epidemiology and organism tracing.

1.5.2.1.1 Bacteriophage type (PT) conversion

For a variety of reasons, *S.* Enteritidis demonstrates the phenomenon of PT conversion. Chart, *et al.* (1989) have shown that the conversion of *S.* Enteritidis PT 4 to PT 7 was due to modification of the LPS. Powell, *et al.* (1995) showed that strains of *S.* Enteritidis PT 4 could independently convert to PT 9a as a result of a mutation affecting phage receptors. Other workers showed that PTs 1, 4 and 6 could be converted to PT 7 due to modification in LPS (Baggesen, *et al.*, 1997). Additionally, (Rankin and Platt, 1995) showed the conversion of PT 6a to PTs 4 and 7. Hickman-Brenner and coworkers also demonstrated PT conversion of PT 14b during subculture, where two colonies remained PT 14b whereas two were converted to PT 8 (Hickman-Brenner, *et al.*, 1991). Consequently, (Baggesen, *et al.*, 1997) proposed that the use of PT 7 as an epidemiological marker, should be treated with caution since genetically unrelated strains of *S.* Enteritidis PT 7 could occur within the same PTs.

Conversion of *S.* Enteritidis PT 4 to PT 24 due to the acquisition of an incompatibility (Inc) group N plasmid has also been demonstrated (Frost, *et al.*, 1989). Furthermore, loss of this plasmid results in reversion of PT 24 to PT 4. Threlfall, *et al.* (1993) also found some interrelationships between *S.* Enteritidis PTs 4, 7, 7a, 8, 13, 13a, 23 and 30 that were associated with the loss or acquisition of an Inc N plasmid. More recently, PT conversion was demonstrated when a conjugative plasmid (pOG670), mediating resistance to ampicillin and kanamycin was transferred to 10 isolates of *S.* Enteritidis. Acquisition of this plasmid resulted in PT conversion (PT 1 to PT 21, PT 2 to PT 22, PT 4 to PT 6, PT 8 to PT 13a, PT 10 to PT 13 and PT 11 to PT 34) (Brown, *et al.*, 1999). PT conversion has also been reported in *S.* Typhimurium due to acquisition or loss of plasmid DNA (Platt, *et al.*, 1987).

Conversion of *Salmonella enterica* serovar Heidelberg was shown to have occurred during an outbreak of *S.* Heidelberg PT 1 (M. W. Heuzenroeder, R. Dalcin, D. Davos and

C. Murray, Abstr. 4th Asia Pacif. Poult. Health Conf., abstr. 144, 1998). Additionally, lysogenisation of *S.* Heidelberg PTs 1 and 3 isolates with *S.* Heidelberg lysogenic phage 4 (Hlp–4), was shown to mediate conversion of PTs 1 and 3 to PT 4 and PT 5 to a non-typable strain (Harvey, *et al.*, 1993).

Phage typing is used in many countries as a routine typing method. Some studies (Hickman-Brenner, *et al.*, 1991) have indicated that phage typing may not necessarily be discriminative enough for epidemiological purposes. Almost half of the 573 cultures of *S.* Enteritidis tested in the USA from a variety of sources, belonged to PT 8 (Hickman-Brenner, *et al.*, 1991). The authors realised that further differentiation of PT 8 was necessary for epidemiological investigation (Stubbs, *et al.*, 1994).

It is possible that due to phage typing, new phages are formed by recombination between infecting phages and residing phages (Schmieger, 1999). This was demonstrated by propagation of the Anderson typing phages on particular hosts of *S.* Typhimurium (also used as reference strains) where the phages gained new plating properties. The same phages with new plating properties when used on different *S.* Typhimurium strains for phage typing, may recombine with residing phages to give a particular phage type (Schmieger, 1999). This observation of formation of new phages is in accordance with the theory of modular evolution of bacteriophages which will be discussed in more detail under section 1.11.

In conclusion, phage typing relies on three factors: (i) host controlled–modification, (ii) adsorption properties of the infecting phage (depending on the base plate) and (iii) susceptibility of the infecting phage to repressors and to various superinfection exclusion systems controlled by residing phages (to be discussed more under section 1.9.1.2) (Schmieger, 1999). Consequently, it may be difficult to re–propagate the set of plating phages in exactly the same way as in the past because propagation of a particular phage on a host carrying a related prophage leads to different recombinants which may lead to

different plating properties (Schmieger, 1999). This implies that the original collection of typing phages may be a historically unique combination of phages and strains and hence repeating the same procedure of Anderson, *et al.* (1977) on different hosts may lead to test phages with completely different plating properties. In short, the unique Anderson phage stocks may never be reproduced, should the phages or the original propagating strains be lost (Schmieger, 1999). It is therefore important to study the newer molecular techniques, which may not only be more reliable, but also more reproducible over time.

Finally, the choice of a typing method depends on professional and personal criteria. Factors such as speed, discriminatory power, reproducibility, cost–effectiveness and technical feasibility should be considered when selecting a typing method (van Belkum, 1997). The method of choice should appeal to users and customers. Furthermore, for epidemiological studies and organism tracing, molecular typing methods such as AFLP, PFGE, ribotyping and plasmid profiling, should be used together with the more established methods such as phage typing.

Having discussed the different typing methods including, bacteriophage typing and its limitations, discussions in this thesis will now focus on the bacteriophage itself, in particular on the well–studied *Salmonella* bacteriophage, P22. It is not the purpose of this literature review to cover all aspects of bacteriophage biology, but to discuss a few points briefly in order to introduce the reader to the result chapters.

1.6 The history of bacteriophage

The bacteriophage was discovered by F. W. Twort in 1915 (cited in Ackermann and DuBow, 1987a). Independently, d'Hérelle in 1915 discovered an invisible agent, which was an obligate parasite of living bacteria. This agent was designated "bacteriophage" (cited in Bradley, 1967). d'Hérelle laid the foundation for phage research

and also described phages of typhoid fever, dysentery, plague, cholera bacteria and *Staphylococci* (Ackermann and DuBow, 1987a). By 1935 many enterobacterial and Gram positive phages were known (Ackermann and DuBow, 1987a). The determination of some phage sizes was achieved in 1932 by filtration through membranes of graded pore size (Ackermann and DuBow, 1987a). Shortly thereafter, subdivision of enterobacterial phages into serogroups was achieved and correlated with particle size, host range and resistance tests. Schlesinger, (1933), demonstrated that phage could be purified by centrifugation and later demonstrated that phages consisted of equal amount of DNA and protein (cited in Ackermann and DuBow, 1987a). The birth of molecular genetics and the visualisation of phages under the electron microscope was brought about by the one–step–growth experiment of Ellis and Delbrück (cited in Ackermann and DuBow, 1987a). The introduction of phage typing of *S.* Typhi (cited in Callow, 1959) and phage therapy of bacterial diseases, in the pre-antibiotic era, were amongst the first major developments. By 1940, at least 560 papers were published on phage therapy (cited in Ackermann and DuBow, 1987a). In the period 1940 to 1965, a sharp rise occurred in the number of publications on epidemiology, reflecting an interest in phage typing, lysogeny and genetics. Concomitant increases in the numbers of morphological and physiological papers and a decline in phage therapy publications occurred as antibiotic use increased (Raettig, 1967).

In their study, Lwoff and Gutmann (1950) demonstrated the ability of lysogenised cells to produce phages over 19 isolations of single cells (Lwoff, 1953). By 1953, the concept of lysogeny was well defined. Furthermore, phage could be induced by ultraviolet (UV) light (Lwoff, 1953). In 1952, Hershey and Chase demonstrated that phage DNA and protein had different roles. Their research work with phage T2, indicated that ^{32}P labelled T2 DNA could enter the bacterial cell and initiate phage synthesis whereas its ^{35}S labelled protein moiety remained outside, fixed to the bacterial cell. This research made it possible

for the structure of DNA to be determined (Watson and Crick, 1953). Additionally, the presence of 5-hydroxymethylcytocine was discovered in the DNA of T-even phages (cited in Ackermann and DuBow, 1987a). Another fundamental discovery was the finding that the genetic material could be a single-stranded DNA (ssDNA) as in bacteriophage ϕX174 or ssRNA as in bacteriophage QB (cited in Ackermann and DuBow, 1987a). Finally, whole phage genomes were sequenced. RNA phage MS2 was the first to be sequenced (Fiers, *et al.*, 1976), followed by many other phages including the more complex members of the group such as phage λ (Sanger, 1982) and more recently, phage P22 (Vander Byl and Kropinski, 2000).

The rapid spread and improvement of electron microscopes, as well as the introduction of negative staining by Brenner and Horne (1959), generated advances in phage morphology and ultrastructure (Ackermann and DuBow, 1987a). Since 1959, approximately 5000 bacterial viruses excluding non–viable and non–propagated viruses without known hosts, have been examined by electron microscopy (Ackermann, H. W. 2000. Frequency of phage descriptions in the year 2000. Millennial Phage Biology meeting, Montreal, Canada). Approximately, 96% of these bacteriophages are tailed, 4% icosahedral, filamentous or pleormorphic. Tailed phages comprise the largest of all virus groups (Ackermann, 1998). In this thesis, discussions will mainly focus on tailed phages, although reference will be made to other phages where appropriate.

1.7 Morphological classification of tailed phages

(Murphy, *et al.*, 1995) classified tailed phages into three families, according to the basic morphological types of Bradley (1967), namely, *Myoviridae* with long contractile tails (A), *Siphoviridae* with long non-contractile tails (B) and *Podoviridae* with short tails (C). These types may be further divided according to head shape into isometric heads

(subtype 1), moderately elongated heads (subtype 2), very long heads with axial ratios of 2.4 to 3.4 (subtype 3) (Ackermann, 1969) (see Figures 1.1 and 1.2). *Myoviridae*, *Siphoviridae* and *Podoviridae* families fall under the order, *Caudovirales*, based on similarities in tailed virus morphology, replication and assembly (Maniloff and Ackermann, 1998). *Siphoviridae* are the most common group at 61%, followed by *Myoviridae* at 24.5% and *Podoviridae* at 13.9% (Krieg and Holt, 1984). The "T" phages from *E. coli* were the first to be well–studied (Demerec and Fanu, 1945). The family *Myoviridae* encompasses T2, T4 and T6 better known as T-even phages. T1 and T5 belong to the *Siphoviridae* family whereas T3 and T7 belong to the *Podoviridae* family. The head capsomers of phage λ were the first to be visualised (Eiserling and Boy de la Tour, 1965). Today, many lambdoid and non-lambdoid phages carried by *E. coli,* as well as other bacteria have been discovered.

The rare A3 and C3 morphotypes are characterised by long narrow heads and tails that are either contractile or short (Ackermann and Eisenstark, 1974). In 1974, both A3 and C3-type phages were isolated from *S.* Newport (Moazamie, *et al.*, 1979). In addition, several C3-type phages were isolated from *Enterobacter cloacae*, *Levinea malonatica* (Markel and Eklund, 1974) and *Proteus mirabilis* (W. C. Schmidt, 1975 PhD thesis, Wayne State University, Detroit, MI (cited in Ackermann and DuBow, 1987a).

The morphology of phages comprising a general *Salmonella* typing set has been studied (Ackermann and Gersham, 1992). The phages belonged to the *Myoviridae*, *Siphoviridae* and *Podoviridae* families and to A1, A3, B1, B2 and C1 morphotypes (Ackermann and Eisenstark, 1974). The tail width in contractile tails was 16-18 nm and 7-11 nm in non-contractile tails. The study identified several staining artifacts observed during electron microscope analysis of phage preparation (Ackermann and DuBow, 1987a). These artifacts included among others, rounded phage heads after staining with phosphotungstic acid and positively stained (deep black and shrunken) uranyl acetate

stained heads (Ackermann and Gersham, 1992). Heads were found to be isometric or elongated. Isometric heads are described as icosahedra because of capsids with hexagonal and pentagonal outlines (Ackermann and Gersham, 1992). Although most phages in this study could be related by their morphology to known enterobacterial phages, two new phage species were discovered. Both phages had an equivalent among the *Rhizobium* phages but had not been observed in the past in enterobacteria. The new species was designated Sen15 after the propagating strain of this phage, *S.* Enteritidis whereas the other species was designated San21 after its usual host, *S.* Anatum (Ackermann and Gersham, 1992).

The well–studied *S.* Typhimurium P22, is a podovirus of the C1 morphotype. Many studies have suggested that P22, despite its morphology, is a member of the lambdoid family. Although phage λ is a siphovirus, about 23 known P22 genes are thought to have homologues in λ. Furthermore, viable λ/P22 hybrids have been formed *in vivo* (Miller, 1992). Recently, based on the morphological difference between P22 and λ, the former has been proposed as a type virus for a new genus which includes phages L (Bezdek and Amati, 1967), ES18 (Kuo and Stocker, 1970), LP7 (Kitamura and Mise, 1970), and *AP*SE-1 (van der Wilk, *et al.*, 1999).

1.8 Induction of bacteriophage

Large numbers of temperate bacteriophage can be released from logarithmic cultures by addition or application of an inducing agent (Bradley, 1967). Inducing agents include hydrogen peroxide (DeMarini and Lawrence, 1992, Figueroa-Bossi and Bossi, 1999, Gille, *et al.*, 1996, Kalinin and Kuznetsova, 1995), UV light (Barnhart, *et al.*, 1976, Berenstein, 1986, Harvey, *et al.*, 1993, Seyedirashti, *et al.*, 1991), mitomycin C (Gemski, *et al.*, 1978, Head, *et al.*, 1988, Walker, 1987, Yee, *et al.*, 1993) and streptonigrin (Muschel

and Schmoker, 1966). Other chemical inducing agents are known, but are often specific for a particular bacterium e.g. chlorophenol induces λ in *E. coli* but not *Salmonella* (DeMarini, *et al.*, 1990). Mitomycin C and UV light are probably the best known of all inducing agents. Mitomycin C damages DNA by cross–linking complementary strands (Lown, *et al.*, 1976) and is likely to activate the SOS response which may result in release of operons from LexA–mediated repression (Walker, 1987).

Induction of bacteriophage has been correlated with the release of toxins (Farkas-Himsley, *et al.*, 1977, Gemski, *et al.*, 1978, Isaacson and Moon, 1975, Takeda, *et al.*, 1979, Yee, *et al.*, 1993). Head, *et al.* (1988) demonstrated that induction of a bacteriophage carrying toxin genes resulted in increased copy number of the toxin genes and a 100– to 200–fold increase in toxin synthesis. It has been shown that the *stx* genes of Enterohaemorhagic *E. coli* are co–expressed with genes of the bacteriophage and an increase in toxin production during bacteriophage induction has been observed (Neely and Friedman, 1998). For this reason, certain antibiotics e.g. quinolones, trimethoprim and furazolodine cannot be used for treatment of haemolytic uraemic syndrome (HUS) as they may increase *stx* gene production (Kimmitt, *et al.*, 2000). However, induction of toxin synthesis without induction of phages has been reported for enterotoxigenic *E. coli* (Farkas-Himsley, *et al.*, 1977) and for *Shigella* isolates (Takeda, *et al.*, 1979). Gemski *et al.*, (1978) suggested that bacteriophage induction by mitomycin C for cellular release may have some applications to the characterisation of toxins.

1.9 Lytic growth *versus* lysogeny

There are two categories of phages on the basis of interaction with their hosts, namely, virulent or temperate. Virulent phages such as the T–even phages, lyse every cell they infect and produce clear plaques (lytic cycle), whereas temperate phages, following

infection, may enter the lytic cycle, or become integrated into the genome of the host cell. The prophages are then replicated as part of the bacterial chromosome (Bradley, 1967, Hayes, 1964a). Temperate phages like λ and P2 can enter either a lytic or lysogenic life cycle upon infection of their host strain, *E. coli*. The phage adsorbs to the cell surface, injects its DNA into the host where it circularises. If the lytic pathway is followed, early transcription begins, leading to middle gene expression. The phage replicative genes are expressed and the phage genome is replicated. The expression of the late genes permits synthesis of phage morphogenic proteins, proteins required for phage assembly and for host cell lysis (Bertani and Bertani, 1971). Occasionally the phage genome becomes integrated into the host cell DNA and is replicated each time the cell divides (Figure 1.3). The prophage state is stable until de-repression occurs. Lysogenic bacteria can potentially produce or release phage as a stable, heritable trait and may be immune to lytic infection by the same or closely related phages (Bertani, 1953). In addition, double or triple lysogeny has been described. For example, one strain of *Staphylococcus* has been reported to carry as many as five different phages (Rountree, 1949).

Most relatives of λ have the same gene order as λ and grow on *E. coli*, although a few grow on *S.* Typhimurium (Campbell, 1994). Like λ, *S.* Typhimurium prophage P22 has been studied extensively. The discovery of generalised transduction by Zinder and Lederberg has made P22 the virus of choice for *Salmonella* genetics (Zinder and Lederberg, 1952). P22 is a relative of phage λ, it encodes an antirepressor and has proved to be an accessible object for studies on morphogenesis (Poteete, 1988).

1.9.1 The life cycle of P22

Like λ, following infection of its host, P22, may enter one of the two pathways, lytic or lysogenic growth. In a lysogen, prophage P22 uses a regulatory system that may lead to the induction of lytic growth or maintenance of the lysogenic state (Poteete, 1988).

1.9.1.1 Lytic growth

P22 adsorbs to the O–antigen of the bacterial lipopolysaccharide (LPS) and attaches to an internal receptor, thereby ejecting a number of minor protein components of the phage which are needed for DNA injection (Poteete, 1988). Three different promoters (P_L, P_R and P_{ant}) are involved in early transcription of the DNA template. For efficient expression of downstream genes, gp24 protein, a λ N homologue, encoded by the P_L transcript, induces RNA polymerase to transcribe past terminators at the end of the short P_L and P_R RNAs. Cro protein, which acts as a repressor and under control of the P_R promoter, binds to operators that overlap P_R and P_L thereby down–regulating transcription from P_L and P_R (Poteete, 1988). Repression of transcription from the P_L and P_R promoters is necessary for maintenance of the lytic pathway otherwise, lysogeny may occur due to overproduction of C1 protein (to be discussed later) (Winston and Botstein, 1981b). The P_{ant} transcript encodes an antirepressor, Ant, as well as the Arc protein (Susskind, 1980, Susskind and Youderian, 1982, Youderian, *et al.*, 1982). Arc protein is a *c2* repressor that turns down transcription by binding to the operator, overlapping P_{ant}. This regulation of the P_{ant} prevents overproduction of antirepressor which would potentially interfere with lytic growth and establishment of lysogeny (Susskind, 1980).

Late gene expression is regulated, as in λ, by an antitermination dependent mechanism involving gp23, a Q homologue (Poteete, 1988). The late genes include gp13 (holin), gp19 (a lysozyme homologue) and the genes involved in morphogenesis. These genes constitute a single operon (Casjens and Adams, 1985). The genes involved in DNA replication, homologous recombination, late transcription regulation and two regulatory genes (*c3* and *c1*) involved in lysogeny, are transcribed by gp24 protein antitermination activity (Poteete, 1988) (Figure 1.4).

1.9.1.1.1 P22 DNA replication and packaging

DNA replication is initiated from an origin (Ori) located within gene *18* (Backhaus and Petri, 1984). Replication requires gp18 (a primase) and gp12 (a helicase) as well as host cell DNA polymerase and replication proteins (Schanda-Mulfinger and Schmieger, 1980). The formation of concatemeric molecules results probably as a result of rolling circle replication (Poteete, 1988). Botstein *et al.* (1970) showed that a circular DNA molecule is necessary for rolling circle replication (Botstein and Matz, 1970).

DNA packaging initially proceeds from a specific sequence designated, *pac*, and proceeds in one direction through the head genes. Subsequently, the next rounds begin where the first round ended. This process continues until the whole phage genome has been packaged (Figure 1.5). Regardless of the genome size, P22 packages a head–full of 43.4 kb DNA which has 1.7 kb direct repeats at each end (Poteete, 1988). Homologous recombination between the heterogeneous repetitious ends of the DNA injected by the phage is necessary for circularisation of the infecting DNA (Botstein and Matz, 1970, Poteete, 1988). P22 can use its own or the host cell's homologous recombination system. It has been demonstrated that the phage homologous recombination system is more active as wild–type P22 replicates well in *rec*A$^+$ cells whereas P22 *erf*$^-$ replicates poorly in *rec*$^-$ cells and nonproductively in *rec*A$^-$ cells. The Erf protein constitutes the P22 homologous recombination system and promotes a high level of inter–molecular phage recombination (Poteete, 1988, Susskind and Botstein, 1978a).

1.9.1.1.2 P22 assembly

Assembly of P22 involves four steps: (1) prohead assembly, (2) DNA packaging, (3) head maturation and (4) tail addition (Poteete, 1988). Prohead assembly involves the

major capsid protein (gp5) which is assembled around a core of scaffolding protein (gp8) and four other proteins present in small amounts (gp1, gp7, gp16 and gp20). When DNA enters the prohead, scaffolding protein leaves intact and the capsid shell expands. This process is dependent on ATP and phage proteins encoded by genes *2* and *3*. Gp1 is necessary for efficient DNA packaging. Although phage assembly can occur in the presence of minor proteins, phage particles assembled without these proteins are unable to inject DNA in subsequent infections. To stabilise the phage heads and prevent loss of DNA in an extracellular environment, gp4, gp10 and gp26 are required. Phage assembly culminates with the addition of six trimers of tail protein (the product of gene *9*) (Poteete, 1988) (Figure 1.6). In order for bacterial lysis to occur, the P22 gp13 protein creates channels through the cytoplasmic membrane of the host, giving gp19 protein, a lysozyme, access to the cell wall, thereby causing lysis of the host cell and consequently, release of mature phage (Poteete, 1988).

1.9.1.2 Lysogeny

Integration of P22 into the chromosome of its normal host, *S.* Typhimurium, is accomplished by recombination between specific sites (*attP* and *attB*) in the phage and bacterial chromosome. Recombinant sites are generated by reciprocal crossing over these two sites giving rise to *attL* (flanks the prophage on the left) and *attR* (flanks the prophage on the right) (Poteete, 1988). P22 always integrates at the same location on the bacterial chromosome and involves an integrase dependent site–specific recombination. In the prophage state, only genes involved in the maintenance of lysogeny are expressed (Susskind and Botstein, 1978a). The phage repressor from the *immC* region, encoded by gene *c2,* inhibits transcription of early genes from the P_L and P_R promoters, simultaneously initiating transcription of its own genes from the P_{RM} promoter. A second repressor from

the *immI* region, encoded by the *mnt* gene is required to turn off expression of an antirepressor (*ant*) by binding to operator overlapping the promoter $P_{ant.}$ The gene *ant* is expressed only when the *mnt* repressor is absent or inactive (Susskind and Botstein, 1978a). Maintenance of lysogeny depends on the availability of these two repressors. If the C2 repressor is inactive, the lytic cycle of the prophage ensues. Furthermore, when the Mnt repressor is inactive, an antirepressor is produced, resulting in inactivation of the C2 repressor, thereby expressing the lytic functions (Susskind and Botstein, 1978a) (Figure 1.4). In addition to inactivating the P22 C2 repressor, the P22 antirepressor protein inactivates the C1 repressor of coliphage λ (Susskind and Botstein, 1975). In contrast, coliphage λ maintains its immunity/repression system with a single repressor. λ repressor binds specifically to two operator sites, preventing all transcription from the λ prophage, except for the transcription of the repressor itself (Susskind and Botstein, 1978a).

1.9.1.2.1 Superinfection Immunity

In addition to its repressor/antirepressor exclusion system (Figure 1.4), prophage P22 expresses three different systems that impede infection by homologous phages (i) the *sieB* gene product which interferes with superinfection by homologous phages (ii) the *sieA* gene product which interferes with DNA injection of homologous phages and (iii) gene *a1* which alters the structure of the bacterial O–antigen (Poteete, 1988, Susskind and Botstein, 1978a, Susskind, *et al.*, 1974b, Young, *et al.*, 1964) resulting in serotype conversion from 4, [5], 12 to 1, 4, [5], 12 (Iseki and Kashiwagi, 1955, Rundell and Shuster, 1975). Figure 1.7 depicts another superinfection exclusion mechanism used by a prophage's immunity system to stop the entry or replication of homoimmune phages.

S. Typhimurium belongs to the *Salmonella* serogroup B that is characterised by the possession of a D–mannose–α1→2–L–rhamnose–α1→3–D–galactose trimeric repeat. However, the mannose residue (α1→3) is substituted with the 3,6–dideoxy hexose abequose (Weintraub, *et al.*, 1992), that is equivalent to O antigen 4. Lysogenisation of cells by P22 results in the addition of α1→6 glucosyl residues on the galactosyl residues leading to the appearance of O antigen 1 (Weintraub, *et al.*, 1992). Recently, the genes involved in this conversion (*gtrA, gtrB* and *gtrC*), were identified and shown to be related to other proteins involved in serotype conversion, for example, those encoded by *Shigella flexneri* phages SfV, SfII and SfX (Vander Byl and Kropinski, 2000).

1.9.1.2.1 Establishment of lysogeny

A number of factors determine whether a P22 infected cell will become a lysogen. These include the nutritional status of the cell as well as the multiplicity of infection (MOI). Low MOI and the availability of a preferred carbon source, e.g. glucose, lead to lytic development of phage (Poteete, 1988, Steinberg and Gough, 1976, Susskind and Botstein, 1978a). It has been shown that more than 90% of cells are lysogenised at MOIs above 5, 45% of cells become lysogens at MOIs of about 3 while at MOIs of 1, only 23% of infected cells become lysogens (Steinberg and Gough, 1976). Thus the frequency of lysogenisation is probably mediated by the rate of synthesis of the C1 protein (Winston and Botstein, 1981b). Gene *c1* is a regulatory protein that induces transcription of the *c2* repressor gene from the promoter P_{RE}. The repressor then promotes transcription of its own gene from P_{RM} promoter. Furthermore, C1 protein delays DNA replication and late gene expression. Additionally, gene *c3* encoded by the P_L transcript, is essential for stabilisation of C1 protein (Poteete, 1988). These two gene products, C1 and C3, stimulate the synthesis of C2 repressor, thus, affecting the switch between lytic growth and lysogeny. However,

eventually C2 synthesis continues without the presence of C1 and C3 gene products since these two genes are ultimately repressed in lysogens (Bronson and Levine, 1971, Bronson and Levine, 1972, Susskind and Botstein, 1978a). Thus the lysogenic state of P22 can be maintained indefinitely.

However, induction of the lysogen by DNA–damaging treatments can cause reversal of repression and excision of the prophage. Excision requires expression of the phage genes, *int* and *xis*. These events lead to a switch from lysogeny to lytic growth of the phage. In this case, the circular phage DNA is generated by excisive recombination and the *mnt* repressor is probably not inactivated, thus interfering with the tail gene (Poteete, 1988).

1.10.1 Homologous recombination system

Homologous recombination requires extensive homology between the two interacting DNA substrates in order to assure faithful inheritance of DNA material, especially under stressful conditions (Quiberoni, *et al.*, 2001). Generally, it is thought that homologous recombination systems are conserved and function in a similar manner. However, a study on *Lactococcus lactis,* has shown that the main components of the double strand break repair system, an exonuclease/helicase and a short modulator sequence Chi, differ between bacteria, particularly when compared to the Gram negative analogues (Quiberoni, *et al.*, 2001). Nevertheless, this study confirmed the conservation of many enzymes across species barriers.

In P22, three genes are involved in homologous recombination. The *abc* (anti–recBCD) genes modulate the activities of the host cell RecBCD protein whereas the *erf* (essential recombination function) gene is involved in *rec*A–independent homologous recombination (Poteete, 1988). Functionally, the *abc* genes are analogues of the λ *gam*

genes (Smith, 1983). Poteete and Fenton, (1984) showed that a plasmid expressing λ *gam* is able to complement P22 *abc*1$^-$ and *abc*2$^-$. Furthermore, the λ analogue of *erf* is *bet* and plasmids that express *bet* are able to complement P22 *erf*$^-$ (Poteete and Fenton, 1984).

1.10.2 Site–specific recombination system

The site–specific recombination pathways of *S.* Typhimurium and *E. coli* are responsible for the integration and excision of the bacteriophage into and out of the host chromosome (Susskind, 1978, Poteete, 1988, Campbell, 1994). Although the *int* genes of all lambdoid phages are related to one another, indicating common ancestry, site–specific recombinations related to *int* are also distributed among non–lambdoid phages, bacterial plasmids and eukaryotes (Anilionis and Riley, 1980, Craig, 1988). Some lambdoid phages such as λ and 434, have a common integration site and are able to interchange all components of the integration system (Campbell, 1994). In contrast, λ and phage 21 have different integration sites and neither Int nor Xis is interchangeable (Schneider, 1992 PhD, Standford Univ, Standford) (cited in Campbell, 1994).

However, λ and HK022 have different integration sites as well as distinct but overlapping specificities of core site recognition (Nagaraja and Weisberg, 1990). The location of *attP*, *int* and *xis* is conserved in both λ and P22. However, both Int and Xis proteins of λ are smaller than those of P22. Another interesting feature of site–specific recombination system is the binding of integrative host factor (IHF) of *E. coli* to the *att* sites. IHF binds to three sites in λ *attP*, two in P22 *attP* and one in P22 *attB* (Leong, *et al.*, 1985). Thus both λ and P22 site–specific recombination systems seem to work by similar mechanisms. However, like many other Int family members, the four *att* sites of P22 share a larger region of homology (46 bp) than is observed in λ (15 bp) (Campbell, *et al.*, 1992,

Leong, *et al.*, 1985). Furthermore, the arrangement, spacing and orientation of the recombination pathways of λ and P22 Int binding sites are different (Smith-Mungo, *et al.*, 1994). The mechanism of integration and excision of bacteriophage λ is shown in Figure 1.8.

1.11 The evolution of bacteriophage

There is general consensus that a bacteriophage genome is originally chimeric. This hypothesis is based on the observation that various segments of the λ genome are derived from separate sources such as their host genome, plasmids or transposons (Campbell, 1988). Illegitimate recombination mechanisms, particularly those effected by transposons, make this evolution mechanism acceptable. There are two extremes: (1) all λ related bacteriophages are direct descendants of the primordial λ phage; (2) new λ–like bacteriophages may be continually assembled from extraneous components. It is likely that, the truth probably lies between these two extremes (Campbell, 1988).

1.11.1 The gene pool

Conservation of the arrangement of major gene clusters is observed in lambdoid bacteriophages. It is likely that in natural settings, lambdoid phages have repeatedly encountered and recombined with each other, therefore, placing all bacteriophages in a common gene pool (Campbell, 1988). Some studies have shown close homology between two lambdoid phages in some parts of the genome while no homology has been shown in other segments of the same genome (Fiandt, *et al.*, 1971, Simon, *et al.*, 1971). These homologies appear to differ from one phage genome to another. For example, the immunity region of bacteriophages P22 and 21 are similar whereas that of λ is completely

different. In addition, the tail cluster genes of bacteriophages λ, 434, 82, PA–2, 21 and 424 share extensive sequence homology while those of P22 are different (Campbell, 1988). Additionally, there seem to be sharp breakpoints between the segments. For example, bacteriophage 434 is homologous to λ in the integrase operon, then becomes homologous to PA–2 to the right (Campbell, 1988). Comparison of the *Rz* genes (λ terminology) of HK97, λ and P22 also indicate two homologous intragenic crossovers for P22 and λ (Casjens, *et al.*, 1989, Casjens, *et al.*, 1992). There is 62% sequence identity between the first 60 bp of P22 and λ, 92% sequence identity for the second 60 bp and 52% sequence identity for the third 60 bp. However, HK97 resembles λ for the first 60 bp, then resembles both λ and P22 for the second 60 bp, then resembles P22 for the rest of the gene. This suggests that HK97 resembles a crossover between a λ–like parent and a P22–like parent in the second 60 bp sequence of its *Rz* gene (Campbell, 1988). Overall differences and similarities of different phage segments, indicate that related phages evolved by recombination with other unrelated phages (Hershey, 1971).

The λ gene pool also includes remnants of related phages found in the chromosome of enterobacteria. For example, *E. coli* K-12 has the Rac prophage which contains an integrase, an origin of replication, recombination genes and a repressor (Kaiser and Murray, 1979). In addition, the qsr' prophage contains genes that are λ *Q* analogues, the lysis genes *S* and *R* and a *cos* site (Highton, *et al.*, 1985). There is a segment homologous to part of the λ *red*X gene in the DNA flanking these elements and part of the λ *Nul* gene (Campbell, 1988). Espion, *et al.* (1983), reported that the qin111 prophage extends almost over the same portion of a λ–related genome as does the qsr' prophage (Campbell, 1988). Furthermore, hybridisation studies with λ probes suggest that the chromosomes of natural strains of *E. coli* and related strains such as *S.* Typhimurium contain λ– related segments (Anilionis and Riley, 1980, Riley and Anilionis, 1980). The availability of sequence

databases for both phages and defective phages substantially contributes to the study of phage evolution.

1.11.1.1 Recombination

In nature, when two phages genomes interact, one or more of the four types of recombination systems may occur (Casjens, *et al.*, 1992): (i) homologous recombination, (ii) site–specific recombination (at the integration site), (iii) recombination at microhomologies of oligonucleotide length, whose sequence are conserved for functional reasons and (iv) illegitimate recombination, where recombining partners have no homology whatsoever. Illegitimate recombination usually results in deletion, duplication, insertion or translocation of a chromosome (Shanado, *et al.*, 1997). Both homologous and site–specific recombinations may occur at high rates, however, with the percent of mismatching between partners, the rate of homologous recombination drops rapidly (Campbell, 1988). Homologous recombination may create new combinations of flanking genes but cannot create new breakpoints between homology and non–homology. Similarly, site–specific recombination cannot create new breakpoints as it may occur between phages with the same integrase specificity, which may be homologous on both sides of the *att* site. In contrast, both microhomologous and illegitimate recombination can generate sharp breakpoints (Campbell, 1988). Due to the exchange points being the same on both partners for the first three mechanisms, recombinants with full complement of phage genes are usually generated. In contrast, illegitimate recombination mechanism generates recombinants, which may not have a complete phage genome. Furthermore, illegitimate recombination is the only mechanism that creates unique breakpoints at arbitrary positions on heterologous partners (Campbell, 1988).

1.11.1.2 Speciation in phages

The lambdoid phages are collectively known as a biological species with its own gene pool (Campbell, 1988). For this to be the case, there must be frequent recombination within the species and probably barriers to block entry of DNA from other sources. However, as previously mentioned, some lambdoid phages are mosaics with genes from other non–lambdoid phages, plasmids, transposons and bacterial chromosomal DNA. Although, homologous recombination is more frequent within the gene pool, illegitimate recombination though rare, may form recombinants with any heterologous DNA, which may come in contact with the phage, as it does not recognise any boundaries. The species concept may then appear to be undermined by the fact that lambdoid phages may have genes from non–lambdoid phages as well as from other sources. However, the species concept is inherently accepted and acquisition of foreign genes may be referred to as lateral transfers (Campbell, 1988).

1.12 Bacterial gene transfer

Horizontal gene transfer is of great importance to the understanding of evolutionary mechanisms as well as the spread of virulence and antibiotic resistance genes (Miller, 1998, Schicklmaier, *et al.*, 1998). Moreover, horizontal gene transfer is very important pertaining to safety of the release of genetically engineered microorganisms. Three forms of horizontal gene transfer are known – conjugation, transformation and transduction. In this thesis, emphasis will be placed on the bacteriophage mediated gene transfer (transduction).

1.12.1 Transformation

Transformation was discovered by Griffith in 1928 and was the very first gene transfer mechanism to be studied. Griffith observed that when non–virulent pneumococcal bacteria were injected into mice with dead virulent *Pneumococcus*, the non–virulent strain picked up the virulence genes from the dead virulent bacteria and became lethal to mice. Gene transfer in both Gram negative and Gram positive bacteria requires that the transforming DNA is stable and the recipient cells display specialised surface proteins that can bind to the DNA (are competent) to internalise it (Miller, 1998). Natural transformation seems to occur only between cells of the same species. Furthermore, few bacterial species are capable of becoming competent for transformation (Miller, 1998), in particular, Gram negative bacteria are rarely naturally competent, however, a few exceptions are known, e.g. *Neisseria* species.

1.12.2 Conjugation

Conjugation was first discovered in *E. coli* by Lederberg and Tatum in 1946. In Gram negative bacteria, a donor cell attaches to a recipient cell by one or more pili, forming a bridge that allows a single strand of plasmid DNA to pass into the recipient cell, where a complementary strand is synthesised. In contrast, in Gram positive bacteria, cells are drawn together by chemical signaling instead of a pilus. Chromosomal DNA can be exchanged by conjugation in extremely rare situations (Miller, 1998). Generally, bacteria exchange plasmids exclusively with members of their own species, however, some "promiscuous" plasmids can transfer DNA between unrelated species, for example, between Gram negative and Gram positive bacteria or from bacteria to yeast cells or even plants (Miller, 1998).

1.12.3 Transduction

As previously mentioned, transduction was discovered in *S.* Typhimurium almost 50 years ago by Zinder and Lederberg, (1952). They demonstrated that genetic material of strain LT22 (with prophage PLT–22) could be transferred to a recipient strain (LT2) by means of a temperate phage, PLT–22 (now known as P22). In the same year, Hershey and Chase discovered that phage infection resulted from the injection of phage DNA into the bacterial cell (Hershey and Chase, 1952). Since then, the presence of transducing phage in many bacteria has been reported, including, *E. coli* by phages λ, P1 and 363 (Lennox, 1955, Morse, 1954), *Shigella* by phage P1 (Adams and Luria, 1958, Lennox, 1955), *Pseudomonas* (Loutit, 1958), *Staphylococcus* (Morse, 1959), *Proteus* (Coetzee and Sacks, 1960) and *Bacillus subtilis* (Takahashi, 1961, Takahashi, 1963). Indeed, bacteriophages are capable of transferring whole plasmids or pieces of chromosome between different hosts, for example, the pathogenicity islands of *S.* Typhimurium which are approximately 40 kb in size (Mills *et al.,* 1995, Shea *et al.,* 1996, Miller, 1998).

Two types of transduction are known. Generalised or unrestricted transduction was demonstrated by Zinder and Liderberg (1952) and subsequently elaborated by Demerec and colleagues, who demonstrated that any small region of donor chromosome can be transferred to a recipient strain (Hayes, 1964b). Morse *et al.* (1956) later discovered that phage λ could mediate transduction in a restricted way (specialised), where, only the genes located near the site of integration could be transferred.

In general, studies of horizontal gene transfer focus on conjugation and transformation and neglect phage mediated transduction (Schicklmaier and Schmieger, 1995). For example, in 1993, at the 4th Symposium on Bacterial Genetics and Ecology held in Wageningen, Germany, there were 46 presentations on horizontal gene transfer where 78% was based on conjugation, 18% on transformation and only 2% on transduction

(Schicklmaier and Schmieger, 1995). Although phage mediated transduction may be regarded as a rare event in nature, and its contribution to genetic exchange neglected, its occurrence may be greater than expected as phages are very common in the environment and relatively stable, as they are protected by the protein coat. Furthermore, phages are also compact and more diffusible than naked DNA (Davison, 1999). Very high concentrations of bacteriophages (100 billion virus particles per milliliter) in fresh and marine waters have been reported (Miller, 1998). Similarly, it was shown that 95% of natural strains of *S.* Typhimurium harbour prophages of which the majority (62%) were found to be generalised transducers (Schicklmaier, *et al.*, 1998). This suggests that transduction may be more common than previously thought.

1.12.3.1 Specialised transduction

Specialised transduction is the process whereby bacterial genes are physically incorporated into the genome of a temperate bacteriophage and are transferred from one host to another (Susskind and Botstein, 1978a). Specialised transduction usually results in gene addition where the transductant carries the transducing material as long as it still carries the prophage. However, if the prophage is lost, the transducing material is usually also lost (Susskind and Botstein, 1978a). In 1956, Morse *et al.* showed that λ particles could transfer galactose (*gal*) fermentation markers to *E. coli* strain K–12 (Morse, *et al.*, 1956). The phage material, together with the adjacent segment of the bacterial chromosome (gal^+) was incorporated as a single unit into a phage particle. When gal^- sensitive cells were lysogenised, the gal^+ marker was inserted into the chromosome of the bacterial cell making it a "heterogenote" for the galactose marker (both gal^- and gal^+ were present in the cell). Induction of these heterogenotes resulted in phage particles which carry the gal^+ marker (Adams, 1959). In rare situations, gene substitution occurs by

recombination when the bacterial genes carried on the phage are homologous to genes in the recipient chromosome. In these circumstances, stable transductants may be produced regardless of the presence or absence of the transducing phage (Susskind and Botstein, 1978a).

In 1962, Campbell proposed that specialised transducing phage were formed by aberrant excision of an integrated prophage, resulting in incorporation of bacterial genes into the phage genome adjacent to the phage attachment site on the bacterial genome (Campbell, 1962). This was shown to apply to certain λ derivatives as well as to some P22 specialised transducing phages (Chan and Botstein, 1976, Hoppe and Roth, 1974, Jessop, 1972, Jessop, 1976). For example, genetic and physical evidence showed that P22 pro–1 and pro–3 phages were formed by aberrant excision. These two phages carry the *proA* and *proB* genes, which are normally located near the P22 attachment site in the *S.* Typhimurium chromosome (Chan and Botstein, 1976, Jessop, 1976).

In addition to Campbell's mechanism of specialised transduction, a second mechanism which involves direct insertion of translocatable elements into the phage genome, has been proposed (Chan and Botstein, 1976, Kleckner, *et al.*, 1975). This mechanism explains the origin of a specialised transducing phage P22Tc–10 which carries genes for tetracycline resistance from an R–factor (Chan, *et al.*, 1972, Watanabe, *et al.*, 1972). P22Tc–10 is defective and could not transduce tet^R upon single infection. However, two defective P22Tc–10 phages could cooperatively transduce tet^R.

During packaging of the Tc–10 or pro–1 concatemers, the DNA is cut to the usual size such that the resulting molecules in the progeny particles are each missing part of the oversize genome and do not have terminal repetition. These DNA molecules are therefore unable to grow or to lysogenise on single infection. However, in a double infection, the two DNA molecules would probably recombine with each other and either lysogenise or

produce progeny particles (Chan, *et al.*, 1972). These findings indicate that P22 can mediate high frequency specialised transduction of very large segments of DNA. In contrast, λ, which packages DNA by a site–specific mechanism producing molecules with unique ends, has a much lower limit of packaging (Feiss, *et al.*, 1977).

1.12.3.2 Generalised transduction

Estimation of the occurrence of generalised transduction may be hampered by the fact that this process is carried out by temperate phages which need a sensitive indicator strain to be detected (Schicklmaier and Schmieger, 1995). It is probable that generalised transducing particles result when the host chromosome serves as a substrate for the sequential head–full packaging mechanism. It has been shown that P22 generalised transducing particles contain DNA molecules of the same molecular weight (MW) as the mature P22 genome which consist primarily of host DNA synthesised before phage infection (Ebel-Tsipis, *et al.*, 1972, Schmieger, 1970). Packaging of host DNA in generalised transduction, is initiated at preferred specific sites. This may explain why P22 has a higher efficiency in transducing certain markers than others (Ebel-Tsipis, *et al.*, 1972, Schmieger, 1972). Although there may be as much host DNA as phage DNA in an infected cell, P22 encapsulates far more phage DNA than host DNA, presumably because P22 DNA has better initiation sites than host DNA (Ebel-Tsipis, *et al.*, 1972, Schmieger, 1970). The properties of phage mutants with increased transduction frequencies (HT mutants) demonstrate the relationship of the phage DNA encapsulation mechanism to generalised transduction (Schmieger, 1972). These mutants were shown to encapsulate as much as 50% host DNA as opposed to 1–5% encapsulated by wild–type P22 (Ebel-Tsipis, *et al.*, 1972, Schmieger, 1972). Furthermore, HT mutants exhibit alterations in co-transduction frequencies for linked markers (Schmieger and Backhaus, 1976).

1.12.3.3 Transduction of plasmids

P22 can also mediate transduction of plasmids (Garzon, *et al.*, 1995, Orbach and Jackson, 1982, Sanderson and Roth, 1988). For example, plasmid pBR322 can be packaged when the vector contains a P22 *pac* site or a *pac* homologue from the bacterial chromosome (Schmidt and Schmieger, 1984, Schmieger, 1982, Vogel and Schmieger, 1986). However, P22HT mutants can mediate transduction of pBR322 without a wild–type *pac* site because of reduced stringency for *pac* recognition (Margolin, 1987). In contrast, transduction of low copy number plasmids such as pSC101 by P22HT, takes place using two separate methods. The first method involves integration into the donor chromosome by homologous recombination where it is subsequently packaged into the transducing particle. The second mechanism is size–dependent and involves formation of putative plasmid multimers, thought to be formed by inter–plasmid recombination (Mann and Slauch, 1997). However, it has been previously demonstrated that a pSC101 derivative could be transduced by phage P1 using homologous recombination. Circularisation of the plasmid occurs in the recipient and the phage then packages the intermediate of a double recombination event between the plasmid and the bacterial chromosome (Trun and Silhavy, 1987). Additionally, transduction of R6K derivative pGP704 by P22HT*int* was achieved by homologous recombination (Mahan, *et al.*, 1993).

1.12.3.4 Transduction of resistance markers

Gene transfer by transformation and conjugation appears to be uncommon in the genus *Salmonella* (Schmieger and Schicklmaier, 1999). Only one case of plasmid mediated transfer of the Apramycin resistance gene in *S.* Typhimurium DT104 strain has been documented (Low, *et al.*, 1997). There is no documentation of transformation in *Salmonella*. However, in the 1950s, transduction of streptomycin, chloramphenicol and

penicillin resistance was demonstrated in *S.* Typhimurium (Banic, 1959, Watanabe and Watanabe, 1959). More recently, transduction of multiple drug resistance of *S.* Typhimurium DT104 has been demonstrated (Schmieger and Schicklmaier, 1999). *S.* Typhimurium DT104 strain is resistant to many antibiotics including ampicillin, chloramphenicol, streptomycin, spectinomycin, sulphonamides and tetracycline. Phage ES18, which is closely related to P22, as well as phage PDT104 carried by all DT104 strains, can mediate transduction of these resistance genes (Schmieger and Schicklmaier, 1999). It is therefore reasonable to consider transduction as a common vehicle for horizontal gene transfer in *Salmonella*.

A similar scenario is found in *Pseudomonas aeruginosa*. Wild–type bacteriophages carried by lysogenic multi–drug resistant hospital strains of *P. aeruginosa,* have been shown to transduce determinants of resistance to aminoglycosides (Knothe, *et al.*, 1981) and β–lactam antibiotics (Blahova, *et al.*, 1993, Seginkova, *et al.*, 1986). Two high frequency transducing phages (HFT), isolated from very ill patients, were able to mediate transduction of antibiotic resistance markers with frequencies of 10^{-5} plaque forming units (pfu)/ml (for phage isolate AP–103) and 10^{-6} pfu/ml (for phage isolate AP–343). The frequency depended on the resistance marker and on the individual recipient strain used (Blahova, *et al.*, 1998a). Furthermore, phage AP–343 was able to transduce multiple–antibiotic resistance markers to all tested *Pseudomonas* recipients. Recently, Blahova *et al.* (2000), demonstrated low–frequency transduction of imipenem resistance and high frequency transduction of ceftazidime and aztreonam resistance by phage AP–151 from *P. aeruginosa*. Although transfer of resistance genes has long been shown to occur by conjugation in *P. aeruginosa,* the carriage of HFT phages by this bacterium isolated from clinical conditions, may contribute to the spread of antibiotic

resistance genes, which may create serious problems when nosocomial infections occur (Blahova, *et al.*, 1998b).

Bacteriophage-mediated transduction of amino acid and drug-resistance markers has also been demonstrated in *Vibrio cholerae* (Ogg, *et al.*, 1981). The frequency of transduction of drug resistance markers was lower compared to that of transduction of amino acid markers, probably because of the recessive nature of many antibiotic resistance alleles. However, a mechanism proposed for P22 which involves the formation of transducing DNA fragments by sequential encapsulation from a small number of preferred starting points in the host chromosome, could account for these differences.

1.12.3.5 Transduction of virulence associated genes

Bacteriophages have played a critical role in the evolution of many bacterial pathogens. Temperate bacteriophages often encode properties that alter the host bacterium following the establishment of lysogeny. The toxins of a number of Gram positive and Gram negative bacteria are encoded in the genomes of bacteriophages (Acheson, *et al.*, 1998, Waldor, 1998). In *S. aureus*, lysogeny can affect the expression of several enzymes and extracellular toxins, some of which may be virulence factors (Parker, 1983). For example, expression of staphylokinase and enterotoxin A may be acquired after lysogenisation (positive phage conversion) (Casman, 1965). In contrast, expression of β–lysin and lipase can be lost following lysogeny (negative phage conversion) (Coleman, *et al.*, 1986, Lee and Iandolo, 1985). Other toxins of Gram positive bacteria encoded in the genomes of bacteriophages include, diptheria toxin, botulinum toxin types C and D as well as streptococcal erythrogenic toxin (Waldor, 1998). The toxin genes, in most of these cases, are located close to the bacteriophage attachment site, suggesting that these genes

have been acquired by transduction (Barksdale and Arden, 1974, Betley and Mekalanos, 1988, Johnson, *et al.*, 1986, Laird and Groman, 1976).

Phages related to λ have been shown to carry the genes encoding shiga like toxins (Stxs) (Huang, *et al.*, 1987, Newland, *et al.*, 1985, O'Brien, *et al.*, 1984). The genes encoding Stxs (Stx1 and Stx2) have been found in resident prophages in most toxigenic *E. coli* (Newland and Neill, 1988). These toxins are believed to play a very important role in the pathogenesis of haemorrhagic colitis, HUS and thrombotic thrombocytopenic purpura that may result from human infection with lysogenic *E. coli* (Griffin, 1995, Neely and Friedman, 1998). An *in vitro* study demonstrated that a lysate of phage H–19B from an *E. coli* strain O26:H11 isolated from an outbreak of infantile diarrhoea, could transfer enterotoxinogenicity to *E. coli* K–12 (Williams Smith and Linggood, 1971). It was later discovered that this strain harboured a bacteriophage H–19B that encoded Stx1 and had DNA homology with λ (Huang, *et al.*, 1987). Another lambdoid phage designated 933W from *E. coli* strain O157:H9 (isolated from a patient with haemorrhagic colitis), encoded Stx2 (Newland, *et al.*, 1985, O'Brien, *et al.*, 1984). Schmidt *et al.* (1999) demonstrated the spread of Stx2–converting bacteriophages among enteric *E. coli* strains using a derivative of an Stx2–converting bacteriophage, ϕ3538. Interestingly, transient Stx production has also been demonstrated in other non–*E. coli* bacterial species such as *Enterobacter cloacae* (Paton and Paton, 1996) and *Citrobacter freundii* (Tschape, *et al.*, 1995). In addition, a bacteriophage encoding Stx genes, isolated from *Shigella sonnei*, was transduced to a non–toxinogenic *S. sonnei* reference strain as well as to *E. coli* laboratory strains (Beutin, *et al.*, 1999). It has been proposed that the dissemination of lambdoid phages encoding Stx1 and Stx2 may account for the spread of these toxins among diverse *E. coli* serotypes and other bacterial species (Acheson, *et al.*, 1998, Schmidt, *et al.*, 1999). In their study, Acheson, *et al.* (1998) demonstrated intra–intestinal transduction of Stx1–encoding phage

H–19B using a mouse model. The detection of transductants in stool samples and caecal homogenates demonstrated that infectious virions produced within the host intestine are capable of transducing other *E. coli* strains within the gastrointestinal tract (Acheson, *et al.*, 1998).

Similarly, using a model with suckling CD1 mice, transduction of CTXϕ (a filamentous *V. cholerae* phage), was demonstrated from a lysogen to a recipient strain in the gastrointestinal tract (Lazar and Waldor, 1998). It may be possible that phage mediated transduction of non–toxinogenic strains within the gastrointestinal tract may account for the new *V. cholerae* serotypes, such as the *V. cholerae* O139 strain (Acheson, *et al.*, 1998). Moreover, it has also been reported that lysogenic conversion of a non–pathogenic *C. diptheria* strain to a toxinogenic strain could occur in the upper respiratory tract (Pappenheimer and Murphy, 1983). Furthermore, other potential virulence factors encoded by bacteriophages include extracellular enzymes such as streptococcal hyaluronidase (Coleman, *et al.*, 1989), the *oac* gene of *S. flexneri* that encodes an LPS O–antigen acetylase (Clark, *et al.*, 1991), outer membrane proteins such as Lom and Bor of phage λ, a putative virulence gene, *glo*, similar to eukaryotic G proteins carried by *V. cholerae* phage K139 (Reidl and Mekalanos, 1995, Waldor, 1998) and genes for capsule production from *S. pneumoniae* (Russel, 1991).

Figueroa–Bossi and Bossi discovered the Gifsy–2 prophage which carries a *sodC* gene from *S.* Typhimurium, coding for a periplasmic superoxide (Cu, Zn)–dismutase previously implicated in *Salmonella* virulence (De Groote, *et al.*, 1997, Farrant, *et al.*, 1997). Using a mouse model, the study demonstrated the ability of Gifsy–2 phage to transfer its virulence traits to formerly attenuated strains (Figueroa-Bossi and Bossi, 1999). Similarly, another *S.* Typhimurium P2-like SopE-encoding phage, designated SopEϕ, has

been discovered (Mirold, *et al.*, 1999). The role of this phage will be discussed later in section 1.12.4.1.1 concerning type III secretion systems and virulence.

1.12.4 The role of bacteriophages in the evolution and transfer of virulence determinants

The genome of *Dichelobacter nodosus*, the causative agent of ovine footrot, carries a virulence–associated region that contains an integrase gene highly related to the integrases of *S. flexneri* phage Sf6 and coliphages P4 and ϕR73. In addition, this integrase gene has homology with *vapB*, *C* and *D* which are related to plasmid borne genes in *E. coli*, *N. gonorrhoeae*, *Treponema denticola* and *Actinobacillus actinomycetecomitans* (Cheetham and Katz, 1995). Furthermore, like bacteriophages P4 and ϕR73, the *vap* region is flanked by putative phage *att* sites and is adjacent to a tRNA gene. Following the discovery of a circular DNA molecule carrying a *vap* region in *a D. nodosus* strain, it was suggested that these *vap* sequences may have been derived from a bacteriophage or from an integrative plasmid which carried a bacteriophage attachment site and integrase (Cheetham and Katz, 1995). *S. pyogenes* prophages designated T12 and CS112, encode exotoxins SpeA and SpeC, respectively (Johnson, *et al.*, 1986, Zabriskie, 1964). The *speA* and *speC* genes are located close to the *attP* site which may suggest that the toxins were imprecisely excised leading to pyrogenic toxin converting phages (Dobrindt and Reidl, 2000). In addition, *S. flexneri* temperate bacteriophages, designated, SfII, Sf6, SfX and SfV, cause O antigen modification, which is an important virulence trait of this organism. Several authors have proposed that the integration of numerous phages close to or into tRNA genes could have a regulatory effect on expression of various genes (Allison and Verma, 2000, Cheetham and Katz, 1995).

1.12.4.1 Pathogenicity islands

A pathogenicity island (PAI) refers to a large chromosomal region in pathogenic bacteria that encodes virulence genes (Lee, 1996). Most PAIs seem to have been acquired because they differ from the bulk of the genome in their G+C content and codon usage. Furthermore, their borders are usually marked by repeated sequences or insertion elements which suggest that they were delivered to the genome by some kind of recombination (Mecsas and Strauss, 1996). PAIs are found in the genomes of pathogens but are absent in the genomes of non–pathogenic bacteria of the same or related species (Dobrindt and Reidl, 2000, Mecsas and Strauss, 1996). Although the majority of PAIs are chromosomally located, PAIs may also be part of bacterial plasmids and bacteriophages (Hacker and Kaper, 1999). However, some authors emphasise that PAIs are only located on the chromosome (Lee, 1996).

Two large PAIs were discovered in uropathogenic *E. coli* (Knapp, *et al.*, 1986). PAI–I (70 kb) and PAI–II (190 kb) recombine within short direct repeats at both ends and can be lost by deletion at a frequency of 10^{-3} to 10^{-4}. A large virulence locus called LEE (locus of enterocyte effacement), was discovered in enteropathogenic *E. coli* (McDaniel, *et al.*, 1995). The LEE locus encodes gene products involved in degradation of the apical structure of enterocytes, resulting in diarrhoea. Although PAI–I and LEE encode totally different virulence genes, their location on the *E. coli* chromosome is exactly the same. Furthermore, PAI–I and LEE share the same DNA sequence at their right hand ends. This suggests that these two elements were introduced into the *E. coli* chromosome via a similar mechanism (Lee, 1996). Interestingly, PAI–I, PAI–II and LEE insert into the chromosome of *E. coli* at sites where bacteriophages are known to insert. PAI–I and LEE insert downstream of the *selC* gene encoding tRNA for selenocysteine (McDaniel, *et al.*, 1995) where a retronphage ϕR73 (a P4–like prophage) also inserts. Similarly, PAI–II and

bactriophage P4 inserts downstream of the *leuX* gene, which encodes a tRNA gene for leucine (Blum, *et al.*, 1994, Pierson and Kahn, 1987). It is interesting to note that PAIs as well as bacteriophages are often associated with tRNA genes, which usually act as targets for the integration of foreign DNA into the genome (Campbell, *et al.*, 1992). Furthermore, PAIs often carry genes or pseudogenes of phage, origin of replication and insertion sequence elements (Dobrindt and Reidl, 2000).

1.12.4.1.1. *Salmonella* pathogenicity islands (SPI)

Five pathogenicity islands designated SPI–1 to SPI–5 have been identified in *S.* Typhimurium. SPI–1 is required for the ability of *Salmonella* bacteria to enter mammalian epithelial cells (Galan, 1996, Mills, *et al.*, 1995) whereas SPI–2 is required for the later stages of systemic disease including survival of these bacteria within macrophages (Ochman, *et al.*, 1996, Shea, *et al.*, 1996). Both islands are approximately 40 kb, a size which can be readily transduced by generalised transducing phages (Mills, *et al.*, 1995, Shea, *et al.*, 1996). The incorporation of SPI–1 and SPI–2 enabled *Salmonella* to invade the host, evade host defense systems and cause systemic infections in mammals (Groisman and Ochman, 1996). SPI–1 encodes 25 genes, including a type III secretion system that is important during the gut associated stages of infection (Galan, 1996). The SPI–1 G+C content is lower (40 to 47%) than the G+C content of the rest of the *S.* Typhimurium chromosome (52%) (Lee, 1996).

1.12.4.1.1 Type III secretion system

The type III secretion system has been found in many Gram negative bacteria that cause disease in both animals and plants. This system is called a contact–dependent

secretion system because it is triggered when a pathogen comes into contact with host cells (Ginocchio, *et al.*, 1994, Watarai, *et al.*, 1995). Effector molecules are transported from the cytoplasm directly to the cell surface where they interact with mammalian cells and modify host cell proteins (Michiels, *et al.*, 1990). Tissue culture studies have shown that the SPI–1 type III system of *Salmonella* is able to translocate a set of at least eight different proteins into the cytosol of host cells (Collazo and Galan, 1997, Galyov, *et al.*, 1997, Jones, *et al.*, 1998). It has been shown that the type III secretion system and most effector proteins are conserved among all *Salmonella* serovars and have probably been acquired by horizontal gene transfer (Groisman and Ochman, 1996).

SopE is one of the effector proteins, translocated into the host cell by the SPI-1 type III secretion system and activates the RhoGTPases CDC42 and Rac1 (Hardt, *et al.*, 1998a). A temperate bacteriophage encoding type III effector protein SopE has been isolated in *S.* Typhimurium (Mirold, *et al.*, 1999). Even though SopEϕ prophage could not be induced from all isolates tested, it was shown to be carried by all natural $SopE^+$ *Salmonella* isolates using Southern blot analysis (Mirold, *et al.*, 1999). Interestingly, *S.* Typhimurium strains DT49, DT204 and DT204c which have a history of causing major outbreaks in the United Kingdom and East Germany during the 1970s and 1980s, were amongst the natural $SopE^+$ isolates tested (Threlfall, *et al.*, 1978a, Threlfall, *et al.*, 1978b, Wray, *et al.*, 1998). Unlike other strains which were also widespread at the time, these three definitive type strains persisted over a long period of time causing massive infections in cattle and humans (Threlfall, *et al.*, 1978a, Threlfall, *et al.*, 1978b, Wray, *et al.*, 1998). The authors speculated that lysogenic conversion with bacteriophage SopEϕ may have contributed to the persistence of *S.* Typhimurium DT49, DT204 and DT204c (Mirold, *et al.*, 1999). Furthermore, the discovery of the SopEϕ prophage suggests the mechanism by which the effector proteins may be transferred to a susceptible bacterial pathogen bearing a type III secretion system.

There is a wealth of evidence that the evolution of pathogenic variants in several bacterial species results from the acquisition of defined elements including, bacteriophages. However, bacteriophages are limited in their capacity to carry exogenous DNA. The release of a phage particle and the formation of a virion, decreases with an increase in the size of their genome which in turn can lead to abortion of phage progeny and formation of a defective phage (Dobrindt and Reidl, 2000). Subsequently, such an event may lead to the formation of PAIs. The role of bacteriophages in the evolution of PAIs is demonstrated by the defective *Staphylococcus* PAIs designated SaPI encoding the toxic shock toxin 1 which can be transferred to other strains by helper phages (Lindsay, *et al.*, 1998).

1.13 Aims and Objectives

S. Typhimurium has been the most common serovar isolated from humans in Australia since 1960 (Australian *Salmonella* Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide, Annual Reports, 1960 – 1999). *S.* Typhimurium definitive phage types (DT) 9, 64 and 135 are among the most common phage types isolated from *S.* Typhimurium and are usually associated with outbreaks in this country, with DT 64 having the highest incidence in human clinical isolates. DT 29 which may be related to DTs 9, 64 and 135 was also included in the study as it caused an outbreak during the course of this study and carries phages related to those examined in this study. This thesis describes the characterisation of *S.* Typhimurium DT 64 temperate phages.

Specific aims and objectives of this study are:

- To harvest temperate bacteriophages from *S.* Typhimurium DTs 9, 64, 135 and 29
- To establish whether these phages can mediate phage type conversion

- To characterise the two *S.* Typhimurium DT 64 bacteriophages both phenotypically and genetically.

Figure 1.1

Basic morphological types of bacteriophages showing tailed (A – C) and non–tailed (D – F) phages. Reproduced from Bradley (1967).

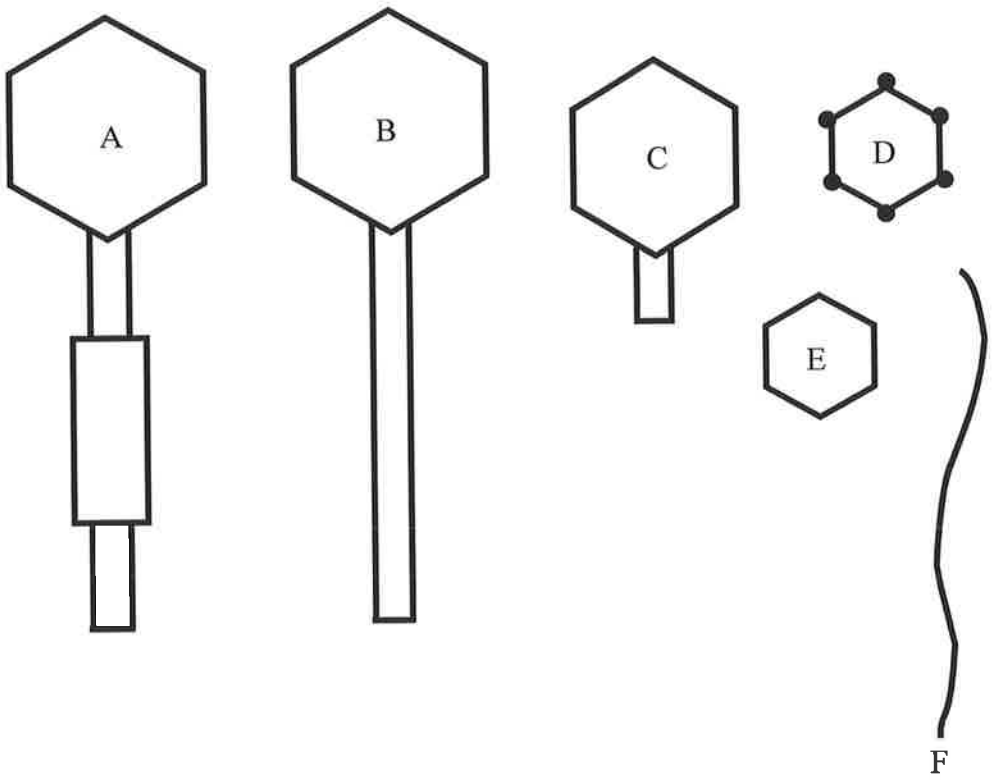
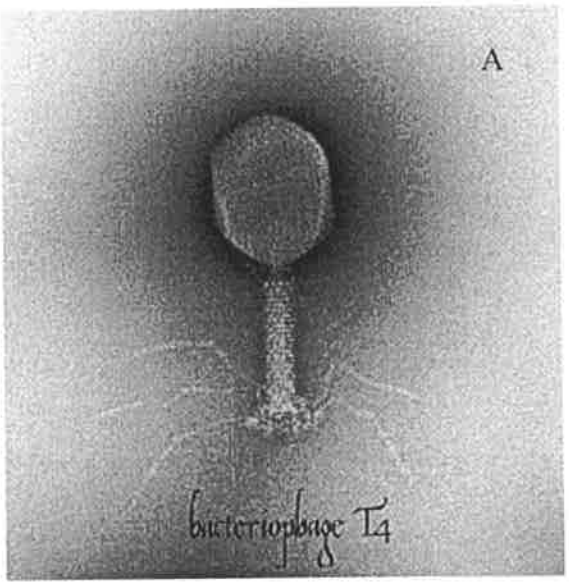
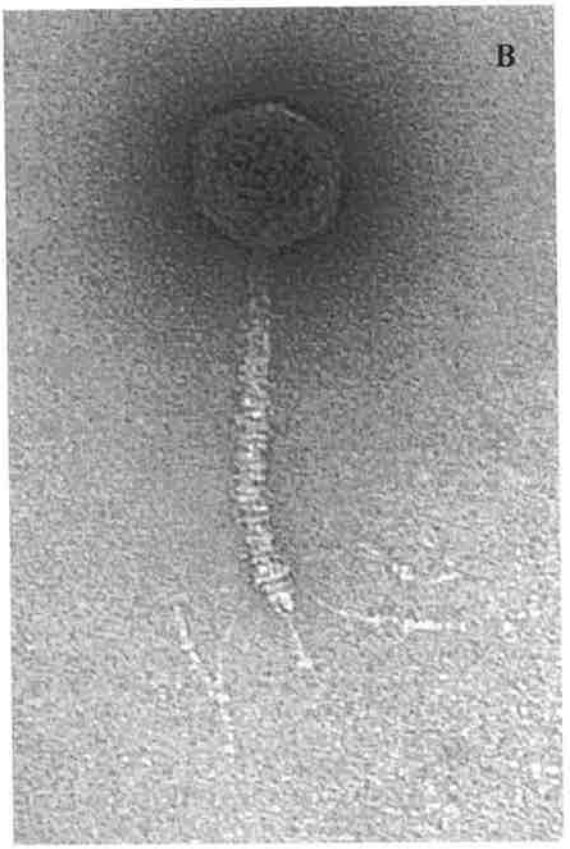

A
B
C
D
E
F

Figure 1.2

Examples of tailed phages according to the basic morphological types of Bradley (1967). A = bacteriophage T4, a member of the Myoviridae (A) family; B = wild–type bacteriophage λ, a member of the Siphoviridae (B) family; C = bacteriophage P22, a member of the Podoviridae (C) family. Reproduced from ASM website with permission.


A
bacteriophage T4


B

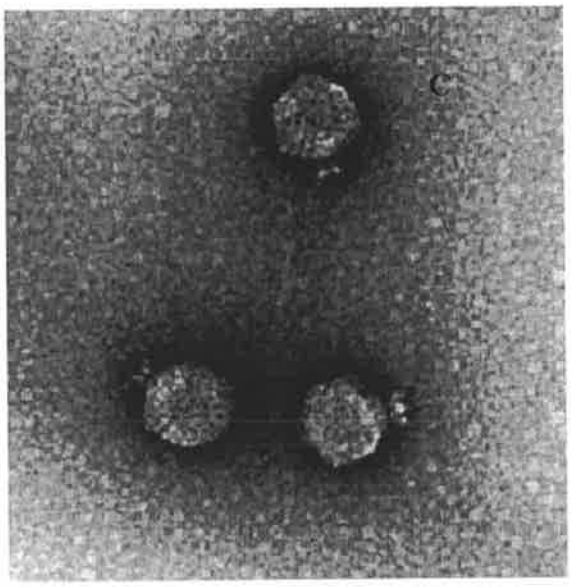

C

Figure 1.3

The life–cycle of temperate integrating bacteriophages

Adsorption of the bacteriophage to the host cell leads to the injection of DNA into the host where it circularises, leading to early transcription. In the event of lytic development of the bacteriophage, the phage DNA is replicated, the structural proteins are synthesised and then assembled into the phage coats. The replicated DNA is then packaged into these coats leading to host cell lysis and the release of the bacteriophage. If the phage is to undergo lysogenic development, lytic functions are repressed and the bacteriophage DNA is integrated into the host chromosome where it becomes a prophage. The prophage is then replicated with the host DNA and remains stably integrated as part of the bacterial chromosome. However, lytic repression is removed by induction of the prophage leading to excision of the prophage from the host chromosome, where the induced bacteriophage re–enters the lytic cycle.

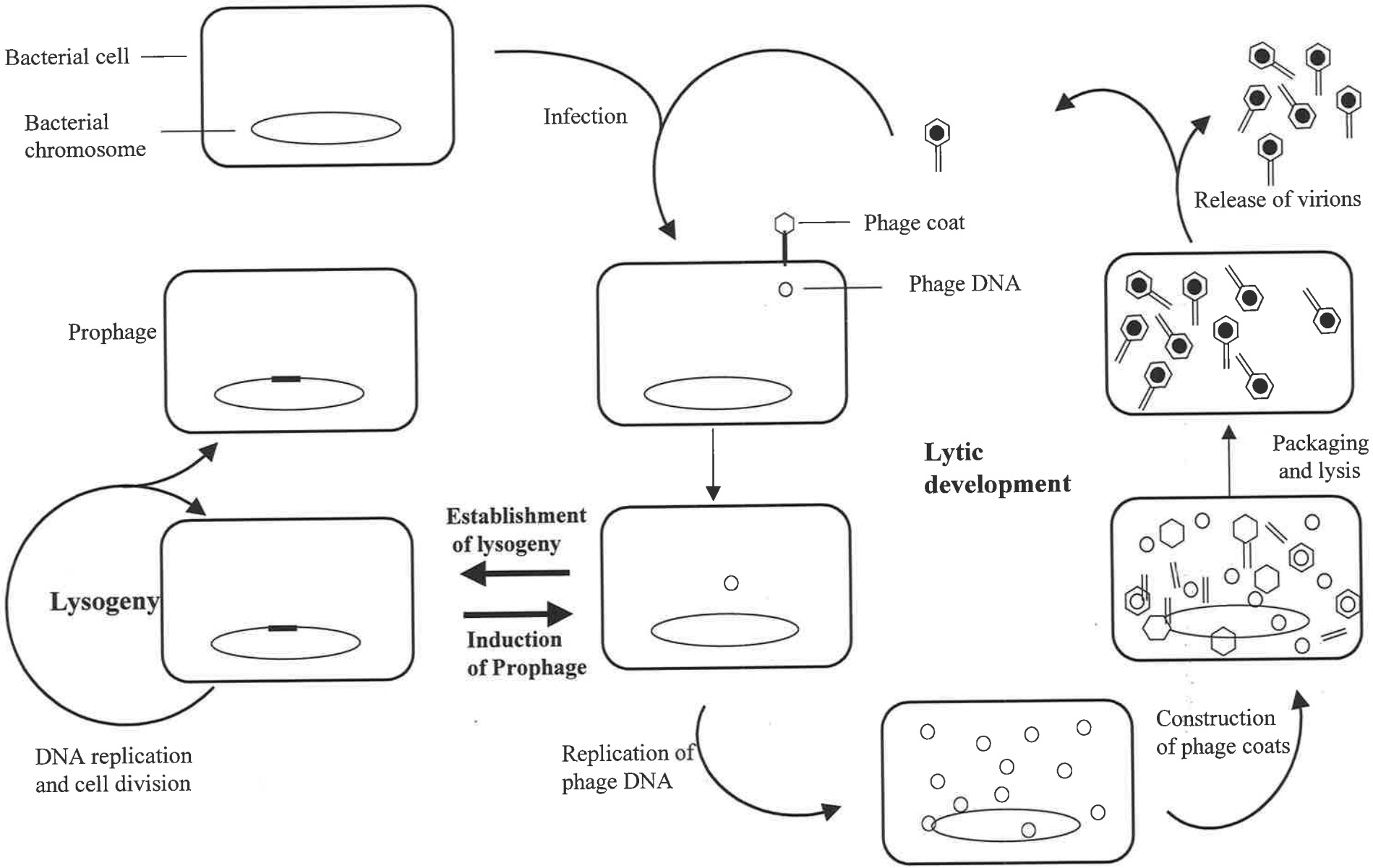

Bacterial cell
Bacterial chromosome
Infection
Phage coat
Phage DNA
Release of virions
Prophage
Lytic development
Packaging and lysis
Lysogeny
Establishment of lysogeny
Induction of Prophage
DNA replication and cell division
Replication of phage DNA
Construction of phage coats

Figure 1.4

A partial genetic map of prophage P22. The start points of transcripts are shown and labelled with the names of the relevant promoters. P_{RM} and P_{RE} transcripts are assumed to terminate between P_L and *c2*. P_L, P_R and P_{mnt} 3' termini are unknown as indicated by the dotted lines. Small genes are indicated by filled (black) blocks, while other genes are shown by arrows, indicating their direction of transcription. Adapted from (Poteete, 1988) and (Vander Byl and Kropinski, 2000)

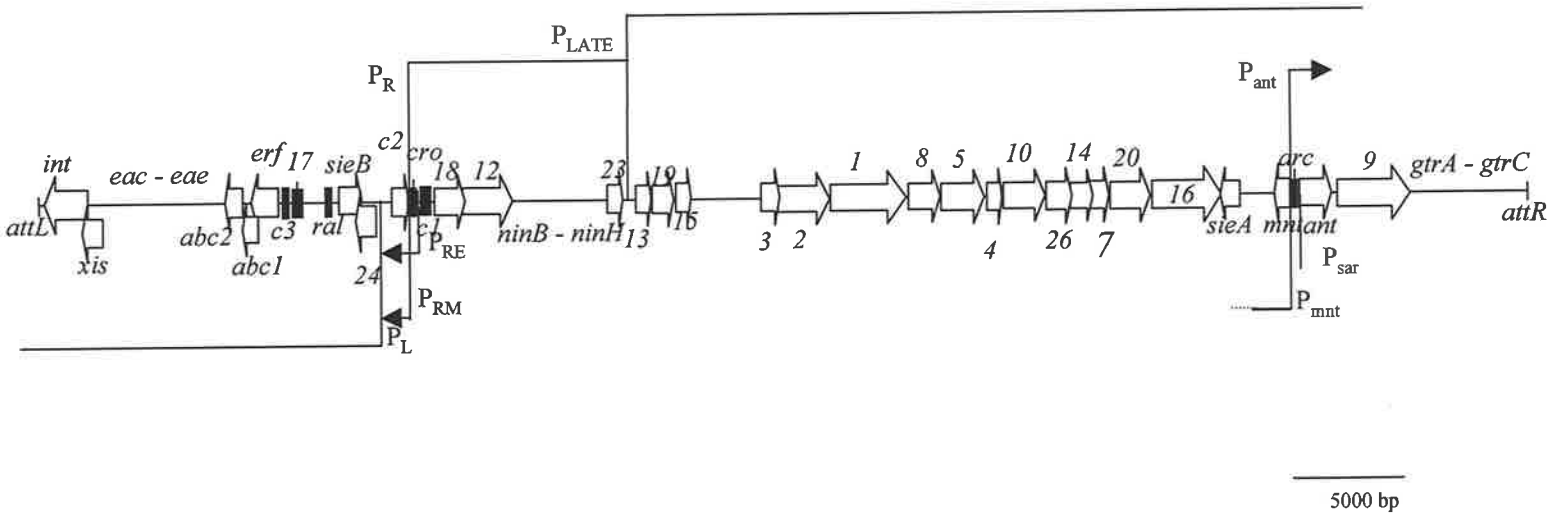

P_LATE
P_R
P_ant
int
eac - eae
erf
17
sieB
c2
cro
18
12
23
19
1
8
5
10
14
20
arc
9
gtrA - gtrC
attL
xis
abc2
abc1
c3
ral
24
c1
P_RE
ninB - ninH
13
1b
3
2
4
26
7
16
sieA
mnt
ant
attR
P_sar
P_RM
P_mnt
P_L
5000 bp

Figure 1.5

A. Sequential DNA packaging of P22–like bacteriophages by the headful packaging mechanism, generated by rolling–circle replication. Packaging is initiated from a *pac* site (which is required only for the start of a series of packaging steps) located within gene *3* in P22. The second round of packaging begins where the first round ends, producing a phage DNA in which a different sequence is repeated. This ultimately generates a molecule with identical sequence at the beginning and at the end (~1.6 kb in P22) leading to terminal redundancy. The terminal ends are used after infection to circularise the molecule by homologous recombination. Consequently, a population of DNA molecules with a terminal redundancy, starting and ending at different points in the genome is produced. However, the gene order and function is conserved and the end result is one circular bacteriophage genome.

Sequential DNA packaging

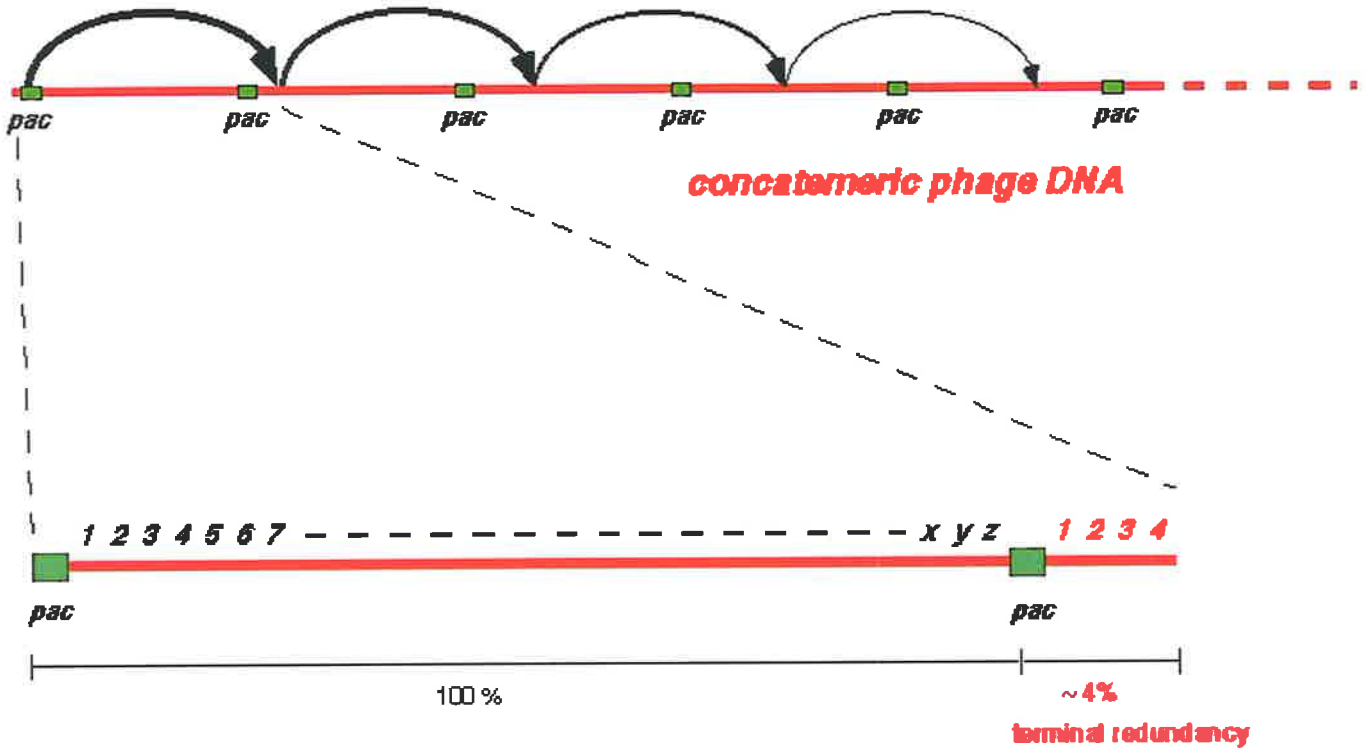


Figure 1.6

Bacteriophage P22 assembly adapted with permission from Prof. J. King's website (http://web.mit.edu/King-lab/www/research/Barrie/P22_asspathB.html).

Assembly Pathway of Bacteriophage P22

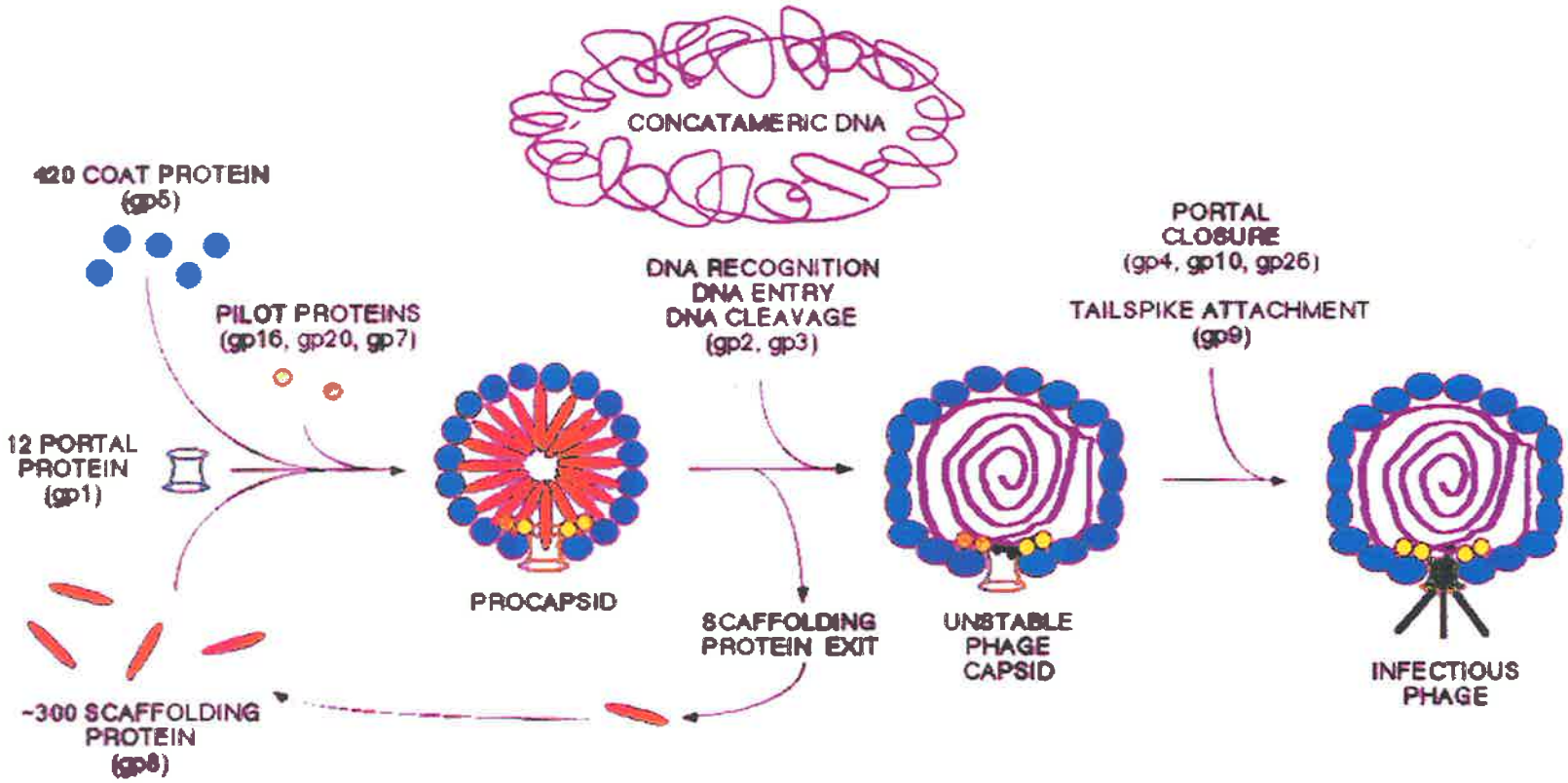


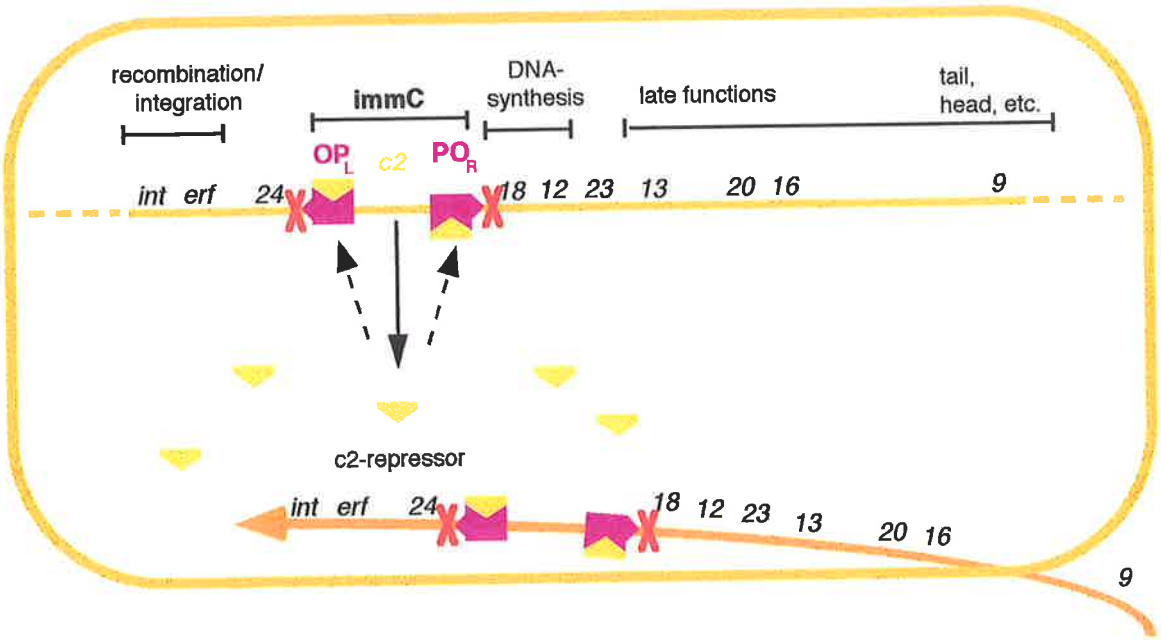
Figure 1.7

A. Superinfection by a homoimmune bacteriophage. The phage is unable to propagate because it possesses the same *imm*C region.

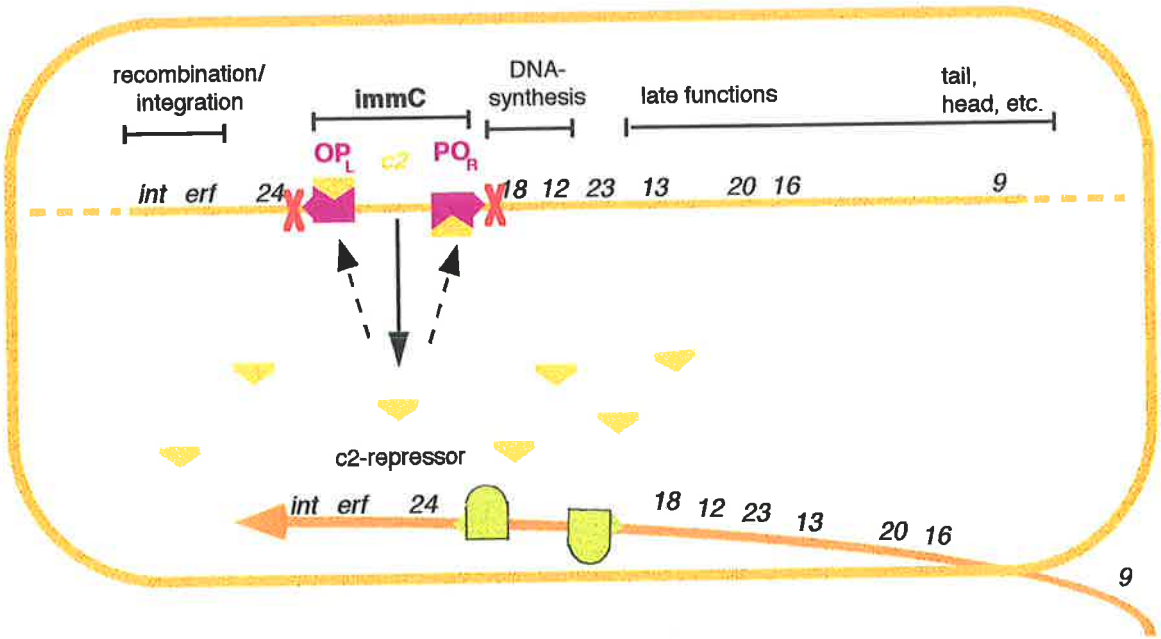
B. Superinfection by a hetereoimmune bacteriophage. The phage propagates because of a different *imm*C region.

C. Superinfection by a homoimmune phage with an antirepressor. The phage is able to propagate because of the presence of an antirepressor which inactivates the repressor of the prophage.

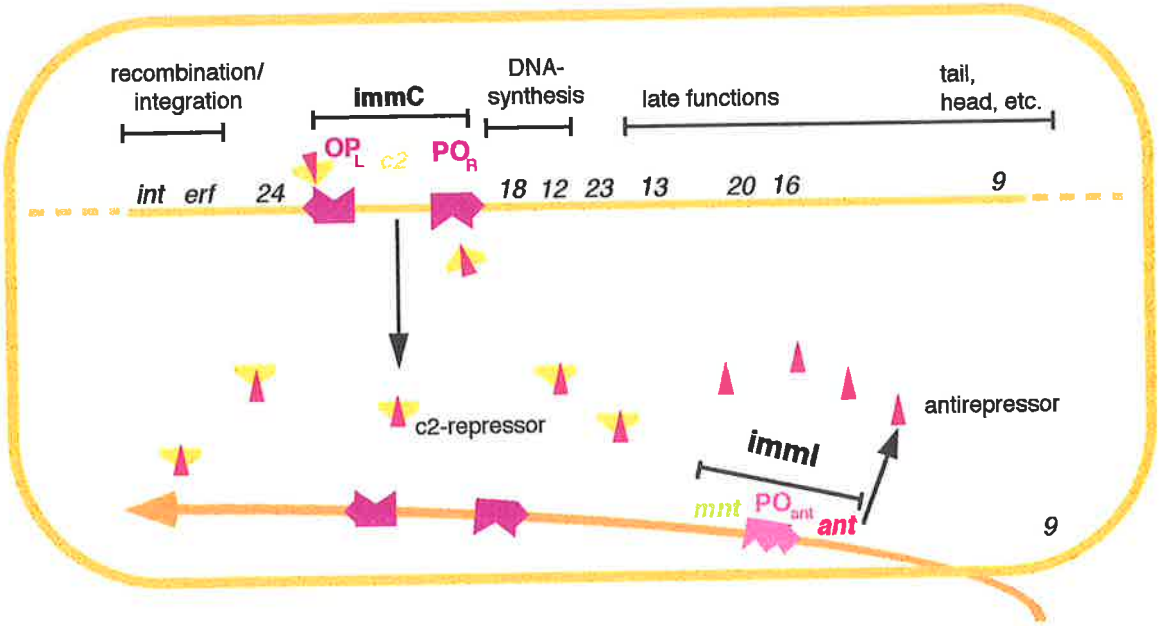
D. Superinfection by a homoimmune bacteriophage with an antirepressor on a lysogen (prophage with an antirepressor). Propagation is not possible as in A, because of the same immunity regions.



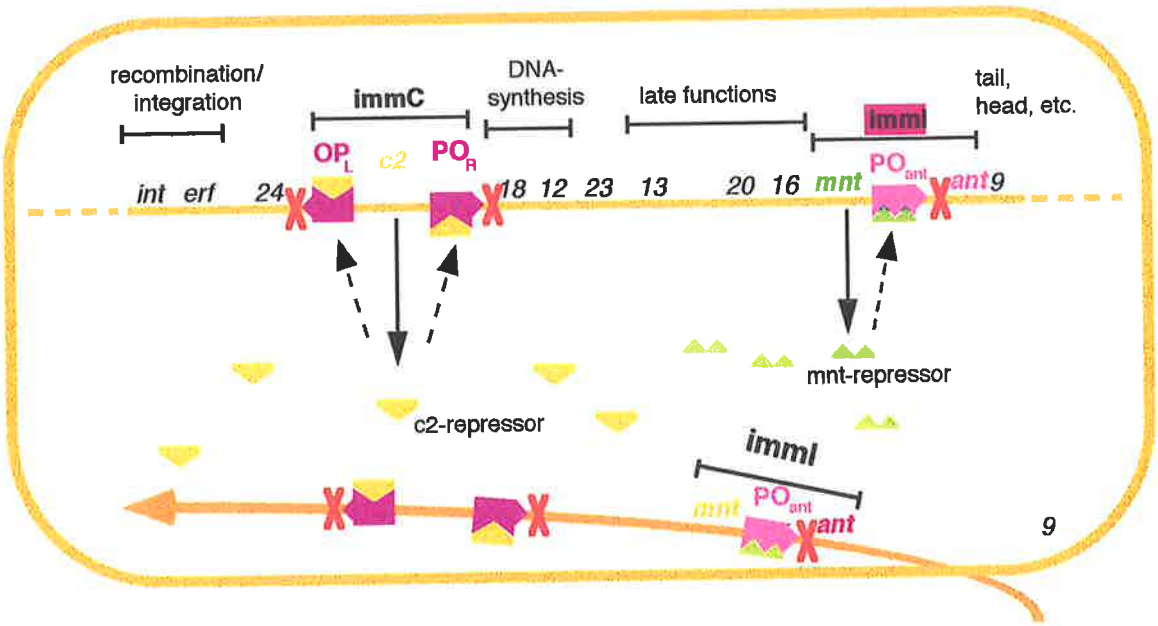
homoimmune superinfecting phage: propagation not possible



heteroimmune superinfecting phage: propagation possible



homoimmune phage with antirepressor: propagation possible



homoimmune phage and prophage with antirepressor: propagation not possible

Figure 1.8

The mechanism of prophage integration and excision of bacteriophage λ. The phage and bacterial attachment sites are denoted *POP'* and *BOB'* respectively. The prophage is flanked by new attachment sites denoted *BOP'* for *attL* and *POB'* for *attR*. Reproduced from (Adams, 1959).

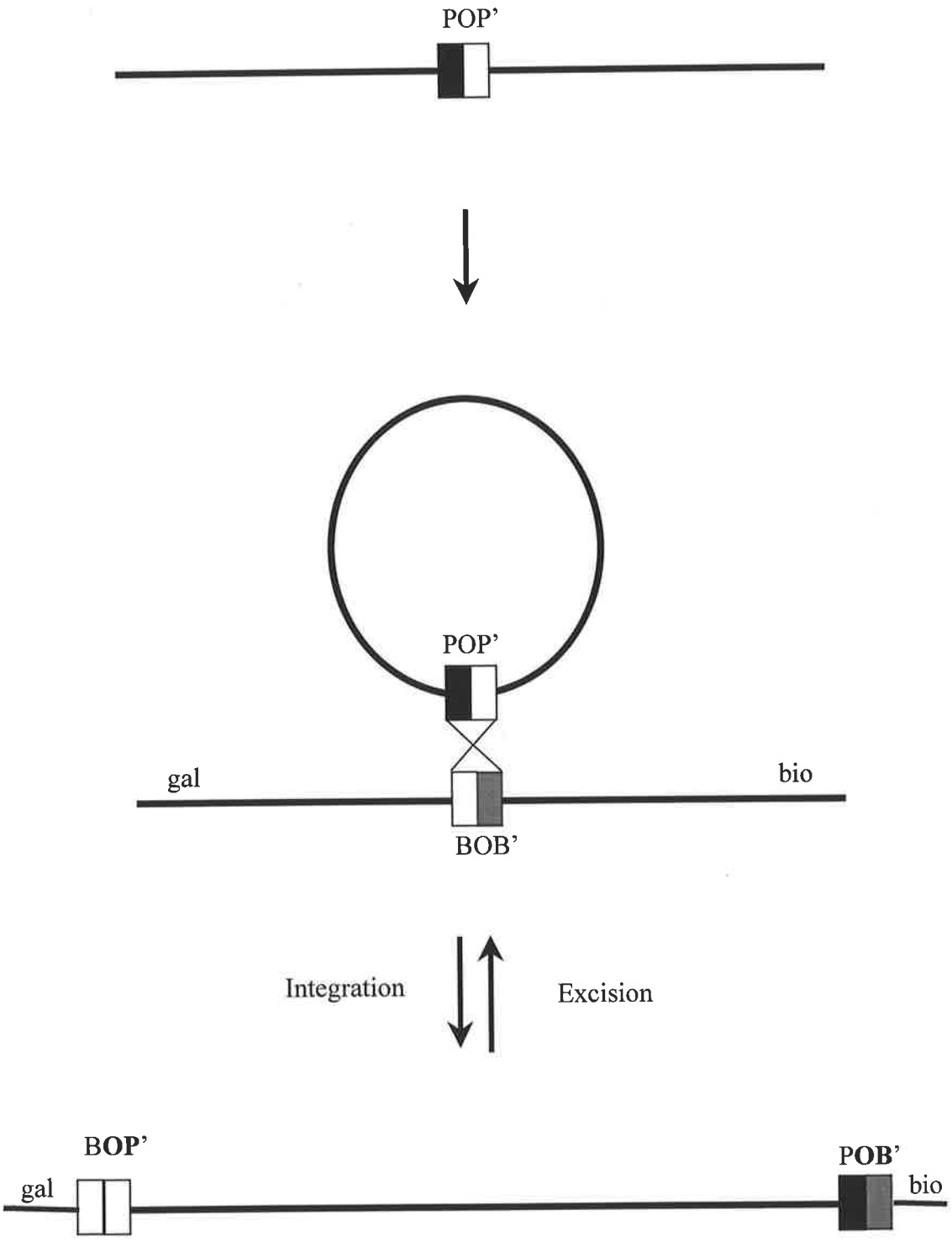

POP'
POP'
gal
bio
BOB'
Integration
Excision
BOP'
POB'
gal
bio

Table 1.1 Survey of the characteristics of several microbial typing methods

Typing method[a]	Typeability	Reproducibility	Discriminatory power	Ease of performance	Ease of intepretation	General availability	Cost
Phenotypic							
Antimicrobial susceptibility	Good	Good	Poor	Excellent	Excellent	Excellent	Low
Manual biotyping	Good	Poor	Poor	Excellent	Excellent	Excellent	Low
Automated biotyping	Good	Good	Poor	Good	Good	Variable	Medium
Serotyping	Variable	Good	Variable	Good	Good	Variable	Medium
Bacteriophage typing	Variable	Fair	Variable	Poor	Poor	Excellent	Medium
MLEE	Excellent	Excellent	Good	Good	Excellent	Variable	High
Genotypic							
Plasmid profiles	Variable	Fair	Variable	Fair	Good	Excellent	Medium
Plasmid REA	Variable	Excellent	Good	Good	Excellent	Excellent	Medium
Chromosomal REA	Excellent	Variable	Variable	Good	Fair	Variable	Medium
Ribotyping	Excellent	Excellent	Good	Good	Good	Variable	High
PFGE	Excellent	Excellent	Excellent	Good	Good	Variable	High
AFLP	Excellent	Good	Excellent	Good	Fair	Low	High
PCR	Excellent	Fair	Excellent	Good	Fair	Good	Medium
DNA sequencing	Optimal	Excellent	Excellent	Poor	Excellent	Low	High

Data obtained from (Farber, 1996, Maslow, *et al.*, 1993, van Belkum, *et al.*, 2001)

[a]MLEE, Multilocus Enzyme Electrophoresis; REA, Restriction Endonuclease Analysis; PFGE, Pulsed–Field Gel Electrophoresis; AFLP, Amplified–Fragment Length Polymorphism; PCR, Polymerase Chain Reaction

Chapter 2

General materials and methods

2.1 Growth media

2.1.1 Solid media

The following solid media were used for bacterial cultivation. *Salmonella* isolates were grown on Xylose–Lysine–Desoxycholate (XLD) medium (Yeast extract 0.3%, L–Lysine HCl 0.5%, Xylose 0.38%, Lactose 0.75%, Sucrose 0.75%, Sodium desoxycholate 0.1%, Sodium chloride 0.5%, Sodium thiosulphate 0.68%, Ferric ammonium citrate 0.08%, Agar 1.25%), Nutrient double strength agar (Difco agar 1.3%, Difco nutrient broth 2%, Sodium chloride 0.85%) and nutrient slopes (Oxoid agar number 3 0.2%, oxoid columbia agar base 3.9%). Mueller-Hinton agar (MH) (30% beef infusion, 1.75% casamino acids, 0.15 % starch and 1.7% Bacto agar) was used for PT conversion studies. Media were purchased from Medvet Science Pty. Ltd., Adelaide, South Australia. Minimal agar (M9–Agar) (10% 10 × M9 salt stock solution [15% Na_2HPO4 × 12 H2O, 3% KH_2PO4, 0.5% NaCl, 1% NH_4Cl], 1% of 1M $MgSO_4$ and 0.1% of 20% glucose, 1.5% agar) and Luria–Bertani medium (1% Oxoid tryptone L42, 0.5% Oxoid yeast L21, 1% NaCl and 1.5% agar) were used for transduction and phage propagation. Selection of plasmid pGEM-7Zf(-) containing bacteriophage ST64B or ST64T DNA inserts transformed into *E. coli* DH5α cells was achieved on Columbia blood agar base (CA) (1% Bacto pantone, 1% Bacto Bitone, 0.3% tryptic digest of beef heart, 0.1% corn starch, 0.5% NaCl and 1.5% Bacto agar) containing 0.007 mM X-gal and 0.1 mM IPTG.

2.1.2 Liquid media

The following liquid media were used for bacterial cultivation. Luria-broth (L-broth) (1% Oxoid tryptone L42, 0.5% Oxoid yeast L21, and 0.5% NaCl) was the general growth medium for all strains. SOC medium containing 2% Bacto tryptone, 0.5% Bacto yeast, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose was used to maximise recovery of *E. coli* transformants following electroporation (Dower, *et al.*, 1988). "Terrific broth" consisting of 1.2% Bacto tryptone, 2.4% Bacto yeast extract, 0.5% glycerol, 1.15% KH_2PO_4 and 6.25% K_2HPO_4 was used for growth of bacterial strains of *E. coli* containing plasmids for extraction (Sambrook, *et al.*, 1989). Brain Heart Infusion (BHI) broth consisting of 1.25% Calf brain infusion solids, 0.5% Beef heart infusion solids, 1% Proteose peptone, 0.2% Glucose, 0.5% Sodium chloride and 0.25% Disodium phosphate was used for growth of *Salmonella* isolates for Pulsed- field Gel Electrophoresis (PFGE).

2.1.3 Antibiotics

Antibiotics were added where appropriate to broth or solid media during cloning experiments to select for or maintain, transformants: ampicillin (Ap) 100 µg/ml and spectinomycin 40 µg/ml. For transduction of resistant markers, 25 µg/ml of tetracycline (Tet) was added.

2.2 Chemicals and reagents

The following AnalaR grade chemicals were used. Calcium chloride was purchased from Ajax Chemicals, NSW, Australia. Butanol, caesium chloride (CsCl), chloroform, ethanol, ethidium bromide, ethylene-diamine-tetra-acetic-acid (EDTA) disodium salt, glacial acetic acid, formamide, glucose, glycerol, glycine, hydrogen peroxide, hydrochloric acid (HCl), iso-amyl alcohol, magnesium chloride ($MgCl_2$),

methanol, magnesium sulphate (Mg_2SO_4), β-mercaptoethanol, phenol, potassium acetate, periodic acid, potassium chloride (KCl) di-potassium hydrogen orthophosphate (K_2HPO_4), ammonium chloride (NH_4Cl,) potassium di-hydrogen orthophosphate (KH_2PO_4), polyethylene glycol $(PEG)_{6000}$, PEG_{8000}, propan-2-ol (isopropanol), sucrose, sodium citrate, sodium dodecyl sulphate (SDS), di-sodium hydrogen orthophosphate (Na_2HPO_4), sodium chloride (NaCl), 4-chloro-1-naphthol, sodium lauroylsarcosine (sarkosyl), Tris(hydroxymethyl) aminomethane and trichloroacetic acid were purchased from BDH Chemicals (Victoria, Australia) or Sigma Chemical Company (St. Louis, MO, US).

Guanidinium isothiocyanate was purchased from Gibco BRL (Gaithersburg, MD, US). Skim milk powder was from Carnation, Australia. Bovine serum albumin was obtained from the Commonwealth Serum Laboratories, Melbourne, Australia. Antibiotics (ampicillin, kanamycin, spectinomycin and tetracycline) were purchased from Boehringer Mannheim GmbH (Mannheim, Germany). Isopropyl-β-D-thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal), Digoxigenin labelling mix, 10 × hexanucleotides, nitroblue tetrazolium chloride (NBT), and brom-4-chlor-3-indolyl-phosphate toluidine salt (X-P) were purchased from Boehringer Mannheim.

Electrophoresis grade reagents were obtained from the following companies as indicated: acrylamide (National Diagnostics, Atlanta, Georgia, US), ammonium persulphate (Bio-Rad, Hercules, California) and N,N,N,N,-tetramethylethylenediamine (TEMED) (Sigma), agarose (Bio-Rad or Progen Industries, Qld, Australia). Low melt point preparative agarose and bromophenol blue were obtained from Bio Rad. Coomassie brilliant blue R-250 for staining PAGE gels was purchased from Sigma.

For PCR, the four deoxyribonucleotide triphosphates (dATP, dTTP, dGTP, dCTP) were purchased from Pharmacia biotech (Quarry Bay, Hong Kong). Adenosine-5'-triphosphate sodium salt (ATP) was purchased from Boehringer Mannheim.

2.2.1 Enzymes

Lysozyme, RNase A and DNase I were obtained from Boehringer Mannheim. Klenow enzyme and T_4 DNA ligase were obtained from Pharmacia or New England Biolabs (Beverly, MA). Proteinase K was obtained from Merck chemicals (Darmstadt, Germany). Restriction endonucleases were purchased from Amersham, Boehringher Mannheim, New England Biolabs or Pharmacia and were used with the appropriate buffer from the same supplier. AmpliTaq DNA polymerase and associated supplied buffers: Buffer II (100 mM Tris and 100 mM KCl) and 25 mM $MgCl_2$ were purchased from Roche (Mannheim, Germany). DynaZyme^{TM} EXT DNA polymerase and associated supplied buffers: 10 x Mg^{2+} – free buffer (50 mM Tris–HCl, 15 mM $(NH_4)_2SO_4$, 0.1% Triton X–100) and 50mM $MgCl_2$ were purchased from Finnzymes (Finland). The XL PCR reagents were obtained from Applied Biosystems, Roche (New Jersey, USA): r*Tth* DNA Polymerase XL, 10 mM dNTP Blend (2.5 mM each dNTP), 3.3 x XL Buffer II, 25 mM $Mg(OAc)_2$ solution and control reagents, control template λ DNA 1µg/ml, XL control primers SC1002 and SC1011 at 25µM each.

2.2.2 Oligonucleotide primers

Synthetic oligonucleotides (primers) were synthesised using reagents purchased from Applied Biosystems or Ajax Chemicals (acetonitrile). Synthesis was performed on an Applied Biosystems 381A DNA synthesiser by the Division of Molecular Pathology, IMVS, Adelaide. To prepare oligonucleotides for experimental use, an aliquot (100 µl) was extracted with butanol (1 ml) to remove residual salts, centrifuged (14,000 x g) for 10 minutes, dried *in vacuo* and resuspended in sterile purified water (100 µl) (Sawadogo and Van Dyke, 1991). Alternatively, purified synthetic oligonucleotides were purchased from Geneworks, Australia. All primer stocks were stored at –20°C

2.3 Bacterial strains and plasmids

Bacterial strains used in this thesis are listed in Table 2-1 and Table 2–2. pGEM-7Zf(-) (cloning vector, ampicillin resistant purchased from Promega), pGB2 (a low copy number plasmid derived from *Escherichia coli* plasmid pSC101 and resistant to spectinomycin 50 μg/ml and streptomycin 100 μg/ml) (Churchward, *et al.*, 1984) and pCL1920 (a low copy number plasmid containing a 580 bp *Bst*UI fragment carrying the *lac* promoter/operator, multiple cloning sites and a *lacZ* fragment of pUC19, resistant to spectinomycin 50 μg/ml and streptomycin 100 μg/ml (Lerner and Inouye, 1990).

Bacterial strains were maintained in snap freeze medium at minus 70°C. Fresh cultures were prepared by removal of a small portion of the frozen stock using a sterile glass pasteur pipette and inoculating onto the appropriate solid media with or without antibiotic selection as required.

2.4 Extraction of chromosomal DNA

Chromosomal DNA was extracted using one of the following two procedures:

Method 1: Genomic DNA extraction was prepared by the method of Dr Renato Morona, University of Adelaide (personal communication). *Salmonella* cultures were inoculated into LB broth (10 ml) and incubated overnight at 37°C. The cultures were pelleted at 3 500 rpm in a Heraeus Christ (model GL) centrifuge for 15 min at room temperature. The pellet was resuspended in 3 ml of 0.85% saline. Proteins were removed from the suspension by extraction with an equal volume of phenol and the tubes were vigorously mixed by vortexing for two minutes at 30 sec intervals and centrifuged again as above. The aqueous layer containing DNA was removed into a clean centrifuge tube and an equal volume of isopropanol was added to the tube to precipitate chromosomal DNA. The precipitated DNA was apparent as a white cloud-like suspension in the isopropanol and was removed using a sterile pipette tip into an

microfuge tube containing 1 ml of 70% (v/v) ethanol. The ethanol was discarded, DNA dried *in vacuo* for 10 minutes and resuspended in 800 µl 1 x TE solution and stored in 4°C.

Method 2: Alternatively, genomic DNA was prepared by the rapid extraction method of Saunders *et al.* (1990). An overnight culture of *Salmonella* having an approximate volume of 100 – 150 µl, was removed from MH plates using a sterile disposable 10 µl inoculating loop (Disposable Products Co., SA. Australia) and placed into a sterile microfuge tube containing 1 ml of sterile water. The cells were pelleted at 17,000 x g for 3 min, resuspended in 200 µl of 2 mg/ml of lysozyme (made up in water) then incubated for 15 min at room temperature. The suspension was lysed by the addition of 400 µl of 5 M guanidinium isothiocyanate in 100 mM EDTA, pH 7.0 by gentle mixing. Proteins were removed from the suspension by extraction with an equal volume of phenol, chloroform, isoamylalcohol mixture (25:24:1). The suspension was mixed vigorously for 30 secs to emulsify and centrifuged at 17, 000 x g for 5 min to separate the phases. Chromosomal DNA in the aqueous phase was precipitated by the addition of 0.56 vol of isopropanol. The precipitated DNA was apparent as a white cloud-like suspension in the isopropanol and was removed using a sterile pipette tip into an microfuge tube containing 80% ethanol. The DNA was washed twice in 80% ethanol by gentle mixing and aspiration. The DNA was dried *in vacuo*, resuspended in 1 x TE solution and stored at 4°C.

2.5 Extraction of Plasmid DNA

Plasmid DNA was isolated by one of the two following procedures:

Method 1: Large-scale plasmid purification was performed by the three step alkali lysis method (Garger, *et al.*, 1983). Cells from a 500 ml culture of *E. coli* (Terrific broth) were harvested at 4,500 x g for 15 min at 4°C and resuspended in 12 ml of

solution 1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). Freshly prepared lysozyme (4 ml of 20 mg/ml in solution 1) was mixed with the cell suspension and incubated at room temperature for 10 min. Addition of 27.6 ml of solution 2 (0.2 M NaOH, 1% (w/v) SDS), followed by a 5 min incubation on ice slurry resulted in total lysis of the cells. After the addition of 14 ml of solution 3 (5 M potassium acetate, pH 4.8) and incubation on ice for 15 min, precipitated protein, chromosomal DNA and high MW RNA were removed by centrifugation (4,500 x g for 15 min at 4°C). The supernatant was then extracted with an equal volume of a TE saturated phenol, chloroform, isoamylalcohol mixture (25:24:1) at 4,500 x g for 15 min at 4°C. Plasmid DNA from the aqueous phase was precipitated with an equal volume of 100% isopropanol at room temperature for 20 min and collected by centrifugation (11, 000 x g for 30 min at 4°C). After washing in 70% (v/v) ethanol (11, 000 x g for 20 min at 4°C), the pellet was dried *in vacuo* and resuspended in 1.6 ml of 1 x TE. Plasmid DNA was further purified from contaminating protein and RNA by centrifugation on a two step CsCl ethidium bromide gradient according to (Garger, *et al.*, 1983). The DNA, CsCl, ethidium bromide mixture was prepared by mixing the 1.6 ml of DNA in 1 x TE with 2.910 g of solid CsCl and 300 µl of 10 mg/ml of ethidium bromide. Final volume of the mixture was made up to approximately 3 ml with 1 x TE solution. The mixture was divided between two 4.7 ml polyallomer tubes filled with 3.2 ml of less dense CsCl in TE solution (1.470 g/ml, Refractive index 1.3780). The tube was filled to the top with the less dense CsCl solution then centrifuged at 372, 000 x g for 3 h at 25°C. The plasmid DNA band was removed by side puncture of the tube with a 19-gauge needle attached to a 1 ml syringe. The ethidium bromide was extracted using isoamylalcohol. CsCl was then removed by dialysis overnight (twice) against 5 litres of 1 x TE at 4°C. Plasmid DNA was stored at 4°C.

Method 2 (small–scale plasmid DNA): Small-scale plasmid purification was performed by the three step alkali lysis method using a modification of Garger, *et al.*

(1983). Overnight bacterial cultures (4 ml) were transferred to a sterile 10 ml centrifuge tube (Disposable Products Co., SA. Australia) and harvested by centrifugation at 3 500 rpm for 10 min in a Heraeus Christ Labfuge (model GL). The pellet was resuspended in 200 µl of solution 1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA) containing 2mg/ml lysozyme. The cells were lysed by the addition of 400 µl of solution 2 (0.2 M NaOH, 1% (w/v) SDS) with gentle mixing followed by a 10 min incubation on ice to lyse the cells. After the addition of 300 µl of solution 3 (5 M potassium acetate, pH 4.8) the suspension was incubated on ice for 5 min. Precipitated protein, chromosomal DNA and high MW RNA were then pelleted by centrifugation (17,000 x g for 5 min). The supernatant was transferred to a fresh tube and extracted once with an equal volume of a Tris saturated phenol, chloroform, isoamylalcohol mixture (25:24:1). Plasmid DNA from the aqueous phase was precipitated with an equal volume of 100% isopropanol at room temperature for 20 min and collected by centrifugation 17, 000 x g for 20 min. After washing in 70% (v/v) ethanol (17,000 x g for 5 min), the pellet was dried *in vacuo* and resuspended in 40 µl of sterile distilled water.

2.5.1 Additional purification of small–scale plasmid DNA

Briefly, plasmid DNA (20 µl) was treated with RNAse A (0.1 mg/ml) for 30 min at 37°C, cooled on ice (1 min), centrifuged at 17,000 x g (1 min) and the supernatant precipitated with 20 µl of 7.5 M NH_4Ac and 40 µl of isopropanol. After centrifugation for 15 mins at 17,000 x g, the pellet was washed once in 75% ethanol – 25% 50mM NaAc and once with 100% ethanol. The pellet was dried *in vacuo* and resuspended in 10 µl of purified water.

2.6 Induction of phages

In order to release phages from putative lysogens, the suspected lysogens were grown from an overnight culture at 37°C in Luria Bertani (LB) broth with shaking for 3 hours or until they reached an OD_{550} of 0.3. Mitomycin C was then added at a concentration of 1 µg/ml and incubated for a further 20 hours with vigorous shaking as described by Gemski, *et al.*, (1978) and Yee, *et al.*, (1993). Chloroform was added to the lysate at a final concentration of 1% (w/v) and incubated a further 15 min with vigorous shaking to lyse the residual bacteria. Bacteriophages were harvested by centrifugation at 11,000 x g for 20 minutes at 4°C.

From this stage, two different methods were used:

Method 1: Whole phage lysate was obtained using this method. To the filtrate, 1 µg/ml of DNase and 5 µg/ml of RNase were added and incubated at 37°C for 30 min with shaking. An equal volume of 20% PEG_{6000}/2M NaCl in SM buffer (0.58% NaCl, 0.2% $MgSO_4$.7H2O, 5% 1 M Tris–HCl pH 7.5 and 0.5% of 2% gelatin solution) was added and incubated on ice slurry for 60 min. Phage particles were recovered by centrifuging at 14 000 x g at 4°C for 45 min. The pellet was resuspended in SM buffer 10 ml of SM buffer per 1 litre of original culture. To release DNA from phage particles, 0.1 volume of both 10% SDS and 0.5 M EDTA (pH 8) were added to the particles and incubated at 68°C for 15 min. An equal volume of phenol-chloroform was then added to the mixture and vortex mixed for 20 sec, then centrifuged at 17 000 x g for 5 min at room temperature and the aqueous layer containing DNA extracted. DNA precipitation was achieved by the addition of 0.1 volume of 3M sodium acetate (pH 5.2) and 0.6 volume isopropanol or 100% ethanol and incubated at –20°C for 1 hour or overnight, respectively. The DNA was pelleted by centrifugation at 11 000 x g for 20 min at 4°C and the pellet washed in 70% (v/v) ethanol at 17 000 x g for 10 minutes at room temperature to remove salts. The pellet was left to air dry for 10 min at room temperature and was resuspended in 1 x TE buffer or purified water.

Method 2: Purification of phage particles was carried out according to Miller, (1987) and Brown, (1991). Briefly, 1 µg/ml of DNase and 5 µg/ml of RNase were added to the lysate followed by incubation at 37°C for 30 min, to remove bacterial nucleic acids. Solid NaCl (40 g per litre) was dissolved gently into the phage lysate followed by the addition of 140 g PEG_{6000} per litre and dissolved by gentle mixing at room temperature. The lysate was then left overnight at 4°C to precipitate, centrifuged at 11 000 x g for 20 min and the supernatant discarded. The pellet was resuspended in 16 ml of SM buffer per litre of original stock. Chloroform (0.5%) was added to the tube and stirred on a magnetic shaker at 37°C for 2 hrs to recover most of the phage particles, centrifuged at 15 000 x g for 20 minutes and the supernatant saved for further analysis at 4°C.

2.6.1 Caesium chloride gradient centrifugation (CsCl)

The amount of solid CsCl added to the lysate was calculated from the weight of the volume of the solution using the formula 45.5/54.5 x weight of solution = g CsCl, to give a density of 1.5 g/ml. After the CsCl had dissolved, the density was checked by weighing 100 µl of the solution which gave 0.15 g, otherwise, more CsCl was added to bring the entire solution to 1.5 g/ml. However, if the solution was too dense, more SM buffer was added to bring the solution down to the correct density (Miller, 1987).

Bacteriophage lysate (2.5 to 3 ml) was added into 4.7 ml polyallomer tubes and the tube filled with the CsCl stock of the same density to the top and was subjected to CsCl gradient centrifugation at 100 000 x g for 24 hrs at 4°C. Phage particles which were clearly evident in two well separated bands in DT 64 (a single band was observed from DTs 9 and 135) were removed by side puncture of the tube with a 19-gauge needle attached to a 1 ml syringe and dialysed in 2 *l* of 1 x TE at 4°C overnight. The lysates were used for transmission electron microscopy, bacteriophage conversion and transduction studies.

2.6.2 Propagation of bacteriophage ST64T

Phage particles were further purified after CsCl centrifugation according to Maniatis *et al.* (1982). Briefly, 10–fold serial dilutions of bacteriophage stocks were prepared in SM buffer and 0.1 ml of each dilution was added to a tube. An aliquot of an indicator bacterium (0.1 ml) DB21 (see Table 2–2 for genotype) was added to the tube containing the phage and incubated at 37°C for adsorption to occur. Melted 0.7% agar was added to the tube and poured onto an LB plate containing hardened agar. After 16 h incubation at 37°C, a turbid plaque was picked, added to 1 ml of SM buffer with 1 drop of chloroform and left standing at room temperature for 2 h to allow phage particles to diffuse out of the agar. The lysate was then stored at 4°C in SM buffer until required.

2.6.2.1 Preparation of phage stocks from single plaques

From the phage stock, 1 in 20 of a resuspended plaque was mixed with 0.1 ml of an overnight culture of DB21, incubated at 37°C for 15 min, added 3 ml of 0.5% top agar, mixed and poured onto a freshly made LB agar plate. The plates were incubated for 8–12 h at 37°C. The soft top agar was scraped off into a sterile centrifuge tube when confluent lysis was achieved. To rinse off the remaining top agar from the plate, 5 ml of SM buffer was added to the plate then added to the tube with top agar together with 0.1 ml of chloroform. After mixing by rotation for 15 min at 37°C, the tubes were centrifuged at 4, 000 x g for 10 min at 4°C. The supernatant was recovered and chloroform added to 0.3% and the stock stored at 4°C. The final titer in most cases was approximately 10^{10}pfu/ml.

2.6.3 Extraction of phage DNA from purified particles

Extraction of phage DNA was prepared according to (Brown, 1991). To the dialysed phage lysate, 0.5 M EDTA was added to a final concentration of 20 mM.

Proteinase K (Boeringer Mannheim) was added to a final concentration of 50 μg/ml after which 10% SDS was added to a concentration of 0.5% (w/v). The tubes were incubated at 65°C for 1 hour, left to cool at room temperature. Proteins were removed by extraction with an equal volume of phenol, then phenol–chloroform and lastly chloroform. For each extraction with phenol, phenol–chloroform and chloroform, the suspension was mixed vigorously for 30 sec and centrifuged at 17,000 x g for 5 min in order to separate the phases. Phage DNA in the aqueous phase was precipitated by the addition of 0.1 vol of 3 M sodium acetate (pH 7) and 100% ethanol and incubated at 4°C overnight. The tube was centrifuged at 11 000 x g for 20 min at 4°C and the pellet washed in 70% (v/v) ethanol at 17 000 x g for 10 min at room temperature to remove salts. The pellet was left to air dry for 10 min resuspended in 1 x TE buffer or purified molecular water.

2.6.4 PEG precipitation purification

Whole phage DNA was further purified for sequencing, following the instructions in the Perkin Elmer Dye Terminator Sequencing Cycle Ready Reaction Kit Instructions (PM 402078). Briefly, 32 µl of bacteriophage DNA was mixed with 8 µl 4 M NaCl followed by 40 µl of 13% autoclaved PEG_{8000}. The samples were incubated at –20°C overnight after vigorous mixing. Precipitated DNA was then centrifuged at 25,000 x g for 30 min at 4°C, washed in 500 µl 70% (v/v) ethanol, centrifuged at 17 000 x g for 5 min and dried *in vacuo*. The pellet was resuspended in 20 µl of water and stored at –20°C.

2.6.5 Transmission Electron Microscopy (TEM)

A crude bacteriophage lysate with a titre of $>10^9$ pfu/ml derived from *S.* Typhimurium DT 64 was centrifuged at 30,000 x g for 60 min at 4°C (Miller, 1987). Pellets were resuspended and washed twice with 0.1 M ammonium acetate (Bradley,

1967). Discrete bacteriophage purified bands from the top and bottom bands of CsCl gradient were also prepared. The crude lysate and purified phages were deposited on copper coated grids, stained with 2% phosphotungstic acid at pH 7.2 or uranyl acetate (2%, pH 4.5) and viewed under a Philips EM 300 (Laval University, Quebec) or 100 electron microscope at 60kV (The University of Adelaide, South Australia).

2.7 Analysis and manipulation of DNA

2.7.1 DNA quantitation

The concentration and purity of DNA in solutions was determined by measurement of absorption using an UV1101 Biotech photometer with wavelengths of 260 nm and 280 nm. The concentration was calculated using the following standards: absorption of 1.0 at A_{260} is equal to 50 µg of double stranded DNA/ml or 20 µg/ml for oligonucleotides (Sambrook *et al.*, 1989).

2.7.2 Restriction endonuclease digestion of DNA

Digestion of DNA with restriction enzymes was performed using the supplied buffer for the enzyme with or without the addition of BSA as specified by the manufacturer's instructions. For restriction digestion of chromosomal DNA, 5-10 µg was incubated with the appropriate enzyme buffer and 20-40 U of each restriction enzyme in a final volume of 50-100 µl at 37°C or 25°C, as specified for a particular enzyme, and incubated overnight. For digestion of plasmid or phage DNA, an aliquot of 0.5-1 µg (unless specified otherwise) was incubated with the appropriate buffer and 4 U of enzyme and incubated for 1-2 hours at the appropriate temperature. For some methods, the restriction digest was terminated by heating at 95°C for 10 min.

Prior to loading onto an agarose gel for electrophoresis, a one tenth volume of tracking dye (15% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 0.1 mg/ml RNase A) was added to the sample or aliquot.

2.7.3 Agarose gel electrophoresis of DNA

Electrophoresis of DNA samples was performed at room temperature on horizontal, 1% or 2% (w/v) agarose gels. Gels were elecrophoresed in a Pharmacia model GNA-200 tank (Pharmacia Biotech Asia Pacific, Ltd., Quarry Bay, HK.) or a BRL model H5 tank (BRL, Gaithersburg, MD, USA). Gels were run in 1 x TAE buffer (40 mM sodium acetate, 40 mM Tris and 2 mM EDTA). After electrophoresis the gels were stained in distilled water containing approximately 2 mg/ml ethidium bromide and DNA was visualised with ultraviolet light on a Model TM-36 transilluminator (UVP, Inc., San Gabriel, CA, USA) and photographed using Polaroid 667 positive film. The sizes of restriction enzyme fragments, PCR products or phage bands were estimated by comparing their relative mobility with that of *Eco*RI digested *Bacillus subtilis* bacteriophage SPP-1 DNA (Geneworks, Thebarton, SA, Australia). The sizes of the fragments of DNA in this commercially available preparation used were: 8.557, 7.427, 6.106, 4.899, 3.639, 2.799, 1.953, 1.882, 1.515, 1.412, 1.164, 0.992, 0.710, 0.492, 0.359 and 0.081 kilobases (kb). Alternatively, Hi–Lo mixed DNA markers (Geneworks, Therbaton, SA, Australia) were used. These markers are a combination of equal amounts of three DNA digests; λ/*Hind*III, λ/*Pvu*I and SPP–I/*Eco*RI resulting in a range of commercially available DNA fragments of the following sizes: 23.130, 14.321, 12.712, 11.936, 9.533, 9.416, 8.510, 7.350, 6.557, 6.110, 4.840, 4.361, 3.590, 2.810, 2.322, 2.027, 1.950, 1.860, 1.510, 1.390, 1.160, 0.980, 0.720, 0.564, 0.480, 0.360, 0.125 kb.

2.7.4 Purification of DNA fragments

Bio-Rad low-gelling-temperature agarose at a concentration of 1.0% (w/v) was used for separation of restriction fragments required for cloning or labelling purposes. The specific fragment/s were recovered by the following method. The DNA band of interest was excised from the gel and the agarose melted at 65°C for 10 min. The sample was then vortexed vigorously with equal volume of Tris saturated phenol. Residual phenol was removed with chloroform and the DNA precipitated with two volumes of ethanol and one-tenth volume of 3 M sodium acetate, pH 5.2. DNA was collected by centrifugation (17, 000 x g for 25 min), washed once with 70% (v/v) ethanol and dried *in vacuo* before being resuspended in purified water. Alternatively, Qiagen gel extraction kit was used according to manufacturer's instructions. Purified DNA was stored at –20°C until required.

2.8 *In vitro* cloning

DNA fragments to be cloned were combined with appropriately digested vector DNA in a ratio of approximately 3-1 or 2-1. The DNA fragments were ligated with 5-10 U of T_4 DNA ligase in ligase buffer (20mM Tris-HCL, pH 7.5, 10mm $MgCl_2$, 10mM dithiothreitol and 0.6 mM ATP) in a final volume of 20 µl and incubated overnight at 4°C. The ligated DNA was purified prior to transformation of *E. coli* strains as follows. An equal volume of Tris saturated phenol, chloroform, isoamyl alcohol (25.24.1) was mixed with the ligation reaction then centrifuged at 17,000 x g for 5 min. The aqueous phase was precipitated in two volumes of ethanol and 0.1 vol of 3 M sodium acetate, pH 5.2. DNA was collected by centrifugation (17,000 x g for 25 min), washed once with 70% (v/v) ethanol and dried *in vacuo* before being resuspended in 10 µl of purified water.

2.8.1 Electroporation

Electrocompetent *E. coli* cells were prepared using the method of Dower, *et al.* (1988). *E. coli* DH5α was made competent for electroporation with plasmid DNA as follows: an overnight shaken culture in (LB broth) was diluted 1:20 into LB broth (250ml) and incubated for 3 h at 37°C with gentle agitation at 150 rpm (A_{600} of 0.6, ca. 4 x 10^8 cells/ml). The log phase culture was chilled on ice for one hour and then harvested at 4, 500 x g for 20 min at 4°C. The pellet was washed twice in 250 ml of ice-cold purified water and once in 5 ml of ice-cold 10% glycerol. The final pellet was resuspended in 800 to 1,000 µl of sterile 10% glycerol and stored at −70°C in 40 µl aliquots for further use.

Electroporation was performed according to the method described by Dower, *et al.* (1988). Purified ligated DNA or CsCl gradient prepared plasmid DNA (total volume 10 µl) was added to the bottom of a chilled 0.2 cm electroporation cuvette (Bio Rad) containing 40 µl of ice-thawed electrocompetent cells. Electroporation of the cells/DNA suspension was performed using the Bio-Rad Gene Pulser with capacitance extender (Model No. 1652078) using the following settings: 200 ohms resistance, 25 µF capacitance and 2.5 kV pulse strength. Following electroporation, approx. 1 ml of SOC medium was added immediately and the cells gently resuspended and transferred to a sterile universal container (Disposable Products Co., SA, Australia). The cell suspension was incubated for 1 h at 37°C with shaking. The culture was then plated onto selection plates directly or concentrated by centrifugation and plated.

2.9 Southern hybridisation

2.9.1 Preparation of Digoxigenin labelled DNA

Labelling of DNA fragments with Digoxigenin-dUTP was performed according to the manufacturers instructions (Boehringer Mannheim). PCR products were purified using a

QIAquick PCR purification kit (Qiagen, Germany) according to the manufacturer's instructions and eluted in 30µl of elution buffer prior to labelling. Purified DNA fragments, restriction digests (10-15 µl) or SPP1 molecular weight marker (1 µg) (up to 3 µg DNA total for each) were labelled in a total volume of 20 µl. Mineral oil was layered over the reaction mix and incubated overnight at 37°C. The labelled DNA was added to 10 ml of hybridisation fluid with 5 µl of labelled SPP1 marker and stored at –20°C. High stringency hybridisation fluid consisted of 50% (v/v) formamide, 7% (w/v) SDS , 1% (w/v) skim milk powder, 5 x SSPE, (1 x SSPE is 0.18M NaCl, 10mM sodium phosphate, 1mM EDTA) and 2.5 mg/ml salmon sperm DNA. Low stringency hybridisation fluid consisted of 18% (v/v) formamide, 5 x SSC (1 x SSC is 0.15M sodium chloride plus 0.015M sodium citrate pH 7.0), 1% (w/v) skim milk powder, 7% (w/v) SDS and 2.5 mg/ml salmon sperm DNA (Cianciotto, *et al.*, 1990). This protocol should allow for the detection of sequences with 10% bp mismatch.

2.9.2 Southern blot

Southern blot analysis was performed according to the method of Southern (1975). Briefly, ethidium bromide stained agarose gel was soaked for 8 to 10 min in 250 mM HCl. The gel was then washed twice in 500 ml of denaturing solution (1.5 M NaCl, 0.5 M NaOH) and twice in 500 ml of neutralising solution (0.5 M Tris, pH 7.4: 1.5 M NaCl). DNA bands from the gel were allowed to transfer by capillary action onto Hybond N^+ membrane (Amersham Pharmacia Biotech, Buckinghamshire, UK) for 18 h at 25°C in 10 x SSC.

2.9.3 Hybridisation and development

Nylon membrane filters were soaked in 5 x SSC solution after transfer and the DNA cross-linked onto nylon under ultraviolet illumination (wavelength approximately

302 nm) for 2 min. The nylon filter was then probed with the appropriate low or high stringency hybridisation fluid, containing the Digoxigenin labelled probe, and incubated at 42°C for 18 hours. For high stringency conditions, unbound Digoxigenin labelled DNA was removed by washing the nylon membrane filter twice in 2 x SSC plus 0.1% (v/v) SDS for 10 minutes at room temperature, followed by 2 washes in 1 x SSC plus 0.1% (v/v) SDS for 25 minutes at 68°C. For low stringency conditions, the membrane was washed twice in 5 x SSC with 0.1% SDS for 10 min at 37°C, followed by 2 washes in 2 x SSC with 0.1 SDS at room temperature.

The filters were then developed according to the manufactures protocol with the exception that the filter was blocked with 5% skim milk in Buffer 1 (100 mM Tris HCl, pH 7.5, 150 mM NaCl) for 90 min at room temperature.

2.9.4 Colony hybridisation

Colony blot hybridisation was performed essentially using the method of Paton, *et al.* (1992). Briefly, colonies to be screened were grown in a 96-well microtitre tray overnight and then centrifuged at 1 500 rpm for 15 min in a Hermle Z300 centrifuge. The supernatant was discarded and the pellet resuspended in 10µl of 1 x TE buffer. After emulsifying the pellet, 5 µl of 10% SDS was added followed by 50 µl of a solution containing 0.5 M NaOH and 1.5 M NaCl. The lysate was transferred to a nylon membrane (Hybond N+) for hybridisation.

2.10 Polymerase chain reaction (PCR)

2.10.1 Standard PCR

Each PCR reaction mix contained 200 µM of each nucleotide dATP, dGTP, dTTP, and dCTP, 20 pmol of each primers, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5

mM $MgCl_2$, 2 U of AmpliTaq DNA polymerase and test DNA (50-100 ng) in a total reaction volume of 50 µl. The amplification protocol varied especially with the annealing temperature, depending on the primer set being used. Reactions were amplified in a Corbett Research Thermal Sequencer Model FTS-960 (Sydney, Australia) and reactions were analysed by electrophoresis on 1% or 2% (w/v) agarose gels and the size of the products estimated by comparing with lanes containing either SPP-1 or Hi–Lo DNA markers.

2.10.2 Extra long (XL) PCR

Occasionally, XL PCR was used for the amplification of sections of ST64T and ST64B large *Sma*I fragments (30,712 bp and 23,650 bp, respectively) for sequencing. It consisted of three layers: the lower mix, the wax and the upper mix. The lower mix comprised of 1x of 3.3x XL Buffer II, 800 µM dNTPs, 1 mM $Mg(OAc)_2$ and 20 pmol of each primer in a total of 20µl. An ampliwax bead (Ampliwax PCR Gem, Perkin Elmer) was added and melted for 5 min at 80°C, left to cool so as to harden the wax. The upper mix comprising of 1x of 3.3x XL Buffer II, 2 – 4 U rTth DNA polymerase XL and sample DNA (50 – 100 ng) in 30µl. The total reaction mixture was 50 µl.

The amplification protocol varied with the annealing temperature, depending on the primer set being used. For amplification of an approximately 12 kb fragment, the amplification protocol consisted of 16 cycles of 1 min at 94°C, 15 sec at 94°C, 5 min at 56°C, another 12 cycles with 15 sec increments of 15 sec at 94°C and 5 min at 60°C. This was followed by a 6 min interval at 72°C and a soak at 4°C. Reactions were amplified in a Corbett Research Thermal Sequencer Model FTS-960 (Sydney, Australia) and analysed by electrophoresis on 1% (v/w) agarose gels and the size of the products estimated by comparing with lanes containing Hi–Lo DNA markers.

2.11 DNA sequencing

The nucleotide sequence of clones was determined by dideoxy chain–termination using the PRISM™ Ready Reaction Dye Primer cycle Sequencing Ready Reaction M13 Kit (Perkin Elmer). Nucleotide primers flanking the cloned insert (M13 forward and reverse primers) were used to sequence the beginning of the insert. The nucleotide sequence of the remainder of the inserts, as well as PCR amplified products (after the amplicons were purified using QIAquick PCR Purification Kit, Qiagen GmbH, Hilden, Germany) was determined by primer walking with dideoxy chain–termination chemistry using the PRISM™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing Kit (FS or Big Dye) (Perkin Elmer). For the large bacteriophage DNA fragments, sequencing was achieved by using PEG_{8000} precipitated whole bacteriophage DNA. This was achieved by sequencing the ends of the cloned bacteriophage DNA fragments using whole bacteriophage genome as a template to determine the fragment order, with primers designed to read outward from the ends of the fragments, thus reading into the adjacent fragments. When sequence from the large *Sma*I fragments (30,712 bp for ST64T and 23,650 bp for ST64B) was determined, sequencing subsequently proceeded by primer walking using dideoxy chain–termination chemistry as above. Sequencing reactions were purified using ethanol/sodium acetate or isopropanol precipitaion according to the manufacturer's recommendations. DNA sequencing was then performed on an ABI 373A or 377 DNA sequence analyser (Applied Biosystems, USA) in the Division of Molecular Pathology, IMVS, Adelaide.

2.11.1 DNA sequence analysis

The nucleotide sequence was corrected and stripped of poor quality data using Chromas, version 1.62 and assembled into contigues using GeneCompar 2.0 and/or

GeneBase 1.0 (Applied Maths, Kotrijk, Belgium). A compendium of online tools from the University of Adelaide, Department of Molecular Biosciences, Discipline of Microbiology and Immunology, South Australia (http://www.microbiology.adelaide.edu.au/links/index.html; and Professor Andrew Kropinski's online Tools, University of Queens, Department of Microbiology and Immunology, Canada (http://www.queensu.ca/micr/faculty/kropinski/online.html) were used for analysis. Open reading frames (ORFs) were analysed using graphical ORF finder at the University of Adelaide web page and WebGene Mark.HMM (Lukashin and Borodovsky, 1998) (http://genemark.biology.gatech.edu/GeneMark/whmm.cgi). Putative ORFs were scanned for homologues by using BlastP (Altschul, *et al.*, 1990, Altschul, *et al.*, 1997), against the non–redundant GeneBank database (http://www.ncbi.nlm.nih.gov/blast/blast.cgi). Molecular weigths (MW) and isoelecteric points (pI) were determined at ProtParam tools (http://www.expansy.ch/tools/protparam.html). The sequences were compared using ClustalW at Bioedit or at the European Molecular Biology Laboratory–European Bioinformatics Institute (http://www.ebi.ac.uk/clustalw/) and OMIGA 1.1 from Oxford Molecular Group (Campbell, California) was employed for comparative hydropathy plots, using the method of Kyte-Doolittle (Kyte and Doolittle, 1982).

Proteins were scanned against Prosite and Protein families (Pfam) databases for conserved motifs at the Swiss Institute for Experimental Cancer Research Profile Scan server (http://www.ch.embnet.org/software/PFSCAN_form.html). Prediction of transmembrane helices in proteins was performed using TMHMM (Sonnhammer, *et al.*, 1998) at the centre for Biological Sequence analysis at the Technical University of Denmark (http://www.cbs.dtu.dk/services/TMHMM-1.0/).

For further DNA sequence analysis, different programs were employed including, GeneCompar 2.0 and/or GeneBase 1.0 (Applied Maths, Kotrijk, Belgium) used to search for inverted and direct repeats as well as restriction sites. For putative tRNA scanning,

tRNAscan–SE (http://www.genetics.wustl.edu/eddy/tRNAscan–SER/) (Eddy and Durbin, 1994) and FASTtRNA (http://bioweb.pasteur.fr/seqanal/interfaces/fastrna.html) (El-Mabrouk and Lisacek, 1996) were employed. Potential IHF–binding sites were scanned using MacTargsearch at SEQSCAN (http://www.bmb.psu.edu/seqscan/seqform1.html). For promoter regions, Martin Reese's Promoter Prediction by Neural Network program (http://www.fruitfly.org/seq_tools/promoter.html) was employed.

2.12 SDS polyacrylamide gel electrophoresis (PAGE)

The protocol for SDS PAGE was modified from the method of Laemmli, (1970). The stacker gel contained 4% acrylamide in 2 mM Tris, pH 6.8 and 0.001% SDS. The separating gel was 12% acrylamide in solution containing 11 mM Tris, pH 8.8, and 0.6% SDS. The gel was run on a Hoeffer Model SE-600 tank (Pharmacia) at 15 mA for 18 to 24 h. All bacteriophage particle samples were mixed with an equal volume of sample buffer (6.25 mM Tris pH 6.8, 2% SDS, 10% glycerol, 5% β–mercaptoethanol) (Lugtenberg, *et al.*, 1975) and heated at approximately 100°C for 5 min prior to loading.

Benchmark^{TM} pre–stained protein ladder (Gibco Life Technologies, CA, USA) containing proteins with apparent molecular weights, 172.6, 111.4, 79.6, 61.3, 49.0, 36.4, 24.7, 19.2, 13.1 kDa, was electrophoresed as a marker track for molecular size estimation.

PAGE gels were stained with 0.275% Coomassie brilliant blue R-250 in 10% (v/v) ethanol, 10% (v/v) methanol and 7.5 % glacial acetic acid for a minimum of 30 min. Gels were de-stained to visualize protein bands by gentle agitation in 10% (v/v) ethanol, 10% (v/v) methanol and 7.5 % glacial acetic acid.

2.13 Bacterial typing methods

2.13.1 Serotyping of *Salmonella* isolates

This procedure was performed by the Australian Salmonella Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide, South Australia, according to the Kauffmann and White scheme (Kauffmann, 1954). Different polyvalent sera were used to detect which serovar one is investigating. For example, polyvalent D was positive if the serovar was *S.* Typhi. For *S.* Typhimurium, polyvalent B containing O–antigens 4 and 5 was used. Positive results using this polyvalent showed that the serovar was Typhimurium. O–antigens 12, 27 and 1 were individually tested to complete the O–antigen group of a Typhimurium serovar. The flagella antigens (H phase 1 and 2) of the isolate were similarly tested. To perform this test, a single colony under test was emulsified on a slide with a drop of saline. A specific polyvalent sera or single antigen was added to the slide and the slide was rocked for 10 seconds. A positive result was shown by agglutination.

2.13.2 Bacteriophage typing

For *S.* Typhimurium, bacteriophage typing was performed using the Anderson scheme of 31 phages (Anderson, *et al.*, 1977). Briefly, cultures to be typed were subcultured onto dry 5 ml nutrient agar slopes and incubated overnight at 37°C. Nutrient double strength agar (NA D/S) plates were also incubated at 37°C overnight with lids on. The following day, the NA (D/S) plates were dried inverted without lids for 1 hr at 37°C before being used. Two NA (D/S) plates were used for *S.* Typhimurium to cover the 31 phages used. For each freshly grown culture on nutrient agar slope, 5 ml of nutrient broth was inoculated and incubated on a shaker for 90 min at 37°C. A pasteur pipette was used to flood the dried NA D/S plates with the broth culture and extra broth was removed from the plates. The plates were inoculated with phages as soon as they were dry (within 10 –

15 min after being flooded with the broth culture). Plates were incubated overnight at 37°C and phage patterns were then read. All negative readings were confirmed with a 10 x magnifying hand lens. Phage typing was done in the same way for all other *Salmonella* serovars except that only one NA (D/S) plate was required.

2.13.3 Pulsed-field gel electrophoresis (PFGE)

Method 1: PFGE was performed according to (Maslow, *et al.*, 1993b). A single colony of *S.* Typhimurium culture was inoculated into 10 ml BHI broth and incubated at 37°C overnight. Low melt point agarose (1.3% w/v) was prepared in PIV buffer (10 mM Tris (pH 7.6), 1M NaCl) and melted by steaming for 25 min. Molten agar (1 ml) was added to the tubes in the heating block at 50°C. The plug mold was chilled on ice in the refrigerator at 4°C. Cold PIV buffer (5 ml) was dispensed into sterile tubes on ice and 3 ml of overnight cultures were dispensed into the tubes with PIV buffer and centrifuged at 1 700 x g for 15 min at 4°C. The supernatant was decanted and 1.5 ml cold PIV buffer was added to the pellet and placed on ice. The pellet was quickly resuspended by vortex mixing and adding 1 ml of the pellet to the tube holding 1 ml of molten agarose on the heating block, followed by a brief vortex mix and filling of the plug mold. The mold was kept on ice and placed in a refrigerator for 30 min at 4°C to completely solidify. To a total of 4 ml of lysis buffer (6mM Tris pH 7.6, 1 M NaCl, 100 nM EDTA pH 7.6, 0.2% sodium deoxycholate and 0.5% sodium lauroyl sarcosine), 80 µl of a 50 mg/ml stock lysozyme and 8 µl of a 10 mg/ml RNAse were added to make up a fresh lysis solution. To each labelled 5 ml screw cap serum tubes, 4 ml of lysis solution was added. The plug mold was opened after 30 min and plugs pushed into their respective tubes with a sterile plastic loop and incubated overnight at 37°C with gentle agitation at 80 rpm.

The tubes were chilled for 15 min to harden the plugs and the lysis solution aspirated. To each tube, 4 ml ESP solution ([ES buffer: 0.5 M EDTA pH 8.0, 10%

sodium lauroyl sarcosine] and 100 µl of proteinase K per ml of ES buffer) was dispensed into each tube and the tube incubated all day at 50°C with gentle agitation in a waterbath. Plugs were chilled at the end of the day for 15 minutes, fresh ESP solution added to the tube and the tube incubated overnight at 50°C. The plugs were again hardened by chilling on an ice slurry, decanted into a small weighing tray, 2–3 mm piece of plug cut from the end of the plug and placed into a 1 ml microfuge tube containing 1 ml of 1 x TE buffer to rinse the excess ESP solution. The first 1 x TE buffer was discarded and 1 ml of fresh 1 x TE added to the plug. The plugs were washed 4 times by gentle shaking in a 37°C incubator or waterbath for 2 h for the first two washes and 1 h each for the last two washes. After removing the last 1 x TE buffer, 100 µl restriction enzyme mix was added to the tube with a plug and incubated at the appropriate temperature. Restriction endonuclease *Xba*I was predominantly used with *Salmonella* strains whereas *Sma*I was used for *Staphylococcus aureus* control strains. A 1% Pulsed-field gel was prepared with CHEF grade agarose in 0.5 x TBE (1 x TBE = 0.089 M Tris–borate, 0.089 M boric acid, 0.002 M EDTA), allowed to set at room temperature for 30 min and left at 4°C overnight to chill.

A 1% low melt point (LMP) agar was prepared in 0.5 x TBE and cooled down to 65°C. Approximately 1.6 *l* of cold 0.5 x TBE buffer was poured into a CHEF tank and cooled to 14°C. After overnight incubation of the plugs with an appropriate restriction enzyme, the enzyme mix was removed from each plug and the plug tipped into a weighing tray. The plug was then lifted with a disposable loop and allowed to slide into the well. Molecular weight markers were included in the outer lanes and one in the centre of the gel to allow accurate sizing of bands and reliable comparisons of patterns between different gels. The wells were then covered with a few drops of molten LMP agarose and allowed to set for 10 min.

Either CHEF DR–II or CHEF DR–III apparatus was used. The gel was loaded into the tank according to the manufacturer's instructions. For CHEF DR–II, after the

chiller reached 14°C, the pulse wave generator was turned on in the following manner: initial time 1 sec, final time 20 sec, start ratio 1.0 and run time 22.5 h. The power supply was then turned on and set at 200V for 22 h. For CHEF DR–III, the power was turned on and the chiller set at 14°C, and the run parameters entered in the following manner: initial switch time 1 sec, final switch time 20 sec, run time 22 h, volts/cm 6, included angle 120°. The current range was 140–150 mA. After electrophoresis, the gel was stained in EtBr for 30 min and photographed under UV light illumination.

Method 2: An adaptation of a rapid PFGE method by Matushek, *et al.* (1996), which reduced the procedure from 6 to 3 days was shown to work well with *Salmonella* isolates. Briefly, *Salmonella* isolates were subcultured in BHI broth as the method above. Plugs were prepared as described in the method above. Plugs were placed in screw–capped serum tubes containing 3 ml of 1 x lysis solution (recipe as above) with 1 x lysozyme and incubated at 37°C for 1–2 h with gentle shaking. The lysis solution was replaced with 3 ml of ESP solution, incubated at 50°C in a shaking water bath for 2 x 1 h with a change of ESP after 1 h. The ESP solution was removed and the plugs transferred to 10 ml centrifuge tubes with 7 ml of 1 x TE and incubated at 50°C with one change of 1 x TE. The plug was then chilled and stored in 5 ml 1 x TE in serum tubes until required. Approximately, 2 mm fragments of plugs were digested in 200 µl of digest mix for 2 h at 37°C for *Xba*I and the gel was loaded and electrophoresis carried out as described above. Staining of the electrophoresed gel was carried out on day 3.

2.13.4 Amplified fragment length polymorphism (AFLP)

This method was adapted from the AFLP Microbial Fingerprinting kit (PE Applied Biosystems, Perkin Elmer, Norwalk, Conn.). Briefly, approximately 200 to 300 ng of genomic DNA was digested using *Eco*RI (5U) and *Mse*I (1U) (New England Biolabs,

Hertfordshire, United Kingdom [NEB]) and simultaneously ligated to specific *Eco*RI (5' –CTCGTAGACTGCGTACC– 3') (3' –CATCTGACGCATGGTTAA– 5') and *Mse*I (5'–GACGATGAGTCCTGAG– 3') (3' –TACTCAGGACTCAT– 5') adaptors (manufactured and supplied by GeneWorks Pty Ltd, Adelaide, South Australia) at room temperature overnight. The ligated fragments were then diluted 1 in 20 and amplified using specific pre–selective primers for *Eco*RI (5'- GACTGCGTACCAATTC -3') (0.5 μM) and *Mse*I (5'- GATGAGTCCTGAGTAA - 3') (0.5 μM) as supplied in the kit. The conditions for thermocycling were as follows: 2 min at 72°C, followed by 20 cycles of 94°C for 20s, 56°C for 30s and 72°C for 2 min and a final soak at 4°C. The amplifications were performed under oil in a DNA Thermal Cycler 480 (Perkin-Elmer Corp., Norwalk, Conn.). The amplicons were then diluted 1 in 20 and used as template for the selective amplification using *Eco*RI primer (*Eco*RI plus A) (5 μM) labelled with a blue fluorescent dye, 5 -carboxyflourescein and *Mse*I primer (*Mse*I plus A) (5 μM). Selective amplification was also performed in the same DNA Thermal Cycler 480 as follows: 94°C for 2 min, 94°C for 20s (30 cycles), a decreasing annealing temperature starting at 66°C in the first cycle, with a decrease of 1 degree for the next 9 cycles then remaining at 56°C for 20 cycles (30s), then 72°C for 2 min. A final extension was performed at 60°C for 30 min followed by holding at 4°C. The PCR products were loaded on a 5% denaturing polyacrylamide gel using 1 x TBE (90 mM Tris, 90 mM boric acid, 2 mM EDTA) running buffer (Long Ranger; FMC Bioproducts, Rockland, Maine) on an ABI Prism 377 automated DNA sequencer (Perkin- Elmer Corp.).

GeneScan Analysis Software v3.1 (PE Biosystems) was employed to accurately size and quantify the fragments. The gel and electropherograms were analysed using GeneScan. The electropherograms were then used in Genotyper Software v3.6NT (PE Biosystems) to accurately identify the characteristic fragments of the convertants tested, and the tabular densitometric values were formatted on a spreadsheet and transferred to GelCompar v4.1 (Applied Maths, Kortrijk, Belgium), via the MWtoGEL conversion tool

(Applied Maths, Kortrijk, Belgium). A dendrogram was generated using the unweighted pair group method (UPGMA) for clustering on a matrix based on the pairwise comparisons of the Dice coefficient.

Table 2.1 Bacterial isolates used in this study

Name of isolate	Number of isolates used	Source
S. Typhimurium DT 9[a]	11	Human, SA and NSW
	4	Chicken meat, NSW and Qld
	2	Bovine, SA
S. Typhimurium DT 64[ab]	5	Chicken litter, NSW
	11	Chicken meat, NSW
	10	Human, SA and NSW
	2	Avian liver, SA
	1	Turkey meat, NSW
	1	Meat pie, SA
S. Typhimurium DT 135[a]	14	Human, SA
	4	Chicken litter, Qld
	1	Ovine, SA
	3	Bovine, SA
S. Typhimurium DT 135a[a]	1	Human, NSW
	2	Chicken meat, NSW
S. Typhimurium DT 29[a]	4	Chicken meat, NSW
	1	Bovine intestine, SA
	3	Human, SA
S. Typhimurium RDNC/A045[a]	8	Human, SA
	1	Bovine, Turkey meat SA
S. Typhimurium DT 8[a]	2	Chicken meat, NSW
	2	Ovine liver, Qld
	1	Human, SA
S. Typhimurium DT 12a	5	Human, SA
S. Typhimurium DT 12	3	Chicken meat/litter, Qld
S. Typhimurium DT 41	3	Human SA, Avian liver NSW, Equine NSW
S. Typhimurium DT 44[a]	4	Human, SA
S. Typhimurium DT 141[a]	2	Chicken litter, NSW
S. Typhimurium DT 108[a]	3	Human, SA
S. Typhimurium DT 126[a]	2	Chicken meat, NSW
S. Typhimurium DT 197	2	Bovine, SA
S. Typhimurium DT 201[a]	3	Human, SA
S. Enteritidis PT 1	1	Chicken meat, NSW
	1	Chicken viscera, Singapore
S. Enteritidis PT 4	7	Human SA and Qld
	6	Chicken litter, Qld
S. Enteritidis PT 5a	1	Human, SA
S. Enteritidis PT 6	2	Human, Qld
S. Enteritidis PT 14	3	Human, Qld
S. Enteritidis PT 26	3	Human, Qld
S. Virchow PT 8	4	Human, NSW and Qld
	1	Meat beef, Qld
	1	Chicken meat, NSW
	1	Macedamia nuts, Qld
	1	Sewage sludge, NSW
S. Virchow PT 11	1	Chicken meat, Singapore

S. Virchow PT 34	1	Crocodile environment, Qld
	1	Chicken meat, SA
S. Virchow, PT 34	3	Human, NSW and SA
S. Virchow, PT 31	1	Human, SA
S. Bovismorbificans PT 32	2	Human, Qld
	1	Bovine intestine, SA
S. Bovismorbificans PT 14	2	Human, NSW and meat processor effluent, W
S. Chester	2	Human, SA
S. Hadar PT 10	2	Chicken layer and viscera, Singapore
S. Hadar PT 22	2	Chicken litter, Singapore
	1	Human, NSW
S. Hadar PT 2	3	Human, Qld and SA
S. Heidelberg PT 5	3	Bovine and Human, Qld
S. Heidelberg PT 15	3	Human and Porcine, Qld
S. Heidelberg PT 16	1	Human, Qld
	2	Equine, Vic
S. Heidelberg PT 1	2	Human, Qld
	1	Chicken meat, NSW
S. II Sofia	3	Chicken, SA

a: lysogenic for phage ST64B (this study), b: lysogenic for phage ST64T (this study)
SA: South Australia, NSW: New South Wales, Qld: Queensland, Vic: Victoria, WA: Western Australia

Table 2.2 Other strains used in this study

***S.* Typhimurium bacterial strains**	**Genotype**	**Reference/Source**
DB21	*S.* Typhimurium LT2, positive for all three modification restriction systems (LT, SA and SB) and lysogenic for Fels 1	(Colson and van Pel, 1974) H. Schmieger
DB21 (*sie*)	*S.* Typhimurium DB21, lysogenic for P22 *sieA sieB*, a P22 mutant defective in both superinfection exclusion systems	(Susskind, *et al.*, 1971)
HisHB22(P22)	*S.* Typhimurium LT2, *hisB22*$^{-}$ and lysogenic for P22	(Schicklmaier, *et al.*, 1998)
DB21(ES18)	*S.* Typhimurium LT2, lysogenic for ES18	(Schicklmaier, *et al.*, 1998)
DB21(Tn*10*)	DB21 *zia*-748::Tn*10*	H. Schmieger
TrpB8(P22)	*Trp–8*$^{-}$, lysogenic for P22	(Schicklmaier, *et al.*, 1998)
Q1	free of cryptic phages Fels 1 and Fels 2	B. Stocker and K. Sanderson
ATCC 14028s[a]	Virulent strain	B. Stocker and K. Sanderson
E. coli DH5α	F^{-}, ø80d*lacZ*ΔM15, *recA1*, *endA1*, *gyrA96*, *thi–1*, *hsdR17* (rk^{-}, mk^{+}), *supE44*, *relA1*, *deoR*, Δ(*lacZYA–argF*)U169	(Hanahan, 1985)

a: American Type Culture Collection, Atlanta, Ga and kindly supplied by Profs. B. Stocker and E. Sanderson, at the *Salmonella* Genetic Stock Center in Calgary.

Table 2.3 Bacteriophages used in this study

Bacteriophage	Genotype	Source or reference
ST64B	*S.* Typhimurium DT 64 temperate phage from CsCl bottom band	This study
ST64T	*S.* Typhimurium DT 64 temperate phage from CsCl top band	This study
P22	Wild type temperate phage isolated from an *S.* Typhimurium LT2 smooth strain	H. Schmieger
P22(H5)	P22 *c2* mutant	(H. Schmieger, 1999)

Table 2.4 Plasmids used in this study

Plasmid	Genotype	Source or reference
PGEM–7Zf(–)	Ap^r cloning vector	Promega, Madison, WI, USA
pGB2	Spe^r and Str^r cloning and sequencing vector	(Churchward, *et al.*, 1984)
pSCL1920	Spe^r and Str^r cloning and sequencing vector	(Lerner and Inouye, 1990)

Chapter 3

Temperate phages in *Salmonella enterica* serovar Typhimurium: Implications for epidemiology

3.1 Introduction

Salmonella enterica serovar Typhimurium has remained the most common cause of food poisoning world-wide (Kariuki, 1999). Epidemiological studies have shown that the bulk of non typhoidal *Salmonella* infections at any one time are caused by only one, or relatively few *Salmonella* strains (Threlfall, *et al.*, 1978). In any given time period a few phage types tend to dominate within a geographical region (Anderson, *et al.*, 1977). In Australia, DTs 9, 64 and 135 are amongst the most common isolates from both humans and poultry (Australian *Salmonella* Reference Laboratory, Annual Reports, Adelaide, 1996 - 1999).

For organism tracing during an outbreak of salmonellosis, in addition to serotyping, phage typing is used to differentiate isolates within the same serovar. The first system of phage typing, designated "Scheme 1", was established in 1943 (Felix and Callow, 1943), later extended to "Scheme 2" (Callow, 1959) which was then refined by Anderson (Anderson, *et al.*, 1977) and is still one of the main identification schemes used worldwide. Phage typing measures the plaquing ability of a set of bacteriophage on the test strains. Phage typing is based on the assumption that epidemiologically related strains will exhibit the same plaquing ability with the test phage and therefore phage type while unrelated strains will have different phage types. For this to be valid, the phage typing system must have a high discriminatory power in the population where the system is applied and the

different types must demonstrate high stability within the population (Hunter, 1990, Olsen, *et al.*, 1993). Although phage typing is still used widely for epidemiological purposes, several studies have shown that some phage types are able to spontaneously convert from one type to another by virtue of infection with temperate phage or plasmid (Baggesen, *et al.*, 1997, Brown, *et al.*, 1999, Chart, *et al.*, 1989, Hickman-Brenner, *et al.*, 1991, Rankin and Platt, 1995); this could potentially have serious implications for epidemiology and organism tracing.

Amplified fragment length polymorphism (AFLP) is one of the newest and promising molecular typing methods (Blears, *et al.*, 2000, Janssen and Dijkshoorn, 1996, Vos, *et al.*, 1995). It is based on the selective amplification of genomic restriction fragments by PCR in order to generate fingerprinting patterns consisting of large numbers of bands (Huys, *et al.*, 1996, Janssen and Dijkshoorn, 1996, Tamada, *et al.*, 2001, Vos, *et al.*, 1995). This method, together with other molecular methods like pulsed-field gel electrophoresis (PFGE) have shown to have high discriminative power as compared to phenotypic methods, such as phage typing.

Many studies have shown that the great majority of *S.* Typhimurium isolates are lysogenic (Callow, 1959, Figueroa-Bossi and Bossi, 1999, Schicklmaier, *et al.*, 1998, Schicklmaier and Schmieger, 1995). It was observed that more than 76.5% of *S.* Typhimurium natural isolates tested, spontaneously released phages during cell propagation (Schicklmaier, *et al.*, 1998). Furthermore, 93.5% of the testable isolates could mediate generalised transduction.

This chapter focusses on temperate bacteriophages induced by mitomycin C from *S.* Typhimurium definitive phage types (DTs) 64, 9 and 135. It supports previous studies in that *S.* Typhimurium strains are often lysogenic for one or more bacteriophages, which may be able to mediate phage type conversion. Two previously unknown temperate phages carried by *S.* Typhimurium DT 64 are described. One of these phages, ST64T, is capable of mediating both generalised transduction and phage type conversion.

3.2 Materials and methods specific to this chapter

3.2.1 Strains and phages

Isolates of *S.* Typhimurium, *Salmonella enterica* serovar Enteritidis, *Salmonella enterica* serovar Virchow, *Salmonella enterica* serovar Heidelberg, *Salmonella enterica* serovar Hadar, *Salmonella enterica* serovar Bovismorbificans, *Salmonella enterica* serovar Chester and *Salmonella enterica* II serovar Sofia used in this study were obtained from the Australian Salmonella Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide, South Australia. They are listed in Table 2–1. *Salmonella* strains and phages used in transduction and immunity tests are listed in Table 2–2. Bacteriophage typing and induction of phages were performed as explained in chapter 2. The phage types in *S.* Typhimurium are designated Definitive Phage Types (DT) whereas other phage types of *Salmonella* serovars other than Typhimurium are designated Provisional Phage Types (PT) (The International Phage Typing Scheme of Colindale) (Anderson, *et al.*, 1977). The RDNC strains show phage reaction that does not conform with phage reaction of known phage types.

3.2.2 Phage type conversion

Bacteriophage type conversion was carried out according to Harvey, *et al.* (1993). Briefly, *S.* Typhimurium DTs 9, 64, 135 and 29 were used in this study as indicator strains and to induce temperate bacteriophages. Additionally, *S.* Typhimurium DTs 41, 44, *S.* Enteritidis PT 26 and *S.* Heidelberg PTs 1, 5, and 15 were also used as indicator strains. Sensitivity to the induced phage was measured by streaking 20 µl of crude high titer lysate ($>10^9$pfu/ml) across a Mueller Hinton agar plate followed by an overnight culture of bacterial strain which was streaked at right angles to the phage streak. The plates were incubated at 37°C overnight. Colonies arising at and after the intersection of the bacterial growth with the phage streak, were further subcultured twice onto Mueller Hinton agar.

These were then phage typed according to the method outlined in chapter 2 to confirm whether phage type conversion had occurred.

3.2.3 Phage immunity studies

DB21, HisHB22(P22), DB21(*sie*) and DB21(ES18) (Table 3.2) (Coetzee, 1987, Rankin and Platt, 1995, Schmieger, 1999), were used as indicator strains and 200 µl of overnight cultures of the above were added to 3 ml of top agar and overlayed onto LB plates. After setting, 10 µl of undiluted, 10^{-2} and 10^{-4} dilutions were spotted onto overlayed plates and incubation was carried out at 37°C overnight. Phage P22 (H5) which is a clear plaque mutant, was used as a control (Schmieger, 1999).

3.2.4 Generalised transduction of phage ST64T

ST64T was propagated on strains DB21 and DB21(Tn*10*). The induced phage lysates were mixed with HisHB22(P22) (*his* deletion and wild–type P22 lysogen), Trp8(P22) (*trp* mutant and wild–type P22 lysogen) and DB21. The multiplicity of infection (moi) did not exceed 10 in each case. Adsorption was allowed to occur for 10 minutes at 37°C without shaking. After adsorption, the auxotrophs that had been infected with phage were plated onto M9 plates. In the case of the lysate from DB21(Tn*10*) and DB21 mixture, 1 ml of LB broth was added and the mixture incubated for a further 2 hours at 37°C to allow for expression of tetracycline resistance gene. Cells were then plated (100 µl) onto LB with tetracycline plates. The efficiency of transduction was determined using the method of Schmieger and Schicklmaier (1999). Briefly, a 100 µl aliquot of the adsorption mixture was added to 3 ml of top agar (0.7% oxoid Bacteriological agar, LII), the mixture was then added to a membrane filter placed onto LB agar plate without added tetracycline. The plate was incubated for 2 hours to allow integration and expression of the resistance gene. The membrane filter was then transferred to LB agar with tetracycline and

incubated at 37°C for 2 days. To test for the presence of spontaneous resistance mutants, 100 µl of DB21, a recipient strain, was cultured on LB agar plate with tetracycline. All phages propagated on different strains were assayed for sterility by spotting 10 µl of phages onto either LB or LB plus tetracycline agar plates.

3.3 Results

3.3.1 Caesium chloride gradient centrifugation

Two bluish bands containing phage particles were observed and extracted from *S.* Typhimurium DTs 64 and 29. A more intense band referred to as the bottom phage band was observed at the interface between the 1.45 and 1.50 g/ml density whereas the top phage band was observed at or above the 1.15 g/ml interface (Maniatis, 1982). However, *S.* Typhimurium DTs 9 and 135 generated only one band (at the interface between 1.45 and 1.50 g/ml density).

3.3.2 Electron microscopy

The crude bacteriophage lysate from *S.* Typhimurium DT 64 showed many isometric heads (mostly icosahedral and a few octagonal) with a diameter of about 63 nm between opposite apices and a tail of approximately 10 x 7-8 nm (Figure 3.1). The CsCl gradient purified two phages were designated ST64T and ST64B (*Salmonella* Typhimurium DT 64 Top or Bottom) to reflect their respective banding density in the CsCl gradient. Both phages had similar morphologies, i.e. icosahedral heads with a short tail. The majority of phages from the CsCl gradient bottom phage (ST64B) band showed icosahedral heads with no visible tails. Surprisingly, sequence analysis of ST64B showed many putative tail genes in contrast to one tail gene found in ST64T. These findings imply that ST64B should have a long tail, if the tail genes are expressed. This is further discussed

in chapter 5. However, phages ST64T and ST64B were indistinguishable by electron microscopy, and both appear to belong to the family, Podoviridae.

3.3.3 Restriction endonuclease analysis of ST64T and ST64B genomic DNA

Genomic DNA extracted from purified ST64T and ST64B phages and restricted with *Sma*I showed two distinct restriction enzyme digestion patterns and indicated that DNA from a crude lysate (phage lysate not run on a CsCl gradient) was a combination of both ST64T and ST64B genomes (Figure 3.2).

3.3.4 Propagation of ST64T and ST64B

It was not possible to purify ST64B from a single plaque. ST64B could not be propagated on any available strain in our collection, hence transduction and phage type conversion could only be shown with ST64T which could be propagated on different strains including DB21, DB21(*sie*), LB5000 (Table 2.1), *S.* Typhimurium DTs 9 and 135, *S.* Heidelberg PT 1, *S.* Enteritidis PT 26 and *S.* Bovismorbificans PT 32 (Table 2.2). Although the two bands containing phages ST64T and ST64B particles were well separated when run on a CsCl gradient, traces of ST64T could still be found in the CsCl purified phage ST64B particles. ST64B particles were thought to plaque on DB21, but when DNA from the purified plaque was extracted and digested with *Sma*I, it was discovered that the genome belonged to contaminating ST64T. Induced phage particles from *S.* Typhimurium DTs 9 and 135 (DTs 9 and 135 carry a prophage with identical *Sma*I restriction sites as ST64B) could not plaque on any strain tested.

3.3.5 Detection of phages ST64T and ST64B in other serovars

Digoxigenin labelled ST64T and ST64B genomic DNA were used as probes in Southern analysis to determine the presence of these phages in other *Salmonella* serovars.

The ST64B genome was found in the majority of all *S.* Typhimurium DTs tested (Figure 3.3 and Figure 3.4). In addition, incomplete ST64B sequences were also found in *S.* Typhimurium DT 12 (Figure 3.4). A complete ST64T genome was only found in *S.* Typhimurium DTs 64 and 29 (Figure 3.5). In other *Salmonella enterica* serovars, fragments hybridising to the ST64B genome were detected in Enteritidis, Heidelberg, Virchow and Sofia II. No bands hybridising to ST64B genome were found in serovars Bovismorbificans, Hadar, and Chester. However, no ST64T sequences were found in any *Salmonella enterica* serovars tested other than Typhimurium.

ST64T and ST64B were unable to propagate on the *S.* Typhimurium Q1 and ATCC 14028s strains. Southern hybridisation analysis indicated a prophage most probably related to ST64B in the *S.* Typhimurium Q1 strain and also in a virulent isolate ATCC 14028s. An incomplete ST64B phage genome was found in the HisHB22(P22) strain (Figure 3.6).

3.3.6 Phage type conversion mediated by phage ST64T

Crude phage lysates were harvested from 6 isolates of each of *S.* Typhimurium DTs 9, 64, 135 and 29 and used in phage type conversion studies (Table 3.1). A total of 150 potential convertants were detected when phage lysates induced from *S.* Typhimurium DTs 64 and 29 were used to infect *S.* Typhimurium DTs 9 and 135. To confirm whether phage type conversion had occurred, the 150 potential convertants were subcultured and phage typed. *S.* Typhimurium DT 9 was converted to DT 64 and DT 135 was converted to DT 16 (Table 3.2). To determine whether the convertants contained the ST64T or ST64B phage DNA, *Sma*I digested chromosomal DNA isolated from these convertants was subjected to Southern hybridisation analysis. Labelled *Sma*I digested ST64T and ST64B genomes were used as probes. In all cases (DT 9 to DT 64 and DT 135 to DT 16) convertants were found to be lysogens of ST64T and ST64B (Figures 3.7). To confirm that ST64T was the phage causing conversion and that it could mediate phage type conversion without ST64B acting

as a helper phage, purified phage particles of ST64T were used for conversion experiment on the same strains (DTs 9 and 135). Similar results were obtained (DT 9 was converted to DT 64 and DT 135 converted to DT 16).

To determine whether ST64T could mediate phage type conversion in other serovars and phage types, it was used to lysogenise *S.* Heidelberg PTs 1, 5, 15, 16; *S.* Enteritidis PT 26, *S.* Typhimurium DTs 41 and 44, *S.* Bovismorbificans PT 32. These isolates were originally shown to be non-lysogens of ST64T by Southern analysis. *S.* Heidelberg PT 1 was converted to PT 4, *S.* Enteritidis PT 26 was converted to PT 6a, *S.* Typhimurium PT 41 was converted to PT 29, *S.* Typhimurium PT 44 converted to an unknown typing reaction and *S.* Bovismorbificans PT 32 converted to an unknown typing reaction. Southern hybridisation analysis indicated that all converted phage types were now lysogens of ST64T (Figure 3.8).

3.3.7 Pulsed-field Gel Electrophoresis

Analysis of *Xba*I digested chromosomal DNA isolated from DTs 9, 64, 135 and 29 by PFGE showed differences between and within these phage types. However, when the phage type convertants were screened and analysed using PFGE, no detectable difference in the PFGE pattern of the parental strains and the phage type convertants was observed. Southern hybridisation analysis of the PFGE fragments on the parental DT 64, indicated that ST64B was found in a large *Xba*I (~700 kb) fragment whereas ST64T was found in both the large (~700 kb) and the second largest (~600 kb) fragments.

3.3.8 Amplified-fragment length polymorphism (AFLP) analysis of phage type convertants

Randomly selected phage type convertants (10 samples of DT 135 to DT 16) and (9 samples of DT 9 to DT 64) were subjected to AFLP analysis. *Eco*RI plus A and the *Mse*I

plus A primer combination generated 45 – 50 scorable amplified fragments evenly distributed over the 1,000 bp-marker range. Analysis of these fragment profiles using Genotyper Software highlighted the fragments characteristic of the genomic conversion (Table 3.3). Two fragments, 28.68 bp and 37.91 bp were present in the parental DT 135 but absent in all DT 16 convertants tested. Additionally, a cluster of fragments (40.53 bp, 44.60 bp, 47.39 bp and 50.13 bp), were detected in all the DT 135 to 16 convertants but absent from the parental samples tested (Figure 3.9). Other fragments indicative of variation in the genomic DNA of the convertants included 192.68 bp, 228.81 bp, 295.90 bp, 335.04 bp, 379.05 bp, 423.91 bp and 478.16 bp. A second type of DTs 135 to 16 convertants was identifiable by a unique 328.22 bp fragment present only in 6 of the 10 samples tested. All the convertants lacked 2 large fragments beyond the 1,000 bp marker, which the Genescan software was unable to size.

The DTs 9 to 64 convertants had unique fragments, 30.22 bp and 39.04 bp present in the parental DT 9 and absent in the convertant DT 64. Additionally, the DT 64 convertants had a fragment cluster consisting of the fragments 40.45 bp, 44.54 bp, 46.55 bp and 49.88 bp which were absent in the parental DT 9 (Figure 3.10). Other fragments characteristic of conversion to DT 64 included 192.44 bp, 195.88 bp, 224.57 bp, 295.85 bp, 328.23 bp, 334.93 bp, 378.96 bp, 423.92 bp, and 478.28 bp, as well as the absence of a fragment at approximately 1,040 bp.

3.3.8.1 Cluster analysis of AFLP results

Using UPGMA dendrogram, three major clusters were identified for DT 9 to DT 64 group (Figure 3.11). One cluster consisted of the parental DT 9 isolate, the second with the DT 64 original sample and the last one with all convertants. There is >92% similarity between all convertants and the original DT 64 isolate with the convertants having 95% similarity amongst themselves. The parental DT 9 isolate exhibit <88% similarity to the convertants (DT 64).

Similarly, a dendrogram was generated for the DT 135 to DT 16 group, using UPGMA (Figure 3.12). Two major clusters were evident where one consisted of the DT 135 original samples which exhibited >97% similarity to each other. The other cluster consisted of all convertants with >94% similarity to each other. The difference in similarity between the convertants and the parental DT 135 cluster was significant at <88%. The two dendrograms clearly demonstrate the conversion from the parental DT 135 to DT 16 as well as DT 9 to DT 64. Unfortunately, we did not have the original DT 16 to include it in the analysis. We propose that the original DT 16 would be similar to the DT 16 convertants as shown with the original DT 64 and the DT 64 convertants.

3.3.9 Immunity to phage ST64T

ST64T was able to plaque on indicator strain DB21(ES18) but not on HisHB22(P22) which is lysogenic for wild–type P22. ST64T could also plaque on DB21(*sie*) which is lysogenic for a P22 mutant defective in both superinfection exclusion systems. ST64T sequence analysis has confirmed that this bacteriophage has a different immunity region to that of P22 (accession number AY052766). Based on the biological results, which are confirmed by sequence analysis, ST64T is heteroimmune to P22 and ES18. The exclusion of ST64T from the P22–lysogenic strain HisHB22(P22) is obviously controlled by one of the two superinfection exclusion systems, A or B.

3.3.10 Demonstration of generalised transduction with ST64T

ST64T was able to transduce transposon Tn*10*, *his*$^+$ and *trp*$^+$ markers. The efficiency of transduction was determined quantitatively. The transductants/plaque forming unit (T/P) ratios for both *his*$^+$ located at position 44.8 min and *trp*$^+$ markers located at position 38.1 min were 2×10^{-6} transductants/pfu. However, ST64T was able to transduce transposon Tn*10* from strain DB21 (zia-748::Tn*10*) located at position 81 min

(Sanderson, *et al.*, 1995) to the tetracycline sensitive DB21 wild-type strain at a lower frequency of 6×10^{-7} transductants/pfu. ST64T would appear to be capable of generalised transduction as it was able to transfer markers located at well–separated positions on the *Salmonella* chromosome.

3.4 Discussion

Crude and CsCl purified ST64T and ST64B phages were indistinguishable by electron microscopy, and both appear to belong to the P22 family. The gross morphology indicates that both ST64T and ST64B phages are podoviruses of the C1 morphotype (Ackermann and Gersham, 1992) and they appear similar to P22. The presence of many putative tail genes in ST64B in contrast to one putative tail gene in ST64T (chapters 5 and 4, repsectively), contradicts the morphological results and suggests that the tail of ST64B should be longer than that of ST64T. This point will be discussed further in chapter 5 together with genomic sequence analysis of ST64B. Restriction and sequence analyses (chapters 4 and 5) have shown that both phages are distinct.

*Sma*I restriction analysis of ST64T and ST64B phage genomes confirmed that DNA from the crude lysate was a combination of genomic DNA from each of the phage bands that could be separated on a CsCl gradient. The dsDNA-phages are thought to be the most abundant group of similar organisms in the biosphere (Coetzee, 1987). They have typically proven to be difficult to classify since phages with similar morphology, mode of replication and genomic architecture may be different at the nucleotide level (Hendrix, *et al.*, 1999).

Southern hybridisation analysis using both phage genomes (ST64T and ST64B) as probes, indicated that ST64B genome was found in most *S.* Typhimurium DTs tested whereas ST64T genome was found only in DTs 64 and 29. Interestingly, when phage ST64B was used as a probe to screen other *Salmonella enterica* serovars, the results

indicated that sequences hybridising to ST64B are found in a number of non-Typhimurium *Salmonella* serovars. These bands may represent hybridisation to host chromosomal sequences carried by the probing phage ST64B, which may be involved in the integration of the phage into the chromosome during lysogeny. However the strains Q1 and ATCC 14028s appear to carry prophages which closely resemble ST64B. Recently, two other phages designated Gifsy–1 and Gifsy–2 were also discovered in strain ATCC 14028s with Gifsy–2 carrying the *sodC* gene which has been shown to potentially influence virulence (Figueroa-Bossi and Bossi, 1999). ST64T and ST64B are distinct from these phages.

Lysogenisation of different serovars (Heildelberg, Enteritidis and Bovismorbificans) confirms that whenever ST64T integrates into the bacterial genome, phage type conversion results. This is in agreement with the work of Anderson and Felix (Anderson and Felix, 1953) who showed that lysogenisation of different host strains with the same temperate phage may produce different phage types. A similar scenario may exist with the ST64B phage, which is carried by a number of different *S.* Typhimurium DTs. Schmieger (Schmieger, 1999) demonstrated a similar phenomenon when using phages from the Anderson phage typing scheme. Some of these phages yielded identical *Eco*RI restriction patterns from their genomic DNA but then produced different lytic patterns on a given type strain. To explain this difference in phage sensitivity, he suggested that the differences in lytic patterns could be the result of different host-controlled modification-restriction systems.

Phage typing has been used for decades to subdivide isolates within a serovar in epidemiological investigations. This approach is based upon a collection of *Salmonella* phages which were propagated on specific strains of *S.* Typhimurium. It is likely that phage type is primarily determined by the carriage of a temperate phage or presence or absence of potential receptors on the bacterial cell surface. A crude lysate containing ST64T and ST64B induced from *S.* Typhimurium DTs 64 and 29 was able to plaque on both DTs 9 and 135. However, crude lysates containing only ST64B induced from DTs 9

and 135 were unable to plaque on DT 64 and DT 29 strains (Table 3.1). This may be explained by the carriage of ST64B by DT 64 and DT 29 strains, which may mediate immunity to superinfection with ST64B phage. It proved impossible to purify ST64B from another host to gauge its ability to mediate phage type conversion. It is likely that ST64B may influence phage type, but until a suitable host strain is found it is impossible to determine its impact.

It appears that ST64T is the phage from the mixed lysate that mediates phage type conversion in these studies. This is supported by the following evidence. Phage type conversion could also be demonstrated in strains sensitive to ST64T where crude phage lysates from DT 64 and DT 29 strains were able to convert DT 9 strains to DT 64 and DT 135 strains were converted to DT 16. Southern analysis confirmed that ST64T was found in the converted strains (Figure 3.7). Colonies that retained the original phage type did not have ST64T present as a prophage (Figure 3.7). A proposition can be made that the integration of ST64T into the chromosome of *S.* Typhimurium DTs 9 and 135 results in conversion of these phage types by changed immunity to the panel of typing phages. These changes are possibly due to recombination with residing prophages because most of the *Salmonella* isolates are lysogenic for at least one prophage (Schicklmaier, *et al.*, 1998).

This study has demonstrated that lysogeny with ST64T can mediate the conversion of *S.* Enteritidis PT 26 to PT 6a. Phage type conversion in this serovar, has been demonstrated by a number of other researchers. Chart and coworkers have shown that the conversion of *S.* Enteritidis PT 7 from PT 4 is due to modification in lipopolysaccharide (LPS) (Chart, *et al.*, 1989). Baggesen, *et al.* (1997) have shown that PTs 1, 4 and 6 could be converted to PT 7 due to modification in LPS. It was proposed that using PT 7 as an epidemiological marker should be treated with caution since genetically unrelated strains of *S.* Enteritidis PT 7 could occur within the same phage type. Rankin and Platt showed the conversion of *S.* Enteritidis PT 6a strains to PTs 4 and 7 (Rankin and Platt, 1995). PT conversion was also demonstrated by Hickmann-Brenner and coworkers when *S.*

Enteritidis PT 14b was subcultured and four colonies phage typed. Two colonies were found to be PT 14b and two PT 8. Conversion of *S.* Heidelberg was shown in our laboratory to have occurred during an outbreak of *S.* Heidelberg PT 1 (Heuzenroeder, M. W., R. Dalcin, D. Davos, and C. Murray, Abstr. 4th Asia Pacif. Poult. Health Conf., abstr. 144, 1998). Additionally, previous work performed by Harvey, *et al.* (1993) demonstrated the conversion of *S.* Heidelberg PTs 1 and 3 to PTs 4 and 5, respectively. Interestingly, when ST64T was grown on Heidelberg PT 1, it was able to integrate into its genome. This integration resulted in the conversion of PT 1 to PT 4.

Several studies have shown that AFLP has the same discriminatory power as pulsed-field gel electrophoresis (Lindstedt, 2000, Tamada, 2001). In this case, however, PFGE showed no banding difference between the parental phage types and the convertants. However, AFLP was able to discriminate between the parental phage types and the convertants. Genotyping revealed two clusters, which were observed in the DT 16 convertants, where cluster 2 had a unique 328.22 bp fragment. However, the DT 64 convertants (DT 9 to DT 64) had a single cluster with 4 fragments (40.45, 44.54, 46.55, 49.88 bp) which were missing in the parental DT 9. The dendrograms generated with UPGMA clearly revealed the differences between parental samples and convertants. Interestingly, DT 9 conversion to DT 64 generated a new DT 64, which is very similar to the original DT 64 (not converted from DT 9) (Figure 3.11). These results confirmed the phage typing results, which demonstrated phage type conversion by virtue of integration of ST64T into the genome of *S.* Typhimurium DTs 9 and 135. To test for reproducibility of the AFLP technique, a DT 64 standard was repeated and generated 50 scorable bands that varied by less than 1 bp between tests.

The lysogenic state in P22 and λ are controlled by C–immunity systems which are identical in function and organisation, but different in specificity. In addition, P22 has a second immunity system, designated, *immI* which encodes an antirepressor (Susskind, 1980). Although ST64T was able to mediate transduction with the same efficiency as P22,

its immunity region is quite distinct from that of P22. Based on the biological and sequence analyses, ST64T is heteroimmune to P22.

Gene transfer by transformation and conjugation appear to be uncommon in the genus *Salmonella* (Schmieger and Schicklmaier, 1999). Only one case of plasmid mediated transfer of Apramycin (resistance gene) in *S.* Typhimurium DT104 strain has been reported (Low, *et al.*, 1997). However, in the 1950s, transduction of streptomycin, chloramphenicol and penicillin resistance was demonstrated in *S.* Typhimurium (Watanabe and Watanabe, 1959). More recently, transduction of multiple drug resistance of *S.* Typhimurium DT104 has been demonstrated (Schmieger and Schicklmaier, 1999). It is therefore reasonable to consider transduction as a common vehicle for horizontal gene transfer in *Salmonella*. This implies that all *S.* Typhimurium DT 64 isolates carry a potent vehicle (ST64T bacteriophage) suitable for horizontal gene transfer. Unfortunately, the ability of ST64B to mediate transduction could not be established due to lack of a suitable host strain. Recently, a host–independent detection method for generalised transducing phages has been developed (Sander and Schmieger, 2001). PCR amplification of 16S ribosomal DNA was performed using primers specific for 16S rRNA genes of most eubacteria and phage–encapsulated DNA as a template (Sander and Schmieger, 2001). Using this method, it was possible to detect generalised transducing phage particles without isolating or cultivating the host cells. This method may elucidate the natural hosts of generalised transducing phages as well as their contribution to horizontal gene transfer among bacterial communities (Sander and Schmieger, 2001).

Finally, this study reports the discovery of ST64T and ST64B in a virulent *S.* Typhimurium DT 64 strain. ST64T is a generalised transducing phage, is heteroimmune to P22 and mediates phage type conversion. ST64B is conserved in many different strains of *S.* Typhimurium, which suggests it may confer a selective advantage on the bacterium.

The phage type conversion mediated by ST64T raises the question of the stability of the bacterial phage types in natural settings and the possibility of phage type conversion

occurring in an outbreak scenario. This may have serious implications for epidemiology and organism tracing. In conclusion, newer molecular methods such as PFGE and AFLP including whole genome sequencing where comparative studies of different genomes are employed, should be used to complement phage typing and other phenotypic typing methods.

Figure 3.1

Transmission Electron Micrographs showing the morphology of temperate phages ST64B and ST64T isolated from *S.* Typhimurium. A: crude lysate, which is a mixture of ST64B and ST64T phages was used. Morphologically, both phages belong to the *Podoviridae* family. Magnification: 297,000. →: tails, ↔ in A: icosahedral heads, ↔ in B: octagonal heads.

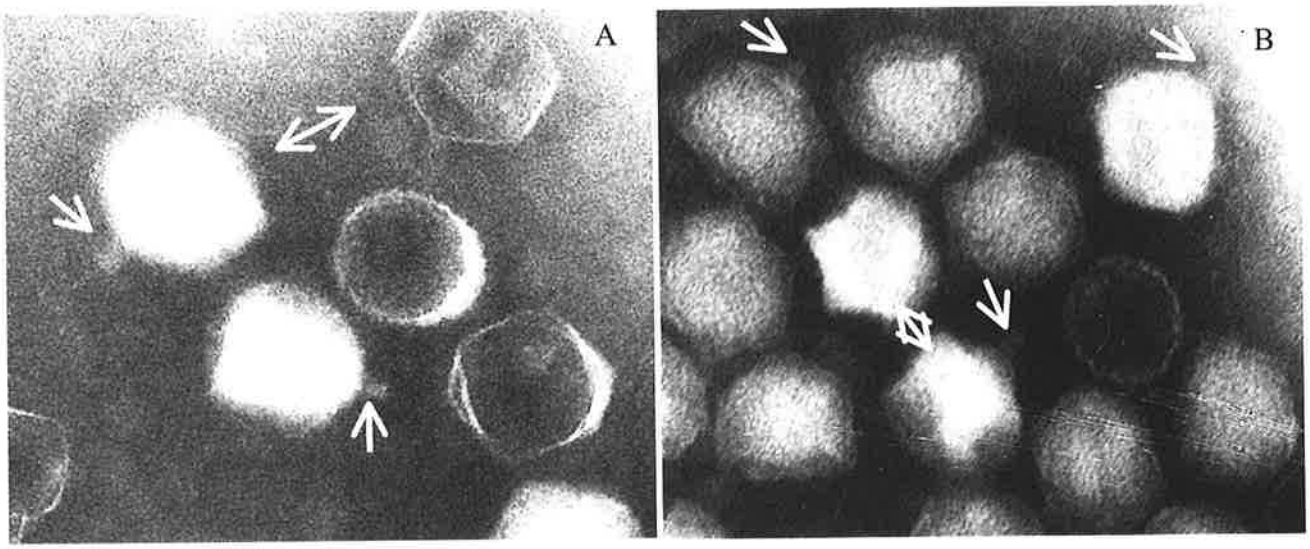

A
B

Figure 3.2

*Sma*I digest of crude and CsCl gradient purified bacteriophage ST64B and ST64T genomes from *S.* Typhimurium DT 64 separated on 1% agarose gel. MWM: Hi–Lo mixed DNA markers (in kb), which are a combination of equal amounts of three DNA digests; λ/*Hind*III, λ/*Pvu*I and SPP–I/*Eco*RI resulting in a range of commercially available DNA fragments (Geneworks, Therbaton, SA, Australia) were used as molecular weight markers. Crude: lysate where bacteriophage particles have not been separated on a CsCl gradient; ST64T and ST64B, genomic DNAs from CsCl gradient purified bacteriophage.

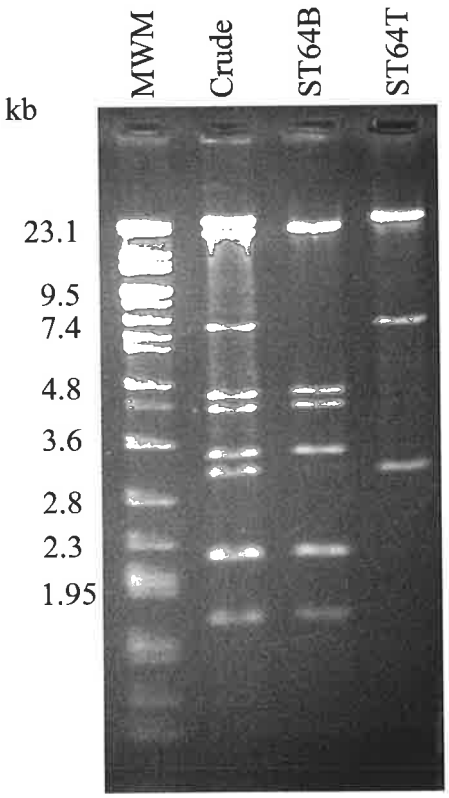

MWM
Crude
ST64B
ST64T
kb
23.1
9.5
7.4
4.8
3.6
2.8
2.3
1.95

Figure 3.3

Southern analysis of *Sma*I digested genomic DNA from *S.* Typhimurium DTs 9, 64, 135, 29 and RDNC/A045 original isolates demonstrating the presence of ST64B in their genomes. Purified ST64B genomic DNA was used as a probe. MWM: *Eco*RI digested *Bacillus subtilis* bacteriophage SPP-1 DNA (Geneworks, Thebarton, SA, Australia) molecular weight marker (in kb).

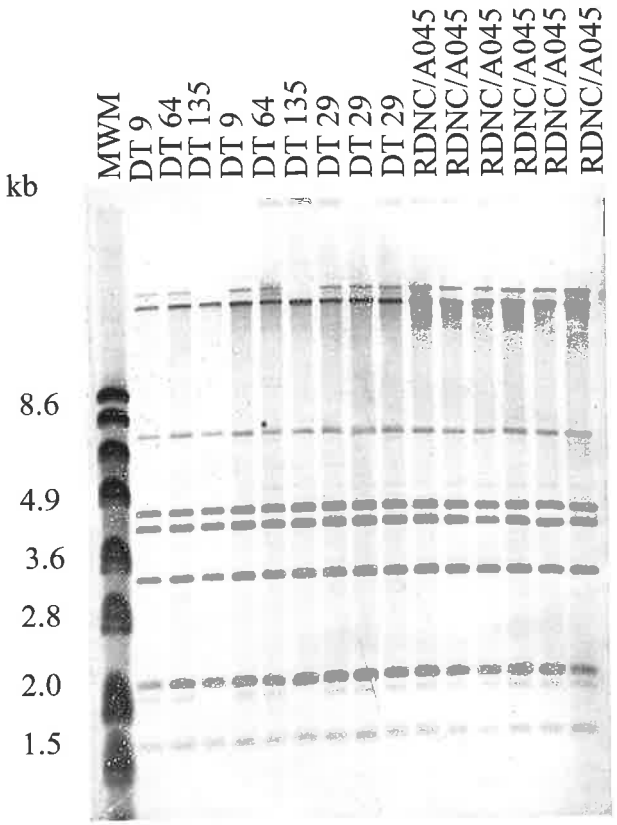

kb
MWM
DT 9
DT 64
DT 135
DT 9
DT 64
DT 135
DT 29
DT 29
DT 29
RDNC/A045
RDNC/A045
RDNC/A045
RDNC/A045
RDNC/A045
RDNC/A045
8.6
4.9
3.6
2.8
2.0
1.5

Figure 3.4

Southern analysis of *Sma*I digested genomic DNA from *S.* Typhimurium DTs 12, 135a, 8, 44 and 201 original isolates demonstrating the presence of ST64B in their genomes. DT 12 exhibited a distinct pattern of banding, however, a few bands were of the same size as those of ST64B. Purified ST64B genomic DNA was used as a probe.

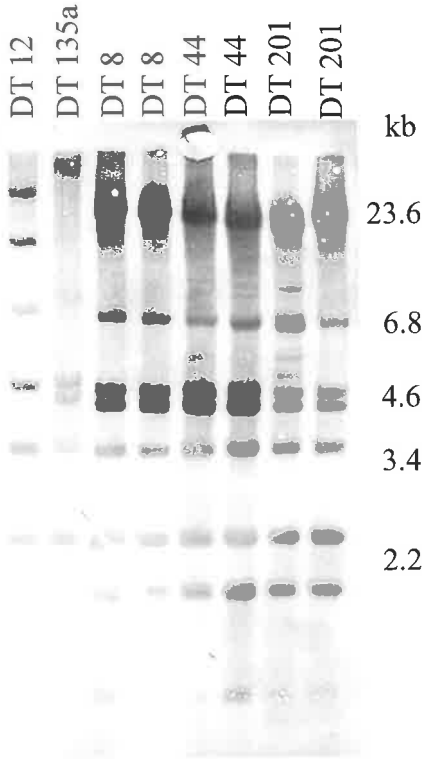

DT 12
DT 135a
DT 8
DT 8
DT 44
DT 44
DT 201
DT 201
kb
23.6
6.8
4.6
3.4
2.2

Figure 3.5

Southern analysis of *Sma*I digested genomic DNA from *S.* Typhimurium phage types demonstrating the presence (DTs 64 and 29) and absence (DTs 9, 135 and RDNC/A045) of ST64T. Purified ST64T genomic DNA was used as a probe. MWM: *Eco*RI digested *Bacillus subtilis* bacteriophage SPP-1 DNA (Geneworks, Thebarton, SA, Australia) molecular weight marker (in kb).

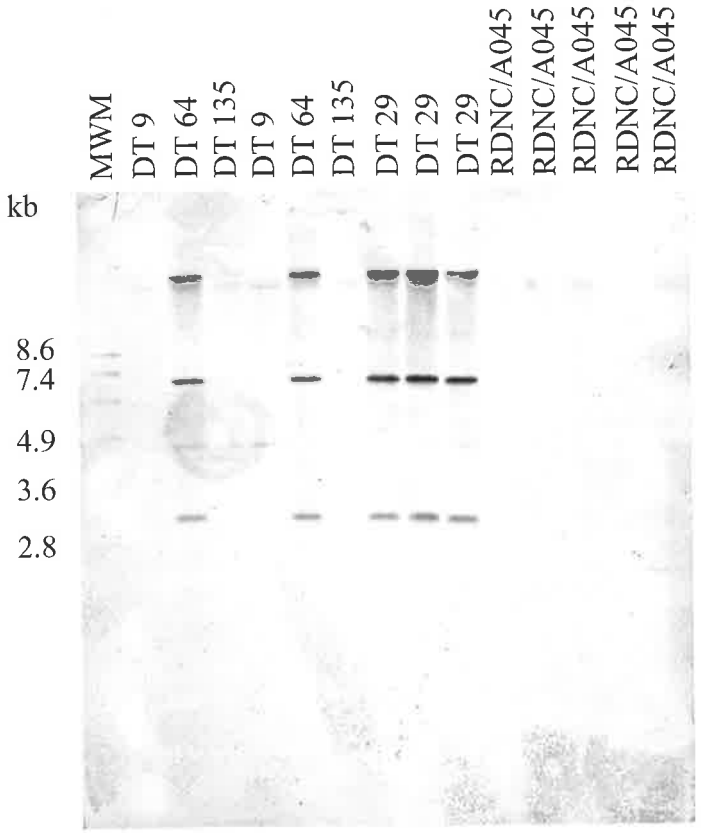

MWM
DT 9
DT 64
DT 135
DT 9
DT 64
DT 135
DT 29
DT 29
DT 29
RDNC/A045
RDNC/A045
RDNC/A045
RDNC/A045
RDNC/A045
kb
8.6
7.4
4.9
3.6
2.8

Figure 3.6

Southern analysis of *Sma*I digested genomic DNA from different LT2 derivatives and Q1 strain to demonstrate the presence or absence of ST64B related prophage. ST64B was isolated from *S.* Typhimurium DT 64 shown in the last lane. Purified ST64B genomic DNA was used as a probe.

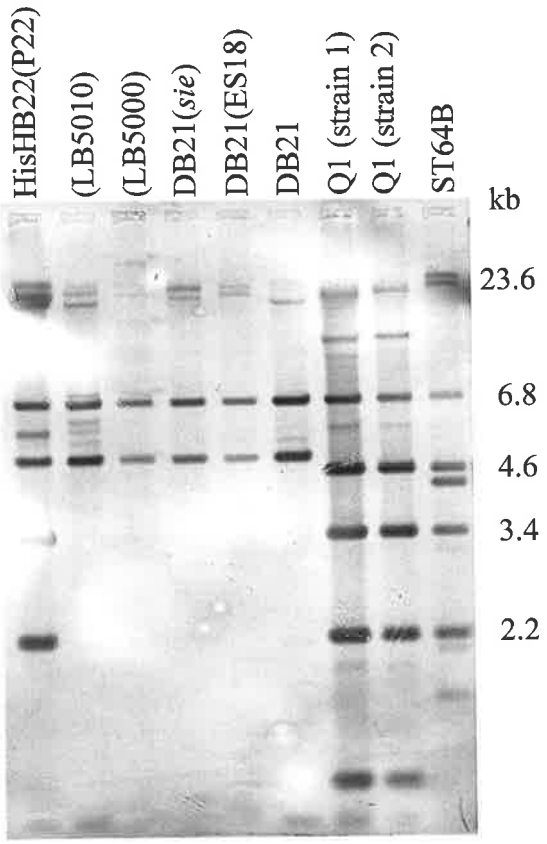

HisHB22(P22)
(LB5010)
(LB5000)
DB21(*sie*)
DB21(ES18)
DB21
Q1 (strain 1)
Q1 (strain 2)
ST64B
kb
23.6
6.8
4.6
3.4
2.2

Figure 3.7

Southern analysis of *Sma*I digested genomic DNA from phage type convertants to demonstrate the presence or absence of ST64T. Original phage type is in brackets ( ). Originally, DTs 9 and 135 do not carry ST64T. Purified genomic DNA from ST64T was used as a probe. MWM: Hi–Lo mixed DNA markers (in kb), which are a combination of equal amounts of three DNA digests; λ/HindIII, λ/PvuI and SPP–I/EcoRI resulting in a range of commercially available DNA fragments (Geneworks, Therbaton, SA, Australia) were used as molecular weight markers (in kb).

kb
MWM
DT64 (DT 9)
DT64 (DT 9)
DT9 (DT 9)
DT9 (DT 9)
DT64 (DT 9)
DT64 (DT 9)
DT64 (DT 9)
DT135(DT 135)
DT64 (DT 9)
DT64 (DT 9)
DT64 (DT 9)
DT16 (DT 135)
DT16 (DT 135)
DT16 (DT 135)
DT16 (DT 135)
DT16 (DT 135)
DT16 (DT 135)
DT16 (DT 135)
DT16 (DT 135)
DT64 (DT 135)
DT64 (DT 9)
23.1
9.5
7.4
3.6

Figure 3.8

Southern analysis of *Sma*I digested genomic DNA from phage type convertants to demonstrate the presence or absence of ST64T in other serovars. H in lanes 2–5 signifies *S.* Heidelberg, E in lanes 6–9 signifies *S.* Enteritidis, T in lanes 10–15 signifies *S.* Typhimurium and B in lanes 16–17 signifies *S.* Bovismorbificans. Unknown in lanes 13–17 means *S.* Typhimurium DT 44 and *S.* Bovismorbificans PT 32 were converted to an unknown phage typing reaction. Original phage type is in brackets ( ). All original phage types do not carry ST64T. Purified genomic DNA from ST64T was used as a probe. MWM: Hi–Lo mixed DNA markers (in kb), which are a combination of equal amounts of three DNA digests; λ/HindIII, λ/PvuI and SPP–I/EcoRI resulting in a range of commercially available DNA fragments (Geneworks, Therbaton, SA, Australia) were used as molecular weight markers (in kb).

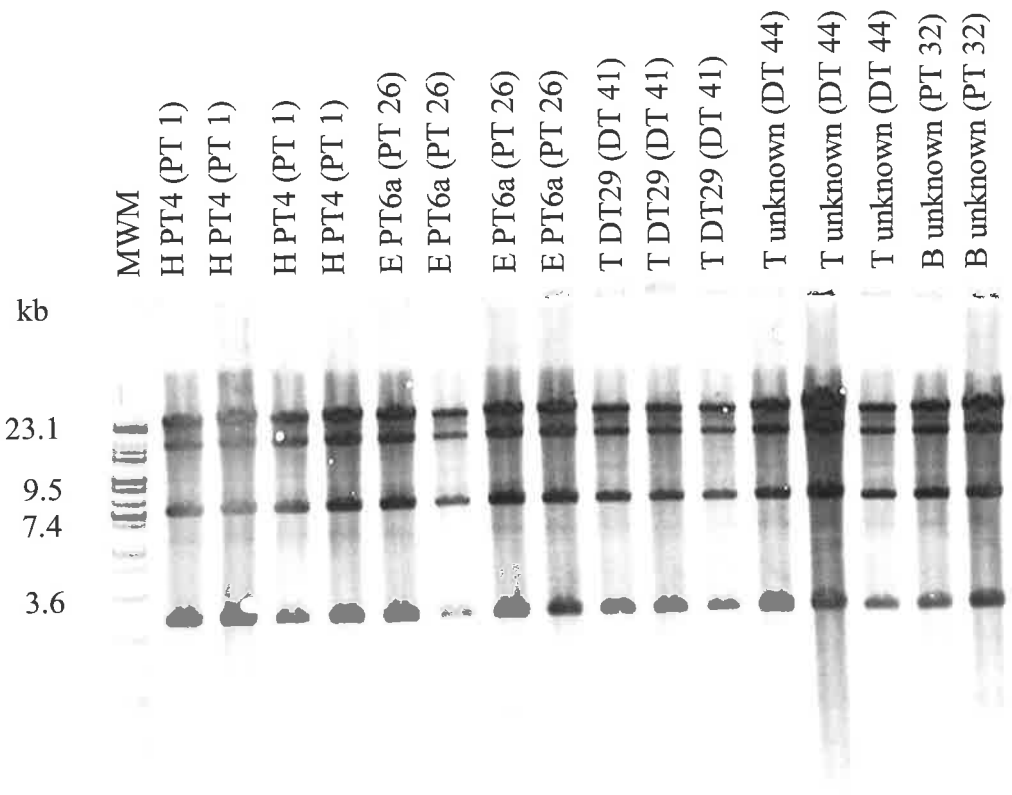

kb
23.1
9.5
7.4
3.6
MWM
H PT4 (PT 1)
H PT4 (PT 1)
H PT4 (PT 1)
H PT4 (PT 1)
E PT6a (PT 26)
E PT6a (PT 26)
E PT6a (PT 26)
E PT6a (PT 26)
T DT29 (DT 41)
T DT29 (DT 41)
T DT29 (DT 41)
T unknown (DT 44)
T unknown (DT 44)
T unknown (DT 44)
B unknown (PT 32)
B unknown (PT 32)

Figure 3.9

Genescan Analysis Software v3.1 software derived electropherograms showing an example of areas of polymorphism within AFLP profile amplifications of parental DT 135 and the convertants from DT 135 to DT 16. The fragment profiles were analyzed by Genotyper software, highlighting the fragments characteristic of genomic conversion. Fragments 28.68 bp and 37.91 bp are present in the parental DTs 135 but absent in the convertant. Fragments 40.53 bp, 44.60 bp, 47.39 bp and 50.13 bp are present in the convertant but absent in the parent DT 135.

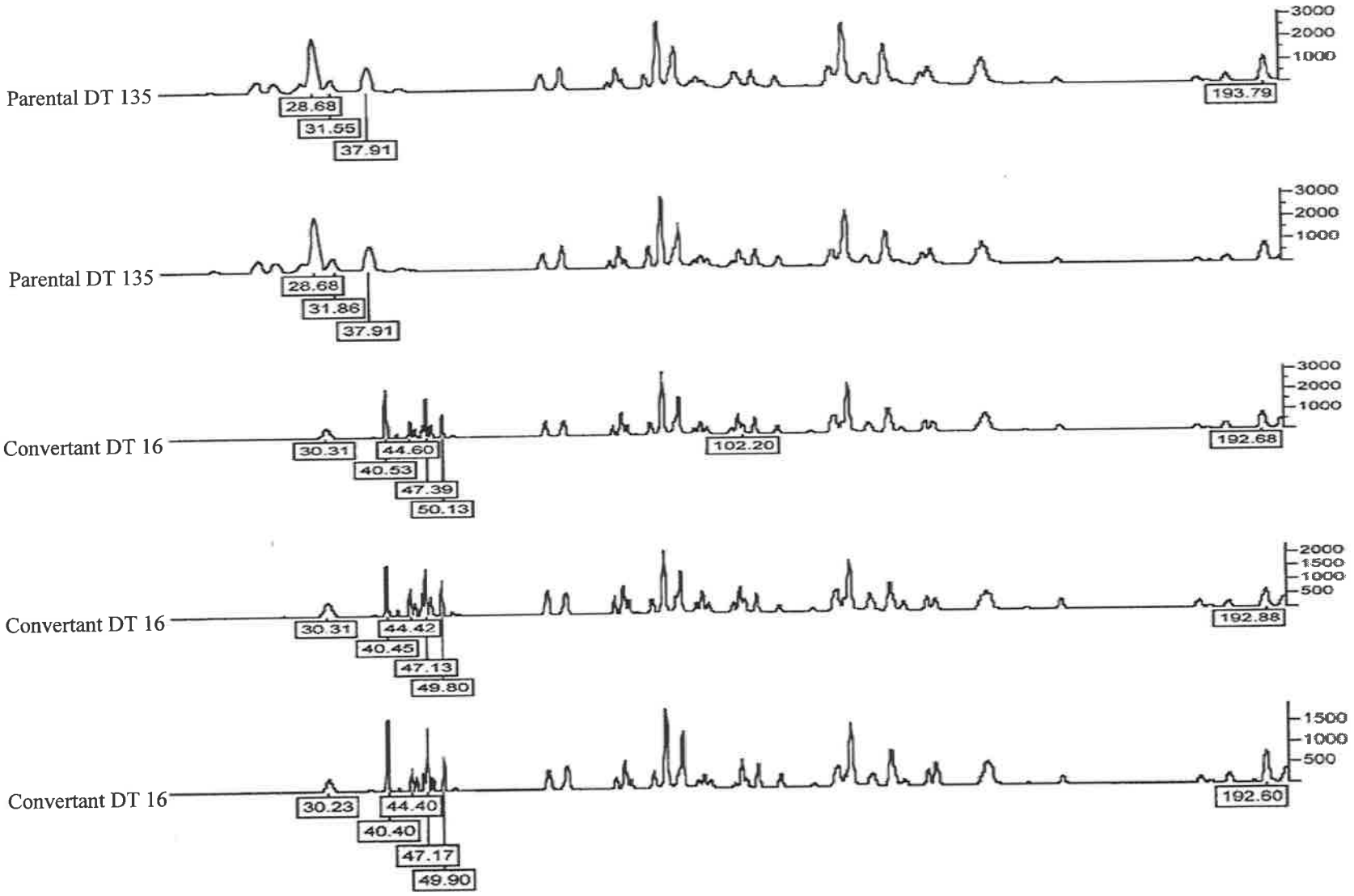

Parental DT 135
28.68
31.55
37.91
193.79
3000
2000
1000
Parental DT 135
28.68
31.86
37.91
3000
2000
1000
Convertant DT 16
30.31
44.60
40.53
47.39
50.13
102.20
192.68
3000
2000
1000
Convertant DT 16
30.31
44.42
40.45
47.13
49.80
192.88
2000
1500
1000
500
Convertant DT 16
30.23
44.40
40.40
47.17
49.90
192.60
1500
1000
500

Figure 3.10

Genescan Analysis Software v3.1 software derived electropherograms showing an example of areas of polymorphism within AFLP profiles amplifications of parental DT 9 and the convertants from DT 9 to DT 64. The fragment profiles were analyzed by Genotyper software, highlighting the fragments characteristic of genomic conversion. Fragments 30.22 bp and 39.04 bp are present in the parental DT9 but absent in the convertants. Fragments 40.45 bp, 44.54 bp, 46.55 bp and 49.88 bp are present in the convertant but absent in the parent DT 9.

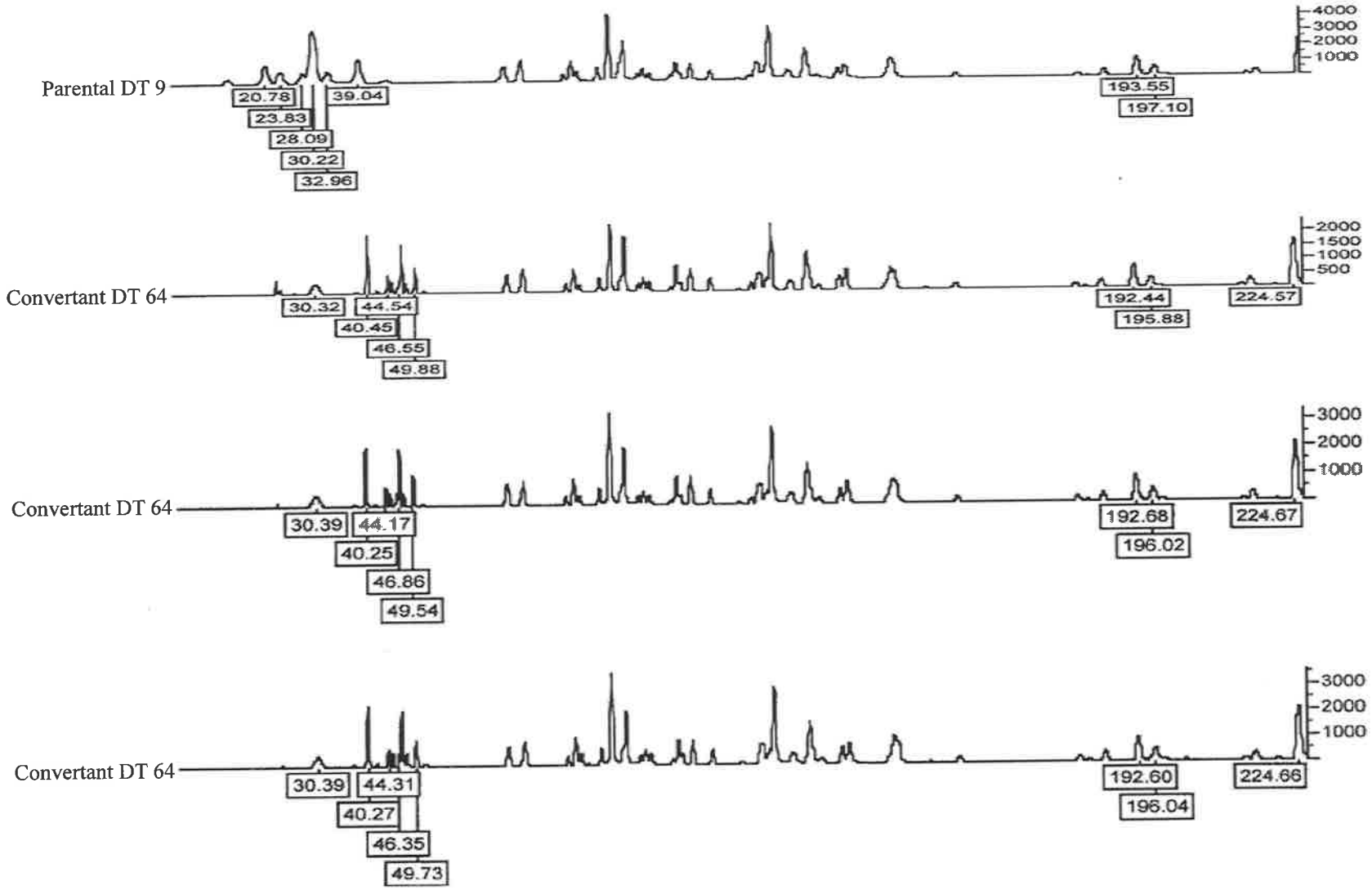

Parental DT 9
20.78
23.83
28.09
30.22
32.96
39.04
193.55
197.10
4000
3000
2000
1000
Convertant DT 64
30.32
40.45
44.54
46.55
49.88
192.44
195.88
224.57
2000
1500
1000
500
Convertant DT 64
30.39
40.25
44.17
46.86
49.54
192.68
196.02
224.67
3000
2000
1000
Convertant DT 64
30.39
40.27
44.31
46.35
49.73
192.60
196.04
224.66
3000
2000
1000

Figure 3.11

An AFLP pattern and dendrogram for the parental DT 9 isolate with its convertants. Patterns are the result of amplification of templates generated after restriction and ligation. The fragments were sized, quantified and analyzed using GeneScan Analysis Software v3.1 (PE Biosystems). The Genotyper Software v3.6NT (PE Biosystems) was used to identify the characteristic fragments and the tabular densitometric values were formatted on a spreadsheet and transferred to GelCompar v4.1 (Applied Maths, Kortrijk, Belgium). A dendrogram was generated using the unweighted pair group method (UPGMA) for clustering on a matrix based on the pairwise comparisons of the Dice coefficient. Percentages of similarity are shown above the dendrogram. Lanes 1–9 (white) are convertants (originally DT 9 but converted to DT 64 after lysogenisation of DT 9 with ST64T bacteriophage). Lane 10 (light grey) is the original DT 64 where ST64T phage was induced and lane 11 is the parent DT 9 (dark grey) which shows <88% similarity to the convertants (lanes 1–9) whereas there is >92% homology between the convertants and the original DT 64 (lane 10).

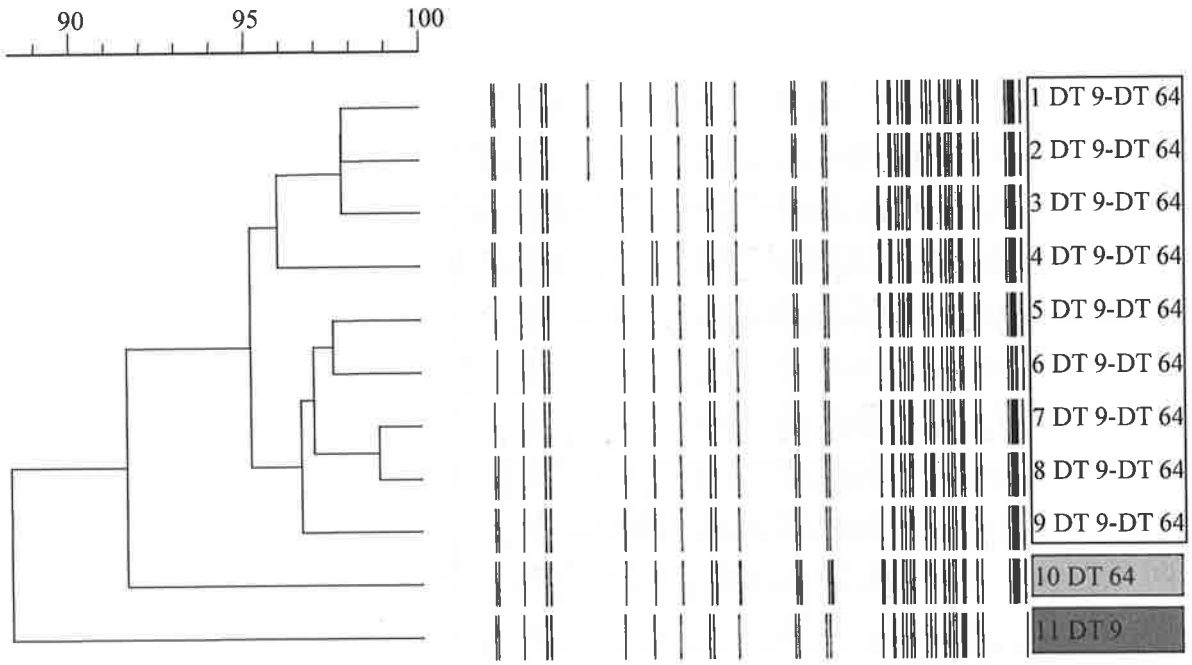

90
95
100
1 DT 9-DT 64
2 DT 9-DT 64
3 DT 9-DT 64
4 DT 9-DT 64
5 DT 9-DT 64
6 DT 9-DT 64
7 DT 9-DT 64
8 DT 9-DT 64
9 DT 9-DT 64
10 DT 64
11 DT 9

Figure 3.12.

An AFLP pattern and dendrogram for the parental DT 135 isolates with the convertants. The fragments were sized, quantified and analyzed using GeneScan Analysis Software v3.1 (PE Biosystems). The Genotyper Software v3.6NT (PE Biosystems) was used to identify the characteristic fragments and the tabular densitometric values were formatted on a spreadsheet and transferred to GelCompar v4.1 (Applied Maths, Kortrijk, Belgium. A dendrogram was generated using the unweighted pair group method (UPGMA) for clustering on a matrix based on the pairwise comparisons of the Dice coefficient. Percentages of similarity are shown above the dendrogram. Lanes 1–10 (white) are convertants (originally DT 135 but converted to DT 16 after lysogenisation of DT 135 with ST64T bacteriophage). Lanes 11 and 12 (grey) the parent DTs 135 isolates exhibiting <88% homology to the convertants (lanes 1–10). The group consisting of convertants, showed >94 % homology.

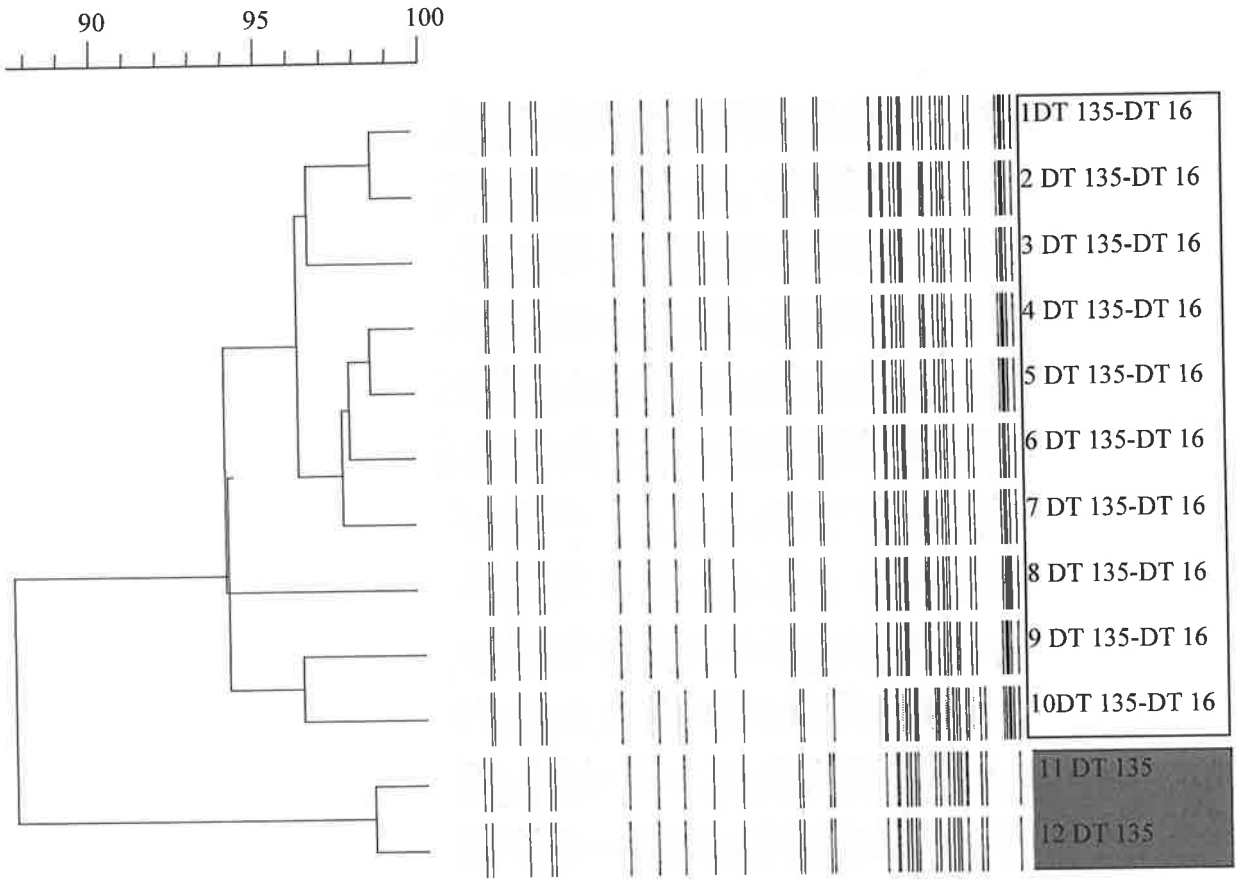

90
95
100
1DT 135-DT 16
2 DT 135-DT 16
3 DT 135-DT 16
4 DT 135-DT 16
5 DT 135-DT 16
6 DT 135-DT 16
7 DT 135-DT 16
8 DT 135-DT 16
9 DT 135-DT 16
10DT 135-DT 16
11 DT 135
12 DT 135

Table 3.1 Test for homo– and heteroimmunity with temperate phages

Crude lysate from *S.* Typhimurium	Plaquing ability on *S.* Typhimurium DTs			
	9	64	135	29
9	–	–	–	–
64	+	–	+	–
135	–	–	–	–
29	+	–	+	–

– = no plaques, + = plagues

Table 3.2 Phage type conversion mediated by induced temperate phage

Crude lysate from *S.* Typhimurium DT	Parental DT[a]	Converted DT[a]
64	9	64
64	135	16
29	9	64
29	135	16

[a]DT = Definitive phage Type

Table 3.3 Fragments in the AFLP profile indicative of genomic change in both the parental samples (DTs 9 and 135) and the convertants (DT 9 converted to DT 64 and DT 135 converted to DT 16)

Parental DT 135	Convertants (from DT 135 to DT 16)	Parental DT 9	Convertants (from DT 9 to DT 64)	Original DT 64*
28.68#	40.53#	30.22#	40.45#	40.05#
37.91#	44.6#	39.04#	44.54#	44.07#
230.15	47.39	193.55	46.55	46.90
297.88	50.13	197.10	49.88	49.83
338.34	192.68	226.26	192.44	192.49
382.51	228.81	297.88	195.88	195.95
426.54	295.90	331.00	224.57	224.64
479.72	328.22	338.39	295.85	295.87
~10.36	335.04	382.51	328.23	328.09
~10.54	379.05	426.52	334.93	334.75
	423.91	479.65	378.96	378.80
	478.16	~10.40	423.92	423.84
			478.28	478.28

* original lysogen where crude phage lysate containing both ST64B and ST64T bacteriophages were obtained for characterisation

smaller fragments approximated

Chapter 4

Characterisation of the genome of phage ST64T

4.1 Introduction

The dsDNA–containing bacteriophages are very likely the most abundant group of organisms in the biosphere with more than 4,500 different dsDNA–bacteriophages capable of infecting a large variety of bacterial hosts described (Coetzee, 1987). Temperate bacteriophage λ is the prototype of a family of phylogenetically related viruses, the lambdoid bacteriophages, which have a double stranded DNA (dsDNA) genome and show a conserved arrangement of regulatory elements (Gilakjan and Kropinski, 1999). Generally, genes of similar function in this family occupy corresponding location and in some instances, are able to form viable recombinants with one another (Campbell, 1994). In addition, the structure and function of a lambdoid phage are more highly conserved than the actual gene sequences (Casjens, *et al.*, 1992) since family members may have only a limited degree of nucleotide sequence similarity to each other (Oberto, *et al.*, 1994). Comparison of the genomes of several newly characterised phages and cryptic prophages, appears to show that most of the dsDNA–tailed phages undergo considerable genetic exchange (Hendrix *et al.*, 1999).

Salmonella serovars are known to harbour many temperate bacteriophages. Most of these belong to the P22 branch of the lambdoid family and are able to facilitate horizontal genetic transfer by transduction (Schicklmaier and Schmieger, 1995). Morphologically, P22 is distinct from λ. Its genomic architecture is similar to that of other lambdoid phages including λ, and viable P22/λ hybrids have been formed *in vivo* (Vander Byl and Kropinski, 2000). However, λ and P22 differ in many ways (Botstein, *et al.*, 1972). P22 DNA is circularly permuted with terminal repetitions (Rhoades, *et al.*, 1968)

which is in contrast to λ DNA which has cohesive ends (Hershey and Burgi, 1965). Phage λ grows in the absence of recombination whereas P22 requires recombination for growth (Botstein and Herskowitz, 1974). In addition to λ *immC* region, P22 has the *immI* region and elaborates two superinfection exclusion systems (Botstein, *et al.*, 1972).

Based largely on the morphological difference between P22 and λ, the former has been proposed as a type virus for a new genus which includes phages L (Bezdek and Amati, 1967), ES18 (Kuo and Stocker, 1970), LP7 (Kitamura and Mise, 1970) and APSE-1 (van der Wilk, *et al.*, 1999). Evidence based on, complementation and hybridisation studies, (Schicklmaier, *et al.*, 1999) has shown that most *Salmonella* bacteriophages are P22–like. Although the *Salmonella* phage ES18 is morphologically similar to λ and the *O* genes of the two phages are 95% similar, approximately 50% of the ES18 genome is identical to that of P22. This observation suggests that full genome sequencing may be the best way to fully classify bacteriophage genomes. This has been suggested before (Vander Byl and Kropinski, 2000).

This chapter presents the complete dsDNA sequence of *S.* Typhimurium bacteriophage ST64T, a member of the P22–like group of phages and its inferred amino acid sequences of gene products. A comparative analysis of these sequences with those of other lambdoid phages, with emphasis on P22 is also presented.

4.2 Results

4.2.1 Sequencing strategy

A 3,137 bp *Sma*I restriction fragment from the ST64T genome was shot–gun cloned into pGEM–7zf(–). A second 6,830 bp *Sma*I fragment was gel–purified before cloning into pGEM –7zf(–), as described in chapter 2. To sequence the ST64T inserts from pGEM–7zf(–), small scale plasmid DNA was first purified and the inserts

initially sequenced using dye primer sequencing chemistry (chapter 2). The remaining sequence was determined using CsCl purified ST64T fragments by dye terminator chemistry with synthetic oligonucleotide primers (Appendix I) as described in chapter 2. PEG_{8000} precipitated whole phage DNA was used to sequence the largest *Sma*I restriction fragment (30,712 kb) as described in chapter 2. Occasionally, PCR products amplified from this fragment (30,712 kb) were used for sequencing. Both the forward and the reverse strands were sequenced.

4.2.2 Nucleotide accession number

The complete nucleotide sequence of ST64T bacteriophage has been deposited with GenBank and has been assigned accession number AY052766. Sequence data is presented on a CD as Appendix III.

4.2.3 Sequence assembly

To determine the fragment order, sequencing primers were designed to read outward from the ends of the 3,137 bp and 6,830 bp *Sma*I fragments using whole ST64T genome as a template, thus reading into the ends of the largest (30,712 bp) *Sma*I fragment. The three *Sma*I fragments were assembled using GeneCompar 2.0 and/or GeneBase 1.0 (Applied Maths, Kotrijk, Belgium).

4.3 Sequence analysis of ST64T genome

ST64T is a linear dsDNA bacteriophage, which is circularly permuted and terminally redundant, resembling P22 in many aspects. Like P22 (Botstein, *et al.*, 1972, Gough and Levine, 1968), the ST64T vegetative map is circular, whereas the prophage map is linear, with unique ends. To exclude the terminally redundant ends of ST64T,

sequencing of the *Sma*I (30,712 bp) large fragment, which includes, among others, putative genes for packaging and integration, was achieved using whole phage genome as a template as outlined in chapter 2. In order to compare ST64T to P22, the ST64T map was opened adjacent to the 15–bp stem–loop structure (AATAAAAATGGGTGTaaACACCCATTTTTATT {bases in the loop are shown in lowercase}) located downstream of the putative tailspike protein, gp9. For this structure, a ΔG of –16.8 kcal/mol was calculated using the Vienna RNA Secondary Structure Prediction at the website http://rna.tbi.univie.ac.at/cgi–bin/RNAfold.cgi#Results (Matthews, *et al.*, 1999). The ST64T genomic sequence was 40,679 bp in size with an overall GC content of 47.5%.

4.3.1 Assignment of probable genes

The P22 nomenclature was adopted in designating ST64T ORFs. A total of sixty five putative open reading frames were identified on both strands of the nucleotide sequence initially using the NCBI graphical ORF finder and WebGene Mark.HMM (Lukashin and Borodovsky, 1998) as outlined in chapter 2. Furthermore, the ORFs were confirmed and the map drawn using DNA Strider 1.3 program with AUG and GUG selected as initiation codons and exported as a PICT file to the graphic program Canvas 2.1 (Figure 4.1). Additionally, putative initiation sites were assigned for genes primarily by visual inspection of six–frame translations of the DNA sequences. In most cases, assignment of probable ORFs was based on the presence of an initiation codon (AUG, except for four cases where it was GUG) with a good ribosome binding sites (Shine Dalgarno sequence) situated upstream of the initiation codon. Only two previously determined genes (*gtrA* and *eac* with 100% and 86% sequence identity to P22 genes, respectively) did not have putative ribosomal binding sites. In about 91% of the ST64T putative genes, the initiation codon was within 14 bp of the putative ribosomal binding site,

with approximately 45% of the genes overlapping the termination codon of upstream genes. A few of these overlaps demonstrate a feature where UGA terminates one gene and AUG initiates the next, in the sequence AUGA. For the majority of the other genes, there was a non–coding space between genes accommodating a putative transcriptional control signal.

4.3.2 Open reading frames present on the genome of bacteriophage ST64T and their functional assignments

The amino acid sequence of the deduced putative proteins in ST64T were used in searches for similarity with other sequences present in databases using BlastP algorithm (Altschul, *et al.*, 1990). The properties of the ORFs showing homology with proteins in NCBI databases will be discussed, and where possible, the function of the ORFs may be inferred. Table 4.1 presents the different 65 ORFs found in ST64T and the graphical map is presented in Figures 4.1 and 4.2.

4.3.2.1 Serotype conversion genes

The three genes (*gtrA*, *gtrB* and *gtrC*) involved in serotype conversion, arranged in the same pattern as in other phages were recently identified in P22 (Vander Byl and Kropinski, 2000). Homologues of these genes transcribed in the complementary strand, were also found in the sequence of ST64T and showed high similarity to P22 genes (Table 4.1). Martin Reese's Promoter Prediction by Neutral Network program (http://www.fruitfly.org/seq_tools/promoter.html) did not reveal any promoter region close to the conversion genes. However, a putative promoter region (**TTGATC**GGTAACAACGATCAAT**TAACGA**GCAAACA) {promoter region for –35 and –10 consensus sequences shown in bold and underlined}, which is identical to the P22 promoter region, except for the –10 region where TAAC**(AT)** in P22 (Vander Byl and

Kropinski, 2000) is replaced by TAAC(**GA**) in ST64T, was identified manually 103 bp upstream of the putative *gtrA* gene (Figure 4.3). The three genes in this conversion cassette, *gtrA*, *gtrB* and *gtrC* overlap each other and are probably transcribed by this promoter. The essential features of each ORF and the inferred amino acid sequences are described below.

GtrC. This 1,458 bp ORF (31.9 %mol GC) begins with a putative GUG codon, which is preceded by a probable prokaryotic ribosomal binding site (RBS) (Shine–Dalgarno box) consensus sequence (UAAGG), 9 bp upstream. The putative gene encodes a 485 amino acid putative protein with a predicted mass of 55,291 Da and pI of 8.7. This ORF is located at the same position and is the same size as the *gtrC* of P22. Alignment of the $gtrC_{P22}$ and $gtrC_{ST64T}$ nucleotide sequences revealed high sequence identity (Figure 4.3). BlastP analysis revealed that $GtrC_{ST64T}$ shared 99% sequence identity with $GtrC_{P22}$. Furthermore, comparison of $GtrC_{ST64T}$ and $GtrC_{P22}$ revealed two mismatches at positions 282 (Val to Ala) and 390 (His to Tyr).

Searches for conserved motifs using Protein Families (Pfam) (Bateman, *et al.*, 1999) at the Washington University server in St. Louis, revealed 4 motifs identified as NADH–Ubiquinone oxidoreductase complex I chain 5N–terminus (PF00662), Integral membrane protein DUF6 (PF00892), Dolichyl–phosphate–mannose–protein mannosyltransferase (PF02366) and Hantavirus glycoprotein G2 (PF01561). TMHMM analysis at the Centre for Biological Sequence Analysis, The Technical University of Denmark revealed 11 transmembrane domains, with its carboxyl terminus probably in the periplasmic region of the cell (Vander Byl and Kropinski, 2000). Furthermore, this gene showed weak structural similarity with the glucosyl transferase gene *bgt* of *S. flexneri* SfX (Verma, *et al.*, 1999) and the chromosomally encoded *gtrI* gene of this bacterium (Adhikari, *et al.*, 1999). Studies have shown that the gene products of *Shigella gtrI* as well as SfX *bgt* can mediate serotype conversion. In *Shigella*, it has been shown that the products of the first two genes involved in serotype conversion are conserved and

interchangeable among serotypes (Guan, 1999, Huan, *et al.*, 1997, Mavris, *et al.*, 1997). However, the third gene appears to be unique to each bacteriophage (Allison and Verma, 2000). This unique protein is referred to as the serotype–specific glucosyl transferase. In addition, homologues were identified in the currently incomplete *Salmonella* genomes in the database. Putative ST64T *gtrC* shares 99% and 98% sequence identity with genes in *S.* Paratyphi and *S.* Dublin A, respectively. The high nucleotide sequence identity of $gtrC_{ST64T}$ to genes in *S.* Paratyphi and *S.* Dublin probably suggests these bacteria carry a functional or defective phage which is closely related to ST64T and P22 in this region of the genome. Furthermore, putative *gtrC* has a striking feature in that its mol% GC (31.9) is considerably lower than that of *gtrAB* cluster (43.4) and also lower than that of the entire ST64T genome (47.5).

GtrB. This 933 bp ORF (41.8 %mol GC) encodes a 312 aa putative protein with a predicted mass of 35,112 Da and a pI of 8.8. A probable RBS (AGGGA) was located 6 bp upstream of the putative start codon. Searches for conserved motifs using Pfam, revealed one conserved motif defined as Glycos_transf_2 19–183 (glycosyl transferase) (PF00535). TMHMM analysis identified two transmembrane domains in the latter two thirds of the protein which may suggest that both the amino and carboxy termini are cytoplasmic (Vander Byl and Kropinski, 2000). $GtrB_{ST64T}$ shares 99% sequence identity with $GtrB_{P22}$ and 96% identity with a putative glycosyltransferase protein of *S.* Typhimurium (accession no. NP_459551). Furthermore, $GtrB_{ST64T}$ showed 88% sequence identity with similar proteins in *Shigella flexneri* phages SfIV, SfV and SfII (Figure 4.4). Two other proteins from *S. flexneri* (accession nos: AAF09026 and AAC39272) shared 88% sequence identity to $GtrB_{ST64T}$. Comparison of the inferred $gtrB_{ST64T}$ gene with the currently incomplete *Salmonella* genomes in the database revealed 96%, 99% and 91% nucleotide sequence identity with genes in *S.* Paratyphi A, *S.* Dublin and *S.* Enteritidis, respectively.

GtrA. This 363 bp ORF (45 %mol GC) encodes a 120 aa putative protein with a predicted mass of 13,500 Da and pI of 9.45. Searches for conserved motifs using Pfam,

revealed one conserved motif identified as glypican 89–110 (PF01153). Similar to P22, $GtrA_{ST64T}$ contains four transmembrane domains as predicted by the TMHMM search algorithm. Alignment of the putative $GtrA_{ST64T}$ with $GtrA_{P22}$ revealed 100% sequence identity, which was confirmed by BlastP analysis. $GtrA_{ST64T}$ also exhibited 93% sequence identity to the bactoprenol–linked glucose translocase of *S.* Typhi (accession no. CAD05041) and to the putative glycosyl translocase of *S.* Typhimurium LT2 (accession no. AAL19511). In addition, 78% and 77% sequence identity was observed with the GtrA proteins of *S. flexneri* bacteriophages SfX and SfV. Furthermore, 79% identity was shown with *E. coli* prophage CPS–53 GtrA homologue. $GtrA_{ST64T}$, like $GtrA_{P22}$ (Vander Byl and Kropinski, 2000), exhibited 77% sequence identity to a product of a defective prophage in *S. flexneri* (Adhikari, *et al.*, 1999) and 79% identity to a hypothetical protein b2350 in *E. coli* (accession no. C65008). Homologues were also identified in the currently incomplete *Salmonella* genomes in the database. The ST64T inferred *gtrA* gene shares 100%, sequence identity with a gene in *S.* Dublin, 93% with genes in *S.* Typhi CT18, *S.* Enteritidis and *S.* Paratyphi A. Like $gtrA_{P22}$, no obvious putative ribosomal binding site could be found upstream of the putative initiation codon of $gtrA_{ST64T}$. However, the three genes overlap each other. Furthermore, it has been shown that the cloned P22 *gtrABC* genes could mediate O–antigen serotype conversion (Vander Byl and Kropinski, 2000). ST64T, like P22, has been shown to be capable of serotype conversion (Personal communication, Carly Tucker, University of South Australia, Honours student). A similar scenario has been observed in *S. flexneri* serotype converting phages, which are capable of mediating glucosylation of the O antigen (Adhikari, *et al.*, 1999, Verma, *et al.*, 1999).

4.3.2.2 Integration and excision

The serotype conversion genes are located immediately downstream of the phage *attP* site, preceded by the *int* and *xis* genes. BlastN revealed a putative ST64T *attP* site

(2,289 – 3,241 nt) which has a very high sequence identity (97%) with the P22 *attP* site. Analysis of the core and arm type Int binding sites as characterised in P22 (Smith-Mungo, *et al.*, 1994), was performed on ST64T nucleotide sequence and found to be almost identical to those of P22 *attP* (Figure 4.5).

Int. This 4,251 bp ORF (48 mol% GC) encodes a putative protein of 387 aa with a predicted mass of 44,862 Da and pI of 9.84. The putative initiation codon is preceded by a RBS (GAGGA) 6 bp upstream. Searches for conserved motifs using Pfam revealed one motif identified as phage integrase family (PF00589). Comparison of the inferred gene product encoded by this ORF with other proteins using BlastP revealed 98% sequence identity to Int_{P22}. Additionally, the predicted molecular weight, pI and size values of putative Int_{ST64T} protein are identical to those of the Int_{P22} protein. These suggest that ST64T Int belongs to the family of site–specific recombinases which catalyse the strand exchange reaction and together with the host encoded protein, integration host factor (IHF), catalyse the recombination event. Using MacTargsearch online program (Goodrich, *et al.*, 1990), a potential IHF–binding site was located at position 3,091–3,117 (CCAGTTAA**ATCAA**ATAC**TTA**CGTATTA) {bases similar to the consensus sequence are shown in boldface}, with a similarity score of 61.7 (see Table 4.2). This potential IHF–binding site is identical to the IHF–binding site of P22 (Vander Byl and Kropinski, 2000). This IHF overlaps with the putative ST64T *int* gene and may be involved in the site–specific recombination of ST64T. Based upon the high sequence homology of the *attP* sites and integrases of P22 and ST64T in this region, it is a possibility that these two bacteriophages integrate at the same location in the *Salmonella* chromosome.

Int_{ST64T} protein also exhibited 89% sequence identity to integrases of *S. flexneri* phages SfV and SfII, 88% to *E. coli* prophage CP–933H, 73% to a cryptic *E. coli* prophage DLP12, 65% with an *Acyrthosiphon pisum* phage *AP*SE–1 which has been proposed to be a member of the family *Podoviridae,* and lastly 54% with *S. flexneri* phage SfX.

Xis. This 351 bp ORF (52 mol% GC) encodes a putative protein with a predicted mass of 12,754 Da and a pI of 9.4. The putative initiation codon is preceded by a probable RBS (UGGAG) 5 bp upstream. The putative *xis* and *int* genes overlap. The Xis_{ST64T} protein is identical to Xis_{P22} protein. This probably suggests that ST64T, like P22, make use of a similar mechanism for integration and excision of the prophage. BlastP analysis revealed that Xis_{ST64T} shared 87% sequence identity with *S. flexneri* SfV phage, which is also a member of the *Podoviridae*.

4.3.2.3 The region encompassing putative Eac to ORF56.

Eac. This 656 bp ORF (49 %mol GC) encodes a putative 218 aa protein with a predicted mass and pI of 24,925 Da and 5.6, respectively. A weak potential RBS (UGAG) was located 14 bp upstream of the initiation codon. BlastP revealed 86% sequence identity to P22 Eac protein.

Eag. This 180 bp ORF (40 %mol GC) encodes a putative 59 aa protein with a predicted mass of 6,644 Da and a pI of 8.31. A potential RBS (UGGAGG) was located 6 bp upstream of the putative initiation codon. The inferred gene product encoded by this ORF is identical (100%) to the Eag protein in P22.

Eaa2. This 546 bp ORF (46 %mol GC) has a GUG as its putative initiation codon with a potential RBS (GGAGG) located 5 bp upstream of the initiation codon. It encodes a putative 181 aa protein with a predicted mass of 20,301 Da and a pI of 4.3. BlastP revealed a 100% sequence identity with the carboxy terminal of P22 Eaa protein (Figure 4.6).

Eaa1. This 806 bp ORF (53 %mol GC) was designated *eaa1* because its inferred gene product has similarity to the amino terminal domain of the P22 Eaa protein. There is a potential RBS (GAGGG) 7 bp upstream of the initiation codon, AUG. BlastP analysis showed 62% sequence identity with P22 Eaa protein. The putative *eaa2* and *eaa1* genes overlap. In order to confirm the separate nature of the putative *eaa1* and *eaa2* genes, this

region of DNA was sequenced several times, in the belief that a sequencing error would explain the insertion in the ST64T nucleotide sequence that potentially encodes two gene products as opposed to a single product in P22. Identical sequence was obtained on all occasions. For this reason, and because of the presence of good potential ribosomal binding sites upstream of each ORF, these two ORFs were designated, *eaa2* and *eaa1*, respectively.

Three ORFs (ORF87, ORF109 and ORF81) with unknown functions were found between the putative *eaa1* and *eae* genes. These will not be discussed because they did not show any similarities with known genes, however, putative ORF81 gene product revealed 96% sequence identity with an unknown protein in *S.* Typhimurium (accession no. T03013). Furthermore, BlastN revealed that ORF81 shared 97% sequence identity with a truncated GTP–binding protein (LepA) found in the left end portion of the Gifsy–1 prophage.

Eae. This 378 bp ORF (51 %mol GC) has a potential RBS (GGAGUGGG) 5 bp upstream of its putative initiation codon. The putative Eae protein showed only 32% sequence identity to P22 Eae protein. Upstream of the gene *eae*, lies ORF56 which is the same size as the ORF56 of P22 and shares 65% sequence identity. ORF56 has a probable RBS (GAGGA) 7 bp upstream of its initiation codon, which overlaps the gene *eae*. The Ea gene products in P22 form part of the P_L operon and are non–essential genes (Susskind, 1980, Youderian and Susskind, 1980). Given sequence similarities of ST64T to P22, it is assumed that these genes in ST64T are also not essential for phage growth.

4.3.2.4 Genes involved in homologous recombination

In P22, three genes are directly or indirectly involved in homologous recombination. These are the *abc* genes (*anti–recBCD*) which modulate the activities of the host cell RecBCD protein and the *erf* (essential recombination function) gene which

promotes *recA*–independent homologous recombination (Poteete, 1988). The following ORFs encoding these genes were identified in bacteriophage ST64T.

Abc2. This 294 bp ORF (49 %mol GC) encodes a putative protein of 97 aa long with a predicted mass of 11,621 Da and pI of 8.95. A potential RBS (GGAGU) was located 5 bp upstream of the putative initiation codon. BlastP revealed that $Abc2_{ST64T}$ shared 94% sequence identity with $Abc2_{P22}$, 91% identity with both $gp39_{HK97}$ and $Abc2_{HK022}$ and 90% identity with an unknown protein encoded by phage HK620. Analysis of this protein with Pfam revealed a conserved motif of the Coesterase 9–51 (carboxylesterase) (PF00135). Studies have shown that the *abc2* gene in P22 is necessary for homologous recombination. P22 lacking *abc2* is grossly defective for replication in a *polA*– host and mildly defective in a wild–type host (Poteete, 1988).

Abc1. This 285 bp ORF (41 mol% GC) has a probable RBS (GGUGG) 4 bp upstream of the putative start codon. Like P22, it encodes a protein of 94 aa with a predicted mass of 10,819 Da and a pI of 10.2. BlastP analysis revealed 95% sequence identity between $Abc1_{ST64T}$ and $Abc1_{P22}$ proteins. The *abc1* gene is thought to contribute to recombination, but unlike *abc2* it is not required for replication in a *polA*– host (Poteete, 1988). The initiation codon of this ORF overlaps the termination codon of its upstream gene, *erf*.

Erf. This 708 bp ORF (47 mol %GC) encodes a 235 aa protein with a predicted mass and pI of 26,816 Da and 6.0, respectively. A probable RBS (GGUGA) was detected 5 bp upstream of the putative initiation codon. BlastP analysis revealed 76% sequence identity between Erf_{ST64T} and Erf_{P22} proteins and 57% sequence identity with unknown proteins encoded by bacteriophages from a Gram positive bacterium, *Lactococcus lactis* (U136 and bIL286). Comparison of Erf_{ST64T} and Erf_{P22} proteins using Align revealed high sequence identity from approximately 180 aa of Erf_{ST64T} to its carboxy terminus (Figure 4.7). However, the amino terminus of putative Erf_{ST64T} is distinct from that of P22 but shares high sequence identity with unknown proteins in phages U136 and bIL286. It has

been shown that Erf is a two domain protein, joined by an inter–domain DNA binding segment (Poteete, 1988).

ORF66. The initiation codon of this ORF is not preceded by any putative consensus RBS sequence. The ORF encodes a protein with a predicted mass of 7,769 Da and pI of 9.58 and is located upstream of the *erf* gene. However, the inferred amino acid sequence is longer than that of Arf of P22 (47 aa) which like ORF66 is positioned immediately upstream of the *erf* gene. Neither BlastN nor BlastP revealed any sequence identity with P22 *arf* gene. However, BlastP revealed 67% sequence identity with an unknown HkaM putative protein of the *E. coli* prophage HK620.

4.3.3.5 Some P_L operon genes

Kil. This 114 bp ORF (38 mol% GC) encodes a 37 aa protein with a predicted mass of 4,409 Da and a pI of 9.3. A probable RBS (AGGA) is located 10 bp upstream of the putative initiation codon. No significant identity was found when this ORF was compared with P22 *kil* gene at both the nucleotide and inferred protein levels using Align. BlastP analysis revealed that putative Kil_{ST64T} shared 58% sequence identity with Kil_{λ}, Kil_{VT2-Sa} and Kil_{933W} (VT2–Sa and 933W are *E. coli* Stx2–converting phages). This inferred protein is smaller in size than Kil_{λ}, Kil_{VT2-Sa} and Kil_{933W} proteins. It has been suggested that this gene may be involved in killing the host cell but its role in the life–cycles of λ and P22 is unknown (Poteete, 1988).

C3. This 168 bp ORF (48 mol% GC) encodes an inferred 55 aa protein with a predicted mass of 6,021 Da and a pI of 9.78. A probable RBS (GAGG) was located 20 bp upstream of the predicted initiation codon, which is beyond the optimal position of the consensus sequence. The initiation codon of the putative *kil* gene overlaps the termination codon of the putative *c3* gene. BlastP revealed 77%, 72% and 70% sequence identity with $C3_{P22}$, $CIII_{\lambda}$ and both $CIII_{933W}$ and $CIII_{HK620}$ proteins, respectively. Searches for conserved

motifs using Pfam revealed 1 motif identified as Lambda_CIII 8–52 (PF02061). C3 is a regulatory protein involved in establishment of lysogeny in P22 and may perform a similar function in ST64T.

gp17. This 315 bp ORF (47 mol% GC) has a probable RBS (UGGAGG) 4 bp upstream of the initiation codon. It encodes a 104 aa inferred product with a predicted mass and pI of 12,263 Da and 4.01, respectively. Using BlastP, $gp17_{ST64T}$ was shown to have 92 and 89% sequence identity to $gp17_{P22}$ and $gp17_{H19-B}$ (H19–B prophage is an *E. coli* prophage encoding the *stx* genes), respectively. This gene product allows P22 to escape Fels–2 prophage exclusion (Poteete, 1988).

ORF232. This 699 bp long ORF (50 mol% GC) has a probable RBS (GAUGGA) 14bp upstream of its initiation codon. It encodes an inferred product of 232 aa with a predicted mass and pI of 25,820 Da and 6.89, respectively. Two almost identical direct repeats (**CAGGTCGGCACCGCG** and **CAGGTTGGTACCGCA**) were identified at position 11,374 – 11,524. A search using Pfam revealed eight motifs identified as Pentapeptide repeats (PF00805). Apparently, these repeats are found in many cyanobacterial proteins and were first identified in *hglK* gene product which is required for localisation of heterocyst–specific glycolipids in cyanobacterium, *Anabaena sp.* strain PCC 7120 (Black, *et al.*, 1995). The structure of this repeat is thought to be a Beta–helix and its significance in ST64T is unknown. Homology search using BlastP showed that ORF232 has 68% sequence identity to a hypothetical protein in *Streptomyces coelicolor* (accession no. CAB66202), 47% to a hypothetical protein in *Synechocystis sp.* strain PCC 680 (accession no. S75470) and 46% identity to a pentapeptide repeat family protein in *Caulobacter crescentus* (accesssion no. AAK24231).

Ral. The inferred product of this ORF is highly similar to the P22 Ral protein. BlastP search revealed that putative Ral_{ST64T} shared 98% and 78% sequence identity with the Ral proteins of P22 and *E. coli* bacteriophages 21 and λ, respectively. This ORF (45 mol% GC) is the same size as that of P22 and a probable RBS (AGGA) was located 9 bp

upstream of the predicted initiation codon. ORF232 overlaps the termination codon of *ral*. The inferred 64 aa protein encoded by this 195 bp long ORF has a predicted mass and pI of 7,336 Da and 8.4, respectively.

gp24 (λ N orthologue). This 339 bp ORF (47 %mol GC) encodes a 112 aa putative protein with a predicted mass of 12,759 Da and a pI of 10.19. A probable RBS (AGGGA) was located 6 bp upstream of the predicted initiation codon. BlastP analysis revealed 100% sequence identity (from position 1 to 76 in both ST64T and phage L) of putative $gp24_{ST64T}$ to *S.* Typhimurium $gp24_{L}$. Additionally, a very high sequence identity (96%) was shown with $gp24_{P22}$. The high degree of amino acid similarity of these three phages (ST64T, L and P22) in their amino–termini (Figure 4.8) differs from other lambdoid phages (λ–like phages propagating on *E. coli*). The N proteins of λ-like phages show conservation in the carboxy–terminal and deviate in the amino terminal domains (Franklin, 1992). The $gp24_{L}$, $gp24_{P22}$ and $gp24_{ST64T}$ proteins showed little homology with N_{λ} protein. P22 gene *24* and λ gene *N* are involved in transcriptional regulation of the early genes, with their gene products acting as antiterminators.

4.3.3.6 Immunity genes of ST64T

A region of immunity control comprising of the genes *c2, cro* and *c1* is found upstream of gene *24* of P22 and other lambdoid phages. The intergenic region (558 bp) between genes *24* and *c2* in ST64T is almost the same length as the spacer region of phage L (530 bp). One of the target sites for gp24, *nutL* which mediates leftward transcriptional control (Schicklmaier and Schmieger, 1997), is located in this intergenic region. The rightward transcriptional control is mediated by *nutR* which is situated between genes *cro* and *c1*. Figure 4.9 presents a comparison of the *nut* sites of ST64T to those of other lambdoid phages (λ, P22 and 21).

C2 (λ CI homologue). This 705 bp ORF (46 mol% GC) encodes a 234 amino acid putative protein with a predicted mass and pI of 26,097 Da and 6.91, respectively. There is a second potential initiation codon (AUG) at position 42 of this ORF, which corresponds to the initiation codon of the *c2* gene of phage L. It is a possibility that ST64T uses the second initiation codon. As predicted by Pfam, two conserved motifs were found and identified as Helix–turn–helix (PF01381) and Peptidase family S24 (PF00717). $C2_{ST64T}$ protein revealed 100% sequence identity to $C2_L$ protein but shares only 54% sequence identity to $C2_{P22}$ protein and 47% sequence identity with the CI proteins of *E. coli* phages, HK620 and 434. Like $C2_L$, $C2_{ST64T}$ showed a relatively high degree of C-terminal sequence identity with $C2_{P22}$ (Figure 4.10). The C– terminus of $C2_{P22}$, is responsible for dimerisation of the two C2–peptides in order to form a functional repressor molecule (Schicklmaier and Schmieger, 1997). However, the N–termini of $C2_{ST64T}$ and $C2_L$ differ from that of P22. The N–terminus of $C2_{P22}$ is responsible for operator recognition and binding. This suggests $C2_{ST64T}$ and $C2_L$ may recognise and bind to different operators.

Cro. This 200 bp ORF (53 mol% GC) encodes a putative protein of 66 amino acid with a predicted mass and pI of 7,370 Da and 8.8, respectively. A probable RBS (GAGGA) was located 8 bp upstream of the putative initiation codon. Cro_{ST64T} protein is almost the same size as Cro_L and like Cro_{P22} and Cro_L, it is transcribed in a rightward direction. BlastP search revealed that putative Cro_{ST64T} shared 98% sequence identity with Cro_L. Furthermore Pfam search revealed one Helix–turn–helix motif (PF01381). Cro is a negative transcriptional regulator, which acts antagonistically to C2.

C1 (λ CII orthologue). Between *cro* and *c1*, there is a non–coding region of 127 bp. A potential C1–binding site is located in this region (**TTGC**GTGTAT**TTGC**), 38 bp upstream of the putative initiation codon of gene *c1* (Table 4.2). The 282 bp ORF (49 %mol GC), encodes an inferred protein of 93 aa with a predicted mass and pI of 10,531 Da and 9.26, respectively. Analysis of this putative protein using Pfam revealed one conserved motif of the Helix–turn–helix family (PF01381). $C1_L$ has been partially

sequenced (62 aa long) (Schicklmaier and Schmieger, 1997). The N–terminus sequence of this ORF gene product is identical to that of $C1_{ST64T}$ with one mismatch at position 58 (Ala to Val). It is probable that $C1_{ST64T}$ and $C1_L$ are the same size and share substantial sequence identity. BlastP analysis revealed 96% sequence identity between $C1_{ST64T}$ and $C1_L$. Furthermore, $C1_{ST64T}$ shares 55% sequence identity with CII_{HK620}, 49% with $C1_{P22}$, and 50% with CII_{434} and CII_λ (**Figure 4.11**). $C1_{P22}$, like CII_λ is a transcriptional activator.

λ CII protein activates transcription from three leftward promoters (P_{RE}, P_I and Pa_Q) by binding to sequences overlapping their "–35" regions. These promoters are inactive in the absence of CII protein (Poteete, 1988). Table 4.2 presents the potential C1-binding sites found in ST64T genome. These results suggest that $C1_{ST64T}$ is probably a transcriptional activator as is the case for $C1_{P22}$ and CII_λ.

4.3.3.7 The replication genes *18* and *12* (λ *O* and *P* homologues)

Two genes involved in DNA replication of P22 and λ (*18* and *12* for P22; *O* and *P* for λ) are situated on the right of *c1* gene. These genes are transcribed in the rightward direction.

gp18. This 900 bp ORF (45 mol% GC) encodes an inferred 299 aa putative protein with a predicted mass of 34,236 Da and a pI of 9.16. A probable RBS (GAGGA) was located 6 bp upstream of the putative initiation codon. The putative $gp18_{ST64T}$ protein exhibited 66% sequence identity with O_{HK022} and 65% sequence identity with both O_{VT2-Sa} and O_{HK620} putative proteins. Comparison of $gp18_{ST64T}$ with gpO_λ and $gp18_{P22}$ proteins, revealed high sequence similarity at their N–termini (Figure 4.12). However, comparison of $gp18_{ST64T}$, O_{HK022} and O_{HK620} revealed striking homology at both their N– and C–termini. All three putative proteins are the same size. It is known that the N–termini of both gpO_λ and $gp18_{P22}$ are responsible for binding to the origin of replication whereas their C–termini cooperate with gpP_λ or $gp12_{P22}$ (Schicklmaier and Schmieger, 1997). The origin of

replication in both P22 and λ is located in the coding region of the gp18 and gpO, respectively. In λ, there are four 20–bp direct repeat sequences that are recognised by the gpO protein to initiate replication. Since $gp18_{ST64T}$ is similar to gpO_{HK022} and because the replication origin of HK022 is unknown, a search for direct repeats within the putative gene *18* of ST64T was performed. Three 20 bp direct repeats (TGTCCAACGGATGTCCAACG) were found at positions 14,851 – 14,870, 14,862 – 14,881 and 14,912 – 14,935. Upstream of this region, there is an A–T rich region **TAAA**c**AATA**ATA**TAAA**t**AATA** with repeats in boldface and underlined. Such DNA repeat sequences are probably associated with the origin of replication of ST64T as is the case for $gp18_{P22}$ and gpO_{λ}.

gp12. This 1,437 bp ORF (50 mol% GC) encodes a putative protein of 478 aa with a predicted mass and pI of 53,481 Da and 5.4, respectively. The putative ORF has a probable RBS (GAGGGU) 7 bp upstream of the putative initiation codon and overlaps the termination codon of gene *18*. Comparison of gpP_{HK022}, gpP_{HK620} and $gp12_{ST64T}$ proteins revealed high sequence identity (Figure 4.13). Furthermore, these inferred gene products were found to be the same size (478 aa). BlastP revealed 98% sequence identity of $gp12_{ST64T}$ with gpP_{HK620} and gpP_{HK022}, respectively. Furthermore, this putative protein showed 98% sequence identity to DnaB protein in *E. coli* O157:H7 (accession no. BAA94127). Comparison of $gp12_{ST64T}$ with $gp12_{P22}$ and gpP_{λ} revealed very little sequence identity. However, Pfam search revealed two conserved motifs of the DnaB–like helicase family (PF00772). It is likely that putative $gp12_{ST64T}$ acts like $gp12_{P22}$ and exerts the helicase function rather than dependence on the bacterial DnaB as it occurs in λ.

4.3.2.8 The region between the replication genes and the late control gene *23*

There are several ORFs between the gene *12* and the late control gene *23* in ST64T, whose functions are, for the most part, unknown. This is also the case for both P22 and λ

where this region encodes the *nin* genes whose functions are unknown (although functions have been inferred for some of them). The distance between gene 12_{ST64T} and $ORF90_{ST64T}$ is 74 bp in comparison with the 75 bp between the genes 12_{P22} and $ninB_{P22}$. One could speculate that similar to P22, the ORF downstream of gene 12_{ST64T} would be *ninB*. However, unlike P22, the ST64T genome encodes 4 unknown putative ORFs, immediately downstream of gene *12*. ORF90, ORF69 and ORF98 have putative RBS (GGGAGA, UGGAGA and AGGAU) 5 bp, 1 bp and 4 bp upstream of their probable start codons, respectively. Pfam search revealed a flagellar hook–length control protein (PF02120) in the gene product encoded by ORF69. ORF86 has a probable RBS (GCGGU) 2 bp upstream of its AUG putative start codon and it shared 30% sequence identity with *orf41* of *Pseudomonas aeruginosa* phage D3.

NinB. This 306 ORF (50 mol% GC), has no probable RBS upstream of the putative initiation codon and potentially encodes a 101 aa putative protein with a predicted mass of 11,918 Da and a pI of 9.42. The first 38 amino acids of the $NinB_{ST64T}$ inferred gene product revealed 92% sequence identity with the gene product encoded by $ORF8_{933W}$, that may be probable $NinB_{933W}$ protein (933W is a bacteriophage carried by *E. coli* serotype O157:H7) and 45% amino acid identity to $NinB_{P22}$ (see other similarities in Table 4.1). The termination codon of this putative gene overlaps the putative initiation codon of the gene *ninD*.

NinD. This 201 bp ORF (49 %mol GC), has a probable RBS (GGAGA) 8 bp upstream of the predicted initiation codon and encodes an inferred gene product of 66 aa with a predicted mass of 7,984 Da and pI of 6.05. $NinD_{ST64T}$ putative protein shares 94% and 75% sequence identity with $NinD_{P22}$ and $NinD_{\lambda}$, respectively.

NinE. This 177 bp ORF (47 %mol GC) has a probable RBS (AGGAGA) 4 bp upstream the predicted initiation codon and encodes an inferred gene product of 58 aa in size with a predicted mass of 7,116 Da and a pI of 10.49. $NinE_{ST64T}$ exhibited high sequence identity (96%) to both $NinE_{P22}$ and $NinE_{\lambda}$.

NinX. This 348 bp ORF (42 %mol GC) has a potential RBS (GAGGA) 2 bp upstream of the predicted initiation codon and encodes a 116 aa putative protein with a predicted mass of 13,166 Da and a pI of 4.64. $NinX_{ST64T}$ exhibited 57% sequence identity with $NinX_{P22}$. Furthermore, $NinX_{ST64T}$ showed a weak sequence identity (32%) with an unknown protein of pMT1 plasmid which is carried by *Yersinia pestis* (accession no. AAC82773).

NinF. This 174 bp ORF (51 %mol GC) has a potential RBS (AGGA) 8 bp upstream of the putative initiation codon and encodes a 57 aa putative protein with a predicted mass and pI of 6,328 Da and 4.34, respectively. ST64T NinF shares 85%, 84%, 75% and 70% sequence identity with related proteins in phages H–19B, P22, HK97 and HK620, respectively. Two additional ORFs (ORF101 and ORF136) of unknown function were found downstream of *ninF* with good putative RBS consensus sequence upstream of their respective predicted initiation codons. The gene product encoded by ORF101 showed 33% sequence identity to ORF80 of phage D3 whereas the gene product encoded by ORF136 showed 96% and 95% sequence identity with an unknown gp66 protein of HK97 and HkbD protein of HK620, respectively.

Rus. This 396 bp ORF with 50 %mol GC, has a probable consensus RBS (GGAGGG) 10 bp upstream of its inferred initiation codon. It potentially encodes a protein of 131 aa with a predicted mass and pI of 15,244 Da and 9.59, respectively. BlastP search revealed 62 and 52% sequence identity with $Rus_{CP-933X}$ and Rus_{82} (both CP-933X and 82 bacteriophages are carried by *E. coli*), respectively.

NinH. This 204 bp ORF (47 %mol GC) has a probable RBS consensus sequence (UGGAGG) 4 bp upstream of the probable initiation codon. It encodes an inferred gene product of 67 aa in size, with a predicted mass and pI of 7,899 Da and 9.65, respectively. Interestingly, $NinH_{ST64T}$ is exactly the same size, as $NinH_{P22}$ exhibiting 100% sequence identity to $NinH_{P22}$. Pfam HMM search revealed one Helix–turn–helix domain of the resolvase family (PF02796).

NinZ. This ORF (55 %mol GC) encodes a NinZ putative protein of 59 aa with a predicted mass and pI of 6,835 Da and 6.73, respectively. A probable RBS (GGAGAGG) was located 4 bp upstream of the putative initiation codon. BlastP search revealed that $NinZ_{ST64T}$ shared 93% sequence identity with $NinZ_{P22}$. The entire *nin* region (*ninB–ninZ*) is characterised by overlapping genes (a total of eight overlapping ORFs). The *ninZ* gene is the last of the *nin* genes and immediately downstream of it, lies the putative late gene antitermination protein, gp23.

gp23 (λ Q homologue). This 519 bp ORF (45 %mol GC) encodes a putative 172 aa protein with a predicted mass and pI of 19,755 and 9.54, respectively. The gene product encoded by this ORF is smaller than Q_{λ} and $gp23_{P22}$ but larger than its orthologue in *E. coli* bacteriophage 21. BlastP revealed that the putative $gp23_{ST64T}$ shares 81% sequence identity to Q_{21} and 33% identity to $P5(Q)_{APSE-1}$ (Figure 4.14). Similar to other lambdoid phages, this putative gene is linked with an upstream gene (*ninZ*) where their initiation and termination codons overlap (AUGA). A potential RBS (GGGAGG) was located 10 bp upstream of the putative start codon. Within 23_{ST64T}, lies a potential C1–binding site **TTGC**GTGGCT**TTGC** which is transcribed in the leftward direction (Table 4.2). This is probably a C1 dependent P_{aQ} promoter, which is inactive in the absence of the C1 protein. Upstream of the termination codon of gene *23*, is a possible rho–independent terminator (T**GCCCAC**CTAGCC**GTGGGC**TTTTCATT {inverted repeats in boldface and underlined}) located at position 21,253–21,279.

4.3.3.9 Genes involved in lysis

Downstream of the putative late transcriptional regulator gene *23*, are the three overlapping genes whose products are predicted to be involved in lysis of the bacterial host. Interestingly, these ORFs are almost identical to those of the *S.* Typhimurium bacteriophage PS3.

Holin (gp13) (λ S homologue). This 318 ORF (48 %mol GC) encodes a putative protein of 105 aa with a predicted mass and pI of 11,389 Da and 9.36, respectively. A putative RBS consensus sequence (GGAGGU) lies 3 bp upstream of its probable initiation codon. BlastP search revealed that putative $gp13_{ST64T}$ shares 99% sequence identity with $gp13_{PS3}$ and 75% identity with a holin protein in *E. coli* phage CP–933X. Alignment of $gp13_{ST64T}$ and $gp13_{PS3}$ inferred amino acid sequences showed one mismatch at position 88 (Val to Gly, respectively) (Figure 4.15). Although $gp13_{ST64T}$ does not resemble that of $gp13_{P22}$ by sequence similarity, it is almost the same size as $gp13_{P22}$ and is located at a comparable map position. The holin two–product motif structure is conserved in most lambdoid phages, with initiation codons at Met–1 and Met–4 or Met–1 and Met–3 producing a lysis inhibitor and effector, respectively (Bläsi and Young, 1996, Nam, *et al.*, 1990). However, putative holin proteins in ST64T and PS3 have the dual start motif at Met–1 and Met–5 positions which would probably result in a 100 and 104 amino acid predicted gene products, respectively (Figure 4.15). TMHMM analysis revealed 3 transmembrane domains. The three transmembrane domains, together with a dual start motif, suggest that ST64T gp13 may be classified as a class I holin.

Endolysin (gp19) (λ R orthologue). This 399 bp ORF (48 %mol GC), encodes a putative protein of 132 aa with a predicted mass of 14,948 Da and pI of 9.64. A probable RBS (GGGAUG) was located 6 bp upstream of the predicted initiation codon. Similar to the holins, $gp19_{ST64T}$ is nearly identical to $gp19_{PS3}$, with only one mismatch at position 81 (Trp to Cys, respectively) (Figure 4.16). BlastP analysis revealed that $gp19_{ST64T}$ shared 99% sequence identity with $gp19_{PS3}$. Furthermore, $gp19_{ST64T}$ showed weak sequence identity (26%) with an endolysin of phage A118 of *Listeria monocytogenes* which acts as an L-alanine-D-glutamate peptidase (Loessner, *et al.*, 1995).

gp15 (λ Rz homologue). This 453 bp ORF (51 %mol GC) encodes a putative protein of 150 aa with a predicted mass and pI of 16,557 Da and 9.27, respectively. A weak potential RBS (UGAGA) was located 7 bp upstream of the probable initiation codon.

The $gp15_{ST64T}$ inferred amino acid sequence exhibited 100%, 68% and 67% sequence identity to $gp15_{PS3}$, Rz_{HK97}, $gp15_{PS34}$ and $gp15_{P22}$, respectively. Although $gp15_{P22}$, $gp15_{PS34}$ and Rz_{HK97} gene products are smaller than the $gp15_{ST64T}$ and $gp15_{PS3}$ gene products, their carboxy–termini show a very high level of identity (Figure 4.17). TMHMM analysis revealed 1 transmembrane domain. This gene is the last of the three lysis genes and appears to be conserved within the lambdoid family (Highton, *et al.*, 1990, Schicklmaier, *et al.*, 1999).

4.3.3.10 ORFs of unknown function between lysis and packaging genes

Upstream of the lysis genes of P22, lie two ORFs (ORFs201 and 80) encoding gene products of unknown function. A similar scenario was observed in ST64T where three ORFs (ORF118, 129 and 134) with unknown functions were located upstream of gene *15*. ORF118 may not be expressed as it has no RBS consensus sequence upstream of its putative start codon. In contrast, both ORF129 and ORF134 have potential ribosomal binding sites (GGAG) 8 bp and (AGGAG) 7 bp upstream of their putative initiation codons, respectively. There is an overlap (UGAUG) of start and stop codons of ORF134 and ORF129, respectively. ORF129 gene product revealed a 37% sequence identity to the inferred amino acid sequence of an unknown protein in HK620 whereas ORF134 revealed a 100% sequence identity with the amino terminus of an unknown protein from *S.* Typhimurium bacteriophage LP7.

4.3.3.11 Genes involved in packaging

The inferred products of genes *3* and *2* most probably represent the small and large ST64T terminase subunits, respectively.

Terminase small subunit (gp3). This 489 bp ORF (53 %mol GC) encodes a putative protein of 162 aa with a predicted mass and pI of 18,713 Da and 5.60,

respectively. A potential RBS (AGGAAG) was located 7 bp upstream of the putative initiation codon. BlastP analysis revealed that $gp3_{ST64T}$ shared 100%, 99%, 98% and 96% sequence identity with $gp3_{PS3}$, $gp3_{LP7}$, $gp3_{PS34}$ and $gp3_{P22}$, respectively. Alignment of the inferred amino acid sequence of $gp3_{ST64T}$ to $gp3_{PS3}$, $gp3_{LP7}$, $gp3_{PS34}$ and $gp3_{P22}$ is presented in Figure 4.18. In P22–like phages, packaging of DNA is initiated by a site–specific endonuclease reaction, at the *pac* site, which is located within gene *3*. A first headful with a terminal redundancy of about 4% is packaged, the DNA is then cut from the concatemer by a non–specific endonuclease, leading to sequential packaging in one direction (Strobel, *et al.*, 1984). It has been shown that the *pac*–recognising protein gp3 is indispensable for *in vitro* encapsulation (Strobel, *et al.*, 1984). When the ST64T genome was digested with *Sma*I restriction enzyme, no under–represented or 'submolar' band was detected. Restriction of the genome with *Eco*RI, yielded 7 bands (A–G) with band D signifying terminal redundancy and circular permutation (Figure 4.19). Comparison of the *Eco*RI restriction sites on both ST64T and P22 genomes, revealed a relationship between these two phages with the A, C and D bands in both phages being of the same size.

Terminase large subunit (gp2). This 1,500 bp ORF (50 mol% GC) encodes a protein of 499 aa with a mass and pI of 57,599 Da and 5.66, respectively. A probable RBS (AGGAG) was located 10 bp upstream of the putative GUG initiation codon. BlastP revealed 98% sequence identity to $gp2_{P22}$ and 78% identity to $gp2_{LP7}$. Additionally, $gp2_{ST64T}$ exhibited 47% sequence identity to a probable phage-like gp2 protein in *Sinorhizobium meliloti* (accession no. NP_386122). This protein (gp2) appears to form a structural complex with gp3 and both may have a role in establishing contact between DNA and a prohead during packaging (Susskind and Botstein, 1978a).

4.3.3.12 Genes involved in morphogenesis

The pathway for capsid morphogenesis of ST64T mimicks that of P22. The genes are tightly packed with little or no intergenic non–coding regions.

Portal (gp1). This ORF (51 %mol GC) has a probable RBS (GAGGA) 5 bp upstream of the putative initiation codon. It overlaps the termination codon of gp2 (UAAUG) where UAA is a stop codon of gp2 and AUG is the start codon of gp1. It encodes a putative protein with a predicted molecular mass and pI of 82,775 Da and 4.75, respectively. BlastP analysis revealed 98% sequence identity between $gp1_{ST64T}$ and $gp1_{P22}$ proteins. Furthermore, putative $gp1_{ST64T}$ shared 32% sequence identity with $P19_{APSE-1}$ (P19 of *AP*SE–1 encodes a putative portal protein). Portal proteins are minor structural proteins, generally with molecular masses of 42 – 60 kDa and variable isoelectric points (λ, 5.7; HK97, 7.7) (Gilakjan and Kropinski, 1999). Although this is considered a general rule in lambdoid phages, both portal proteins of P22 and ST64T are larger at 82.8 kDa. Alignment of $gp1_{ST64T}$ inferred amino acid sequence with that of $gp1_{P22}$ revealed only one mismatch, at position 87 (Ser to Arg) (Figure 4.20). In P22, this gene product (gp1) is one of the minor proteins found in the prohead. It has been shown that proheads formed in the absence of gp1, are unable to encapsulate DNA (Susskind and Botstein, 1978a). This gene product (gp1) together with gp2 and gp3 enable proheads to encapsulate and cut headfuls of DNA from the bacteriophage DNA concatemer (Susskind and Botstein, 1978a).

Scaffolding protein (gp8). This 912 bp ORF (53 %mol GC), has a weak probable RBS (GAGAGA) 6 bp upstream of the putative initiation codon and encodes a putative protein of 313 aa residues with a predicted mass and pI of 33,612 Da and 5.31, respectively. Alignment of putative $gp8_{ST64T}$ with $gp8_{P22}$ amino acid sequences revealed identical or similar sequences except for one mismatch at position 48 (Tyr to Asp). Additionally, BlastP results revealed a 99% sequence identity between the two sequences. In P22, gp8, a scaffolding protein, and gp5 are the two major prohead proteins which form most of the mass of the prohead, however, gp8 is not found in the mature phage (Susskind

and Botstein, 1978a). Based on high amino acid sequence identity to $gp8_{P22}$, it is probable that $gp8_{ST64T}$ is a scaffolding protein.

Major capsid protein (gp5). This 1,293 bp ORF (53 %mol GC) has a probable RBS (AGGA) 8 bp upstream of the putative initiation codon. This ORF overlaps the termination codon of putative gene *8* (UAAUG) where UAA and AUG are the stop and start codons of genes *8* and *5*, respectively. Gene *5* encodes a putative protein of 430 aa with a predicted mass and pI of 46,744 Da and 4.96, respectively. BlastP analysis revealed 99% sequence identity to $gp5_{P22}$. Alignment of the inferred amino acid sequence of $gp5_{ST64T}$ with that of $gp5_{P22}$, revealed a high similarity, with neutral changes at positions 102 (His to Arg), 154 (Leu to Ile), 267 (Leu to Met) and 276 (Thr to Ala) (Figure 4.21). In P22, gp5 has been shown to be the major protein of the mature phage. Based on high sequence identity to $gp5_{P22}$, $gp5_{ST64T}$ is probably the major head protein of mature ST64T phage. This was confirmed by an SDS PAGE gel where both P22 and ST64T structural proteins were compared. The major capsid proteins of both these phages were shown to have similar molecular weights of approximately 46 kDa (Figure 4.22).

In P22, between genes *5* and *4*, there is an ORF69 with an unknown function. A similar scenario is observed in ST64T where a larger ORF, designated ORF186 is located between genes *5* and *4*. This ORF will not be discussed.

Stabilisation protein (gp4). This 501 bp ORF (50 %mol GC) encodes a 166 aa putative protein with a predicted mass and pI of 17,994 Da and 4.89, respectively. A probable RBS (AGGCGG) was located 5 bp upstream of the putative initiation codon. BlastP analysis revealed that $gp4_{ST64T}$ shared 96% sequence identity with $gp4_{P22}$. Furthermore, this gene product had 36% and 37% sequence identity with putative stabilisation proteins in HK620 and *AP*SE–1, respectively. Comparison of $gp4_{ST64T}$ and $gp4_{P22}$ using Align revealed 97% amino acid similarity with five neutral changes at positions 10 (Ala to Val), 28 (Ile to Val), 59 (Ile to Val), 64 (Asp to Glu) and 121 (Ala to Ser), respectively (Figure 4.23). In P22, and probably in ST64T, gp4, gp10 and gp26, are

stabilisation proteins necessary for complete head formation. Mutants defective in one of these three gene products, often release encapsulated DNA, forming empty heads which are distinguishable from proheads based on morphology and sedimentation properties (Susskind and Botstein, 1978a).

Stabilisation protein (gp10). This 1,419 bp ORF (50 %mol GC) has a probable RBS (GGAGA) 5 bp upstream of its initiation codon and encodes a 472 aa putative protein with a predicted mass and pI of 52,268 Da and 5.77, respectively. This putative gene overlaps the putative gene *4*. A homology search using BlastP revealed 95% sequence identity of $gp10_{ST64T}$ to $gp10_{P22}$, 92% to $HkbV_{HK620}$ and 60% to $P28_{APSE-1}$ (HkbV and P28 are putative stabilisation proteins resembling gp10 of P22). A very high similarity of $gp10_{ST64T}$ to $gp10_{P22}$ and $HkbV_{HK620}$ was observed when amino acid sequences were aligned (Figure 4.24). Based on its sequence similarity to $gp10_{P22}$, $gp10_{ST64T}$ may be involved in the stabilisation of the packaged DNA.

Stabilisation protein (gp26). Like P22, ST64T gene *26* is found downstream of gene *10*. It is 639 bp long (43 mol% GC), has a probable RBS (GGAGU) located 5 bp upstream of its putative initiation codon. It encodes a 212 aa putative protein with a predicted mass and pI of 23,883 Da and 5.05, respectively. In contrast to the other putative proteins involved in morphogenesis, $gp26_{ST64T}$ has a low sequence identity to $gp26_{P22}$. Alignment of $gp26_{ST64T}$ with $gp26_{P22}$ and $HkbW_{HK620}$, revealed a deletion in ST64T of 21 amino acids (from position 92 to 102) (Figure 4.25). However, both the N– and carboxy–domains of gp26 are very similar in ST64T and P22. BlastP analysis revealed only 62% sequence identity of $gp26_{ST64T}$ to $gp26_{P22}$, 54% to $HkbW_{HK620}$ and 34% to $P30_{APSE-1}$. This gene product is also required for packaged DNA stabilisation in P22 (Susskind and Botstein, 1978a).

gp14. This 456 bp ORF (49 %mol GC) has a probable RBS (UGGG) 10 bp upstream of the putative initiation codon which overlaps the termination codon of gp26 (UAAUG) where UAA is a stop codon of gp26 and AUG is the start codon of gp10. It

encodes a 151 aa putative protein with a predicted mass and pI of 17,371 Da and 6.95, respectively. BlastP analysis revealed that $gp14_{ST64T}$ shared sequence identity of 96% and 92% with $HkbX_{HK620}$ and $gp14_{P22}$, respectively. The function of the gene product encoded by this ORF in P22 and hence ST64T is unknown.

DNA transfer protein (gp7). This 690 bp ORF (53 %mol GC) encodes a 229 aa putative protein with a predicted mass and pI of 23,227 Da and 8.83, respectively. A probable RBS (GGGAGA) was located 5 bp upstream of the inferred initiation codon. BlastP analysis revealed 66%, 64% and 59% sequence identity of $gp7_{ST64T}$ to $HkbY_{HK620}$, $gp7_{P22}$ and $P32_{APSE-1}$, respectively (both the putative proteins HkbY of HK620 and P32 of *AP*SE–1 phages are related to the DNA transfer protein of P22).

DNA transfer protein (gp20). This 1,335 bp ORF (50 %mol GC) has a probable RBS (AGGAG) 5 bp upstream of the putative GUG initiation codon. It encodes a putative protein of 433 aa with a predicted mass and pI of 46,621 Da and 5.72, respectively. BlastP analysis revealed that $gp20_{ST64T}$ shared 66%, 45% and 24% sequence identity with $gp20_{P22}$, $P33_{APSE-1}$ and $HkbZ_{HK620}$, respectively. The N–terminus of $gp20_{ST64T}$ is different from the N–termini of $gp20_{P22}$ and $HkbZ_{HK620}$. There is a considerable difference in the size of the gene products ($gp20_{ST64T}$, 433 aa; $gp20_{P22}$, 471 aa; $HkbZ_{HK620}$, 449 aa) (Figure 4.26).

DNA transfer protein (gp16). The gene product that is specified by the last of these genes (*16*), has a low sequence identity to $gp16_{P22}$ (31%), $P35_{APSE-1}$ (33%) and $HkcA_{HK620}$ (31%). Comparison of the sequence data indicated ST64T putative gp16 protein has an insertion towards its carboxy–terminus and appears to be similar to $HkcA_{HK620}$ at its N–terminus (Figure 4.27). This ORF is 1,977 bp in length (51 %mol GC) and overlaps the putative gene *20*. It has a probable RBS (UGGGGUGA) 3 bp upstream of its initiation codon and encodes a putative 658 aa protein with a predicted mass and pI of 69,745 Da and 9.32, respectively.

4.3.3.13 Probable Immunity I region

Mnt. This small 249 bp ORF (43 mol% GC) has a probable RBS (UGAGG) 8 bp upstream of its putative initiation codon and lies downstream of gene *16*. It encodes a putative protein of 82 aa with a predicted mass and pI of 9,606 Da and 7.95, respectively and is transcribed in the opposite direction to the morphogenesis genes. BlastP analysis revealed a 60% sequence identity to P22 Mnt protein. Both Mnt_{ST64T} and Mnt_{P22} have similar hydropathy profiles (Kyte and Doolittle, 1982) (Figure 4.28). Based upon this comparison, map position and direction of transcription, this putative gene was designated *mnt*. In P22, Mnt is a regulatory protein in the immunity I (*immI*) region which includes two other genes involved in control of lysogeny, a repressor (*arc*) and an antirepressor (*ant*), as well as the *sieA* gene which is involved in superinfection exclusion. In the prophage state, P22 antirepressor synthesis is prevented by the activity of the product of gene *mnt* (Sauer, *et al.*, 1983). The *arc* gene product negatively regulates the level of antirepressor during lytic infection (Susskind, 1980). In this region of P22, two genes (*sieA* and *mnt*) are expressed in a lysogenic state and two (*ant* and *arc*) are expressed early in the lytic cycle (Adams, *et al.*, 1985). Surprisingly, in the ST64T genome, an ORF encoding a putative Mnt homologue was found without the other gene products, Ant and Arc. Furthermore, there were no ORFs encoding unknown proteins, which may be distinct but analogous to the genes *ant* and *arc* of P22 or of any other phage.

4.3.3.14 Tail protein

gp9. This 2,004 bp ORF (43 %mol GC) encodes a 667 aa putative protein with a predicted mass and pI of 71,995 Da and 5.4, respectively, and is positioned upstream of the putative Mnt protein. A probable RBS (UGGAG) was located 8 bp upstream of the putative initiation codon and is transcribed in the opposite direction to the putative *mnt*. BlastP search revealed 98%, 73% and 64% sequence identity of putative $gp9_{ST64T}$ to the tail spike proteins of P22, SfIV and HK620, respectively. Alignment of $gp9_{ST64T}$, $gp9_{P22}$

and $gp9_{SfVI}$ revealed identical or similar sequences for $gp9_{ST64T}$ and $gp9_{P22}$ except for two mismatches at position 583 (Val to Gly) and position 591 (Trp to Tyr). Even though the tail spike protein of *S. flexneri* phage SfVI showed very low sequence similarity to the P22 and ST64T tail spike proteins, there is a striking similarity at the N–termini of these three gene products (Figure 4.29). A C1–binding site (**TTGC**GAGAGG**TTGT**) was detected at position 39,763 – 39,750 within this ORF. An identical C1–binding site within gene *9* was also shown in P22 (Vander Byl and Kropinski, 2000). This ORF, similarly to P22, probably encodes a putative tail spike protein, which is an endorhamnosidase, capable of hydrolysis of rhamnosyl–1.3–galactose linkages in the O–antigen part of the LPS of *Salmonella* serotypes A, B and D (Iwashita and Kanegasaki, 1973).

4.3.3 Potential C1 binding sites

The P22 C1–binding site recognises a motif TTGC(N_6)TTGY, while its homologue, λ CII protein recognises TTGC(N_6)TTGC. A search for TTGC(N_6)TTGY motif in ST64T nucleotide sequence revealed 11 sites (Table 4.2) which include a probable P_{RE} site (**TTGC**TTTTAG**TTGC**) at position 14,025 – 14,012, located within the *c1* putative gene and transcribed in the leftward direction and a probable P_{a23} (**TTGC**GTGGCT**TTGC**) positioned at 20,783 – 20,770 within the putative *23* gene and also transcribed in the leftward direction. However, examination of the DNA sequence upstream of both the ST64T putative *int* and *xis* genes, did not reveal any TTGCN_6TTGY motif. Nevertheless, there are two C1–binding sites upstream of integrase, which are 100% identical to those of P22 (**TTGC**ATCGGC**TTGC** and **TTGC**AAGGCT**TTGC**). These sites may function in the regulation of integrase expression (Vander Byl and Kropinski, 2000). The significance of the other potential C1–binding sites, primarily located in the morphogenesis genes, is not known. This was also observed in P22 (Vander Byl and Kropinski, 2000). Lambda protein CII stimulates transcription from three

promoters: P_{RE}, (promoter for repressor establishment) P_I (integrase promoter), and P_{aQ} (anti–Q promoter). It is a DNA–binding protein that binds in the homologous "–35" regions of all three promoters. The P22 homologue, C1 protein, stimulates transcription from P_{RE} and P_{a23} (the λ P_{aQ} homologue) (Ho, *et al.*, 1992). Like P22, the ST64T putative C1–binding dependent promoters (P_{RE} and P_{a23}) are located within putative genes *c1* and *23*.

4.3.4 Integration Host Factor (IHF)

Large numbers of pontential IHF–binding sites were identified with the online program MacTargsearch (Goodrich, *et al.*, 1990). These are presented in Table 4.2. The list is restricted to those with MacTargsearch scores of ≥48. A large number of IHF–binding sites was also observed in P22 where the presented list was restricted to the MacTargsearch scores of 50 (Vander Byl and Kropinski, 2000). As in P22, the relevance of these many IHF–binding sites in ST64T is unknown.

4.3.6 Rho–independent terminators and other stem–loop structures

A total of 11 potential rho–independent terminators were identified using the search algorithm of Brendel and colleagues (Brendel, *et al.*, 1986, Brendel and Trifonov, 1984) (Table 4.2). Only three of these terminators are intergenic {(2,871 to 2,900 nt **TTGATCG**TTGTTAC**CGATCAA**TTTTTATT) downstream of *int*, (10,952 to 10,926 nt A**GCCGC**ACTCAGG**CGGCG**GCTGTTGTTTCTT) downstream of *orf232* and (10,956 to 10,928 nt T**GCCGC**TCTATATGG**GCGGC**ATTCTTTTT) downstream of *orf232* and transcribed in the opposite direction}, whereas in P22, most of the rho–independent terminators are intergenic (Vander Byl and Kropinski, 2000). Most of the remaining eight putative rho–independent terminators overlap the 3'end of putative genes.

Using Genebase 1.0 (Applied Maths, Kotrijk, Belgium) program, the following stem–loop structures (inverted repeats) were identified. TGGTTCAGcgtcCTGAACCA (–9.8 kcal/mol) at position 5,734 – 5,753 inside gene *eaa2*, AACAGCCGCaCtcagGcgGCGGCTGTT (–14.8 kcal/mol) at position 10,589 – 10,615 downstream of gene *17*, ATAGCAGAcTcgtaAaTCTGCTAT (–8.5 kcal/mol) at position 11,034 – 11,057, immediately upstream of *orf232*, TCTCTAATACtaacGTATTAGAGA (–13.10 kcal/mol) at position 14,966 – 14,989 and located within gene *18*, ATGAAAATAtcgctTATTTTCAT (–5.60 kcal/mol) at position 17,630 – 17,652 located within *orf98*. The minimum free energy (ΔG) was calculated using the Vienna RNA Secondary Structure Prediction at the website http://rna.tbi.univie.ac.at/cgi–bin/RNAfold.cgi#Results (Matthews, *et al.*, 1999). Several direct repeats (19) were found using Genebase program. A list is presented in Table 4.3.

4.4 Discussion

4.4.1 The O-antigen conversion genes

The genome of ST64T is very similar to that of P22. The putative ST64T O–antigen conversion genes are almost identical to those of P22. Form variation of O antigens have been well characterised in *S.* Typhimurium (Zinder, 1957). *S.* Typhimurium belongs to the *Salmonella* serogroup B which is characterised by the possession of a D–mannose–α1→2–L–rhamnose–α1→3–D–galactose trimeric repeat where the mannose residue (α1→3) is substituted by the 3,6–dideoxy hexose abequose (Weintraub, *et al.*, 1992), an equivalent of O antigen 4. Lysogenisation of cells by P22 results in the addition of α1→6 glucosyl residues on the galactosyl residues leading to the appearance of O antigen 1 (Weintraub, *et al.*, 1992). Smith and Parsell (1974) have shown that O–antigen

conversion from 6_27 to 6_17 in *Salmonella cholera–suis* by a seroconverting phage had an effect on virulence.

A similar phenomenon has been shown in the genus *Shigella,* where integration of different bacteriophages into the host genome can mediate serotype conversion with a consequent increase in virulence (Adhikari, *et al.*, 1999, Allison and Verma, 2000, Guan and Verma, 1998, Mavris, *et al.*, 1997). Serotype conversion has also been shown in *Pseudomonas aeruginosa* (Bergan and Midtvedt, 1975, Kuzio and Kropinski, 1983, Madhubala and Sharma, 1981). Bacteriophage D3 of *P. aeruginosa* has also been shown to mediate serotype conversion (Kuzio and Kropinski, 1983). Studies on P22 showed that the three genes (*gtrA*, *gtrB* and *gtrC*) are involved in serotype conversion, with emphasis on the gene *gtrC* which was suggested to be the main gene involved in the conversion and the *gtrA* and *gtrB* acting as accessory genes (Vander Byl and Kropinski, 2000). It was interesting to note the variation (45 to 31.9%) in the GC content of the three genes involved in seroconversion. The low GC content of *gtrC* has also been observed in the third gene of the conversion cassettes found in P22 (Vander Byl and Kropinski, 2000), *Shigella* and *E. coli* (Allison and Verma, 2000). ST64T has been shown to be capable of mediating O–antigen serotype conversion (Personal communication, Carly Tucker, University of South Australia, BSc Honours project). Furthermore, a 15 bp stem–loop structure, which may act as a rho–independent terminator, was located downstream of the *gtrC* gene.

4.4.2 Integration and excision

In P22, the transition between lytic and lysogenic states is catalysed by the Int protein which belongs to the 30–member Int family of recombinases, facilitating site–specific recombination pathway (Smith-Mungo, *et al.*, 1994). ST64T putative *attP* region, shares high similarity (97% identity) with P22 *attP*. Furthermore, sequences of the

core and arm type Int binding sites as characterised in P22 (Smith-Mungo, *et al.*, 1994), were found almost identical to those of P22 *attP* site. The similarities and the map positions of these two *att* sites, may suggest that their *attB* sites are identical, meaning that ST64T may integrate at the same position in the *Salmonella* chromosome as P22, between *proA* and *proC*.

Unlike P22 P_{aI} and λ P_I, ST64T seems to lack a promoter (P_{aI}) which would be transcriptionally activated by a putative ST64T C1 protein. However, using Martin Reese's Promoter prediction by Neutral Network program, a sequence:

(AG**TTGTTA**TCCCTGACAGCTTCAAGAA**AATAAC**TTACGG**G**AAATT)
−35 −10 +1

with similarity to σ70 promoters in *E. coli* was identified 94 bp upstream of gene *xis*. This promoter is probably responsible for the expression of *int* and *xis* genes. Additionally, there are two other C1–binding dependent promoters upstream of the *int* gene, which may play a role in the expression of integrase and excisionase. Moreover, these two genes overlap each other. The almost 100% (98%) sequence identity between ST64T and P22 Int proteins and the 100% sequence identity of the Xis proteins of these two phages indicate a close phylogenetic relationship between these phages and the possibility that ST64T behaves similarly to P22 in expression of these proteins. Therefore, like P22, the ST64T Int protein probably belongs to the type II Int family recombinases with the unusual feature of being a heterobivalent DNA binding protein (Smith-Mungo, *et al.*, 1994).

4.4.3 Gene *eaa*

The identification of two putative genes *eaa2* and *eaa1* in ST64T instead of a single *eaa* gene as in P22, was unexpected. However, the putative genes *eaa2* and *eaa1* overlap each other and are at a similar map position as the P22 gene *eaa*. The inferred Eaa gene products identified in ST64T may not have any effect on the ST64T bacteriophage life

cycle, because the function of these *eaa–eae* genes in P22 is unknown and may not be required for phage growth.

4.4.4 Genes involved in homologous recombination

The ST64T *abc* genes are very similar to P22 *abc* genes. The products of these genes enable P22 to replicate in the presence of the host's RecBCD function (Murphy, *et al.*, 1987b). The products of P22 genes *abc1* and *abc2* act independently and an increase in expression of one does not suppress the growth defect resulting from lack of the other although the contribution of Abc2 is greater than that of Abc1 (Murphy, *et al.*, 1987b). Furthermore, *abc2* was shown to be analogous to λ *gam* gene (Zissler, *et al.*, 1971) as it is required for P22 growth in a *polA*$^-$ host (Poteete, 1988). The *abc* genes are located downstream of *erf* gene with *abc1* overlapping the termination codon of gene *erf* in both P22 and ST64T. Similarly, the λ gene *exo* is located immediately downstream of *bet* with its initiation codon overlapping the termination codon of *bet.* Although P22 gene *erf* and λ gene *bet* are not homologous on the basis of their amino acid sequences, they are homologues as shown by complementation studies (Poteete and Fenton, 1984) and their map positions. Furthermore, the λ *gam* gene is able to complement P22 *abc*$^-$ function. However, no recombination deficiency could be observed in P22 lacking both or one of the *abc* genes.

4.4.5 Some P_L operon genes

A putative promoter region:

AGTACATT**TTGTAT**TGACGATATTAAG**TACATT**TTGTATT**G**TTTAGCCAT
–35 –10 +1

was identified 98 bp downstream of the *c2* gene using Martin Reese's Promoter Prediction by Neutral Network program (http://www.fruitfly.org/seq_tools/promoter.html). This may

represent the P_L promoter of ST64T, which controls genes expressed by the phage early in infection. The P_L operon in P22 includes genes *24*, *sieB*, *esc*, *17*, *c3*, *arf*, *erf*, *abc1*, *abc2*, *xis* and *int* (Semerjian, *et al.*, 1989, Vander Byl and Kropinski, 2000). Although ST64T is very similar to P22, two genes, *sieB* and *esc* are missing from this region. Like P22, ST64T possesses the gene *17* homologue, which allows the phage to escape exclusion mediated by *Salmonella* prophage Fels–2. A number of *Salmonella* phages are sensitive to superinfection exclusion mediated by the *sieB* gene product. In P22 this prevents growth of superinfecting phages such as L, MG178 and MG40, but does not exclude P22 because of the presence of a small gene *esc,* which makes P22 insensitive to exclusion by *sieB* (Poteete, 1988, Susskind, *et al.*, 1974). Studies have shown that expression of *sieB* leads to abortive infection rather than exclusion of superinfecting phage (Susskind and Botstein, 1978a). Bacteriophage λ has an *sieB* gene mapped in an analogous position in the P_L operon of λ (Susskind and Botstein, 1980). Although there is no significant sequence homology between the P22 and λ *sieB* genes, they both exclude the same *Salmonella* phages (Ranade and Poteete, 1993). P22 also specifies a second exclusion system (A exclusion system) which interferes with development of superinfecting phage. Although the *sieB* gene is located in the P_L operon, it is not functionally part of the operon. It is expressed in the lysogen, unlike all other operon genes (Semerjian, *et al.*, 1989). The A exclusion system which is specified by *sieA* gene, is thought to block entry of superinfecting phage DNA into the cell (Susskind, *et al.*, 1974), thus superinfecting phages cannot express any functions in $sieA^+$ lysogens. In contrast to the B exclusion system, the A system is active not only against superinfecting phages but also against transducing particles (Susskind, *et al.*, 1971). Gene *sieA,* which is positioned adjacent to the *immI* region in P22, is absent at a similar position in ST64T. Unlike *sieB*, there is no λ–analogous *sieA* function known although the *sim* of phage P1 and *imm* of phage T4 seem to be related to *sieA* of P22 since they interfere with the successful injection of phage DNA in the cell (Hofer, *et al.*, 1995). The absence of both genes (*sieB and sieA*) involved

in superinfection exclusion in ST64T, supports the ST64T biological studies (chapter 3). When ST64T was grown on DB21(*sie*), an LT2 derivative which carries a P22 mutant defective in both *sieA* and *sieB* superinfection exclusion systems, there was no growth inhibition. However, P22 could not propagate on the same strain.

P22 and phage L have 95% and 99% nucleotide sequence identity to ST64T at the region between the putative P_L promoter of ST64T and the *abc* genes, respectively. It has been demonstrated that gene *24* encodes an anti–termination protein that acts at terminators within the operon in a way that allows transcription through to the distal genes. This gene occupies a position at the beginning of the P_L operon and has a λ homologue, the gene *N*, which is located at the same map position as gene *24*. The amino terminus of N protein of λ recognises and binds to DNA sites, called *nut*–sites (N–utilisation). The high degree of similarity in the amino terminus of gp24 of phages P22, L and ST64T may indicate that their corresponding *nut*–sites may be identical. This has been proposed for phages P22, ES18 and L (Schicklmaier and Schmieger, 1997). Indeed, when the ST64T nucleotide sequence between the putative *cro* and *c1* genes was analysed, the *nutR* region was located. A search for the *nutL* region between genes *c2* and *24*, led to a region with identical sequence to that described in P22. Both *nutR* and *nutL* are identical to the *nutR* and *nutL* of P22 except for a deletion of one nucleotide (C) in ST64T genome immediately after *boxA* (Figure 4.9)

Transcriptional antitermination in λ requires the phage *nut* sites. The N protein attaches to RNA polymerase at the *nut* sites to form a termination–resistant complex (Friedman and Gottesman, 1983). The *nut* sites can be divided into a promoter–proximal sequence, *boxA* which is conserved amongst the lambdoid phages and *boxB* which is not conserved (Friedman and Gottesman, 1983). Based on the limited C–termini similarity between N_λ, $gp24_{P22}$, $gp24_L$ and $gp24_{ST64T}$, the function of this domain (contact with the RNA polymerase and the NusA protein of the host cell) may occur in a different manner in P22, L and ST64T.

4.4.6 The immunity region

Bacteriophage lysogens are immune to superinfection by homoimmune phages because of the presence of their repressors. In addition to its superinfection exclusion systems A and B discussed above, P22 uses its repressors to exclude homoimmune phages. It has been shown that P22 specifies a bipartite immunity system, with two repressors specified by two different immunity systems, the *c2* repressor gene found in *immC* region and the *mnt* repressor gene located in *immI* region. The gene product of *mnt* prevents the expression of the *ant* gene during lysogeny (Sauer, *et al.*, 1983). Another gene designated *arc*, positioned in the *immI* region, negatively regulates the level of antirepressor during lytic infection (Susskind, 1980). In contrast, phage L possesses the *immC* region and does not have an *immI* region which encodes an antirepressor, gene *ant* and a repressor, the *mnt* gene (Hayden, *et al.*, 1985, Susskind and Botstein, 1978b). Like other lambdoid phages, the C–immunity system of ST64T is located upstream of the anti–termination gene *24*. ST64T is heteroimmune to P22. Its *immC* region is almost 100% similar to that of *Salmonella* phage L. The repressors (C2) of these two phages are 100% identical, with the Cro and C1 proteins having 98 and 96% sequence identity to those of L, respectively. Although the amino termini of P22, L and ST64T C2 proteins are very similar, their N–termini are very different (N–terminus in P22 is responsible for operator recognition and binding). This may explain the heteroimmunity of L and ST64T to P22 and why the promoter/operator regions of both L and ST64T cannot be identified by sequence similarities to those of P22. This explanation has also been proposed for phage L (Schicklmaier and Schmieger, 1997).

Surprisingly, although *immC* region of ST64T is like that of L, and L has no *immI* region, ST64T has a putative *mnt* gene, located upstream of the putative gene *9*, and its inferred gene product shares a significant sequence identity (60%) to the P22 Mnt gene product. P22 antirepressor is capable of inducing any resident lambdoid phage by

neutralising the repressor already present in the cell. However, a P22 prophage is not induced by this antirepressor because of the presence of another repressor, *mnt*, whose gene product prevents transcription of the antirepressor *ant* (Sauer, *et al.*, 1983). Although ST64T lacks the *ant* gene, its Mnt gene product may be capable of neutralising the P22 *ant* gene so as to avoid lytic induction. When ST64T particles were spotted on DB21(ES18) and HisHB22(P22), plaques were seen on DB21(ES18) but not on a P22 lysogen (HisHB22(P22). Similarly, P22 could plaque on an ES18 lysogen but not on ST64T lysogen. It has been shown that ES18 is homoimmune to P22 because their repressors are identical (Schicklmaier and Schmieger, 1997). Therefore, ES18 is unable to propagate in a P22 lysogen. However, P22 is able to propagate in an ES18 lysogen because of the *ant* gene product, which inactivates the repressor of ES18 or any homoimmune phage (Susskind and Botstein, 1978b). It is surprising that ST64T evolved to acquire the *mnt* gene without the whole *immI* region (genes *ant* and *arc*). Alternatively, like P22, ST64T may have possessed the complete set of *immI* genes but somehow lost the other two genes (*ant* and *arc*) due to genetic exchange probably as a result of illegitimate recombination.

Although the putative C1 protein of ST64T has only 49% sequence identity to P22 C1 protein, a few identical TTGC(N_6)TTGY motifs were found in both P22 and ST64T. Even though ST64T and P22 have a very high sequence similarity around the integrase region, a P_{aI} consensus sequence could not be found in ST64T. This may be due to the differences between the putative $C1_{ST64T}$ as opposed to $C1_{P22}$ protein.

4.4.7 The replication genes

Putative genes *18* and *12* (λ homologues *O* and *P*), which are inferred to be involved in the replication of ST64T were found downstream of *c1* gene. This region resembles that of *E. coli* prophage HK022. The conserved 5' terminal nucleotide sequence of O_{HK022}, 18_{P22}, O_{λ} as well as 18_{ST64T} suggests that these phages probably use the same mechanism for binding to the origin of replication. The N–termini of $gp18_{P22}$ and O_{λ} have

been shown to bind to the origin of replication (Schicklmaier and Schmieger, 1997). DNA replication in λ, in addition to the O and P proteins, involves a variety of host proteins, including a primase (DnaG) and a helicase (DnaB) (Taylor and Wegrzyn, 1995). This region also contains DnaA binding sites (Szalewska-Palasz, *et al.*, 1998). In contrast, P22 encodes its own helicase function (Schanda-Mulfinger and Schmieger, 1980) and does not require the binding of the host DnaB. Although $gp12_{ST64T}$ revealed no significant sequence similarity to $gp12_{P22}$, Pfam HMM search revealed conserved motifs of the DnaB–like family of helicases. This strongly suggests that the putative $gp12_{ST64T}$ functions similarly to $gp12_{P22}$, and supplies the helicase function.

4.4.8 The *nin* region

As observed previously in other lambdoid phages, the region between the origin of replication and the late control gene *23* (λ *Q* homologue) is occupied by overlapping genes (*nin*) with unknown function (Karch, *et al.*, 1999). However, the conservation of this region in most phages suggests it may serve some global function. Some genes in this region in λ (*nin5* and *d2*) have been deleted without loss of viability (cited in Karch, *et al.*, 1999). However, phages with these deletions were defective in late gene regulation and the replication efficiency was also reduced (Karch, *et al.*, 1999). In most lambdoid phages, this region consists of at least 9 genes. However, the same region in ST64T had 4 putative genes upstream of the putative *ninB* (*orf101*). Nevertheless, the feature of the overlapping *nin* genes (Kroger and Hobom, 1982) was observed in this region of ST64T.

4.4.9 Gene 23 and the lysis genes

Gene *23*, encodes a delayed early function and is located downstream of the *nin* region. The gp23 protein functions as a second transcriptional antiterminator that regulates expression of the late phage genes by modifying transcription complexes initiating at the

late promoter (Neely and Friedman, 1998). Upstream of λ, there is a $P_{R'}$ promoter with the $t_{R'}$ terminator located directly upstream of the coding region of the lysis proteins (Neely and Friedman, 1998). ST64T showed a similar arrangement. A putative strong promoter:

AAGTC**TTGACA**TGCAATTATCCATTTAGC**CATAAT**TAGAG**G**GTAAGCT
−35 −10 +1

was identified downstream of gene *23* using Martin Reese's Promoter Prediction by Neutral Network program (http://www.fruitfly.org/seq_tools/promoter.html). A rho−independent terminator was also identified immediately upstream of gene *13* (Table 4.2).

Interestingly, the lysis genes of ST64T, which are located immediately downstream of gene *23*, share the same lysis modules with PS3. Indeed, the ST64T lysis cassette resembles that of PS3. The lysis gene cassette in λ and P22 consist of genes *S*, *R*, R_Z and *13*, *19*, *15*, respectively. Most lambdoid phages have homologues of these genes with ST64T possessing *13*, *19* and *15* genes similar to those found in PS3. In λ, the *S* and *R* genes encoding a holin and an endolysin, are required for host lysis (Zhang and Young, 1999). The *R* gene encodes an endolysin that degrades peptidoglycan of the cell wall. The R protein lacks a signal sequence but it is transported across the cytoplasmic membrane to the periplasm by the holin (S), which creates membrane lesions (Neely and Friedman, 1998, Zhang and Young, 1999). The holins are conserved and share an unusual feature of the dual start motif. It has been proposed that the shorter protein forms the holin ring in the membrane whereas the longer protein inhibits ring formation. This may imply that the relative concentration of the two gene products would determine timing of lysis during lytic phage growth (Neely and Friedman, 1998). Although ST64T encodes a dual start motif similar to many other lambdoid phages, these motifs occur at Met−1 and Met−5 unlike the Met−1 and Met−4 or Met−1 and Met−3 motifs observed in other phages. Most lambdoid phages share this dual motif feature, although the holin of phage H−19B phage, is an exception (Neely and Friedman, 1998). It was proposed that this phage may use a

different mechanism to modulate the action of its holin (Neely and Friedman, 1998). Nevertheless, in their micro–review, Bläsi and Young stated that the dual motif is separated by one or two codons, with one of the codons specifying Arg or Lys (Bläsi and Young, 1996). Both Arg and Lys are found between the two potential Met initiation codons of ST64T holin protein. If ST64T modulates the action of its gene *13* by the same mechanism used by P22 and λ, then it would mean that both ST64T and PS3 represent lambdoid phage holins whose dual–start motifs are separated by three codons.

4.4.10 Genes involved in DNA packaging

One of the major differences between P22 and λ is their DNA packaging processes. In λ, the multimeric DNA binds to the products of genes *A* and *Nu1* and the phage prohead. Concatemeric DNA is cleaved by a terminase at specific sites resulting in cohesive 5' extended termini. The DNA is condensed into the phage head with the help of the phage packaging proteins as well as host factors (Hohn, 1983). In contrast, DNA packaging in P22 and P22–like phages, proceeds from a unique site called *pac* which is located within gene *3*. This results in a head–full packaging of a series of terminally redundant, circularly permuted DNA (Španová, 1992, Strobel, *et al.*, 1984, Vander Byl and Kropinski, 2000). P22 packages about 43.4 kb of DNA with 1.7 kb terminal direct repeats (Poteete, 1988) and is 5 to 8% circularly permuted (Vander Byl and Kropinski, 2000). However, a 2.2% terminal redundancy has been reported (Španová, 1992). Interestingly, $gp3_{ST64T}$ is identical to $gp3_{PS3}$ and is also similar to $gp3_{L7}$, $gp3_{PS34}$ and $gp3_{P22}$ with more than 95% sequence identity. PS3, LP7 and PS34 belong to the lambdoid phages and may further be classified as P22–like phages. Furthermore, $gp2_{ST64T}$, which is also involved in DNA packaging is closely related to $gp2_{P22}$ with 98% amino acid sequence identity. It is documented that in P22, DNA packaging by head–full mechanism occasionally may initiate at a *pac*–like site in the host cell chromosome, that may lead to the production of

sequential series of transducing particles (Schmieger, 1982). Based upon the high relatedness of both the packaging genes of ST64T and those of P22 and P22–like phages, it is likely that ST64T uses a head–full packaging mechanism of a series of terminally redundant, circularly permuted DNA. This is further supported by the presence of the under–represented band in the ST64T genome (Figure 4.20). Moreover, ST64T has been shown in this study (chapter 3) to be capable of generalised transduction.

4.4.11 Genes involved in morphogenesis

Most dsDNA phages consists of several copies of a single major protein subunit in addition to a small number of minor proteins (Becker, *et al.*, 1997). The assembly of phage heads constitutes a multistep pathway model for icosahedral viral morphogenesis which includes, assembly of proheads, recognition and cleavage of concatemeric DNA, maturation of proheads to heads, normally in parallel with energy–dependent DNA encapsidation and ultimately head completion (Casjens and Hendrix, 1988). In the initial step of P22 assembly, 420 copies of coat protein (gp5) assemble into a precursor prohead together with gp8, which is a scaffolding protein (Tuma, *et al.*, 1998). Both the coat and the scaffolding proteins form the mass of the prohead (Becker, *et al.*, 1997, Susskind and Botstein, 1978a). Transformation of the prohead (diameter 58 nm) into a mature capsid (diameter 63 nm) involves the loss of the scaffolding protein (gp8), leaving gp5 as the major mature capsid protein (Susskind and Botstein, 1978a, Tuma, *et al.*, 1998). In ST64T, most of the gene products involved in capsid morphogenesis showed high sequence identity (92 to 99%) to P22 morphogenesis proteins, except for the three minor proteins involved in DNA transfer which showed lower sequence identity to P22 proteins (gp7, 66%; gp20, 66%; gp16, 31%). To estimate the size of the major capsid protein of ST64T, SDS–PAGE of phage particle proteins was carried out. This indicated that both ST64T and P22 have approximately 46 kDa proteins, which correspond with the predicted

molecular weight of these gene products (46.7 kDa). Morphogenesis studies in P22 have shown that proheads may form normally and all other steps in morphogenesis may proceed normally in mutants defective in genes *7*, *20* and *16*, however, after adsorption of these particles to a host cell, DNA injection fails (Susskind and Botstein, 1978a). Although particles lacking gp16 could be rescued by co–adsorption of normal P22 virions, the same could not be done for particles lacking gp20 and gp7. It is likely that capsid assembly in ST64T shares a similar mechanism to that employed by P22.

Many studies have shown that the phage protein responsible for adsorption of P22 to host cells is the product of gene *9* (Botstein, *et al.*, 1973, Susskind and Botstein, 1978a). This gene product encodes a tail spike protein which is an endorrhamnosidase, capable of hydrolysis of rhamnosyl–1.3–galactose linkages in the O–antigen part of the LPS of *Salmonella* serotyes A, B and D (Iwashita and Kanegasaki, 1973). It has been demonstrated that heads lacking gp9 cannot catalyse this hydrolysis (Iwashita and Kanegasaki, 1973, Susskind and Botstein, 1978a). Moreover, gene *9* is located adjacent to the three O–conversion genes *gtrA*, *gtrB* and *gtrC*, recently identified in P22 (Vander Byl and Kropinski, 2000).

The high similarity between the gp9 proteins of both ST64T and P22 (98% amino acid sequence identity) including putative proteins involved in morphogenesis and putative proteins involved in the O–antigen conversion, suggests that ST64T uses a similar mechanism for adsorption onto the host cell.

4.4.12 Conclusion

The ST64T genomic sequence data strongly suggest that *S.* Typhimurium ST64T belongs to the P22–like lambdoid family. Some of the differences observed between ST64T and P22 include the genome size of these phages. The genome of ST64T is 40,679 bp whereas that of P22 is 41,724 bp probably because ST64T lacks the other *immI* genes

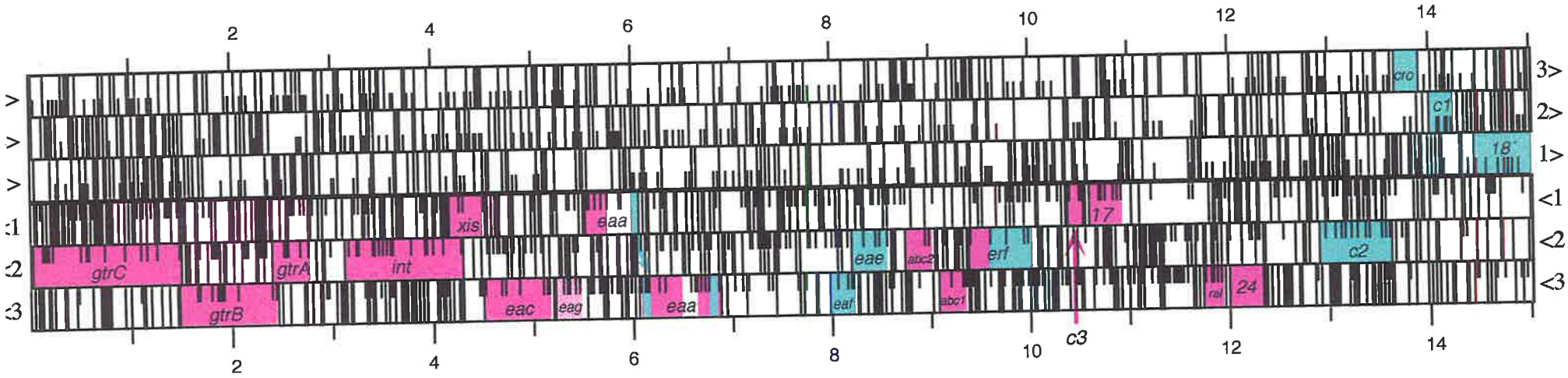
(*ant* and *arc*) as well as both superinfection exclusion genes *sieB* and *sieA*. Interestingly, ST64T genome size is much closer to that of phage L (40,650 ± 0.400 kb) (Hayden, *et al.*, 1985) (and/or 40,500 kb) (Karlovsky, *et al.*, 1984) based on restriction endonuclease digestion of the L genome. The remarkable high sequence similarity in the *immC* region between these two phages (L and ST64T) suggests that they are closely related.

Similarly, the genomic architecture of ST64T is similar to that of P22. A number of regions are very similar to those of P22, including a region encompassing the O–antigen conversion genes and the genes involved in integration and excision that are transcribed in the opposite direction to the structural genes and have 98 to 100% sequence identity to P22. Other inferred gene products predicted to be involved in homologous recombination, packaging as well as many of the gene products involved in morphogenesis were >90% identical to similar proteins of P22. In addition, most other regions, including the region encoding the *nin* genes, were very similar to the P22 counterparts. In contrast, the immunity and the lysis genes were similar to those of phages L and PS3, respectively. However, both L and PS3 are also P22–like phages carried by *S.* Typhimurium. In addition, L has been shown to be a close relative of P22 (Hayden, *et al.*, 1985). Nevertheless, the high sequence similarity of the putative replication genes and the late gene regulator *23* with analogous genes in the "true" lambdoid bacteriophages (HK022 and 21), demonstrates that ST64T, like most of the lambdoid phages, is a mosaic, made up of different "modules" or groups of genes which together perform a particular function. Evolutionary considerations of ST64T as well as that of ST64B, a second bacteriophage isolated from the same *S.* Typhimurium DT 64 (see chapter 5) will be discussed in chapter 6. In conclusion, ST64T genomic sequence suggests that it is a close relative of P22.

Figure 4.1

The open reading frame (ORF) map of phage ST64T prepared with the program DNA Strider 1.3 with AUG and GUG used as start codons. Alignment of ST64T and P22 was carried out using the Compare program of the DNA analysis package MacMolly 3.5. The coloured histogram shows the degree and locations of sequence similarity between ST64T and P22. Numbers below the histograms preceded by the symbol Δ or + indicate positions of deficiency or surplus respectively, of the indicated number of base pairs in ST64T relative to the P22 sequence.

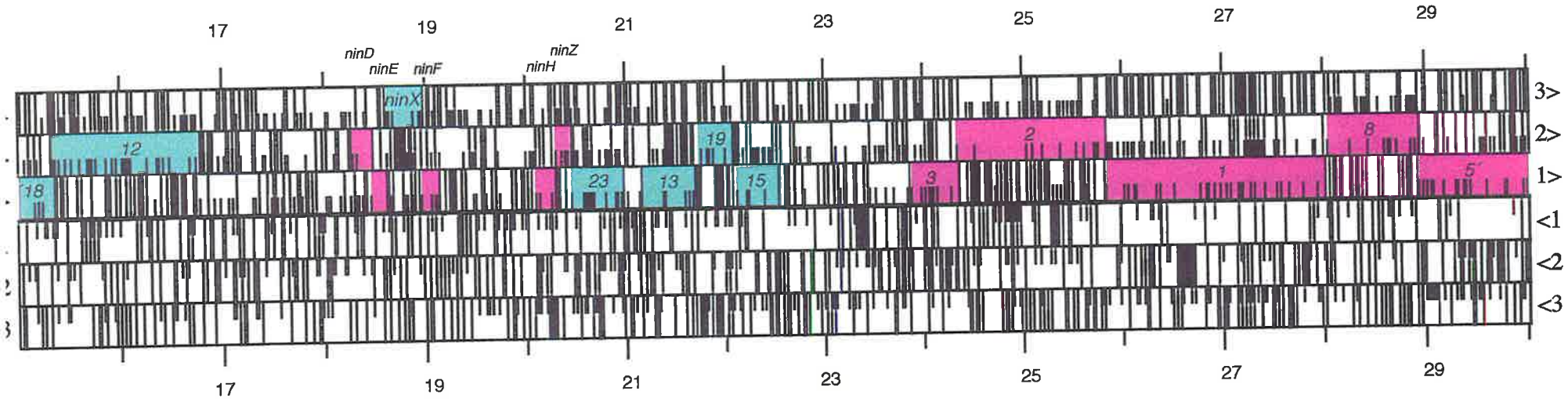
ST64T: 1 - 15 kb

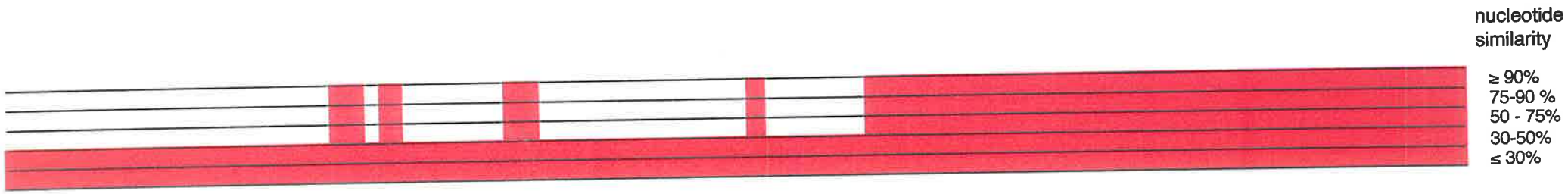


> 85% similarity with P22

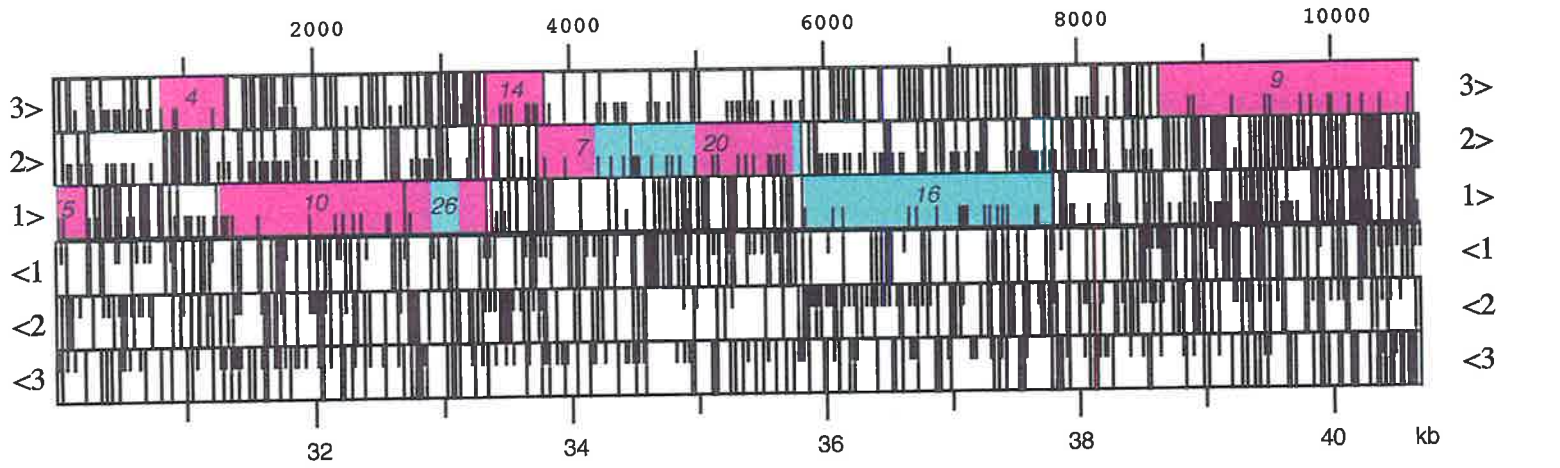
gene assigned according to position and size, weak or no similarity with P22

ST64T: 15 - 30 kb





ST64T: 30 kb - End



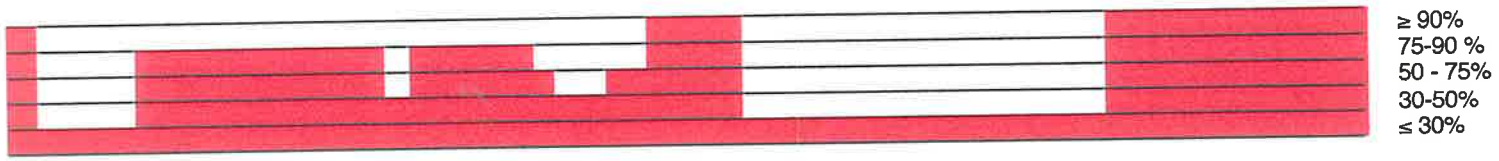


Figure 4.2

Schematic representation of phage ST64T genome showing putative ORFs. The ORFs for which no genetic designation had been previously made are labelled based upon the number of amino acid residues.


High sequence similarity (>90%) with phage P22


Lower sequence similarity (<90%) with phage P22


Sequence similarity with phage L


Sequence similarity with phage HK022


Sequence similarity with phage 21


Sequence similarity with phage PS3


Unknown putative genes or sequence similarity with other phages

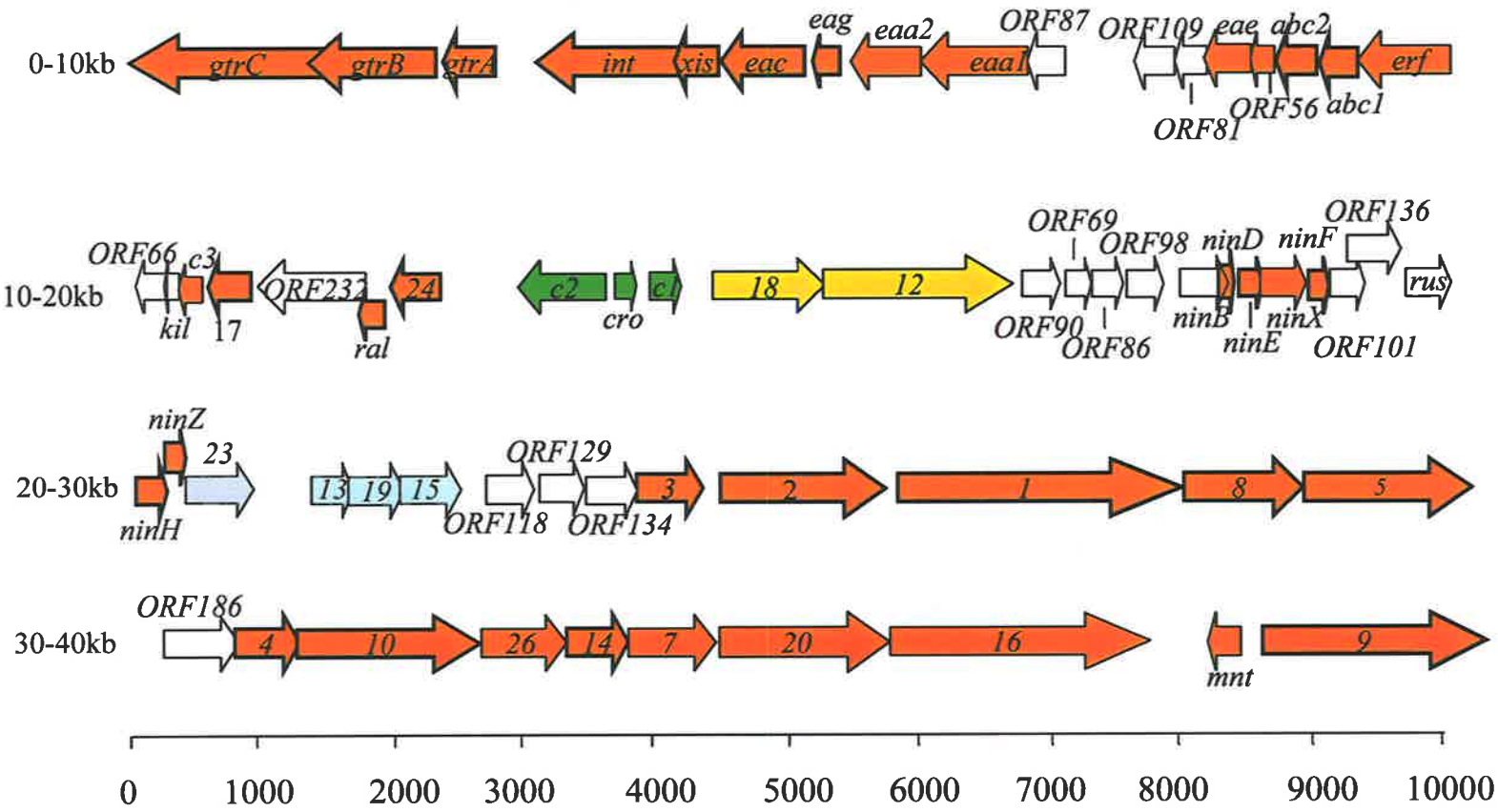

0-10kb
gtrC
gtrB
gtrA
int
xis
eac
eag
eaa2
eaa1
ORF87
ORF109
eae
abc2
erf
ORF81
ORF56
abc1
10-20kb
ORF66
c3
kil
17
ORF233
24
ral
c2
cro
c1
18
12
ORF69
ORF98
ninD
ninF
ORF136
rus
ORF90
ORF86
ninB
ninE
ninX
ORF101
20-30kb
ninZ
23
ninH
13
19
15
ORF129
ORF118
ORF134
3
2
1
8
5
30-40kb
ORF186
4
10
26
14
7
20
16
mnt
9
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

Figure 4.3

Comparison of the region encompassing the putative O–antigen conversion genes *gtrA* to *gtrC* of ST64T with that of P22. The sequence is shown in the sense orientation. The putative promoter region of ST64T is almost identical to that of P22 except for two mismatches in the –10 region. High sequence identity is observed in this region of ST64T and P22.


P22 gtr TGGGAGTCTATCGAAATGGTTCGCCGATATGCTCATCTGGCACCAAATCACCTGACTG
ST64T gtr TGGGAGTCTATCGAAATGGTTCGCAGATATGCTCATCTGGCACCAAATCACCTGACTG
P22 gtr AACATGCTCGACAAATTGACTCGATTTTTGGTACTTCTGTCCCAAATATGTCCCACAG
ST64T gtr AACATGCTCGACAAATTGACTCGATTTTTGGTACTTCTGTCCCAAATATGTCCCACAG
int
P22 gtr TAAAAATAAGGAAGGCACGAATAATACGTAAGTATTTGATTTAACTGGTGCCGATAAT
ST64T gtr TAAAAATAAGGAAGGCACGAATAATACGTAAGTATTTGATTTAACTGGTGCCGATAAT
P22 gtr AGGAGTCGAACCTACGACCTTCGCATTACGAATTATAAGAACTACCTTTTAAGTCAAC
ST64T gtr AGGAGTCGAACCTACGACCTTCGCATTACGAATTATAAGAACTACCTTTTAAGTCAAC
P22 gtr AACATACCACGTCATACCTGCGCTCACACGTCCCATCTTCGAAAGACATGCAAAGCCT
ST64T gtr AACATACCGCGTCATACCTGCGCTCACACGTCCCATCTTCGAAAAACATGCAAAGCCT
P22 gtr TGCAAACCGATGCAAAGATTTGTATGTCCCATTTTTGTCCCAAACCACTTAGCAATCA
ST64T gtr TGCAAGCCGATGCAAAGCTTTGTGTGTCCCGTTTTTGTCCCAAACCACTTAGCAATCA
-35
-10
P22 gtr GCAATAAAAATTGATCGGTAACAACGATCAATTAACATGCATTATATAGATAAAAACT
ST64T gtr GCAATAAAAATTGATCGGTAACAACGATCAATTAACGAGCAAACAATAACTTTAAAACT
P22 gtr ATCAAATTATACATTATTGATCGCTTGTATCGATCAAAACAATTTGTAGTGCTACACT
ST64T gtr ATCAAAGTTTACATTATTGATCGTTTATATCGATCAAAGCAATTTGTAGTGCTACACT
gtrA
P22 gtr TCAGACCTTTCCGAATCCGCTGATTTTCATAATGTTGAAGTTATTCGCTAAGTACACA
ST64T gtr CCAGACCTTTCCGAATCCGCTGATTTTCATAATGTTGAAGTTATTCGCTAAGTACACA
P22 gtr TCGATCGGTGTTCTTAACACGCTTATTCATTGGGGCGTATTTGCTTTTTGTGTGTATG
ST64T gtr TCGATCGGTGTTCTTAACACGCTTATTCATTGGGGCGTATTTGCTTTTTGTGTGTATG
P22 gtr GGATGCATACGCATCAGGCGCTGGCGAACTTTTCCGGTTTTGTTATCGCCGTATCGTT
ST64T gtr GGATGCATACGCATCAGGCGCTGGCGAACTTTTCCGGTTTTGTTATCGCCGTATCGTT
P22 gtr CAGCTTCTATGCCAATGCGCGTTTCACCTTTAATGCCACCACCACCACGCTTCGCTAC
ST64T gtr CAGCTTCTATGCCAATGCGCGTTTCACCTTTAATGCCACCACCACCACGCTTCGCTAC
P22 gtr ATGATGTATGTGGGATTCATGGGAACACTGAGCGCTGTTGTTGGCTGGATGGCTGACC
ST64T gtr ATGATGTATGTGGGATTCATGGGAACACTGAGCGCTGTTGTTGGCTGGATGGCTGACC
gtrB
P22 gtr AATGTTCCTTGCCACCATTGATTACCCTTATCACTTTCTCGGCAATTAGCCTGGTATG
ST64T gtr AATGTTCCTTGCCACCATTGATTACCCTTATCACTTTCTCGGCAATTAGCCTGGTATG
P22 gtr CGGCTTTATCTATTCCAGATTCATTGTTTTCAGGGATATAAGATGAAAATCTCTCTTG
ST64T gtr CGGCTTTATCTATTCCAGATTCATTGTTTTCAGGGATATAAGATGAAAATCTCTCTTG
P22 gtr TCGTTCCAGTTTTTAACGAAGAAGACACGATACCGATTTTCTATAAAACGGTACGTGA
ST64T gtr TCGTTCCAGTTTTTAACGAAGAAGACACGATACCGATTTTCTATAAAACGGTACGTGA
P22 gtr GTTTAATGAGCTAAAAGAATATGAAATTGAGATCGTTTTTATTAATGATGGTAGTAAA
ST64T gtr GTTTAATGAGCTAAAAGAATATGAAATTGAGATCGTTTTTATTAATGATGGTAGTAAA
P22 gtr GATGCGACTGAATCAATAATTAACAAAATAGCTGCATCTGATCCGCTCGTTATTCCGC
ST64T gtr GATGCGACTGAATCAATAATTAACAAAATAGCCGCATCTGATCCGCTCGTTATTCCGC
P22 gtr TTTCGTTTACGCGAAACTTCGGTAAAGAACCTGCTCTTTTCGCGGGTCTCGACCATGC
ST64T gtr TTTCGTTTACGCGAAACTTCGGTAAAGAACCTGCTCTTTTCGCGGGTCTCGACCATGC
P22 gtr AACCGGGGATGCGGTCATTCCTATTGATGTCGATTTACAGGATCCGATAGAAGTTATC
ST64T gtr AACCGGAGATGCGGTCATTCCTATTGATGTCGATTTACAGGATCCGATAGAAGTTATC
P22 gtr CCTCATCTCATTGAGAAGTGGCAGGCTGGCGCGGATATGGTGCTGGCTAAGCGCTCAG
ST64T gtr CCTCATCTCATTGAGAAGTGGCAGGCTGGCGCGGATATGGTGCTGGCTAAGCGCTCAG
P22 gtr ACCGCTCAACTGACGGGCGCATGAAGCGTAAGACAGCTGAGTGGTTTTATAAGCTGCA
ST64T gtr ACCGCTCAACTGATGGGAGGTTGAAGCGTAAGACAGCTGAGTGGTTTTATAAGCTGCA
P22 gtr CAATAAAATCAGCAATCCAAAAATCGAAGAAAATGTTGGCGACTTTCGGTTAATGAGC
ST64T gtr CAATAAAATCAGCAATCCAAAAATCGAAGAAAATGTTGGCGACTTTCGGTTAATGAGC

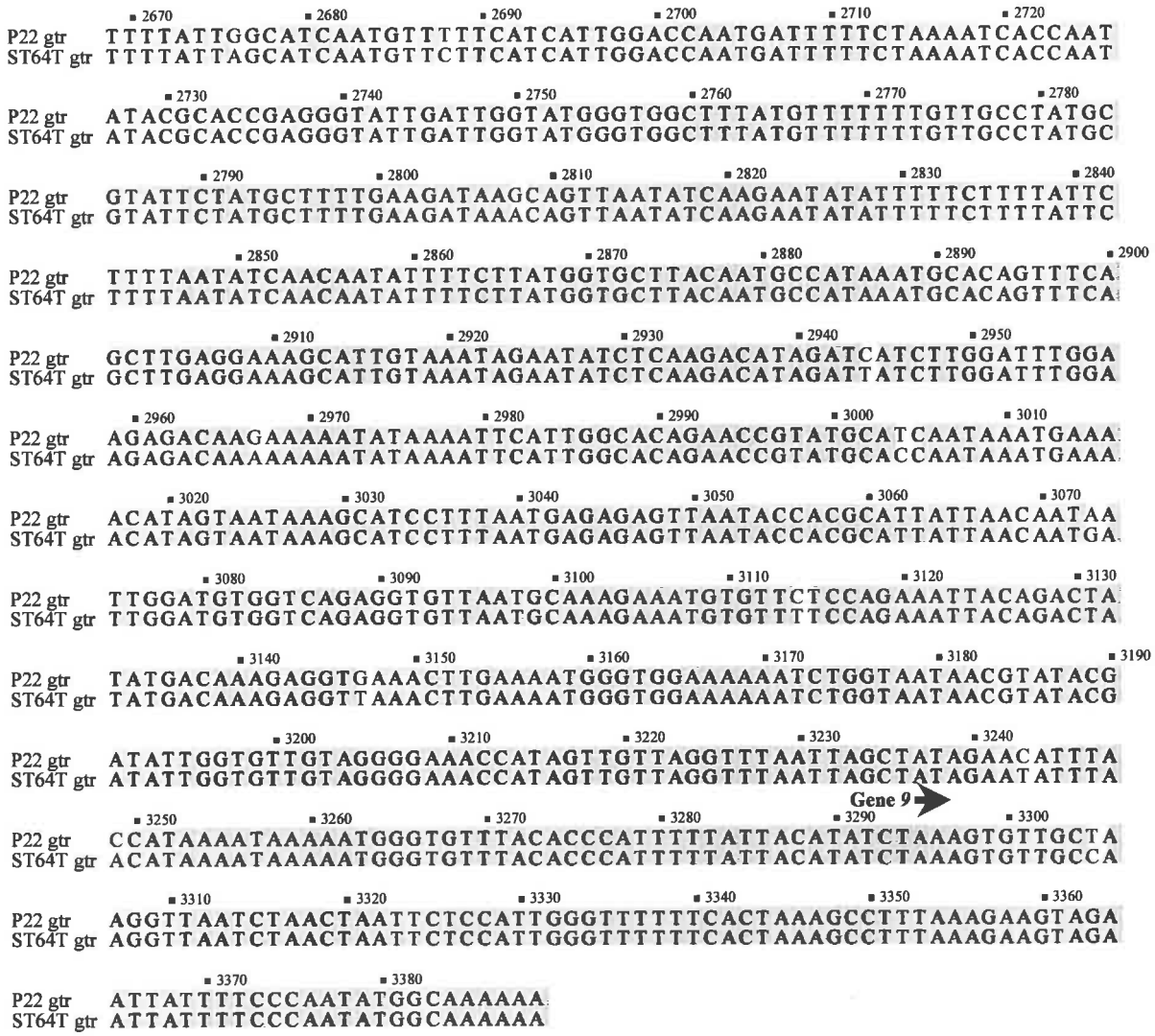

2670 2680 2690 2700 2710 2720
P22 gtr TTTTATTGGCATCAATGTTTTTCATCATTGGACCAATGATTTTTCTAAAATCACCAAT
ST64T gtr TTTTATTAGCATCAATGTTCTTCATCATTGGACCAATGATTTTTCTAAAATCACCAAT
2730 2740 2750 2760 2770 2780
P22 gtr ATACGCACCGAGGGTATTGATTGGTATGGGTGGCTTTATGTTTTTTTGTTGCCTATGC
ST64T gtr ATACGCACCGAGGGTATTGATTGGTATGGGTGGCTTTATGTTTTTTTGTTGCCTATGC
2790 2800 2810 2820 2830 2840
P22 gtr GTATTCTATGCTTTTGAAGATAAGCAGTTAATATCAAGAATATATTTTTCTTTTATTC
ST64T gtr GTATTCTATGCTTTTGAAGATAAACAGTTAATATCAAGAATATATTTTTCTTTTATTC
2850 2860 2870 2880 2890 2900
P22 gtr TTTTAATATCAACAATATTTTCTTATGGTGCTTACAATGCCATAAATGCACAGTTTCA
ST64T gtr TTTTAATATCAACAATATTTTCTTATGGTGCTTACAATGCCATAAATGCACAGTTTCA
2910 2920 2930 2940 2950
P22 gtr GCTTGAGGAAAGCATTGTAAATAGAATATCTCAAGACATAGATCATCTTGGATTTGGA
ST64T gtr GCTTGAGGAAAGCATTGTAAATAGAATATCTCAAGACATAGATTATCTTGGATTTGGA
2960 2970 2980 2990 3000 3010
P22 gtr AGAGACAAGAAAAATATAAAATTCATTGGCACAGAACCGTATGCATCAATAAATGAAA
ST64T gtr AGAGACAAAAAAAAATATAAAATTCATTGGCACAGAACCGTATGCACCAATAAATGAAA
3020 3030 3040 3050 3060 3070
P22 gtr ACATAGTAATAAAGCATCCTTTAATGAGAGAGTTAATACCACGCATTATTAACAATAA
ST64T gtr ACATAGTAATAAAGCATCCTTTAATGAGAGAGTTAATACCACGCATTATTAACAATGA
3080 3090 3100 3110 3120 3130
P22 gtr TTGGATGTGGTCAGAGGTGTTAATGCAAAGAAATGTGTTCTCCAGAAATTACAGACTA
ST64T gtr TTGGATGTGGTCAGAGGTGTTAATGCAAAGAAATGTGTTTTCCAGAAATTACAGACTA
3140 3150 3160 3170 3180 3190
P22 gtr TATGACAAAGAGGTGAAACTTGAAAATGGGTGGAAAAAATCTGGTAATAACGTATACG
ST64T gtr TATGACAAAGAGGTTAAACTTGAAAATGGGTGGAAAAAATCTGGTAATAACGTATACG
3200 3210 3220 3230 3240
P22 gtr ATATTGGTGTTGTAGGGGAAACCATAGTTGTTAGGTTTAATTAGCTATAGAACATTTA
ST64T gtr ATATTGGTGTTGTAGGGGAAACCATAGTTGTTAGGTTTAATTAGCTATAGAATATTTA
Gene 9
3250 3260 3270 3280 3290 3300
P22 gtr CCATAAAATAAAAATGGGTGTTTACACCCATTTTTATTACATATCTAAAGTGTTGCTA
ST64T gtr ACATAAAATAAAAATGGGTGTTTACACCCATTTTTATTACATATCTAAAGTGTTGCCA
3310 3320 3330 3340 3350 3360
P22 gtr AGGTTAATCTAACTAATTCTCCATTGGGTTTTTTCACTAAAGCCTTTAAAGAAGTAGA
ST64T gtr AGGTTAATCTAACTAATTCTCCATTGGGTTTTTTCACTAAAGCCTTTAAAGAAGTAGA
3370 3380
P22 gtr ATTATTTTCCCAATATGGCAAAAAA
ST64T gtr ATTATTTTCCCAATATGGCAAAAAA

Figure 4.4

Comparison of the putative $GtrB_{ST64T}$ with GtrB proteins in SfII, P22 and SfV using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. $GtrB_{ST64T}$ shares 99% sequence identity with $GtrB_{P22}$ and 88% identity with the *Shigella flexneri* phages, SfII, SfV and SfIV.

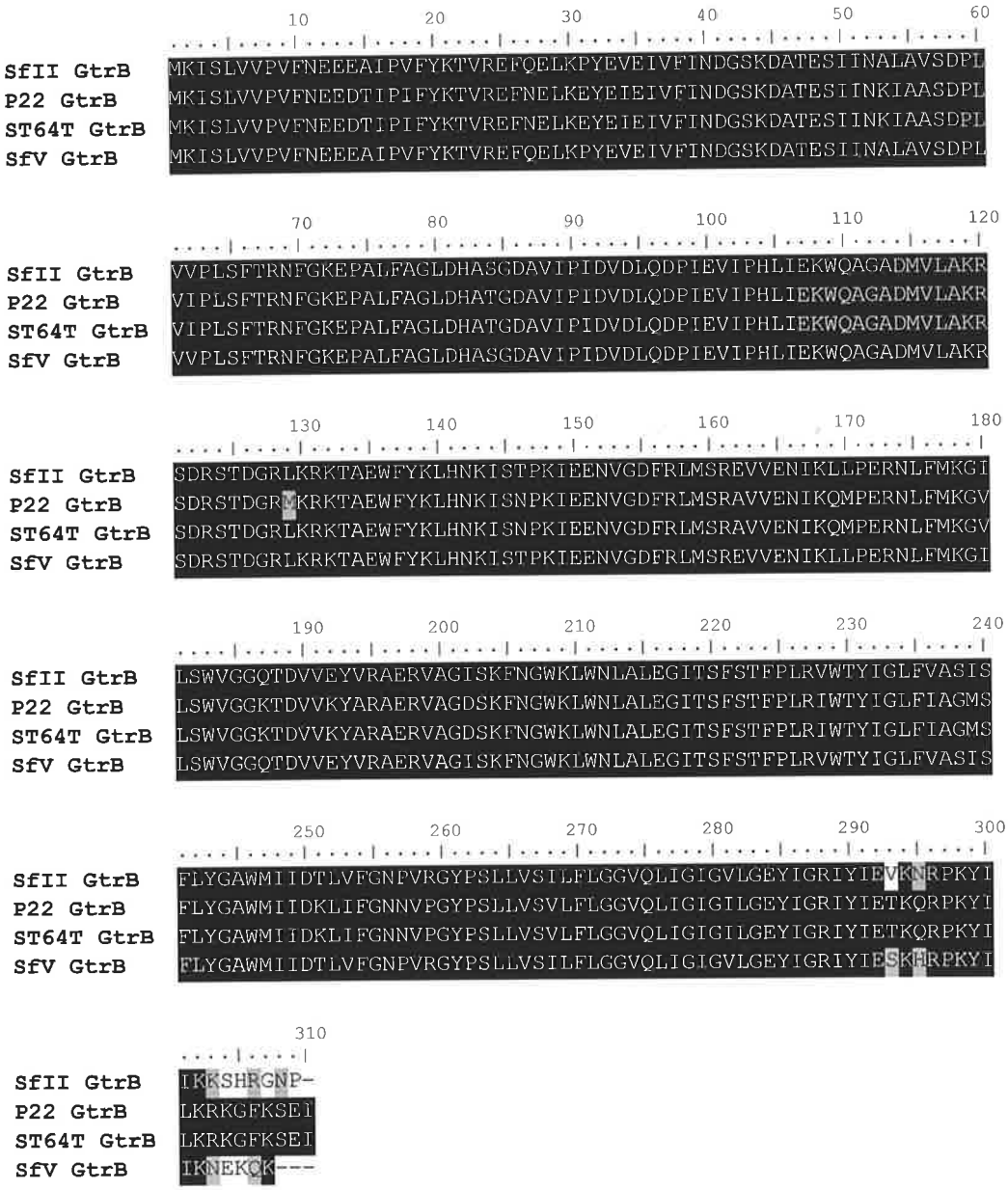

10 20 30 40 50 60
SfII GtrB MKISLVVPVFNEEEAIPVFYKTVREFQELKPYEVEIVFINDGSKDATESIINALAVSDPL
P22 GtrB MKISLVVPVFNEEDTIPIFYKTVREFNELKEYEIEIVFINDGSKDATESIINKIAASDPL
ST64T GtrB MKISLVVPVFNEEDTIPIFYKTVREFNELKEYEIEIVFINDGSKDATESIINKIAASDPL
SfV GtrB MKISLVVPVFNEEEAIPVFYKTVREFQELKPYEVEIVFINDGSKDATESIINALAVSDPL
70 80 90 100 110 120
SfII GtrB VVPLSFTRNFGKEPALFAGLDHASGDAVIPIDVDLQDPIEVIPHLIEKWQAGADMVLAKR
P22 GtrB VIPLSFTRNFGKEPALFAGLDHATGDAVIPIDVDLQDPIEVIPHLIEKWQAGADMVLAKR
ST64T GtrB VIPLSFTRNFGKEPALFAGLDHATGDAVIPIDVDLQDPIEVIPHLIEKWQAGADMVLAKR
SfV GtrB VVPLSFTRNFGKEPALFAGLDHASGDAVIPIDVDLQDPIEVIPHLIEKWQAGADMVLAKR
130 140 150 160 170 180
SfII GtrB SDRSTDGRLKRKTAEWFYKLHNKISTPKIEENVGDFRLMSREVVENIKLLPERNLFMKGI
P22 GtrB SDRSTDGRMKRKTAEWFYKLHNKISNPKIEENVGDFRLMSRAVVENIKQMPERNLFMKGV
ST64T GtrB SDRSTDGRLKRKTAEWFYKLHNKISNPKIEENVGDFRLMSRAVVENIKQMPERNLFMKGV
SfV GtrB SDRSTDGRLKRKTAEWFYKLHNKISTPKIEENVGDFRLMSREVVENIKLLPERNLFMKGI
190 200 210 220 230 240
SfII GtrB LSWVGGQTDVVEYVRAERVAGISKFNGWKLWNLALEGITSFSTFPLRVWTYIGLFVASIS
P22 GtrB LSWVGGKTDVVKYARAERVAGDSKFNGWKLWNLALEGITSFSTFPLRIWTYIGLFIAGMS
ST64T GtrB LSWVGGKTDVVKYARAERVAGDSKFNGWKLWNLALEGITSFSTFPLRIWTYIGLFIAGMS
SfV GtrB LSWVGGQTDVVEYVRAERVAGISKFNGWKLWNLALEGITSFSTFPLRVWTYIGLFVASIS
250 260 270 280 290 300
SfII GtrB FLYGAWMIIDTLVFGNPVRGYPSLLVSILFLGGVQLIGIGVLGEYIGRIYIEVKNRPKYI
P22 GtrB FLYGAWMIIDKLIFGNNVPGYPSLLVSVLFLGGVQLIGIGILGEYIGRIYIETKQRPKYI
ST64T GtrB FLYGAWMIIDKLIFGNNVPGYPSLLVSVLFLGGVQLIGIGILGEYIGRIYIETKQRPKYI
SfV GtrB FLYGAWMIIDTLVFGNPVRGYPSLLVSILFLGGVQLIGIGVLGEYIGRIYIESKHRPKYI
310
SfII GtrB IKKSHRGNP-
P22 GtrB LKRKGFKSEI
ST64T GtrB LKRKGFKSEI
SfV GtrB IKNEKQK---

Figure 4.5

Comparison of the putative *attP* of ST64T with the *attP* of P22. The underlined sequences are regions apparently protected against Dnase I cleavage by Int and IHF. The consensus sequences for the Int core type (C and C'), arm type binding (P1, P2, P3, P'1 and P'2) and IHF binding are indicated by upper case letters. The sites of Int cleavage are indicated by open triangles. The ST64T sequence in this region is almost identical to the P22 sequence, implying that the two phages probably integrate into the bacterial genome at the same site. The P22 sequence was adapted from (Smith-Mungo, *et al.*, 1994).

ST64T

−160 P1 −90 P2
CTGATTGCTAAGTGGT**TTGGGACAAAAA***C***GGGACA***C***AC**AAA*G*CTTTGCATCGGCTTGCAA

−60 P3 IHF
GGCTTTGCATGTTTTTCGAAGA**TGGGACGTGT**GAGCGCAGGTATGA*C*GC*G*GT**ATGTTGT**

C ∇ C' +10
TGACTTAAAAGGTAGTT**CTTATAATT**CGT**AATGCGAA**GGTCGTAGGTTCGACTCCTATTA
Δ

IHF P'1
TCGGCACCAGTTAA**ATCAAATACTT**ACGTATTATTCGTGCCTTCCTTATTTTTACTG**TGGG**A

P'2 +100
CATATT**TGGGACAGAA**GTACCAAAAATCG

P22

−160 P1 −90 P2
CTGATTGCTAAGTGGT**TTGGGACAAAAAA*T*GGGACA*T*AC**AAA*T*CTTTGCATCGGTTTGCAA

−60 P3 IHF
GGCTTTGCATGTCTTTCGAAGA**TGGGACGTGT**GAGCGCAGGTATGACG*T*GGT**ATGTTGTT**

C ∇ C' +10
GACTTAAAAGGTAGTT**CTTATAATT**CGT**AATGCGAAG**GTCGTAGGTTCGACTCCTATTAT
Δ

IHF P'1
CGGCACCAGTTAA**ATCAAATACTT**ACGTATTATTCGTGCCTTCCTTATTTTTACTG**TGGGA**

P'2 +100
CATATT**TGGGACAGAA**GTACCAAAAATCG

Figure 4.6

Comparison of the putative Eaa2 and Eaa1 proteins of ST64T with the Eaa protein of P22 using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. The alignment shows that both putative Eaa1 and Eaa2 proteins of ST64T share sequence identity with Eaa protein of P22, with Eaa1 having high sequence identity with the amino terminus and Eaa2 covering the carboxy–terminus of Eaa_{P22}.

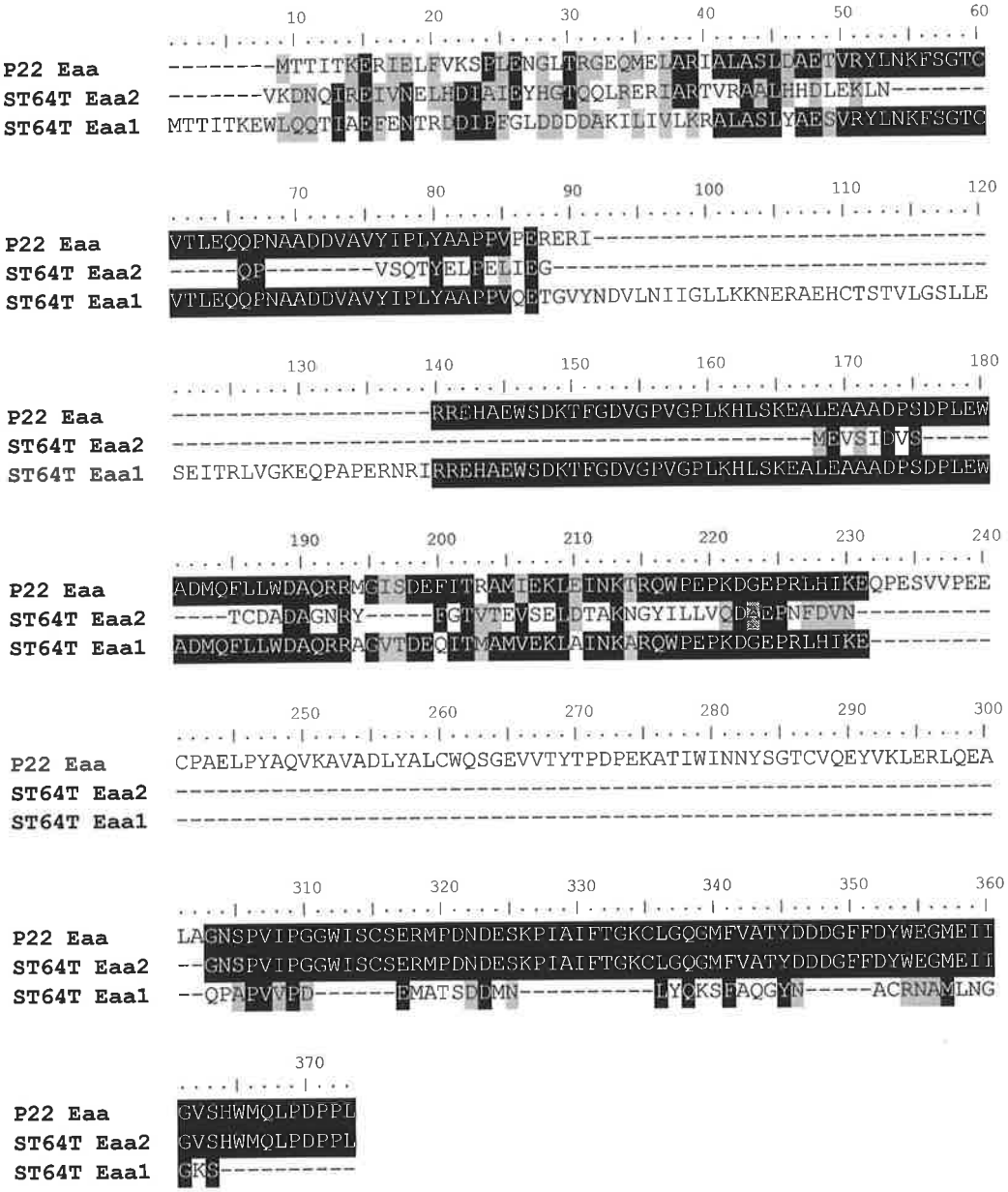

P22 Eaa
ST64T Eaa2
ST64T Eaa1

Figure 4.7

Comparison of the putative Erf_{ST64T} with Erf_{P22} proteins using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align-guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons whilst similar residues are indicated by periods.

Erf

```
                10        20        30        40        50
ST64T  MDLNKFDAPFNPEDIEWRIQQSGKTRDGKVWAMVLAYVTNRAIMKRLDDVCGKE--GWRN
       :. ..: : .   . .    ..  .  ::       : . . :.. .  .      .  .
P22    MS-KEFYARLAEIQEHLNAPKNQYNSFGKY-----KYRSCEDILEGVKPLLKGLFLSISD
                10        20             30        40        50

       60        70        80        90       100       110
ST64T  EYRDIPNNGGVECGISIKIDSEWVTKWDAAENTQVEAVKGGRSGAMKRAAVQWGIGRYLY
       :   : .   :.   .:  :.:  .. .:.  .. :  : : ..:.  .:..    .:
P22    EIVLIGDRYYVKATATIT-DGE--NSHSASAIAREEENKKGMDAAQVTGATSSYARKYCL
           60        70            80        90       100       110

      120       130       140       150       160       170
ST64T  NLEEGFAQISSDKKQGWHRAKLKDGTGFYWLPPSLPDWAMPASCNQPSPENTNQKSPSVD
       :   :.  :.. :    .. : ...             : ::. .. ::     .::.
P22    N---GLFGIDDAKDADTEEHKQQQN------------AAPAKQTKSSP-----SSPAP-
                120       130                    140

      180       190       200       210       220       230
ST64T  CEQILKDFSDYAATETDKKKLIERYQHDWQLLAGHDDAQTRCVQVMNIRINELKQVA
        ::.:: ::.:::::::::::::::::::::::.:::: ::.:::::::::::::::::
P22    -EQVLKAFSEYAATETDKKKLIERYQHDWQLLTGHDDEQTKCVQVMNIRINELKQVA
      150       160       170       180       190       200
```

Figure 4.8

Comparison of the putative $gp24_{ST64T}$ with $gp24_{P22}$ and $gp24_{L}$ using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.


10
20
30
40
50
60
P22 gp24
L gp24
ST64T gp24
MTVITYGKSTFAGNAKTRRHERRRKLAIERDTICNIIDSIFGCDAPDASCEVKAKRIDRV
MTVITYGKSTFAGNAKTRRHERRRKLAIERDTICNIIDSIFGCDVPDASHEVKAKRIDRV
MTVITYGKSTFAGNAKTRRHERRRKLAIERDTICNIIDSIFGCDVPDASHEVKAKRIDRV
70
80
90
100
110
P22 gp24
L gp24
ST64T gp24
TKAISLAGTRQKEVEGGSVLLPGVALYAAGHRKSKQITAR
TKAISLAGTRQKKVEVQR
TKAISLAGTRQKKVEVTAVKKNRIYYRDVNPLGNKIHAVQKQRGKSIPAYYD

Figure 4.9

Comparison of the *nut* sites of phages λ, P22, 21 and ST64T with inverted repeats underlined. The figure shows conservation of *boxA* in the phages indicated. The dashes in the λ sequence were included for better alignment with other sequences. Both the *boxA* and *boxB* of ST64T and P22 are identical.

	boxA	*boxB*
λ *nutR*	**CGCTCTTA**– – –cacattcca	GCCCT**GAAAA**AGGGC
λ *nutL*	**CGCTCTTA**– – –aaaattaa	GCCCT**GAAGA**AGGGC
		boxB
P22 *nutR*	**CGCTCTTTA**cccaatctga	ACCGC**CGACAAC**GCGGT
P22 *nutL*	**CGCTCTTTA**acttcgatgat	GCGCT**GACAA**AGCGC
		boxB
ST64T *nutR*	**CGCTCTTTA**ccaatctga	ACCGCC**GACAAC**GCGGT
ST64T *nutL*	**CGCTCTTTA**acttcgatgat	GCGCT**GACAA**AGCGC
		boxB
21 *nutR*	**TGCTCTTTA**acagttctggcctt	TCACCT**CTAACC**GGGTGG
21 *nutL*	**GGCTCTTTA**acatcgacggac	TCTCAAC**CTAACC**GTTGAGA

Figure 4.10

Comparison of the putative $C2_{ST64T}$ protein with $C2_{P22}$ and $C2_{L}$ proteins using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

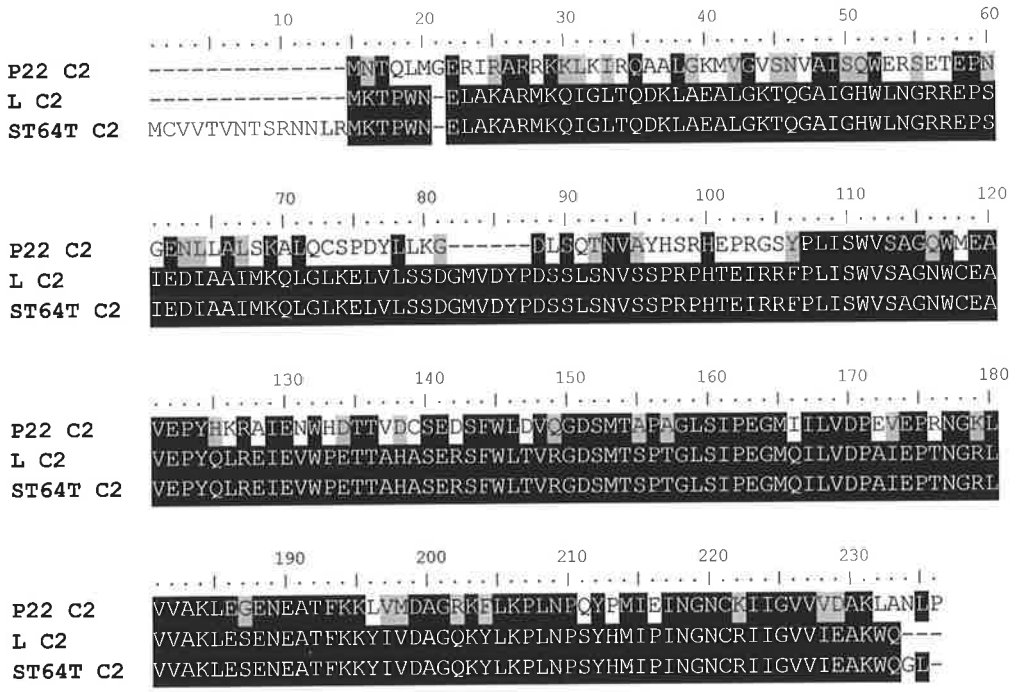

10 20 30 40 50 60
P22 C2 ----------MNTQLMGERIRARRKKLKIRQAALGKMVGVSNVAISQWERSETEPN
L C2 ----------MKTPWN-ELAKARMKQIGLTQDKLAEALGKTQGAIGHWLNGRREPS
ST64T C2 MCVVTVNTSRNNLRMKTPWN-ELAKARMKQIGLTQDKLAEALGKTQGAIGHWLNGRREPS
70 80 90 100 110 120
P22 C2 GENLLALSKALQCSPDYLLKG------DLSQTNVAYHSRHEPRGSYPLISWVSAGQWMEA
L C2 IEDIAAIMKQLGLKELVLSSDGMVDYPDSSLSNVSSPRPHTEIRRFPLISWVSAGNWCEA
ST64T C2 IEDIAAIMKQLGLKELVLSSDGMVDYPDSSLSNVSSPRPHTEIRRFPLISWVSAGNWCEA
130 140 150 160 170 180
P22 C2 VEPYHKRAIENWHDTTVDCSEDSFWLDVQGDSMTAPAGLSIPEGMIILVDPEVEPRNGKL
L C2 VEPYQLREIEVWPETTAHASERSFWLTVRGDSMTSPTGLSIPEGMQILVDPAIEPTNGRL
ST64T C2 VEPYQLREIEVWPETTAHASERSFWLTVRGDSMTSPTGLSIPEGMQILVDPAIEPTNGRL
190 200 210 220 230
P22 C2 VVAKLEGENEATFKKLVMDAGRKFLKPLNPQYPMIEINGNCKIIGVVVEAKLANLP
L C2 VVAKLESENEATFKKYIVDAGQKYLKPLNPSYHMIPINGNCRIIGVVIEAKWQ---
ST64T C2 VVAKLESENEATFKKYIVDAGQKYLKPLNPSYHMIPINGNCRIIGVVIEAKWQGL-

Figure 4.11

Comparison of the putative $C1_{ST64T}$ protein with $C1_{P22}$, $C1_{L\ and}$ CII_{HK620} proteins using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. The amino termini of these proteins show high sequence identity.

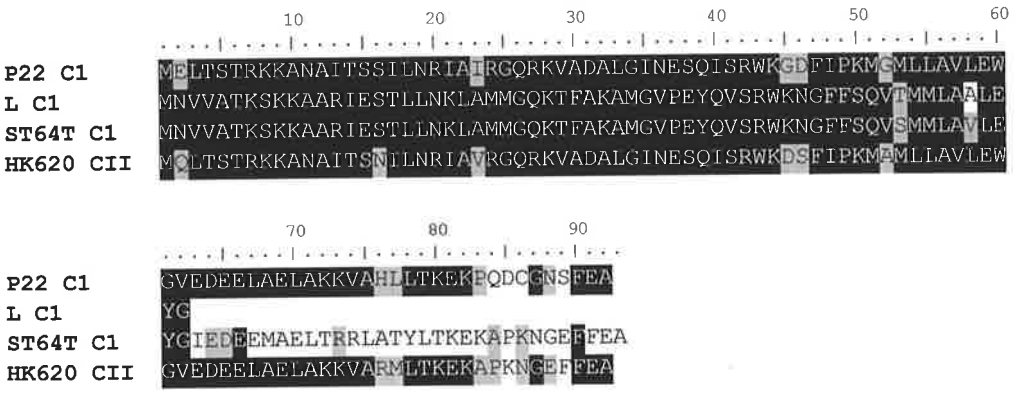

10
20
30
40
50
60
P22 C1
L C1
ST64T C1
HK620 CII
70
80
90
P22 C1
L C1
ST64T C1
HK620 CII

Figure 4.12

Comparison of the putative gp18 protein of ST64T with $gp18_{P22}$, and O_{λ}, O_{HK022} and O_{HK620} proteins using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

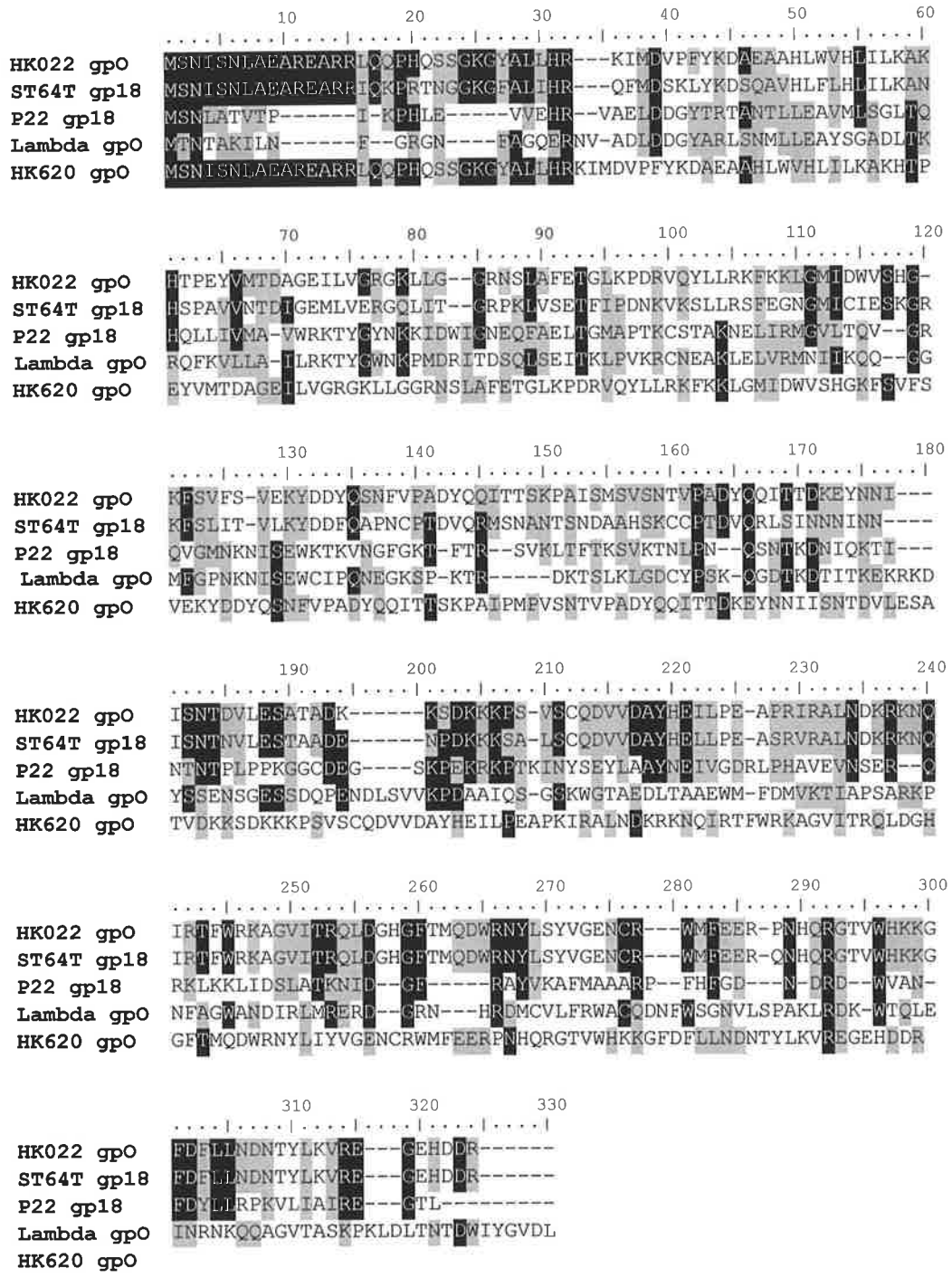

HK022 gpO
ST64T gp18
P22 gp18
Lambda gpO
HK620 gpO

Figure 4.13

Comparison of the putative $gp12_{ST64T}$ protein with gpP_{HK022} and gpP_{HK620} proteins using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. The three proteins share 98% sequence identity.

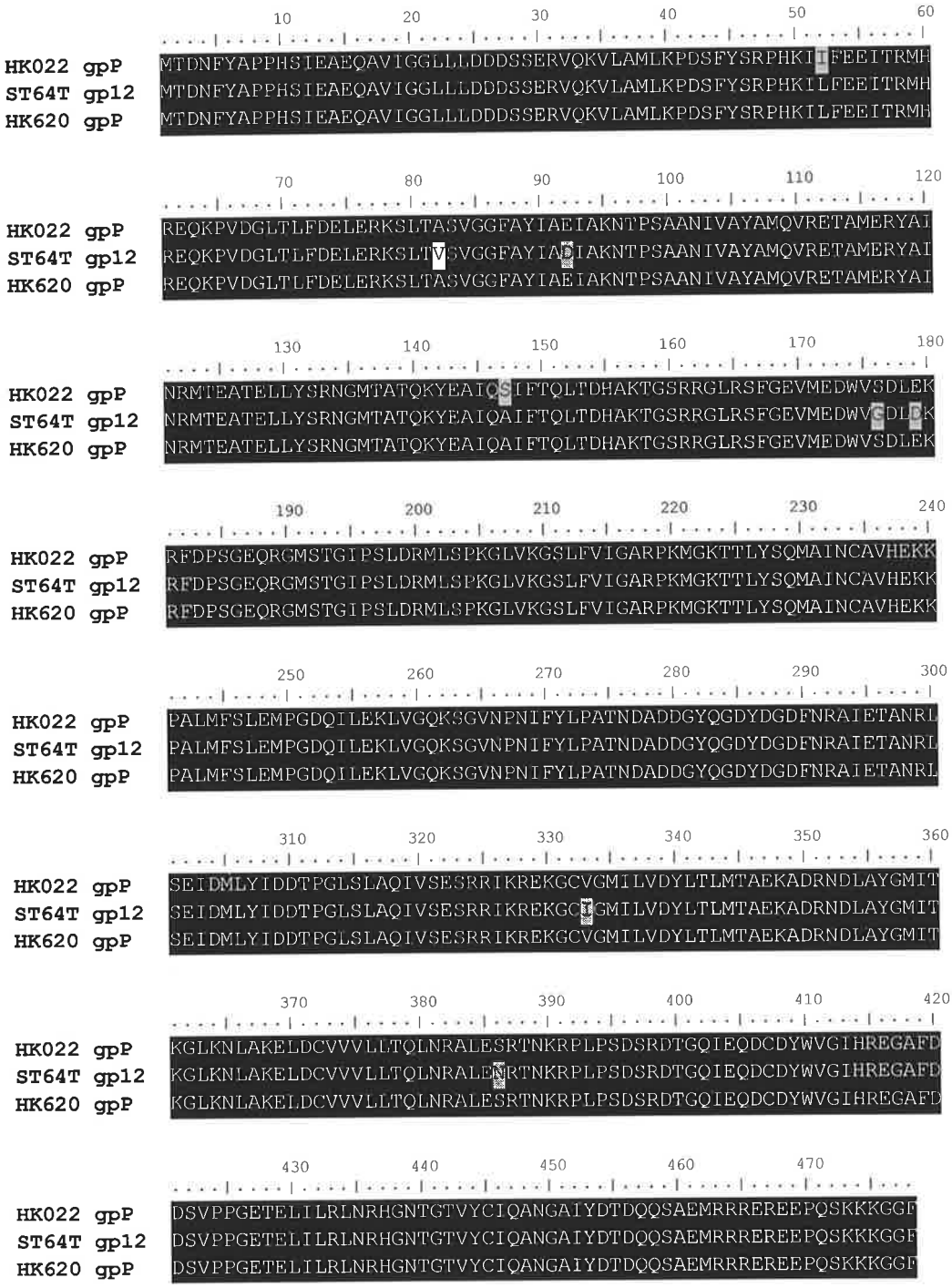

10 20 30 40 50 60
HK022 gpP MTDNFYAPPHSIEAEQAVIGGLLLDDDSSERVQKVLAMLKPDSFYSRPHKILFEEITRMH
ST64T gp12 MTDNFYAPPHSIEAEQAVIGGLLLDDDSSERVQKVLAMLKPDSFYSRPHKILFEEITRMH
HK620 gpP MTDNFYAPPHSIEAEQAVIGGLLLDDDSSERVQKVLAMLKPDSFYSRPHKILFEEITRMH
70 80 90 100 110 120
HK022 gpP REQKPVDGLTLFDELERKSLTASVGGFAYIAEIAKNTPSAANIVAYAMQVRETAMERYAI
ST64T gp12 REQKPVDGLTLFDELERKSLTVSVGGFAYIADIAKNTPSAANIVAYAMQVRETAMERYAI
HK620 gpP REQKPVDGLTLFDELERKSLTASVGGFAYIAEIAKNTPSAANIVAYAMQVRETAMERYAI
130 140 150 160 170 180
HK022 gpP NRMTEATELLYSRNGMTATQKYEAIQSIFTQLTDHAKTGSRRGLRSFGEVMEDWVSDLEK
ST64T gp12 NRMTEATELLYSRNGMTATQKYEAIQAIFTQLTDHAKTGSRRGLRSFGEVMEDWVSDLEK
HK620 gpP NRMTEATELLYSRNGMTATQKYEAIQAIFTQLTDHAKTGSRRGLRSFGEVMEDWVSDLEK
190 200 210 220 230 240
HK022 gpP RFDPSGEQRGMSTGIPSLDRMLSPKGLVKGSLFVIGARPKMGKTTLYSQMAINCAVHEKK
ST64T gp12 RFDPSGEQRGMSTGIPSLDRMLSPKGLVKGSLFVIGARPKMGKTTLYSQMAINCAVHEKK
HK620 gpP RFDPSGEQRGMSTGIPSLDRMLSPKGLVKGSLFVIGARPKMGKTTLYSQMAINCAVHEKK
250 260 270 280 290 300
HK022 gpP PALMFSLEMPGDQILEKLVGQKSGVNPNIFYLPATNDADDGYQGDYDGDFNRAIETANRL
ST64T gp12 PALMFSLEMPGDQILEKLVGQKSGVNPNIFYLPATNDADDGYQGDYDGDFNRAIETANRL
HK620 gpP PALMFSLEMPGDQILEKLVGQKSGVNPNIFYLPATNDADDGYQGDYDGDFNRAIETANRL
310 320 330 340 350 360
HK022 gpP SEIDMLYIDDTPGLSLAQIVSESRRIKREKGCVGMILVDYLTLMTAEKADRNDLAYGMIT
ST64T gp12 SEIDMLYIDDTPGLSLAQIVSESRRIKREKGCIGMILVDYLTLMTAEKADRNDLAYGMIT
HK620 gpP SEIDMLYIDDTPGLSLAQIVSESRRIKREKGCVGMILVDYLTLMTAEKADRNDLAYGMIT
370 380 390 400 410 420
HK022 gpP KGLKNLAKELDCVVVLLTQLNRALESRTNKRPLPSDSRDTGQIEQDCDYWVGIHREGAFD
ST64T gp12 KGLKNLAKELDCVVVLLTQLNRALENRTNKRPLPSDSRDTGQIEQDCDYWVGIHREGAFD
HK620 gpP KGLKNLAKELDCVVVLLTQLNRALESRTNKRPLPSDSRDTGQIEQDCDYWVGIHREGAFD
430 440 450 460 470
HK022 gpP DSVPPGETELILRLNRHGNTGTVYCIQANGAIYDTDQQSAEMRRREREEPQSKKKGGF
ST64T gp12 DSVPPGETELILRLNRHGNTGTVYCIQANGAIYDTDQQSAEMRRREREEPQSKKKGGF
HK620 gpP DSVPPGETELILRLNRHGNTGTVYCIQANGAIYDTDQQSAEMRRREREEPQSKKKGGF

Figure 4.14

Comparison of the putative gp23 protein of ST64T with the Q protein of phage 21 and P5 protein of *AP*SE–1 using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. ST64T gp23 shares 81% sequence identity with phage 21 Q and only 33% identity with *AP*SE–1 putative Q protein.

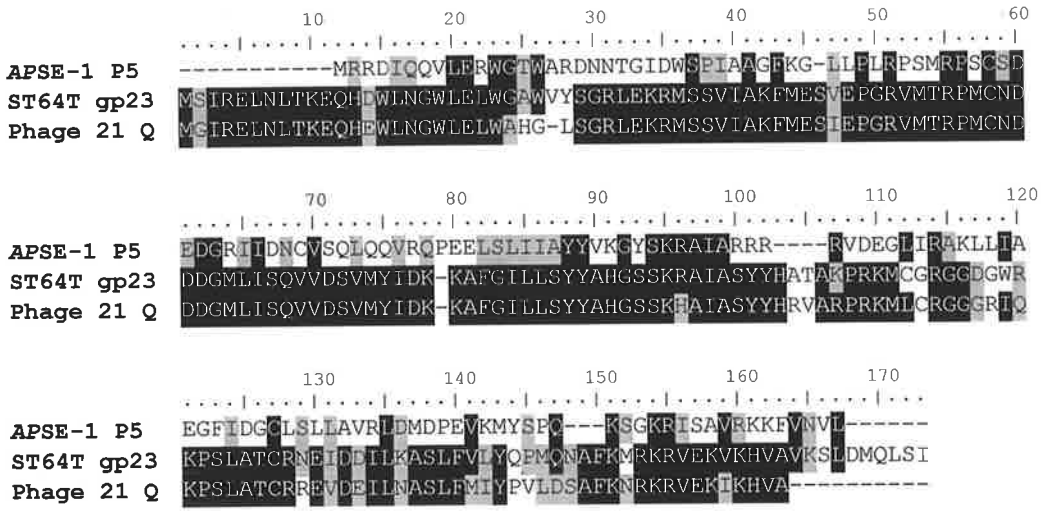

10
20
30
40
50
60
APSE-1 P5
ST64T gp23
Phage 21 Q
70
80
90
100
110
120
APSE-1 P5
ST64T gp23
Phage 21 Q
130
140
150
160
170
APSE-1 P5
ST64T gp23
Phage 21 Q

Figure 4.15

Comparison of the putative gp13 protein of ST64T with the gp13 protein of PS3 using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align-guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons whilst similar residues are indicated by periods.

gp13

```
                10        20        30        40        50        60
ST64T   MAKRMNDDHKIVGLSWLILLGIACWGGLVRYLIDVKQNKATWSWINAFAQIAVSGFTGLI
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
PS3     MAKRMNDDHKIVGLSWLILLGIACWGGLVRYLIDVKQNKATWSWINAFAQIAVSGFTGLI
                10        20        30        40        50        60

                70        80        90       100
ST64T   GGLISVEGGLSFHMILVTSGISGAMGSVALTYFWERLTGMKNANQ
        ::::::::::::::::::::::::::: ::::::::::::::::::
PS3     GGLISVEGGLSFHMILVTSGISGAMGSGALTYFWERLTGMKNANQ
                70        80        90       100
```

Figure 4.16

Comparison of the putative $gp19_{ST64T}$ with the $gp19_{PS3}$ and $gp19_{P22}$ using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. A high sequence identity (99%) is shown between the putative $gp19_{ST64T}$ and $gp19_{PS3}$ with one mismatch at position 81, Trp to Cys, respectively.


10
20
30
40
50
60
PS3 gp19 MQTNKFKFSQRSEKNLNGVNPDLVKVIRRALEITPVDFIVIEGVRTQARQKDMVATGKSQ
P22 gp19 MMQISSNGITRLKREEGERLKAYSDSRGIPTIGVGHTGKVDGNSVASGMTITAEKSSELL
ST64T gp19 MQTNKFKFSQRSEKNLNGVNPDLVKVIRRALEITPVDFIVIEGVRTQARQKDMVATGKSQ
70
80
90
100
110
120
PS3 gp19 TMNSRHLSGNAVDIIPVNTTCKIEEFKPLLKAVKQAADEQGMKLRFGINWKHDPSLPIET
P22 gp19 KEDLQWVEDAISSLVRVPLNQNQYDALCSLIFNIGKSAFAGSTVLRQLNLKNYQAAADAF
ST64T gp19 TMNSRHLSGNAVDIIPVNTTWKIEEFKPLLKAVKQAADEQGMKLRFGINWKHDPSLPIET
130
140
PS3 gp19 KFIDAPHVEIPA
P22 gp19 LLWKKAGKDPDILLPRRRRERALFLS
ST64T gp19 KFIDAPHVEIPA

Figure 4.17

Comparison of the putative $gp15_{ST64T}$ with the $gp15_{PS3}$ and $gp15_{P22}$, $gp15_{PS34}$, Rz_{HK97} using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. A very high level of identity is shown at the carboxy termini of these proteins.

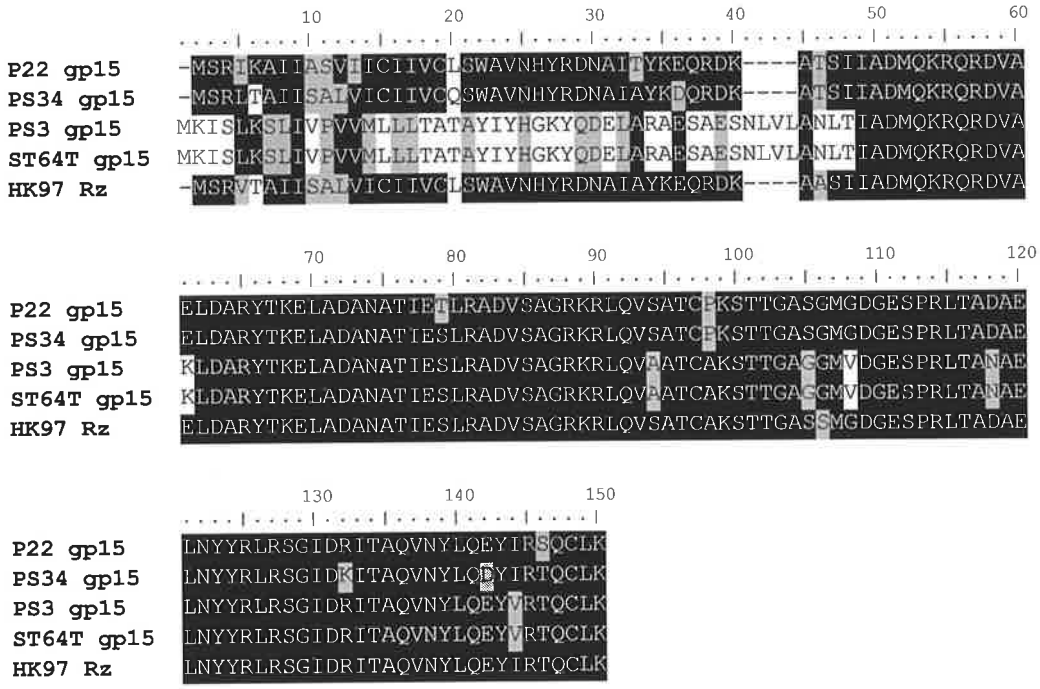

10
20
30
40
50
60
P22 gp15
PS34 gp15
PS3 gp15
ST64T gp15
HK97 Rz
70
80
90
100
110
120
130
140
150

Figure 4.18

Comparison of the putative inferred amino acid sequence of $gp3_{ST64T}$ to $gp3_{PS3}$, $gp3_{LP7}$, $gp3_{PS34}$ and $gp3_{P22}$ using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

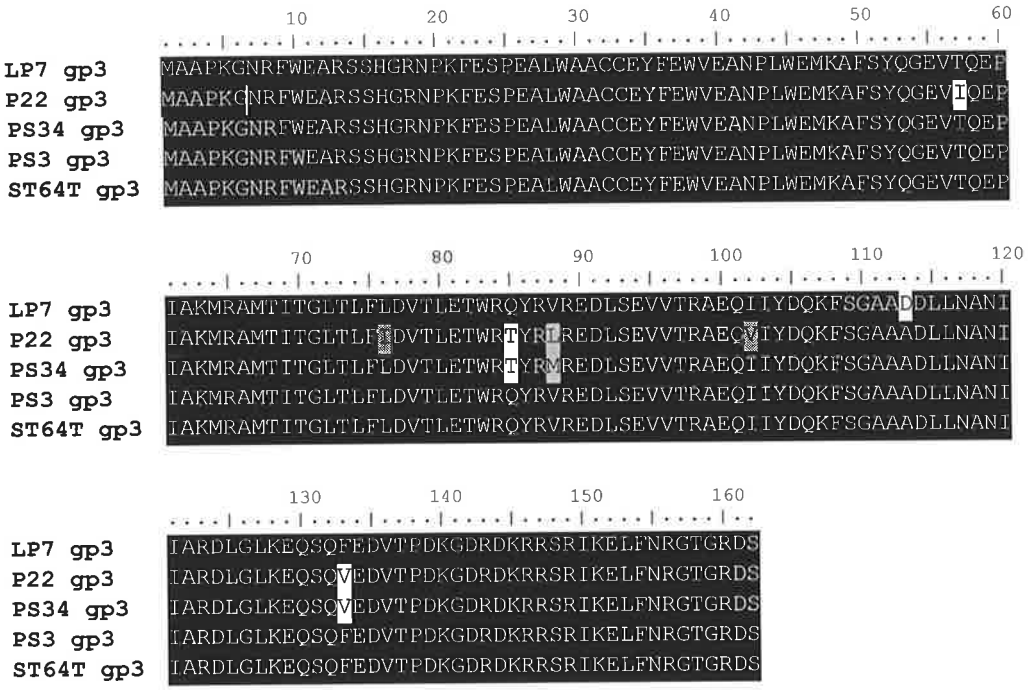

10 20 30 40 50 60
LP7 gp3 MAAPKGNRFWEARSSHGRNPKFESPEALWAACCEYFEWVEANPLWEMKAFSYQGEVTQEP
P22 gp3 MAAPKGNRFWEARSSHGRNPKFESPEALWAACCEYFEWVEANPLWEMKAFSYQGEVIQEP
PS34 gp3 MAAPKGNRFWEARSSHGRNPKFESPEALWAACCEYFEWVEANPLWEMKAFSYQGEVTQEP
PS3 gp3 MAAPKGNRFWEARSSHGRNPKFESPEALWAACCEYFEWVEANPLWEMKAFSYQGEVTQEP
ST64T gp3 MAAPKGNRFWEARSSHGRNPKFESPEALWAACCEYFEWVEANPLWEMKAFSYQGEVTQEP
70 80 90 100 110 120
LP7 gp3 IAKMRAMTITGLTLFLDVTLETWRQYRVREDLSEVVTRAEQIIYDQKFSGAADDLLNANI
P22 gp3 IAKMRAMTITGLTLFLDVTLETWRTYRLREDLSEVVTRAEQVIYDQKFSGAAADLLNANI
PS34 gp3 IAKMRAMTITGLTLFLDVTLETWRTYRMREDLSEVVTRAEQIIYDQKFSGAAADLLNANI
PS3 gp3 IAKMRAMTITGLTLFLDVTLETWRQYRVREDLSEVVTRAEQIIYDQKFSGAAADLLNANI
ST64T gp3 IAKMRAMTITGLTLFLDVTLETWRQYRVREDLSEVVTRAEQIIYDQKFSGAAADLLNANI
130 140 150 160
LP7 gp3 IARDLGLKEQSQFEDVTPDKGDRDKRRSRIKELFNRGTGRDS
P22 gp3 IARDLGLKEQSQVEDVTPDKGDRDKRRSRIKELFNRGTGRDS
PS34 gp3 IARDLGLKEQSQVEDVTPDKGDRDKRRSRIKELFNRGTGRDS
PS3 gp3 IARDLGLKEQSQFEDVTPDKGDRDKRRSRIKELFNRGTGRDS
ST64T gp3 IARDLGLKEQSQFEDVTPDKGDRDKRRSRIKELFNRGTGRDS

Figure 4.19

*Eco*RI digest of P22 and ST64T genomes separated on 1% agarose gel showing a relationship between the two phages with the A, C and D bands in both phages shown to be of the same size. The under–represented band D signifies terminal redundancy in both ST64T and P22.

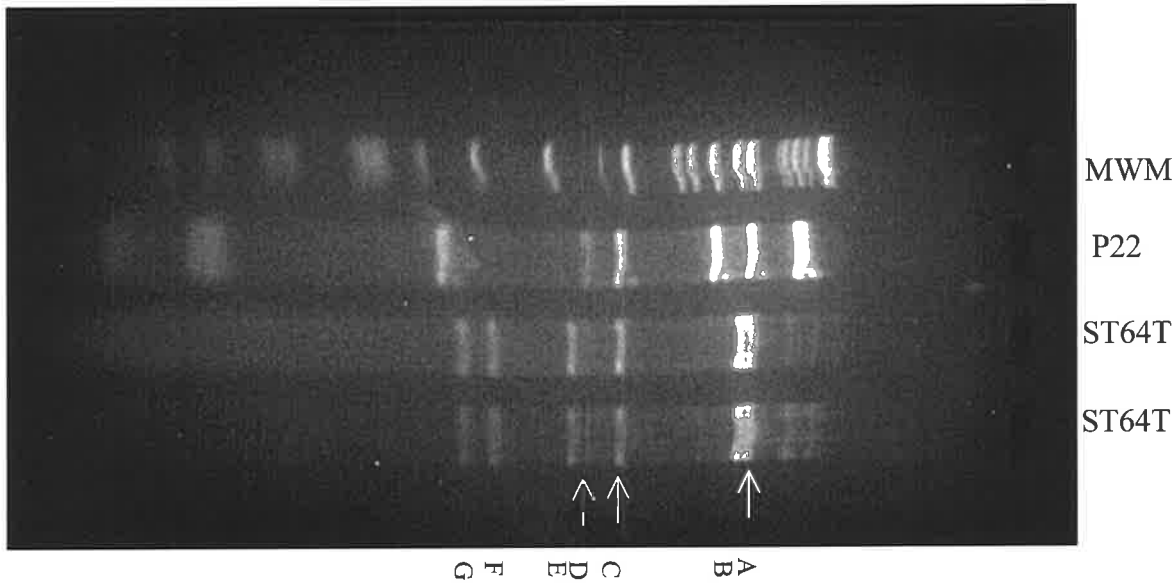

MWM
P22
ST64T
ST64T
A
B
C
D
E
F
G

Figure 4.20

Comparison of the putative inferred amino acid sequence of $gp1_{ST64T}$ with that of $gp1_{P22}$ using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align-guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons whilst similar residues are indicated by periods. Part of the sequence which shows one mismatch at position 87 is presented, otherwise all other residues are identical or similar.

gp1

```
                70        80        90       100       110       120
ST64T   RPVVRKLVSEMRQNPIDVLYRPKDGASPDAADVLMGMYRTDMRHNTAKIAVNVAVREQIE
        :::::::::::::::::::::::::: ::::::::::::::::::::::::::.:::::::
P22     RPVVRKLVSEMRQNPIDVLYRPKDGARPDAADVLMGMYRTDMRHNTAKIAVNIAVREQIE
                70        80        90       100       110       120

               130       140       150       160       170       180
ST64T SGVGAWRLVTDYEDQSPTSNNQVIRREPIHSACSHVIWDSNSKLMDKSDSRHCTVIHSMS
       .:::::::::::::::::::::::::::::::::::::::::::::::::::.:::::::::::
P22    AGVGAWRLVTDYEDQSPTSNNQVIRREPIHSACSHVIWDSNSKLMDKSDARHCTVIHSMS
               130       140       150       160       170       180

               190       200       210       220       230       240
ST64T   QNGWEDFAEKYDLDADDIPSFQNPNDWVFPWLTQDTIQIAEFYEVVEKKETAFIYQDPVT
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     QNGWEDFAEKYDLDADDIPSFQNPNDWVFPWLTQDTIQIAEFYEVVEKKETAFIYQDPVT
               190       200       210       220       230       240

               250       260       270       280       290       300
ST64T   GEPVSYFKRDIKDVIDDLADSGFIKIAERQIKRRRVYKSIITCTAVLKDKQLIAGEHIPI
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     GEPVSYFKRDIKDVIDDLADSGFIKIAERQIKRRRVYKSIITCTAVLKDKQLIAGEHIPI
               250       260       270       280       290       300

               310       320       330       340       350       360
ST64T   VPVFGEWGFVEDKEVYEGVVRLTKDGQRLRNMIMSFNADIVARTPKKKPFFWPEQIAGFE
        :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22      VPVFGEWGFVEDKEVYEGVVRLTKDGQRLRNMIMSFNADIVARTPKKKPFFWPEQIAGFE
               310       320       330       340       350       360

               370       380       390       400       410       420
ST64T   HMYDGNDDYPYYLLNRTDENNGEMPTQPLAYYENPEVPQANAYMLEAATAAVKEVATLGV
        :::::::::::::::::::.:..:::::::::::::::::::::::::::.:::::::::::
P22     HMYDGNDDYPYYLLNRTDENSGDLPTQPLAYYENPEVPQANAYMLEAATSAVKEVATLGV
               370       380       390       400       410       420

               430       440       450       460       470       480
ST64T   DAEAVNGGQVAYDTVNQLNMRADLETYVFQDNLATAMRRDGEIYQSIVNDIYDVPRNVTI
        :.:::::::::.:::::::::::::::::::::::::::::::::::::::::::::::::
P22     DTEAVNGGQVAFDTVNQLNMRADLETYVFQDNLATAMRRDGEIYQSIVNDIYDVPRNVTI
               430       440       450       460       470       480

               490       500       510       520       530       540
ST64T   TLEDGSEKEVQLMAEVVDLATGERQVLNDIRGRYECYTDVGPSFQSMKQQNRAEILELLG
        ::::::::.:::::::::::::::.:::::::::::::::::::::::::::::::::::::
P22     TLEDGSEKDVQLMAEVVDLATGEKQVLNDIRGRYECYTDVGPSFQSMKQQNRAEILELLG
               490       500       510       520       530       540

               550       560       570       580       590       600
ST64T   KTPQGTPEYQLLLLQYFTLLDGKGVEMMRDYANKQLIQMGVKKPETPEEQQWLVEAQQAK
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22      KTPQGTPEYQLLLLQYFTLLDGKGVEMMRDYANKQLIQMGVKKPETPEEQQWLVEAQQAK
               550       560       570       580       590       600

               610       620       630       640       650       660
ST64T   QGQQDPAMVQAQGVLLQGQAELAKAQNQTLSLQIDAAKVEAQNQLNAARIAEIFNNMDLS
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     QGQQDPAMVQAQGVLLQGQAELAKAQNQTLSLQIDAAKVEAQNQLNAARIAEIFNNMDLS
               610       620       630       640       650       660
```

```
                670       680       690       700       710       720
ST64T   KQSEFREFLKTVASFQQDRSEDARANAELLLKGNEQTHKQRMDIANILQSQRQNQPSGSV
        :::::::::::::::::::::::::::::::::.:::::::::::::::::::::::::
P22     KQSEFREFLKTVASFQQDRSEDARANAELLLKGDEQTHKQRMDIANILQSQRQNQPSGSV
                670       680       690       700       710       720


ST64T   AETPQ
        :::::
P22     AETPQ
```

Figure 4.21

Comparison of the putative $gp5_{ST64T}$ and $gp5_{P22}$ proteins using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align-guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons whilst similar residues are indicated by periods.

gp5

```
                10        20        30        40        50        60
ST64T   MALNEGQIVTLAVDEIIETISAITPMAQKAKKYTPPAASMQRSSNTIWMPVEQESPTQEG
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     MALNEGQIVTLAVDEIIETISAITPMAQKAKKYTPPAASMQRSSNTIWMPVEQESPTQEG
                10        20        30        40        50        60

                70        80        90       100       110       120
ST64T   WDLTDKATGLLELNVAVNMGEPDNDFFQLRADDLRDETAYRHRIQSAARKLANNVELKVA
        ::::::::::::::::::::::::::::::::::::::::::.::::::::::::::::::
P22     WDLTDKATGLLELNVAVNMGEPDNDFFQLRADDLRDETAYRRRIQSAARKLANNVELKVA
                70        80        90       100       110       120

               130       140       150       160       170       180
ST64T   NMAAEMGSLVITSPDAIGTNTADAWNFVADAEELMFSRELNRDMGTSYFFNPQDYKKAGY
        ::::::::::::::::::::::::::::::::::.::::::::::::::::::::::::::
P22     NMAAEMGSLVITSPDAIGTNTADAWNFVADAEEIMFSRELNRDMGTSYFFNPQDYKKAGY
               130       140       150       160       170       180

               190       200       210       220       230       240
ST64T   DLTKRDIFGRIPEEAYRDGTIQRQVAGFDDVLRSPKLPVLTKSTATGITVSGAQSFKPVA
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     DLTKRDIFGRIPEEAYRDGTIQRQVAGFDDVLRSPKLPVLTKSTATGITVSGAQSFKPVA
               190       200       210       220       230       240

               250       260       270       280       290       300
ST64T   WQLDNDGNKVNVDNRFATVTLSATTGLKRGDKISFTGVKFLGQMAKNVLAQDATFSVVRV
        ::::::::::::::::::::::::::::.::::::::.::::::::::::::::::::::::
P22     WQLDNDGNKVNVDNRFATVTLSATTGMKRGDKISFAGVKFLGQMAKNVLAQDATFSVVRV
               250       260       270       280       290       300

               310       320       330       340       350       360
ST64T   VDGTHVEITPKPVALDDVSLSPEQRAYANVNTSLADAMAVNILNVKDARTNVFWADDAIR
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     VDGTHVEITPKPVALDDVSLSPEQRAYANVNTSLADAMAVNILNVKDARTNVFWADDAIR
               310       320       330       340       350       360

               370       380       390       400       410       420
ST64T   IVSQPIPANHELFAGMKTTSFSIPDVGLNGIFATQGDISTLSGLCRIALWYGVNATRPEA
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     IVSQPIPANHELFAGMKTTSFSIPDVGLNGIFATQGDISTLSGLCRIALWYGVNATRPEA
               370       380       390       400       410       420

               430
ST64T   IGVGLPGQTA
        ::::::::::
P22     IGVGLPGQTA
               430
```

Figure 4.22

Proteins from ST64T particles separated on an SDS–PAGE gel with a separating gel of 12% acrylamide and 4% acrylamide stacker gel (Laemmli, 1970), to approximate the molecular size of the major head protein. Proteins from P22 particles were run concurrently with the proteins from ST64T as a control. The gel was stained in Coomassie brilliant blue stain. A band of approximately 46 kDa is observed in both ST64T and P22, corresponding to the putative gp5.

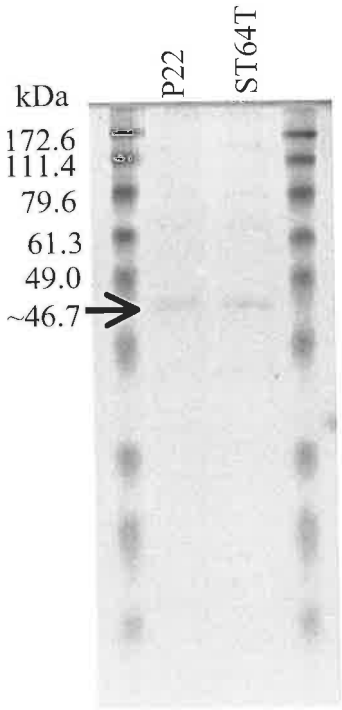

P22
ST64T
kDa
172.6
111.4
79.6
61.3
49.0
~46.7

Figure 4.23

Comparison of the putative $gp4_{ST64T}$ and $gp4_{P22}$ proteins using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align-guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons whilst similar residues are indicated by periods.

gp4

```
                10        20        30        40        50        60
ST64T   MQIKTKGDLARAALRKLGVASDATLTDIEPQSMQDAVDDLEAMMAEWYQDGKGIITGYIF
        :::::::::.::::::::::::::::.:::::::::::::::::::::::::::::::.:
P22     MQIKTKGDLVRAALRKLGVASDATLTDVEPQSMQDAVDDLEAMMAEWYQDGKGIITGYVF
                10        20        30        40        50        60

                70        80        90       100       110       120
ST64T   SDDDNPPAEGDDHGLRSSAVSAVFHNLACRIAPDYALEATAKIIATAKYGKELLYKQTAI
        :::.::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
P22     SDDENPPAEGDDHGLRSSAVSAVFHNLACRIAPDYALEATAKIIATAKYGKELLYKQTAI
                70        80        90       100       110       120

               130       140       150       160
ST64T   ARAKRAPYPSRMPTGSGNSFANLNEWHYFPGEQNADSTTPHDEGNG
        .:::::::::::::::::::::::::::::::::::::::::::::::
P22     SRAKRAPYPSRMPTGSGNSFANLNEWHYFPGEQNADSTTPHDEGNG
               130       140       150       160
```

Figure 4.24

Comparison of $gp10_{ST64T}$, $gp10_{P22}$ and $HkbV_{HK620}$ using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

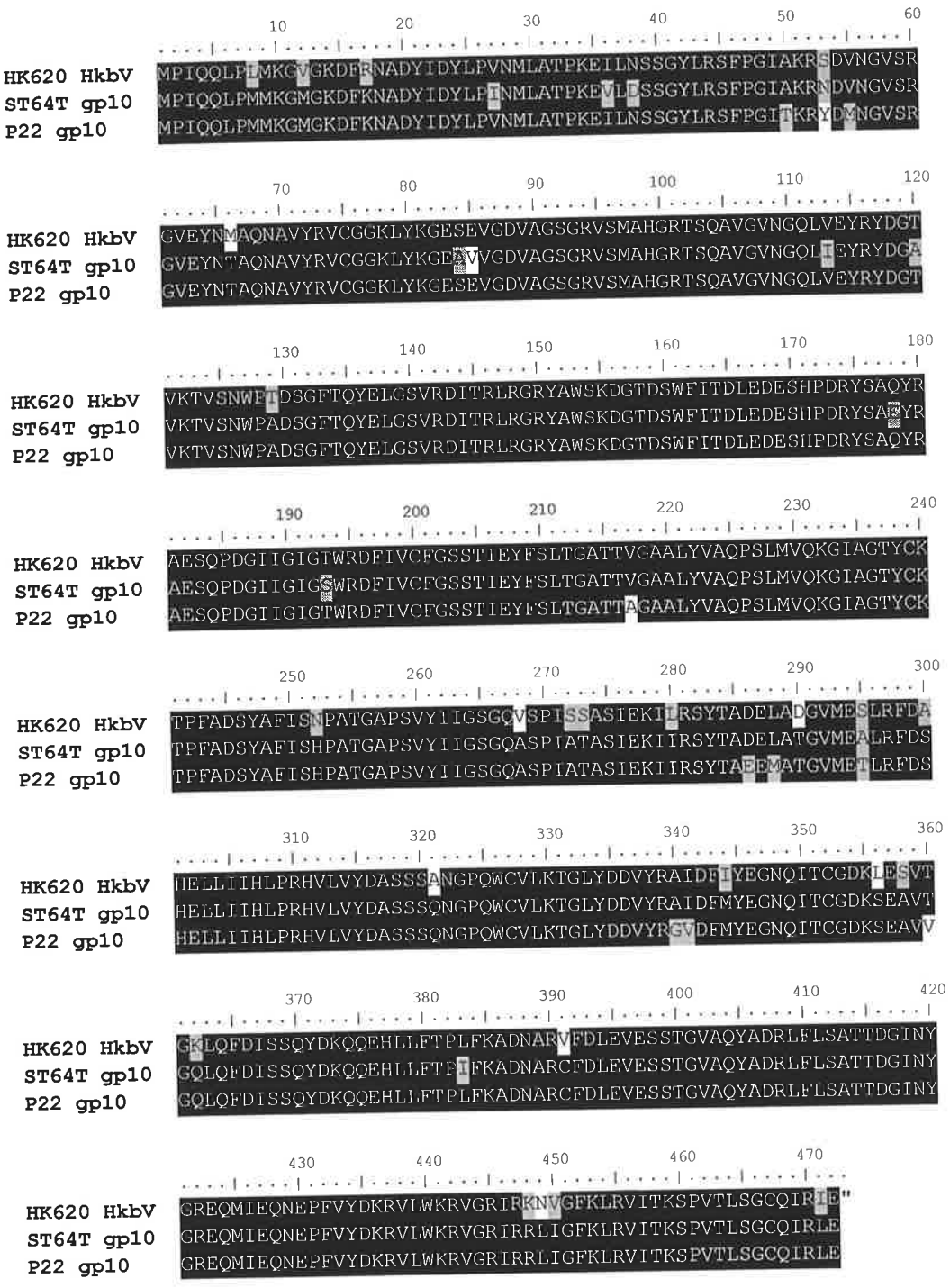

10 20 30 40 50 60
HK620 HkbV MPIQQLPMMKGMGKDFKNADYIDYLPVNMLATPKEILNSSGYLRSFPGIAKRSDVNGVSR
ST64T gp10 MPIQQLPMMKGMGKDFKNADYIDYLPINMLATPKEVLESSGYLRSFPGIAKRNDVNGVSR
P22 gp10 MPIQQLPMMKGMGKDFKNADYIDYLPVNMLATPKEILNSSGYLRSFPGITKRYDVNGVSR
70 80 90 100 110 120
HK620 HkbV GVEYNMAQNAVYRVCGGKLYKGESEVGDVAGSGRVSMAHGRTSQAVGVNGQLVEYRYDGT
ST64T gp10 GVEYNTAQNAVYRVCGGKLYKGEDVVGDVAGSGRVSMAHGRTSQAVGVNGQLIEYRYDGA
P22 gp10 GVEYNTAQNAVYRVCGGKLYKGESEVGDVAGSGRVSMAHGRTSQAVGVNGQLVEYRYDGT
130 140 150 160 170 180
HK620 HkbV VKTVSNWETDSGFTQYELGSVRDITRLRGRYAWSKDGTDSWFITDLEDESHPDRYSAQYR
ST64T gp10 VKTVSNWPADSGFTQYELGSVRDITRLRGRYAWSKDGTDSWFITDLEDESHPDRYSAQYR
P22 gp10 VKTVSNWPADSGFTQYELGSVRDITRLRGRYAWSKDGTDSWFITDLEDESHPDRYSAQYR
190 200 210 220 230 240
HK620 HkbV AESQPDGIIGIGTWRDFIVCFGSSTIEYFSLTGATTVGAALYVAQPSLMVQKGIAGTYCK
ST64T gp10 AESQPDGIIGIGSWRDFIVCFGSSTIEYFSLTGATTVGAALYVAQPSLMVQKGIAGTYCK
P22 gp10 AESQPDGIIGIGTWRDFIVCFGSSTIEYFSLTGATTAGAALYVAQPSLMVQKGIAGTYCK
250 260 270 280 290 300
HK620 HkbV TPFADSYAFISNPATGAPSVYIIGSGQVSPISSASIEKILRSYTADELADGVMESLRFDA
ST64T gp10 TPFADSYAFISHPATGAPSVYIIGSGQASPIATASIEKIIRSYTADELATGVMEALRFDS
P22 gp10 TPFADSYAFISHPATGAPSVYIIGSGQASPIATASIEKIIRSYTAEEMATGVMETLRFDS
310 320 330 340 350 360
HK620 HkbV HELLIIHLPRHVLVYDASSSANGPQWCVLKTGLYDDVYRAIDFIYEGNQITCGDKLESVT
ST64T gp10 HELLIIHLPRHVLVYDASSSQNGPQWCVLKTGLYDDVYRAIDFMYEGNQITCGDKSEAVT
P22 gp10 HELLIIHLPRHVLVYDASSSQNGPQWCVLKTGLYDDVYRGVDFMYEGNQITCGDKSEAVV
370 380 390 400 410 420
HK620 HkbV GKLQFDISSQYDKQQEHLLFTPLFKADNARVFDLEVESSTGVAQYADRLFLSATTDGINY
ST64T gp10 GQLQFDISSQYDKQQEHLLFTPIFKADNARCFDLEVESSTGVAQYADRLFLSATTDGINY
P22 gp10 GQLQFDISSQYDKQQEHLLFTPLFKADNARCFDLEVESSTGVAQYADRLFLSATTDGINY
430 440 450 460 470
HK620 HkbV GREQMIEQNEPFVYDKRVLWKRVGRIRKNVGFKLRVITKSPVTLSGCQIRIE"
ST64T gp10 GREQMIEQNEPFVYDKRVLWKRVGRIRRLIGFKLRVITKSPVTLSGCQIRLE
P22 gp10 GREQMIEQNEPFVYDKRVLWKRVGRIRRLIGFKLRVITKSPVTLSGCQIRLE

Figure 4.25

Comparison of $gp26_{ST64T}$, $gp26_{P22}$ and $HkbW_{HK620}$ using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. The figure shows conservation of sequence at both the N– and carboxy–termini of these proteins.

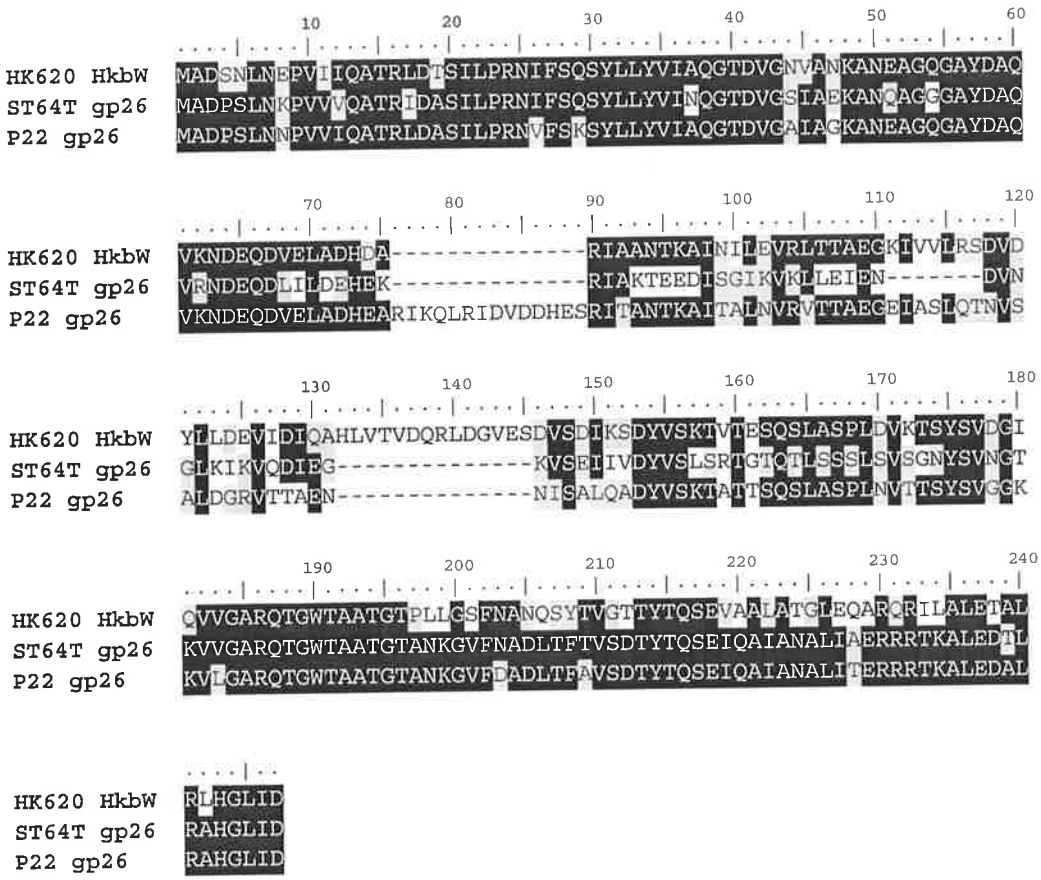

HK620 HkbW
ST64T gp26
P22 gp26
MADSNLNEPVIIQATRLDTSILPRNIFSQSYLLYVIAQGTDVGNVANKANEAGQGAYDAQ
MADPSLNKPVVVQATRIDASILPRNIFSQSYLLYVINQGTDVGSIAEKANQAGGGAYDAQ
MADPSLNNPVVIQATRLDASILPRNVFSKSYLLYVIAQGTDVGAIAGKANEAGQGAYDAQ
VKNDEQDVELADHDA----------RIAANTKAINILEVRLTTAEGKIVVLRSDVD
VRNDEQDLILDEHEK----------RIAKTEEDISGIKVKLLEIEN-------DVN
VKNDEQDVELADHEARIKQLRIDVDDHESRITANTKAITALNVRVTTAEGEIASLQTNVS
YLLDEVIDIQAHLVTVDQRLDGVESDVSDIKSDYVSKIVTESQSLASPLDVKTSYSVDGI
GLKIKVQDIEG----------KVSEIIVDYVSLSRTGTQTLSSSLSVSGNYSVNGT
ALDGRVTTAEN----------NISALQADYVSKTATTSQSLASPLNVTTSYSVGGK
QVVGARQTGWTAATGTPLLGSFNANQSYTVGTTYTQSEVAALATGLEQARQRILALETAL
KVVGARQTGWTAATGTANKGVFNADLTFTVSDTYTQSEIQAIANALIAERRRTKALEDTL
KVLGARQTGWTAATGTANKGVFDADLTFAVSDTYTQSEIQAIANALITERRRTKALEDAL
RLHGLID
RAHGLID
RAHGLID

Figure 4.26

Comparison of the putative $gp20_{ST64T}$ protein with $gp20_{P22}$ and $HkbZ_{HK620}$ proteins using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. The N–terminus of $gp20_{ST64T}$ is identical to that of $gp20_{P22}$ whilst its carboxy–terminus (last 7 aa) is identical to $HkbZ_{HK620}$.

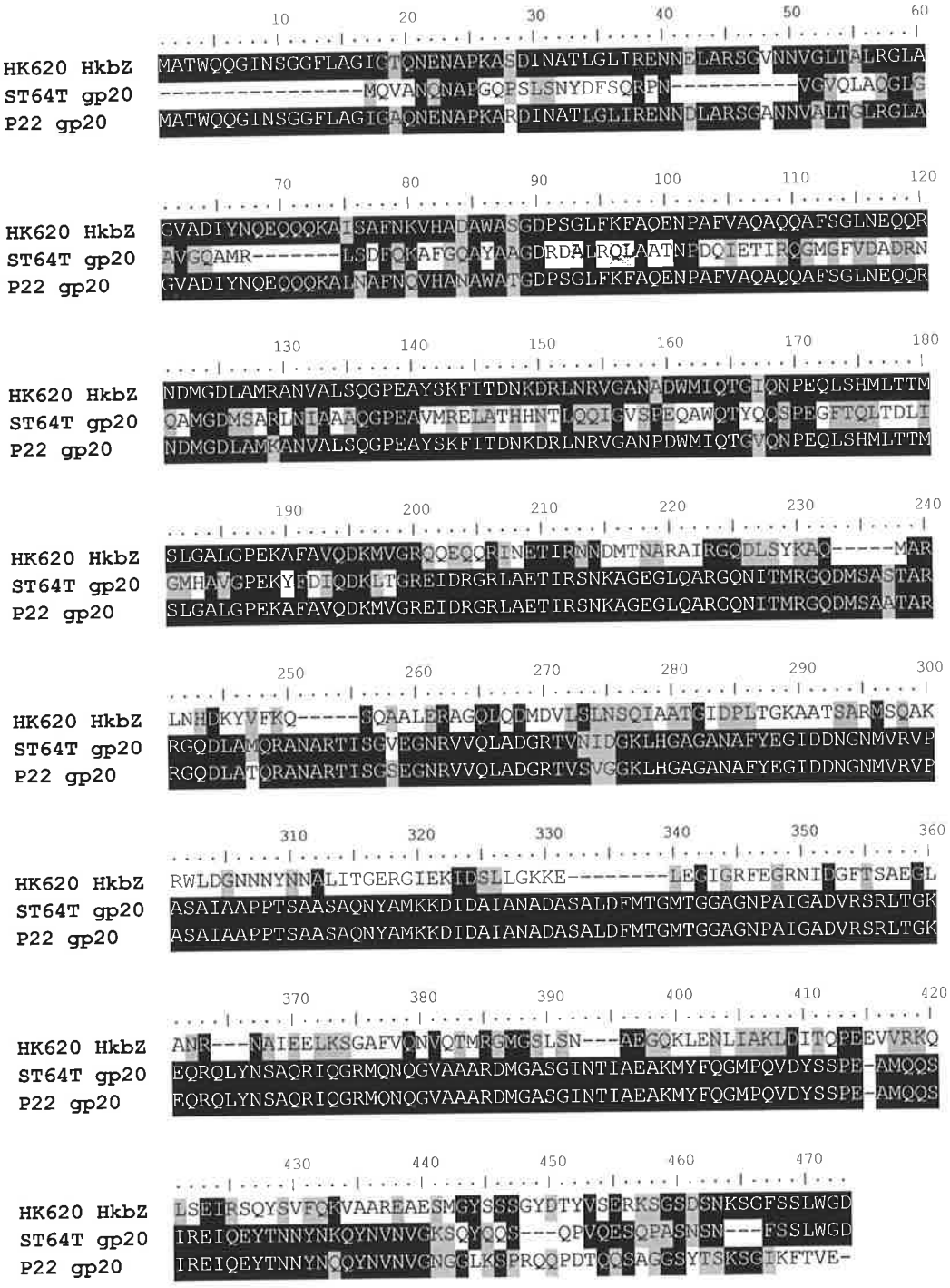

HK620 HkbZ
ST64T gp20
P22 gp20
MATWQQGINSGGFLAGIGTQNENAPKASDINATLGLIRENNDLARSGVNNVGLTALRGLA
MQVANQNAPGQPSLSNYDFSQRPN
VGVQLAQGLG
MATWQQGINSGGFLAGIGAQNENAPKARDINATLGLIRENNDLARSGANNVALTGLRGLA
GVADIYNQEQQQKALSAFNKVHADAWASGDPSGLFKFAQENPAFVAQAQQAFSGLNEQQR
AVGQAMR
LSDFQKAFGQAYAAGDRDALRQLAATNPDQIETIRQGMGFVDADRN
GVADIYNQEQQQKALNAFNQVHANAWATGDPSGLFKFAQENPAFVAQAQQAFSGLNEQQR
NDMGDLAMRANVALSQGPEAYSKFITDNKDRLNRVGANADWMIQTGIQNPEQLSHMLTTM
QAMGDMSARLNIAAAQGPEAVMRELATHHNTLQQIGVSPEQAWQTYQQSPEGFTQLTDLI
NDMGDLAMKANVALSQGPEAYSKFITDNKDRLNRVGANPDWMIQTGVQNPEQLSHMLTTM
SLGALGPEKAFAVQDKMVGRQQEQQRINETIRNNDMTNARAIRGQDLSYKAQ
MAR
GMHAVGPEKYFDIQDKLTGREIDRGRLAETIRSNKAGEGLQARGQNITMRGQDMSASTAR
SLGALGPEKAFAVQDKMVGREIDRGRLAETIRSNKAGEGLQARGQNITMRGQDMSAATAR
LNHDKYVFKQ
SQAALERAGQLQDMDVLSLNSQIAATGIDPLTGKAATSARMSQAK
RGQDLAMQRANARTISGVEGNRVVQLADGRTVNIDGKLHGAGANAFYEGIDDNGNMVRVP
RGQDLATQRANARTISGSEGNRVVQLADGRTVSVQGKLHGAGANAFYEGIDDNGNMVRVP
RWLDGNNNYNNALITGERGIEKIDSLLGKKE
LEGIGRFEGRNIDGFTSAEGL
ASAIAAPPTSAASAQNYAMKKDIDAIANADASALDFMTGMTGGAGNPAIGADVRSRLTGK
ASAIAAPPTSAASAQNYAMKKDIDAIANADASALDFMTGMTGGAGNPAIGADVRSRLTGK
ANR
NAIEELKSGAFVQNVQTMRGMGSLSN
AEGQKLENLIAKLDITQPEEVVRKQ
EQRQLYNSAQRIQGRMQNQGVAAARDMGASGINTIAEAKMYFQGMPQVDYSSPE
AMQQS
EQRQLYNSAQRIQGRMQNQGVAAARDMGASGINTIAEAKMYFQGMPQVDYSSPE
AMQQS
LSEIRSQYSVFQKVAAREAESMGYSSSGYDTYVSERKSGSDSNKSGFSSLWGD
IREIQEYTNNYNKQYNVNVGKSQYQQS
QPVQESQPASNSN
FSSLWGD
IREIQEYTNNYNQQYNVNVGNGTLKSPRQQPDTQQSAGGSYTSKSGIKFTVE

Figure 4.27

Comparison of the putative $gp16_{ST64T}$ protein with $gp16_{P22}$ and $HkcA_{HK620}$ proteins using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. The N–terminus of $gp16_{ST64T}$ is similar to that of $HkcA_{HK620}$.

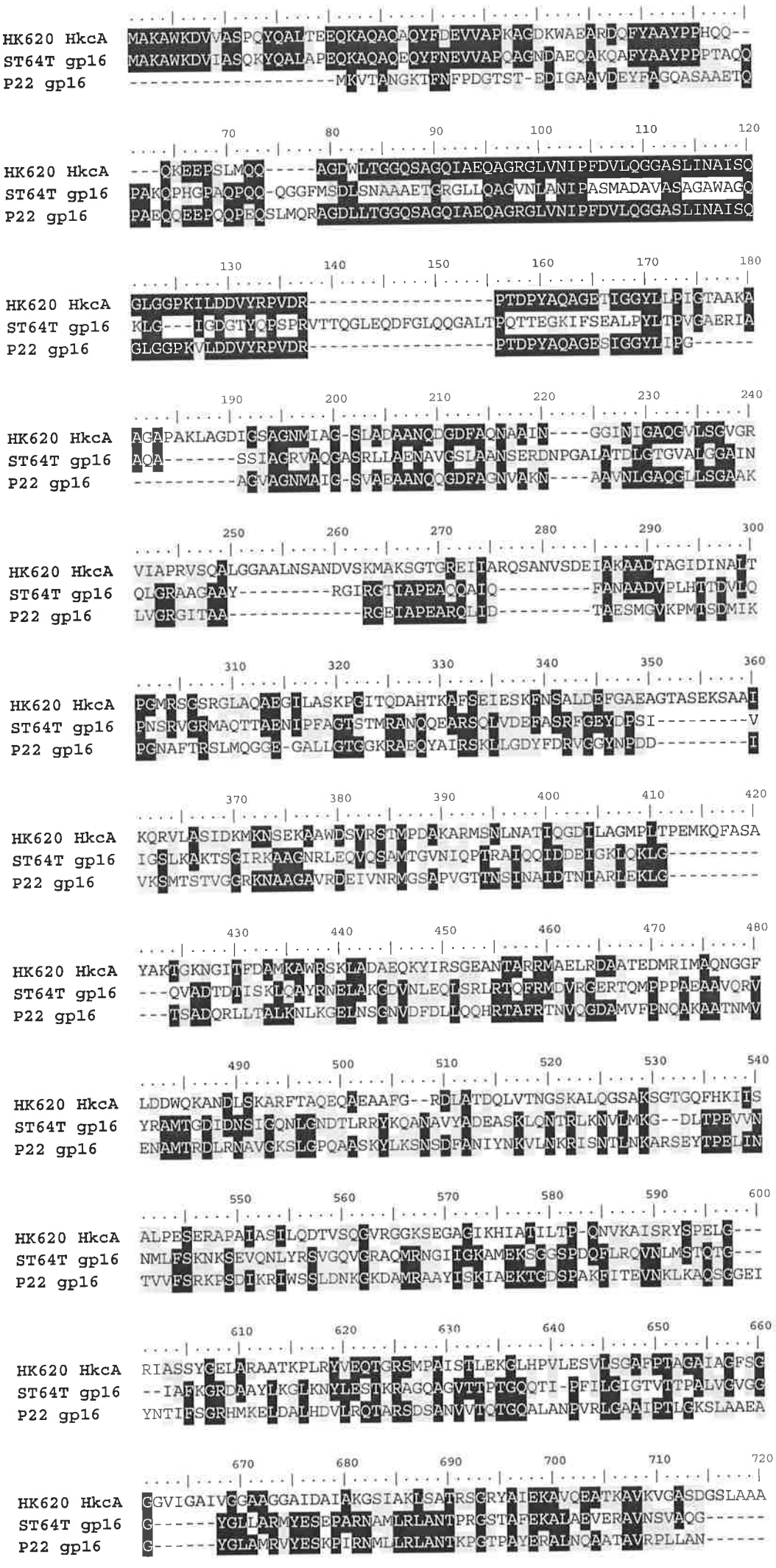

HK620 HkcA
ST64T gp16
P22 gp16

Figure 4.28

Hydropathy prediction of the putative Mnt protein of ST64T compared to the Mnt protein of P22 using Kyte–Doolittle hydropathy (Kyte and Doolittle, 1982). The plot was produced by OMIGA 1.1.

ST64T Mnt

P22 Mnt

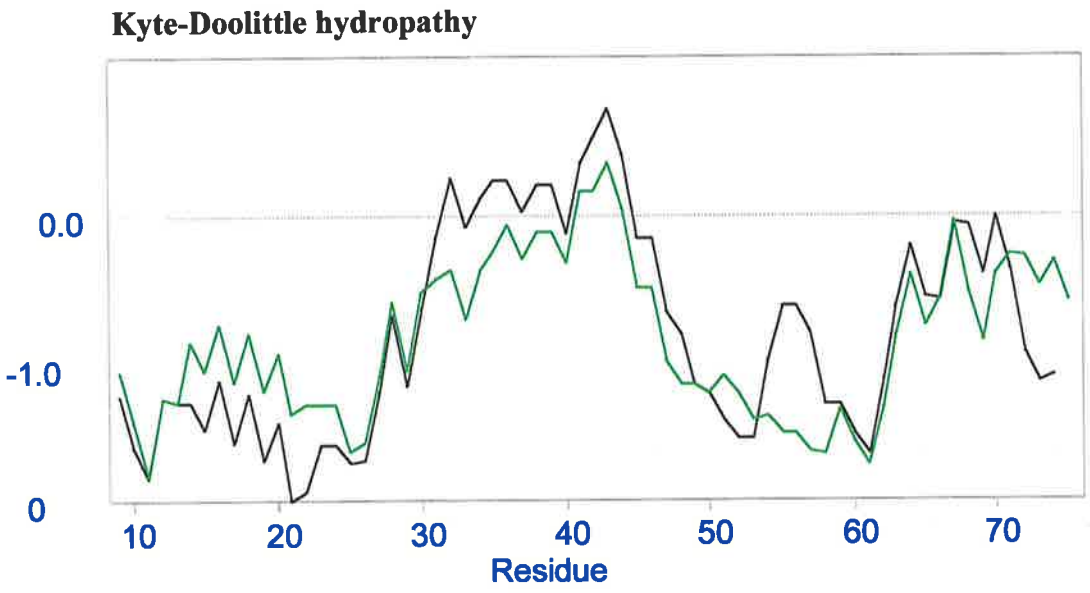

Kyte-Doolittle hydropathy
0.0
-1.0
0
10
20
30
40
50
60
70
Residue

Figure 4.29

Comparison of the putative $gp9_{ST64T}$ protein with the $gp9_{P22}$ and $gp9_{SfVI}$ proteins using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively. A high sequence identity (98%) is shown between the $gp9_{ST64T}$ and $gp9_{P22}$ proteins.

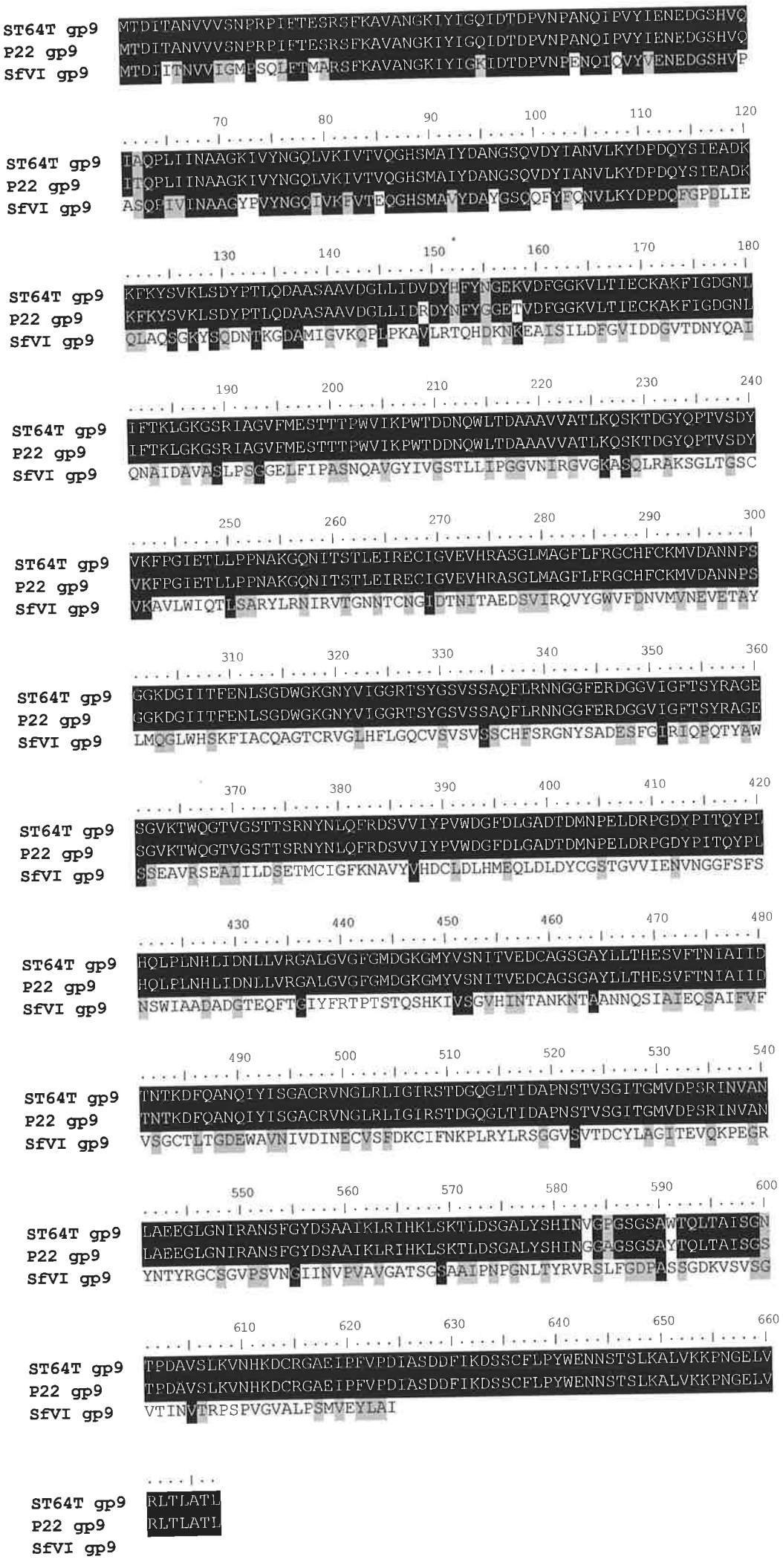

ST64T gp9
P22 gp9
SfVI gp9
MTDITANVVVSNPRPIFTESRSFKAVANGKIYIGQIDTDPVNPANQIPVYIENEDGSHVQ
MTDITANVVVSNPRPIFTESRSFKAVANGKIYIGQIDTDPVNPANQIPVYIENEDGSHVQ
MTDITNVVIGNPSQIFTMARSFKAVANGKIYIGKIDTDPVNPENQICVYVENEDGSHVP
70 80 90 100 110 120
IAQPLIINAAGKIVYNGQLVKIVTVQGHSMAIYDANGSQVDYIANVLKYDPDQYSIEADK
IAQPLIINAAGKIVYNGQLVKIVTVQGHSMAIYDANGSQVDYIANVLKYDPDQYSIEADK
ASQPIVINAAGYPVYNGQIVKEVTEQGHSMAVYDAYGSQQFYFQNVLKYDPDQFGPDLIE
130 140 150 160 170 180
KFKYSVKLSDYPTLQDAASAAVDGLLIDVDYHFYNGEKVDFGGKVLTIECKAKFIGDGNL
KFKYSVKLSDYPTLQDAASAAVDGLLIDRDYNFYCGETVDFGGKVLTIECKAKFIGDGNL
QLAQSGKYSQDNIKGDAMIGVKQPLPKAVLRTQHDKNKEAISILDFGVIDDGVTDNYQAI
190 200 210 220 230 240
IFTKLGKGSRIAGVFMESTTTPWVIKPWTDDNQWLTDAAAVVATLKQSKTDGYQPTVSDY
IFTKLGKGSRIAGVFMESTTTPWVIKPWTDDNQWLTDAAAVVATLKQSKTDGYQPTVSDY
QNAIDAVASLPSGGELFIPASNQAVGYIVGSTLLIPGGVNIRGVGKASQLRAKSGLTGSC
250 260 270 280 290 300
VKFPGIETLLPPNAKGQNITSTLEIRECIGVEVHRASGLMAGFLFRGCHFCKMVDANNPS
VKFPGIETLLPPNAKGQNITSTLEIRECIGVEVHRASGLMAGFLFRGCHFCKMVDANNPS
VKAVLWIQTLSARYLRNIRVTGNNTCNGIDTNITAEDSVIRQVYGWVFDNVMVNEVETAY
310 320 330 340 350 360
GGKDGIITFENLSGDWGKGNYVIGGRTSYGSVSSAQFLRNNGGFERDGGVIGFTSYRAGE
GGKDGIITFENLSGDWGKGNYVIGGRTSYGSVSSAQFLRNNGGFERDGGVIGFTSYRAGE
LMQGLWHSKFIACQAGTCRVGLHFLGQCVSVSVSSCHFSRGNYSADESFGIRIQPQTYAW
370 380 390 400 410 420
SGVKTWQGTVGSTTSRNYNLQFRDSVVIYPVWDGFDLGADTDMNPELDRPGDYPITQYPL
SGVKTWQGTVGSTTSRNYNLQFRDSVVIYPVWDGFDLGADTDMNPELDRPGDYPITQYPL
SSEAVRSEAIILDSETMCIGFKNAVYVHDCLDLHMEQLDLDYCGSTGVVIENVNGGFSFS
430 440 450 460 470 480
HQLPLNHLIDNLLVRGALGVGFGMDGKGMYVSNITVEDCAGSGAYLLTHESVFTNIAIID
HQLPLNHLIDNLLVRGALGVGFGMDGKGMYVSNITVEDCAGSGAYLLTHESVFTNIAIID
NSWIAADADGTEQFTGIYFRTPTSTQSHKIVSGVHINTANKNTAANNQSIAIEQSAIFVF
490 500 510 520 530 540
TNTKDFQANQIYISGACRVNGLRLIGIRSTDGQGLTIDAPNSTVSGITGMVDPSRINVAN
TNTKDFQANQIYISGACRVNGLRLIGIRSTDGQGLTIDAPNSTVSGITGMVDPSRINVAN
VSGCTLTGDEWAVNIVDINECVSFDKCIFNKPLRYLRSGGVSVTDCYLAGITEVQKPEGR
550 560 570 580 590 600
LAEEGLGNIRANSFGYDSAAIKLRIHKLSKTLDSGALYSHINVGPGSGSAWTQLTAISGN
LAEEGLGNIRANSFGYDSAAIKLRIHKLSKTLDSGALYSHINGGAGSGSAYTQLTAISGS
YNTYRGCSGVPSVNCIINVPVAVGATSGSAAIPNPGNLTYRVRSLFGDPASSGDKVSVSG
610 620 630 640 650 660
TPDAVSLKVNHKDCRGAEIPFVPDIASDDFIKDSSCFLPYWENNSTSLKALVKKPNGELV
TPDAVSLKVNHKDCRGAEIPFVPDIASDDFIKDSSCFLPYWENNSTSLKALVKKPNGELV
VTINVTRPSPVGVALPSMVEYLAI
RLTLATL
RLTLATL

Table 4.1 ST64T ORFs and putative function

Gene	From[1]	To[1]	Strand	Mass (kDa)	Putative function	Related phage sequences[2]	BlastP e–value	% identity according to BlastP
gtrC	17	1474	–	46.9	Unknown protein	P22 *orf485*; AAF74999	0.0	98
gtrB	1464	2441	–	37.0	Bactoprenol glucosyl transferase; O–antigen conversion	P22 *gtrB;* AAF75000	e–175	99
						CPS-53 *gtrB* homolog; P77293	e–155	88
						SfIV *orf5*; AAK50573	e–157	88
						SfV *bgt*; AAB72133	e–155	88
						SfII *bgt*; AAC39272	e–155	88
						SfX *bgt*; AAF22454	e–153	86
gtrA	2393	2755	–	13.5	Translocase; O–antigen conversion	P22 *gtrA*; AAF75001	2e–47	100
						SfX *gtrA*; AF056939	6e–39	78
						SfV *orf6*; AAB72134	8e–38	77
int	3104	4267	–	44.9	Integrase	P22 *int*; AAF75002	0.0	98
						SfV*int*; AAB72135	0.0	89
						CP–933H *int*; AAG54569	0.0	88
						DLP12 *int*; INTD_ECOLI	0.0	73
						APSE–1 *int*; F157835_38	0.0	65
						SfX	e–141	54
xis	4144	4494	–	12.8	Excisionase	P22 *xis*; AAF75003	1e–52	100
						SfV *xis*; AAB72136	8e–47	87
eac	4497	5132	–	24.9	Unknown	P22 *eac*; AAF75004	7e–97	86
eag	5227	5406	–	6.6	Unknown	P22 *eag*; AA75005	7e–20	100
eaa2	5503	6048	–	20.3	Unknown	HK620	5e–43	67
						P22 *eaa*; AA75006	1e–39	100
eaa1	6045	6860	–	30.3	Unknown	P22 *eaa*; AA75006	1e–68	62

ORF87	6871	7134	–	9.5	Unknown	CP–933R; AAG56433	1e–18	56
ORF109	7664	7993	–	12.3	Unknown			
ORF81	7974	8219	–	8.7	Unknown			
eae	8183	8560	–	13.8	Unknown	P22 *eae*; AAF75010	3e–13	32
ORF56	8557	8727	–		Unknown	HK97 gp38; AAF31114	6e–13	75
						HK620; AAK28857	1e–12	75
						P22 *orf56*; AAF75011	1e–10	65
abc2	8738	9031	–	11.6	Anti-RecBCD	P22 *abc2*; AAF75012	3e–43	94
						HK97 gp39; AAF31118	2e–41	91
						HK022 *abc2*; AAF30380	4e–40	91
						HK620 unknown; AAK28858	1e–39	90
abc1	9078	9362	–		Anti–RecBCD	P22 *abc1*;	1e–39	95
ORF235	9362	10069	–	26.8	Recombination protein	P22 *erf*; AAF75014	8e–24	76
						U136 *orf252*; AAF74078	3e–44	57
						bIL286 *orf14*; AAK08301	3e–44	57
ORF66	10078	10278	–	7.8	Unknown	HK620; AAK28861	2e–14	67
kil	10263	10376	–	4.4	Unknown	VKIL_LAMBD; P03758	8e–06	58
						VT2–Sa; NP_050515	8e–06	58
						933W; NP_049476	1e–05	58
c3	10369	10536	–	6.0	Regulatory protein	P22 *c3*; AFF75017	9e–13	77
						λ *cIII*; RPC3_LAMBD	1e–10	72
						HK620 *cIII*; AKK28863	3e–10	70
						933W *cIII*; AAD25422	7e–10	70
17	10600	10914	–	12.3	Superinfection exclusion	P22 *17*; AAF75018	2e–45	92
						H-19B *17*; AAD04642	8e–45	89

ORF232	11086	11784	–	25.8	Unknown			
ral	11724	11918	–	7.3	Antirestriction	P22 r*al*; AAF75021	3e–22	98
						21 *ral*; PO7235	1e–17	78
						λ *ral*; VRAL_LAMBD	3e–17	78
24	11997	12335	–	12.8	Antitermination/Regulatory protein	P22 *N*; S10069	3e–30	96
						L *24*; CAA63998	5e–30	100
						P22 *24*; AAF75023	7e–30	96
c2	12893	13597	–	26.1	Repressor	L *c2*; CAA63999	e–122	100
						P22 *c2*; AAF75024	6e–53	54
						HK620 *cI*; AAK28868	3e–48	47
						434 *cI*; S32822	3e–42	47
cro	13674	13889	+	7.4	Antirepressor	L *cro*; CAA64000	1e–22	69
c1	14000	14281	+	10.5	Transcriptional activator	L *cI*; CAA64001	1e–21	96
						HK620 *cII*; AF335538	1e–19	55
						P22 *cI*; AAF75026	1e–16	49
						434 *cII*; RPC2_434 cII;	2e–14	50
						PC2_LAMBD	3e–14	50
18	14455	15354	+	34.2	Replication protein gp18	HK022 *O*; AAF30383	e–109	66
						VT2–Sa; BAA84309	e–108	65
						HK620; AAK28872	e–106	65

12	15344	16780	+	53.5	Replication protein gp12	HK620 *P*; AAK28873	0.0	98
						HK022 *P*; AAF30384	0.0	98
ORF90	16855	17127	+	10.2	Unknown	VT2–Sa; BAA84361	9e–11	41
						CP–933N; AAG55882	9e–10	41
ORF69	17137	17346	+	84.1	Unknown	933W; CAB39288	3e–04	44
ORF86	17346	17606	+	10.2	Unknown	D3 *orf41*; AAF80800	0.17	30
ORF–98	17609	17905	+	11.7	Unknown			
ninB	18003	18308	+	11.8	Unknown	VT2-Sa; BAA84314	3e–44	91
						CP933V; AAG57234	1e–42	91
						933W; AAD25438	1e–42	91
						933W *orf8*/*ninB*; CAB39290	2e–14	92
ninD	18278	18478	+	7.9	Unknown	P22 *ninD*; AAF75031	3e–22	94
						λ *ninD*	2e–16	75
ninE	18445	18621	+	7.1	Unknown	P22 *ninE*; AAF75032	1e–17	96
						λ *nin60*; NINE_LAMBD	1e–17	96
						H–19B *orf58*; AAD04650	2e–17	94
						HK620; AAK28876	2e–16	96
ninX	18618	18965	+	13.2	Unknown	P22 *ninX*; AFF75033	2e–27	57
ninF	18958	19131	+	6.3	Unknown	P22 *ninF*; AAF75034	5e–15	84
						H–19B *orf58B*; AAD04651	6e–15	85
						HK–97; AAF31139	2e–14	75
						HK620; AAK28877	8e–12	70

ORF101	19121	19426	+	11.2	Unknown	D3 *orf80*; AAF80836	0.003	33
ORF136	19299	19709	+	15.4	Unknown	HK97 gp66; AAF31141	7e–47	96
						HK620; AAK28879	2e–46	95
rus	19706	20101	+	15.2	Cross over junction (endodeoxyribonuclease)	CP–933X *rus*; AAG55970	1e–26	62
						82 r*usA*; CAA63330	2e–26	52
ninH	20098	20301	+	7.9	Unknown	P22 *ninH*; AAF75037	3e–26	100
						H-19B *orf59*; AAD04654	1e–16	70
						λ *nin*; Q1BP0L	2e–16	71
ninZ	20282	20461	+		Unknown	P22 *ninZ*; CAA55166	1e–13	93
23	20458	20976	+	19.8	Antitermination	21 *Q*; CAB39993	2e–67	81
						APSE-1 P5 *Q*; AAF03999	6e–04	33
13	21376	21693	+	11.4	Lysis	PS3 gp13; CAA09700	8e–44	99
						CP-933X ; AAG55972	1e–29	75
19	21680	22078	+		Lysis	PS3 gp19; CAA09701	1e–65	99
15	22075	22527	+	16.6	Lysis	PS3 gp15; CAA09702	1e–67	100
						PS34 gp15; CAA09707	1e–41	67
						HK97 *Rz*; AAF31146	5e–41	68
						P22gp15; AAF75041	4e–39	67
ORF118	22709	23065	+	12.9	Unknown	P22; AAF75061	0.001	35
ORF129	23069	23458	+	14.5	Unknown	HK620; AAK28904	7e–17	37
ORF134	23458	23862	+	14.1	Unknown	LP7 3 5' region	9e–21	100
3	23866	24354	+	18.7	Packaging protein (terminase small subunit)	PS3 gp3; CAA09703	1e–85	100
						LP–7 gp3; P16937	4e–84	99
						PS34 gp3; CAA09708	4e–76	98
						P22 gp3; AAF75043	2e–83	96
2	24332	25831	+	50.6	Terminase large subunit	P22 gp2; AAF75044	0.0	98
						LP–7 gp2; P16938	0.0	78
1	25831	28008	+	82.8	Portal protein	P22 1; AAF75075	0.0	98
						APSE–I; AAF03962		

							2e–90	32
8	28022	28933	+	33.6	Scaffolding protein	P22 gp8; AAF75046	e–136	99
5	28933	30225	+	46.7	Coat protein	P22 gp5; AAF75047	0.0	99
ORF186	30266	30826	+	21.0	Unknown			
4	30810	31310	+	18.0	DNA stabilisation protein	P22 gp4; AAF75049	3e–86	96
						HK620; AAK28896	1e–18	36
						APSE–1 P27; AAF03974	9e–11	37
10	31270	32688	+	52.3	Packaged DNA stabilisation protein	P22 gp10; AAF75050	0.0	95
						HK620; AAK28897	0.0	92
						APSE–1 P28; AAF03971	e–166	60
26	32692	33330	+	7.4	Packaged DNA stabilisation protein	P22 gp26; AAF75051	2e–65	62
						HK620; AAK28898	3e–53	54
						APSE–1 P30; AAF03973	2e–09	34
14	33330	33785	+	17.4	Head assembly/ Unknown protein	HK620; AAK28899	1e–80	96
						P22 unknown; AAF75052	1e–77	92
7	33788	34477	+	23.2	DNA transfer protein	HK620; AAK28900	8e–47	66
						P22 gp7; AAF75053	2e–41	64
						APSE–1 P32; AAF03975	3e–36	59
20	34487	35821	+	47.7	Unknown protein/ DNA transfer protein	P22 gp20; AAF75054	e–118	66
						APSE–1 P33; AAF03976	1e–40	45
						HK620; AAK28901	4e–19	24
16	35821	37797	+	69.7	DNA transfer protein	P22 gp16; AAF75055	2e–57	31
						APSE–1 P35; AAF03978	1e–55	33
						HK620; AAK28902	2e–17	31
mnt	38250	38498	–	9.6	Regulatory protein	P22 *mnt*; AAF75057	2e–19	60
9	38634	40367	+	72.0	Endorrhamnosidase (tail spike protein)	P22 gp9; AAF75060	0.0	98
						SfIV tsp; AAD33394	7e–73	64
						HK620 tsp; AAK28905	2e–41	73

[1]Nucleotide coordinates corresponding to the first nucleotide of the initiation codon and the last nucleotide of the termination codon[2]Where similar sequences exist, the phage designation preceded the gene designation (where possible) and the GenBank accession number. Only phage related sequences are shown, other sequences related to bacterial genes are cited in text.

Table 4.2 Potential C1–, IHF–binding sites and rho–independent terminators in ST64T genome[a]

Location	Similarity score	Sequence
Potential C1 binding sites		
2946 – 2959		**TTGC**ATCGGC**TTGC**
2956 – 2969		**TTGC**AAGGCT**TTGC**
13949 – 13962		**TTGC**GTGTAT**TTGC**
{14025 – 14012}		**TTGC**TTTTAG**TTGC**
{20783 – 20770}		**TTGC**GTGGCT**TTGC**
{25000 – 24987}		**TTGC**GTGGCT**TTGC**
{26957 – 26944}		**TTGC**CCGTAT**TTGT**
{27652 – 27639}		**TTGC**CGGGTC**TTGT**
{35061 – 35048}		**TTGC**TGCGGA**TTGT**
34700 – 34713		**TTGC**GTCAAC**TTGC**
{39763 – 39750}		**TTGC**GAGAGG**TTGT**
Rho–independent terminators		
2871 – 2900		A**TTGATCG**TTGTTAC**CGATCAA**TTTTTATT
5146 – 5173		A**CCGCC**ATCA**GGCGG**CTTGGTGTTCTTT
{5161 – 5138}		G**CCGCC**TGAT**GGCGG**TTTTTTATT
10592 – 10626		A**GCCGC**ACTCAGG**CGGCG**GCTGTTGTTTCTTC TTT
{10612 – 10585}		A**GCCG**CCGCCTGAGTG**CGGC**TGTTTTAT
{10956– 10928}		T**GCCGC**TCTATATGG**GCGGC**ATTCTTTTT
{14005 – 14042}		T**GCC**T**CGAAGAA**TTCGCCG**TTCTTCG**G**GGC**T TTTTCTTTT
21253 – 21279		T**GCCCA**CCTAGCCG**TGGGC**TTTTCATT
{23463 – 23439}		T**GCCA**TTACT**TGGC**TCCTTCTTTTT
38235 – 38259		**GCCCGG**TTCG**CCGGGC**TATTTTTTT
39198 – 39220		T**CCCG**CATTGC**CGGG**GTTTTTAT
Potential IHF–binding sites		
427 – 453	58.0	TTCTTGAT**ATTAA**CTGC**TTA**TCTTCAA
{607 – 581}	48.5	TTATTGGC**ATCAA**TGTT**TTT**CATCATT
{1133 – 1107}	48.3	GAAAATCT**ATCATA**TA**GAT**ATGATTCA
1650 – 1676	48.1	TATCAATA**ATCAT**CCA**TG**CACCGTAAA
{3036 – 3010}	50.9	CCTTTTA**AGTCAA**CAACA**TA**CCGCGTC
3091 – 3117	61.7	CCAGTTAA**ATCAA**ATAC**TTA**CGTATTA
{14785 – 14759}	49.8	CGATACAA**ATCA**TTCC**ATT**CCCTTCAA
31340 – 31366	48.1	ACCTACCA**ATCAA**CATG**TTG**GCCACAC
Lambda *cro*/c*II*	61.5	TGCATACA**TTCAAT**CAA**TTG**TTATCTA

Putative C1–binding sites (boldface) were determined by scanning the sequences for the consensus sequence (TTGCN6TTGY). IHF–binding sites were scanned using MacTargsearch (Goodrich, *et al.,* 1990) which displays sites with scores of >29. [a]Similarity scores were included in the case of IHF–binding sites. Only those with scores of greater than 48 were included following Vander Byl and Kropinski (2000) who included scores of ≥50. For comparison, a characterised IHF–binding site of phage λ was included and bases identical to the consensus sequence are in boldface. Rho–independent terminators were determined by scanning the sequence fro a stem–loop structure (boldface and underlined) followed by a region rich in thymine residues (Brendel, *et al.,* 1986). The sites in brackets are those associated with the complementary strand.

Table 4.3 Potential stemloop structures and/or inverted repeats and direct repeats in ST64T

Location	Sequence (ΔG values for stemloop sequences in kcal/mol)	
Stemloop structures/ Inverted repeats		
543 – 559	ATCATTGGtCCAATGAT	
5734 – 5753	TGGTTCAGcgtcCTGAACCA	−9.8
10589 – 10615	AACAGCCGCaCtcagGcgGCGGCTGTT	−14.80
11034 – 11057	ATAGCAGAcTcgtaAaTCTGCTAT	−8.50
14166 – 14989	TCTCTAATACtaacGTATTAGAGA	−13.10
17630 – 17652	ATGAAAATAtcgctTATTTTCAT	−5.70
30555 – 30573	TAAGACCACcGTGGTCTTA	−7.90
32639 – 32670	AATAAAAATGGGTGTaaACACCCATTTTTATT	−16.80
Direct repeats		
404 – 419	**TAAAAGAATAAAAGAA**	
1018 – 1033	**ATGATATTATGATATT**	
7425 – 7453	**CGTTAATT**AGTTCGTCCAGTA**CGTTAATT**	
11449 – 11512	**CAGGTCGGCACCGCGCAGGTCGGCACCGCGCAG GTCGGCACCGCGCAGGTCGGCACCGCGCAGGTC GGCACCGCGCAGGTCGGCACCGCGCAGGT**	
12757 – 12799	**CAATACAAAATGTACTTA**ATATCGT**CAATACAAAA TGTACTTA**	
14851 – 14881	**TGTCCAACGGATGTCCAACGGATGTCCAACG**	
14950 – 14966	**AATAATATAAATAATAT**	
17734 – 17762	**TTTTGCGAG**ATGGCTTGAGAGT**TTTGCGA**	
20893 – 20911	**CAAAATGC**TTT**CAAAATGC**	
22832 – 22857	**ATTACAGCA**TTGACAGA**ATTACAGCA**	
29652 – 28676	**ATGGCAAC**TGGATAACG**ATGGCAAC**	
31040 – 31057	**AGCGCAGTCAGCGCAGT**	
32637 – 32652	**AATCACCAAATCACCA**	
35428 – 35445	**CATGACTGGCATGACTGG**	
35948 – 35964	**AACAGGCTAAACAGGCT**	
35996 – 36015	**AACAACCA**GCAAAA**CAACCA**	
40052 – 40080	**TACCAATA**TAGCCATAATTGA**TACCAATA**	

Stem–loop structures are presented with the calculated ΔG values (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi#Results) (Matthews, *et al.*, 1999) with bases in the loop shown in lowercase. Inverted repeats are in boldface. In the case of repeats without internal bases, the start of a repeat is underlined.

Chapter 5

Characterisation of the genome of phage ST64B

5.1 Introduction

Comparison of lambdoid phage genome organisation and relationships was first carried out by DNA heteroduplex mapping (Simon, *et al.*, 1971). Since then, a number of studies have been carried out using genetic and DNA sequencing methods (Juhala, *et al.*, 2000). These analyses have shown the diversity of most phages, mostly due to mosaicism arising by homologous and illegitimate recombination between members of a phage family such as the lambdoid family (Casjens, *et al.*, 1992).

Natural relatives of λ have been isolated from various sources, with most of them growing on *E. coli* and a few on *S. Typhimurium* (P22, L, LP7 and ES18) (Campbell, 1994). The best studied of the temperate phages is bacteriophage λ which infects *E. coli* strains. It is the archetype of the lambdoid phages. P22 is the best–studied *Salmonella* infecting phage and it is a relative of λ. The major aspects of the life–cycles of λ and P22 are used as examples for comparative studies with newly isolated bacterial viruses.

Upon establishment of the lysogenic state, temperate bacteriophages are able to transfer foreign DNA into a host bacterial cell. For example, the toxins of a number of both Gram positive and Gram negative bacteria are encoded in the genomes of bacteriophages (Acheson, *et al.*, 1998, Waldor, 1998). In *S. aureus*, lysogeny can affect the expression of several enzymes and extracellular toxins, some of which may be virulence factors (Parker, 1983). Until recently, the contribution of *Salmonella* phages to virulence has not been known. Virulence factors such as the SopE gene product, an effector translocated by the SPI–1 type III secretion system and the superoxide dismutase (SodC) are encoded by the *S.* Typhimurium phages SopEϕ and Gifsy–2, respectively

(Figueroa-Bossi and Bossi, 1999, Mirold, *et al.*, 1999). Furthermore, *S.* Typhimurium phage P22 like *Shigella flexneri* phages SfII, Sf6, SfV and SfX, cause O–antigen modification in the lysogenic state (Allison and Verma, 2000, Vander Byl and Kropinski, 2000).

The genome of ST64B bacteriophage, isolated from a clinical isolate *S.* Typhimurium DT 64 was determined. The discovery that ST64B–like sequences were found in almost all *S.* Typhimurium isolates tested (using hybridisation analysis), led to the sequencing of its genome and for comparative analysis of its genome with that of ST64T, which is also a prophage in the same isolate, *S.* Typhimurium DT 64. The ST64B genome was also compared with two previously determined and well characterised lambdoid family genomic sequences, those of *E. coli* phage λ and *S.* Typhimurium phage P22. Currently, there are more than 30 completed phage genome sequences in the databases (van Belkum, *et al.*, 2001), a tiny fraction of the total estimated number of phages in the world ($\sim 10^{30}$) (cited in Smith, 1999). In this chapter, a comparative analysis of ST64B genomic sequence with other dsDNA phage sequences is presented.

5.2 Results

5.2.1 Sequencing strategy

*Sma*I restriction fragments (4,564 bp, 4,253 bp, 3,394 bp, 2,168 bp, 1,641 bp) from ST64B genome were shot–gun cloned into pGEM–7zf(–), as described in chapter 2. To sequence the ST64B inserts from pGEM–7zf(–), small–scale plasmid DNA was first purified and the inserts were initially sequenced using dye primer sequencing chemistry (chapter 2). The remaining sequence was determined using CsCl purified ST64B fragments by dye terminator chemistry with synthetic oligonucleotide primers (Appendix II) as described in chapter 2. PEG_{8000} precipitated whole phage DNA as well as amplified DNA fragments by PCR were used to sequence the largest *Sma*I restriction fragment

(23,650 bp). An additional 479 bp *Sma*I fragment was discovered by using whole phage DNA (PEG_{8000} precipitated) and sequencing outward from the 1,641 bp and the 4,564 bp fragments.

5.2.2 Nucleotide accession number

The nucleotide sequence of ST64B bacteriophage has been deposited with GenBank and has been assigned accession number AY055382. Sequence data is presented on a CD as Appendix III.

5.2.3 Sequence assembly

To determine the fragment order, primers were designed to read outward from the ends of the genomic *SmaI* fragments (4,564 bp, 4,253 bp, 3,394 bp, 2,168 bp, 1,641 bp), using PEG_{8000} precipitated whole ST64B genome, thus reading into adjacent fragments. The sequence of the ends of the largest fragment (23,650 bp) and the smallest fragment (479 bp) were detected using this strategy. The sequence data for the seven *Sma*I fragments were assembled using GeneCompar 2.0 and/or GeneBase 1.0 (Applied Maths, Kotrijk, Belgium) and the fragment order was sequentially found to be 23,650 bp, 4,253 bp, 4,564 bp, 479 bp, 1,641 bp, 3,394 bp and 2,168 bp.

5.3 Sequence analysis of ST64B genome

For ease of comparison with other dsDNA phage genomes, the ST64B map was opened 16 bp upstream of a putative terminase small subunit. The ST64B genomic sequence was 40,149 bp with an overall GC content of 51.3%.

5.3.1 Assignment of probable genes

The λ nomenclature was adopted where possible, in designating ST64B ORFs. Open reading frames were initially identified using a graphical ORF finder and WebGene Mark.HMM version 2 (Lukashin and Borodovsky, 1998). In most cases, assignment of probable genes was based on the presence of an initiation codon (AUG or GUG) with a good ribosome binding sites (Shine Dalgarno sequence) situated upstream of the initiation codon. However, a few ORFs did not show any putative ribosomal binding sites. A potential ribosomal binding site could not be found in two previously determined genes encoding putative terminase large subunit and DNA invertase pin. Finally, similarity of the ORFs to the sequences in the database was also taken into account in the assignment of putative genes. ST64B open reading frames were designated SB1 to SB56 (*Salmonella* ST64B phage ORF1 to ORF56). A total of 56 putative open reading frames were identified on both strands of the nucleotide sequence.

5.3.2 Open reading frames present on the genome of ST64B and their functional assignments

The amino acid sequence of the deduced putative proteins in ST64B were used in searches for similarity with other sequences present in databases using BlastP algorithm (Altschul, *et al.*, 1990). Occasionally, BlastN was also used. In general, ST64B putative genes showed only weak similarities to genes in the NCBI database. The properties of selected ORFs showing similarities with genes in the database will be discussed with emphasis on those encoding proteins that are similar to proteins of characterised phages. However, it was not easy to infer functions of most putative genes because ST64B showed weak similarities with unknown genes of phages from diverse groups of bacteria including bacterial genes, mostly from *E. coli*. Table 5.1 presents the different 56 ORFs found in ST64B and the graphical map is presented in Figure 5.1. Figure 5.2 indicates the ST64B general map architecture and compares it to the map of phage λ.

5.3.2.1 Genes involved in packaging

SB1 (Terminase small subunit). This 495 bp ORF (52 %mol GC) encodes a putative 164 amino acid protein with a predicted mass of 17,951 Da and pI of 9.37. The putative initiation codon is preceded by a RBS (AGGAG) 8 bp upstream. SB1 gene product showed low sequence identity to putative terminase small subunits of *Lactococcus* bacteriophages BK5-T (34%), prophage pi3 (29%) as well as 29% identity to ORF40 of bIL286, and 36% identity to an unknown protein of the endosymbiont bacteriophage GMSE-1. BK5–T and bIL286 are *cos*–recognising phages (Desiere, *et al.*, 1999). The map position of this putative protein and its low, but significant sequence identity to putative small terminase subunits in bacteriophages carried by a *Lactococcus lactis* bacterium, suggests that it is the terminase small subunit protein of ST64B.

SB2 (Terminase large subunit). This 1,734 bp long ORF (51 %mol GC) encodes a 577 aa putative protein with a predicted mass of 65,361 Da and pI of 6.03. This ORF overlaps with SB1 in a sequence AUGA where AUG is the start codon of SB2 and UGA the stop codon of SB1. Although the terminase large subunit genes in λ and P22 revealed RBS (GGAGAGG) 5 bp and (GGAG) 11 bp upstream of the initiation codon, respectively, no potential RBS upstream of the putative initiation codon of ST64B SB2 could be found. Searches for conserved motifs using Protein Families (Pfam) (Bateman, *et al.*, 1999), revealed two motifs defined as Ets-domain (PF00178) and Antenna complex alfa/beta subunit (LHC) (PF00556). The SB2 inferred gene product showed 93% sequence identity to a 50.9 kDa hypothetical protein, the product of the *E. coli ymfN* gene (accession no. P75978). Furthermore, the ST64B putative terminase large subunit protein exhibited 46% sequence identity to the terminase large subunit of *Pseudomonas aeruginosa* phage D3. In addition, sequence identity to a phage D3 terminase–like protein in *H. influenzae* (accession no. AAF27357) (31%), terminase large subunit in *S. aureus* phage phiSLT (30%), *E. coli* phage CP–933C (30%) and phage–like sequences in *Clostridium acetobutylicum* (accession no. NP_348518) (29%) was shown. Moreover, 24 – 28%

sequence identity was observed with many terminase large subunit proteins in other phages including, *Staphylococcus aureus* phiPV83, *Lactococcus casei* phage A2, *L. lactis* bIL285, HK022 and HK97. Comparison of the putative terminase large subunit of ST64B, terminase large subunit of phage D3 as well as the unknown YMFN protein in *E. coli*, revealed significant similarity with the carboxy-terminal domain of ST64B putative terminase and YMFN proteins exhibiting 100% sequence identity (Figure 5.3). Similar to the putative terminase small subunit, the putative terminase large subunit of ST64B also showed similarities to terminase large subunits of the *cos*–recognising phages.

5.3.2.2 Genes involved in morphogenesis

In lambdoid phages, a portal protein is expected downstream of the terminase large subunit. However, a small ORF upstream of SB4, encoding a 60 aa putative protein with 96% amino acid identity to a hypothetical protein in *E. coli*, the product of the *ymfR* gene (accession no. P75979) was found. This 60 aa gene product contains one transmembrane domain and one conserved motif of the glycosyl hydrolase family 9 (PF00759).

SB4 (portal protein). This 1,242 bp ORF (56 %mol GC) begins with a probable GUG initiation codon and is preceded by a putative RBS (GAGG) 7 bp upstream. The initiation codon overlaps with the termination codon of SB3 in a sequence UAGUG. SB4 encodes a 414 aa putative portal protein with a predicted mass and pI of 46,094 Da and 6.13, respectively. This putative protein revealed 32% sequence identity to portal proteins of *E. coli* phages HK022 and HK97. Furthermore, the SB4 gene product revealed 31% sequence identity to a *Streptomyces* phage phi-C31 portal–like protein found in *Haemophilus influenzae*, 27%, 29% and 27% sequence identity to portal proteins in *Bacillus* phage phi-105, *E. coli* O157:H7 prophage CP–933C and *Pseudomonas* phage D3, respectively. Additionally, a very high sequence identity (99%) was observed with a 17.6 kDa hypothetical protein in inte–pin intergenic region of *E. coli*, the product of the *E. coli*

ymfO gene (accession no. P75980). Sequence identity of the SB4 inferred gene product and the hypothetical YMFO protein was evident at the N–termini of both proteins from position 1 – 138 aa (Figure 5.4). Each portal protein serves as the entrance and exit port for the DNA, the site for DNA assembly and the attachment site for the tail (Gilakjan and Kropinski, 1999).

SB5 (protease). This 687 bp (53 %mol GC) ORF encodes a putative 228 aa with a predicted mass and pI of 25,370 Da and 5.88, respectively. BlastP search revealed a 35% sequence identity to prohead protease of HK022 and HK97 phages and a 37% identity to a *Brevibacterium flavum* bacteriophage BFK20. Furthermore, 43% sequence identity was shown with a putative prohead protease of *Rhodobacter capsulatus* (accession no. AAF13181) and *Caulobacter crescentus* (accession no. AAK24750). This ORF encodes the ST64B putative prohead protease and it overlaps the upstream ORF encoding the putative portal protein. However, a probable RBS (GGUGG) was located beyond the optimal positioning for the consensus sequence.

SB6 (major capsid protein). This 1,206 bp ORF (55 %mol GC) encodes a 401 aa putative protein with a predicted mass of 44,243 Da and pI of 5.41. The putative initiation codon is preceded by a probable RBS (AGGAG) 7 bp upstream. Searches for conserved motifs using Pfam, revealed two motifs described as bZIP transcriptional factor (PF00170) and DNA gyrase/topoisomerase IV, subunit A (PF00521). BlastP analysis revealed 43% and 37% sequence identity of SB6 inferred gene product to a major capsid protein of *Streptomyces* phi-C31 and *E. coli* CP-933C, respectively. A 24% sequence identity was shown with major head proteins of HK97 and HK022. Furthermore the SB6 gene product revealed 41%, 36% and 40% identity to the major capsid protein of phages from *Agrobacterium tumefaciens* (accession no. AAK86763), *M. loti* (accession no. NP_108602) and *Caulobacter crescentus* (accession no. AAK24747). The SB6 gene product also showed 21 to 25% sequence identity to putative major head proteins in *Lactococcus* bacteriophages bIL285, BK5–T and pi3 as well as *Streptococcus*

thermophilus phage DT1. High sequence similarity was shown at the carboxy termini when ST64B putative major capsid protein was aligned with related proteins from the database (Figure 5.5). However, an SDS-PAGE analysis (Figure 5.6) of whole ST64B phage indicated the major capsid protein had a molecular weight of approximately 28 kDa. This contrasts with a predicted molecular weight of 44 kDa. This difference may be due to post translational processing of the pre–protein before maturation as a major head protein.

SB13 (Tail protein). This 1,497 bp ORF (55 %mol GC), encodes a 498 aa putative protein with a predicted mass and pI of 53,239 Da and 5.31, respectively. A potential RBS (AGGU) was located 7 bp upstream of the putative initiation codon. Pfam revealed one conserved motif identified as a prokaryotic glutathione synthetase, N terminal domain (GSH-S_N) (PF02951). BlastP search revealed 37% sequence identity to a *H. influenzae* phage Mu gpL. The gpL product is one of the major proteins in phage Mu detected by SDS-PAGE and is a tail sheath protein (Admiraal and Mellema, 1976).

SB14 (Tail protein). This 354 ORF (51 %mol GC) encodes a putative protein of 118 aa long with a predicted mass and pI of 12,897 Da and 4.78, respectively. Upstream of the putative initiation codon (7 bp), a probable RBS (AGGAGG) was detected. The gene product encoded by this ORF showed 24% sequence identity to a tail tube protein (gpM) in Mu.

SB16 (Tail protein). This 1,928 bp ORF (56 %mol GC) encodes a 642 aa putative protein with a predicted mass and pI of 67,708 Da and 8.37, respectively. A probable RBS (GAGGA) was located 9 bp upstream of the putative initiation codon. Pfam search revealed two conserved motifs identified as major outer sheath protein C-terminal region (MOSP C) (PF02722) and an AP endonuclease family 2 (PF01261). Furthermore, one transmembrane domain was identified using the TMHMM algorithm. BlastP analysis revealed 32% sequence identity to Mu gpN and 19% identity to a phage related protein in P2. In addition, 25% sequence identity was shown with phage related proteins in the *Wolbachia* endosymbiont of *Drosophila melanogaster* (accession no. AAK85310) and

Xylella fastidiosa (accession no. F82769). Phage Mu gpN is one of the minor tail proteins with a mass of 60 kDa (Giphart-Gassler, *et al.*, 1981). However, BlastN analysis revealed that SB16 shared 99% identity at nucleotides 903 – 1,232 with the 3' end of a region flanking an oxygen–regulated gene required for *Salmonella* internalisation (*orgA*) (accession no. L33855) (Jones and Falkow, 1994).

SB18 (Tail protein). This 1,059 bp ORF (55 %mol GC) encodes a putative 352 aa protein with a predicted mass and pI of 38,188 Da and 4.89, respectively. A probable RBS (GGAGG) was located 8 bp upstream of the putative initiation codon. Pfam revealed a conserved motif of the family, nitrile hydratase alpha chain (PF02979). Furthermore, the SB18 inferred gene product showed 31% sequence identity to the tail protein (gpP) of Mu and 30% identity to a hypothetical protein in *E. coli* (accession no. BAB38407).

SB19 (Base plate assembly). This 534 bp ORF (52 %mol GC), encodes a 177 aa putative protein with a predicted mass and pI of 18,854 Da and 7.82, respectively. A probable RBS (GGAGG) was located 11 bp upstream of the putative initiation codon. The SB19 inferred gene product exhibited 30% sequence identity to a Mu gp45, 33% and 30% identity to a base plate assembly protein in phages of *Yersinia pestis* (accession no. CAC90082) and *N. meningitidis* (accession no. AAF41502), respectively. A search for conserved motifs using Pfam revealed a topoisomerase DNA binding C4 zinc finger (PF01396). The initiation codon of this ORF overlaps with the termination codon of SB18.

SB20. This 414 bp (55 %mol GC) ORF encodes a 137 aa putative protein with a predicted mass and pI of 15,750 Da and 9.04, respectively. No RBS consensus sequence was found upstream of the initiation codon. The inferred gene product encoded by SB20 exhibited 42% and 37% sequence identity to an unknown phage protein in *Y. pestis* and gp46 protein in phage Mu, respectively.

SB21. This 546 bp (53 %mol GC) encodes 181 aa inferred gene product with a predicted mass and pI of 19,689 Da and 6.72, respectively. A probable RBS (GGGACGGGG) was located 19 bp upstream of the putative initiation codon. BlastP

analysis revealed that the SB21 inferred gene product had 50% sequence identity with a hypothetical protein HI1520 of *E. coli* (accession no. G64034).

SB22. This ORF with no known putative function lies downstream of putative SB21. A potential RBS (GUGG) was located 13 bp upstream of the putative initiation codon. Pfam search revealed a nitrogen regulatory protein P-II conserved motif (PF00543). SB22 which is a 522 bp long ORF (55 %mol GC) would encode a 173 aa putative protein with a mass and pI of 18,380 Da and 5.9, respectively. BlastP revealed 57% sequence identity of this gene product to a hypothetical 28.2 kDa YMFP protein of the defective *E. coli* phage E14. This inferred gene product appears to be truncated, aligning only with the hypothetical YMFP protein of cryptic phage E14 from position 99 (Figure 5.7).

SB23. This 588 bp ORF (51 %mol GC) has a potential RBS (GGGAG) 11 bp upstream of the putative initiation codon and encodes a putative protein of 195 aa with a predicted mass of 22,126 Da and pI of 4.36. BlastP revealed 28% and 23% sequence identity of SB23 gene product with proteins in *H. influenzae* phages FluMu gp48 and Mu gp48, respectively. A higher level of sequence identity of 59% was shown with a 21.6 kDa hypothetical protein in *E. coli* (accession no. NP_415671) and 54% identity to ORF1 of *S. flexneri* phage SfV.

SB24 (Tail). This 1,563 bp (48 %mol GC) encodes a 520 aa putative protein with a predicted mass and pI of 55,364 Da and 5.05, respectively. A potential RBS (AGGA) was located 8 bp upstream of the putative initiation codon. The inferred gene product encoded by this ORF exhibited 99% sequence identity with a probable phage–related tail fiber protein in *S.* Typhimurium LT2 (AAL21137). Furthermore, 72% and 60% sequence identity with tail–like protein of P2 found in Fels–2 prophage, a λ related protein found in plasmid pMT1 (accession no. NP_047853) and a probable phage–like tail fiber protein in *Y. pestis* (accession no. CAB55186). Pfam search revealed two conserved motifs, which were identified as peptidase A4 family (PF01828) and trypsin (PF00089). However,

BlastN analysis revealed 82% sequence identity of SB24 to a small section (nt 305 – 377) of the *sopE* gene in *S. Typhimurium* which is implicated in the pathogenesis of *Salmonella*. It has been shown that *sopE* is flanked by sequences resembling tail and tail–fiber genes of P2–like phages (Mirold, *et al.*, 1999). There was an overlap between this ORF and the ORF upstream.

SB25 (Tail). This 568 bp (45 %mol GC) ORF encodes a 189 aa putative protein with a predicted mass and pI of 21,561 Da and 4.83, respectively. A probable RBS (GGGA) was located 3 bp upstream of the putative initiation codon. The inferred gene product encoded by SB25 resembles the probable tail fiber assembly protein (gpU) of Mu which is a minor protein with a mass of 20.5 kDa (Giphart-Gassler, *et al.*, 1981). BlastP analysis revealed that the product encoded by SB25 shared 63% identity with an unknown protein in Fels–2 prophage. Furthermore, 54% sequence identity was shown with putative tail fiber assembly proteins of both Gifsy–2 and Fels–1, 40% and 38% identity shown with tail fiber proteins in Mu and P2, respectively. Additionally, 37% identity was shown with an unknown protein in *E. coli* phage 186. Pfam search revealed a conserved motif identified as a domain of unknown function DUF144 (PF02413). BlastN analysis revealed additional sequence (nt 409 – 557) within SB25 with 89% sequence identity to the *sopE* gene. SB25 marks the end of the morphogenetic genes in ST64B.

SB26. This 1,008 bp (38 %mol GC), encodes a putative protein of 335 aa with a predicted mass and pI of 37,890 Da and 6.7, respectively. A probable RBS (AGGGU) was located 7 bp upstream of the putative initiation codon. BlastP revealed that SB26 inferred gene product shared a 47% sequence identity to an unknown protein encoded by a shiga-like toxin encoding phage CP-933K. Furthermore, there was 75% sequence identity to a probable cytoplasmic protein in *S.* Typhimurium (accession no. AAL21040). In addition, there was 60% and 58% sequence identity to unknown proteins in *S.* Typhimurium LT2 (accession no. AAF33527) and *E. coli* O157:H7 Z4328 (accession no. AAG58112), respectively.

SB27. The product deduced from this small 218 bp ORF (52 %mol GC) exhibited 83% and 81% sequence identity with 73 aa carboxy termini of the DNA invertase pin in prophages CP–933H and E14, respectively (Figure 5.8). TMHMM search revealed one transmembrane domain located in this ORF. Furthermore, a search for conserved motifs using Pfam revealed a Helix–turn–helix domain of resolvase. However, comparison of the gene product of SB27 (73 aa) to the gene products of cryptic E14 and CP–933H (184 aa and 324 aa, respectively), suggests this gene product is truncated. This together with the absence of a potential RBS consensus sequence upstream of the initiation codon, strongly suggests that this inferred product is not expressed.

5.3.2.3 Integration and excision

SB28 (Integrase). This 990 bp ORF (45 %mol GC) encodes a putative protein of 329 aa with a predicted mass and pI of 38,017 Da and 9.68, respectively. A weak but probable RBS (UGAGAG) was located 3 bp upstream of the putative initiation codon. Searches for conserved motifs using Pfam revealed one motif belonging to the phage integrase family (PF00589). Furthermore, BlastP search revealed considerable sequence identity of 57% and 49% to integrases of CP–933U and CP–933M phages, respectively. A considerable number of integrases from the lambdoid phages (HK022, HK97, H19J, 434 and λ) shared 25% sequence identity with ST64B putative integrase. A 30% sequence identity was also shown with a phage related integrase in *X. fastidiosa* strain 9a5c (accession no. F82654). Based upon sequence similarities to integrases of other phages and the conserved motif of the integrase family, the inferred gene product encoded by SB28 is probably an integrase which catalyses the site–specific integration of ST64B prophage. Figure 5.9 compares the hydropathy profiles of the inferred amino acid sequence of the ST64B Int and the putative Int proteins of the *E. coli* cryptic phages CP–933M and CP–933U. Kyte–Doolittle hydropathy (Kyte and Doolittle, 1982), Karplus protein flexibility (Karplus and Schultz, 1985), Parker antigenicity (Parker, *et al.*, 1986)

and von Heijne hydrophylicity (von Heijne, 1981) are displayed. Each of the plots confirm the significant conservation of sequence and likely conservation of function.

SB29. Upstream of the putative integrase, lies a 243 bp ORF (46 %mol GC) which encodes a putative protein of 80 aa with a predicted mass and pI of 9,168 Da and 5.0, respectively. A potential RBS (GGGAG) was located 11 bp upstream of the putative initiation codon. This putative protein showed no significant similarity to any protein in the database. In general, lambdoid phages encode excisionase genes, which lie upstream of the integrase genes. Alignment of the SB29 inferred gene product with several lambdoid excisionase proteins, showed some similarity at the carboxy termini of these proteins (Figure 5.10). Based upon map position, size and similarity to other excisionase proteins, it is likely that the SB29 encoded gene product is an excisionase (Xis).

5.3.2.4 A region probably encoding the *ea* genes

SB30. This 570 bp ORF (51 %mol GC) encodes a putative protein of 189 aa with a predicted mass and a pI of 20,974 Da and 4.89, respectively. A potential RBS (AGGGGA) was located 11 bp upstream of the putative initiation codon. The SB30 gene product shared 58% and 56% sequence identity with the endodeoxyribonucleases of phages CP–933O, CP–933P, CP–933M and CP–933U, respectively. Comparison of the inferred amino acid sequence of ST64B exonuclease with the analogous sequences in the CP–933 phages, revealed a high similarity within approximately the last 200 C–terminal amino acids of these proteins. Pfam search revealed one motif identified as an exonuclease (PF00929). However, the inferred SB30 product is much smaller than the exonuclease gene products of the CP–933 phages (Figure 5.11).

SB31. This 834 bp ORF (50 %mol GC) encodes an unknown protein of 277 aa with a predicted mass and pI of 30,779 Da and 4.38, respectively. BlastP search revealed a 60% sequence identity of the gene product encoded by SB31 to P22 Eaa protein. In

addition, the gene product encoded by SB31 revealed high sequence identity to unknown proteins encoded by prophages H19–J (73%), 933W (71%) and VT2–Sa (71%). Alignment of the SB31 gene product with P22 Eaa protein showed a marked conservation of sequence at the amino–terminus (Figure 5.12).

SB32. This 618 bp ORF (51, %mol GC) has a strong potential RBS (GGAGG) 8 bp upstream of its putative initiation codon and is located at a similar map position as the *ead–eai* genes in P22. Transcription is also in the same direction. However, the SB32 gene product showed 68% sequence identity to the gp45 of a linear bacteriophage, N15, and to an unknown protein in CP–933R (61%). Alignment of the SB32 gene product with Ead_{P22} and $gp45_{N15}$ showed a significant level of sequence identity in the case of $gp45_{N15}$ (Figure 5.13).

SB33. This 516 ORF (49 %mol GC) encodes a 171 aa putative protein with a predicted mass and pI of 19,756 Da and 5.81, respectively. A probable RBS (AGGAG) is located 6 bp upstream of the putative initiation codon. The inferred gene product encoded by SB33 did not have any orthologues in the database. In P22, genes involved in homologous recombination are found upstream of the genes *eaa – eai*. However, none of the ORFs in ST64B showed any similarity to any of the analogous genes involved in homologous recombination. Furthermore, no significant sequence identity was observed when SB33 putative gene product was compared to both the P22 Erf and λ Bet proteins.

SB34. This 231 bp ORF (48 %mol GC) encodes a 76 aa putative protein with a predicted mass and pI of 8,276 Da and 4.68, respectively. A potential RBS (GGCGGU) was located 7 bp upstream of the putative initiation codon. The inferred gene product encoded by SB34 did not have any orthologues in the database. The immunity region of both λ and P22 has the *cIII* and *c3* genes, respectively, located downstream of the repressor gene. Based on the size, direction of transcription and map position of SB34, it is possible that it encodes a putative CIII-like gene product, which would be involved in establishment of lysogeny in ST64B. However, alignment of this sequence with the probable analogous

sequences in P22 and λ, did not reveal any similarities. Furthermore, no conserved motifs that would confirm the inferred function of this gene product could be detected.

SB35. This 539 bp (48 %mol GC) encodes a putative protein with a predicted mass and pI of 20,285 Da and 4.4, respectively. Using Pfam, one conserved motif of the family T–antigen specific domain (PF02380) was found. BlastP analysis revealed 67% sequence identity of ST64B gene product to the unknown b2361 protein in *E. coli* (accession no. F65009). A probable RBS (AGGAG) was located 9 bp upstream of the putative initiation codon.

SB36. This 828 bp ORF (51 %mol GC) encodes a 275 aa putative protein with a predicted mass and pI of 30,670 Da and 5.0, respectively. BlastP search revealed that the SB36 gene product shared 53% and 35% sequence identity with a 30.5 kDa hypothetical protein in *E. coli* (accession no. P76513) and to a conserved hypothetical protein in *X. fastidiosa* (accession no. NP_298938), respectively.

SB37. This 639 bp ORF (50 %mol GC) encodes a putative protein of 123 aa with a mass and pI of 13,471 Da and 7.93, respectively. A probable RBS (GGGU) was located 6 bp upstream of the putative initiation codon. This putative protein has a significant amino acid identity (50%) to a 16.5 kDa unknown protein in *E. coli* and 37% sequence identity to an unknown protein in *X. fastidiosa*. In λ and P22, the genes N_{λ} and 24_{P22}, lie downstream of the repressor genes cI_{λ} and $c2_{P22}$. Based on the map position, size and direction of transcription of SB37, the product of this gene may be an early gene transcriptional regulator. However, when aligned with N_{λ} or $gp24_{P22}$ there was little sequence identity. Nevertheless, it has been shown that N_{λ} and $gp24_{P22}$ themselves share little sequence identity (Franklin, 1985a). It is thus possible that the product encoded by SB37 may be a transcriptional regulator.

5.3.2.5 Immunity genes

SB38 (CI). This 696 bp (44 %mol GC) encodes a putative protein of 231 aa with a predicted mass of 25,665 Da and pI of 6.33, respectively. This ORF is preceded by a GGUU 11 bp upstream of the putative start codon, which resembles the consensus for a RBS (GGAGGUU). BlastP analysis revealed that the product encoded by SB38 shared 42% and 26% sequence identity to the CI repressor proteins of HK022 and phage D3, respectively. Additionally, 30%, 28%, 27% and 26% identity was observed to phage like repressor proteins found in *X. fastidiosa* (accession no. XF1650), *H. influenzae* (accession no. A64031), *N. meningitidis* (accession no. CAB71959) and *Streptococcus pyogenes* (accession no. NP_269132), respectively. Based upon similarity to several phage and bacterial repressors, the product specified by SB38 is probably a CI–like repressor. This hypothesis is supported by the presence of a helix–turn–helix (PF01381), peptidase family S24 (P00717) and a signal peptidase S26 (PF00461) conserved motifs. Furthermore a putative rho–independent terminator was located downstream of the ST64B putative *cI* gene and a promoter region

CGCAATT**TTACA**CTTGCGCAATGTGCAAT**TTAAAT**GTAAT**T**TGCACTCAT
–35 –10 +1

was identified using Martin Reese's Promoter Prediction by Neutral Network program (http://www.fruitfly.org/seq_tools/promoter.html), and may represent a putative P_L.

SB39 (Cro). Unlike the putative repressor CI encoded on the complementary strand, this putative ORF is transcribed in the rightward direction and is 225 bp long (46 %mol GC). A strong potential RBS (GGAGG) was detected 5 bp upstream of the initiation site. The inferred gene product encoded by the SB39 ORF is 74 aa with a predicted mass of 8,406 Da and and pI of 7.92, respectively. Pfam search revealed one motif of the family helix–turn–helix 3 (PF01381). In addition, BlastP analysis revealed 34% sequence identity of this gene product to a probable DNA–binding protein of

Streptomyces coelicolor (accession no. CAB39703) and 32% identity to a putative protein, similar to the family of transcriptional regulators, MerR and LacI in *Sinorhizobium meliloti* (accession no. CAC49293). It has been shown that CI and Cro proteins of λ as well as C2 and Cro proteins of P22 have helix–turn helix motifs. Based upon the fact that the SB38 and SB39 gene products have this motif and their relative map positions, suggest that they probably are CI and Cro orthologues of ST64B.

SB40. This 282 bp ORF (50 %mol GC), has a potential RBS (GCUGG) 6 bp upstream of the putative initiation codon. It potentially encodes a 93 aa inferred gene product with a predicted mass and pI of 10,363 kDa and 4.5, respectively. BlastP search indicated that the inferred gene product encoded by SB40 shared 71% sequence identity with a hypothetical protein encoded by a gene of unknown function, *ymfL* in the cryptic lambdoid prophage E14. The map position and size of SB40 suggests that it encodes a CII-like (λ CII orthologue) gene product. However, no potential CII–binding site ($TTGCN_6TTGC$) was found associated with this ORF. The absence of the CII–binding site found in λ probably indicates that the putative CII protein of ST64B is distinct from both the P22 C1 and λ CII proteins.

SB41. This 1,158 bp ORF (50 %mol GC) encodes a putative protein of 385 aa with a predicted mass and pI of 42,432 Da and 9.18, respectively. A potential RBS (UGGA) was located 11 bp upstream of the putative initiation codon. BlastP analysis revealed 58% sequence identity to ORF199 of a satellite bacteriophage P4. ORF199 was thought to encompass the whole immunity region of P4 (Ghisotti, *et al.*, 1992). Furthermore, 52% and 53% sequence identity was observed with a Rha protein from the lambdoid phage phi80 and ORF179 of the bacterium *S. flexneri*, respectively. BlastN analysis revealed a portion of an insertion sequence IS*911*–like at position 506 to 507 (67 nt) with 92% identity to an IS*911* element of *S. flexneri.* The discovery of an IS*911*–like insertion sequence within this ORF probably suggests that the product of this gene may not be expressed and may not have an effect on the immunity of ST64B (see discussion).

5.3.2.6 The replication genes

SB42. This 819 bp ORF (49 %mol GC), has a strong probable RBS (GGAGG) 7 bp upstream of the putative initiation codon. The inferred gene product encoded by SB42 is 272 aa in size with a predicted mass and pI of 29,075 Da and 8.93, respectively. BlastP analysis revealed that the inferred gene product encoded by SB42 shared a 37% sequence identity with an unknown protein in phage CP–933M and with a hypothetical protein in *E. coli* (accession no. NP_309100). However, the carboxy terminus of SB42 inferred gene product (249 to 272 aa) showed 83% identity to a λ–like replication protein O in *E. coli* (accession no. BAA16223). Furthermore, 36% amino acid sequence identity was observed at position 19 to 88 of the SB42 gene product to a putative λ–like replication protein in *E. coli* O157:H7 (accession no. NP_310303). Pfam search revealed a bacterial regulatory protein of the ArsR family (PF01022) and an uncharacterised protein family UPF0074 (PF02082). A search for direct repeats revealed two 11 bp repeats **(GAACCAGTCAAaGAACCAGTCAA)** located at positions 31,415 – 31,425 and 31,427 – 31,437. In many lambdoid phages including λ, P22 and ES18, direct repeats are associated with the origin of replication. Consequently, it is likely that SB42 encodes a gpO orthologue protein in ST64B.

SB43. This 525 bp ORF (52 %mol GC) has a potential RBS (GGGA) 1 bp upstream of the putative initiation codon. SB43 would potentially encode a 174 bp putative protein with a predicted mass and pI of 19,653 Da and 9.73, respectively. The SB43 gene product showed 69% sequence identity to an 18.8 kDa hypothetical protein in the INTC–DSDC intergenic region of *E. coli* (accession no. P76510). Additionally, 41% sequence identity was shown with a putative transcriptional activator in *E. coli* O157:H7 over a short span of sequence (position 110 – 157). The SB43 gene product and the hypothetical protein in *E. coli* showed very good sequence similarity, most notably at the carboxy terminus. A second potential AUG start was detected at position 43 of the SB43 ORF. If this second AUG is indeed the probable initiation codon, the inferred gene

product would be shorter and thus have a higher overall similarity to the hypothetical protein in *E. coli* (Figure 5.14). Although Pfam did not reveal any conserved motifs, ScanProsite at Expasy identified an ATP/GTP–binding site motif A (P–loop).

SB44. The inferred gene product encoded by this 894 bp ORF (54 %mol GC), is 297 aa long with a predicted mass of 33,375 Da and pI of 4.90. A potential RBS (AGGG) was located sub–optimally 16 bp upstream of the putative initiation codon. BlastP revealed 75% sequence identity to both a 10.3 kDa hypothetical protein, the product of the *E. coli yfdN* gene (accession no. P76509) and to another hypothetical protein, b2356 in *E. coli* (accession no. A65009). Furthermore, 33% sequence identity was shown to a putative methyltransferase encoded within phages CP–933V and VT2–Sa. A search for conserved motifs using Pfam revealed one isocitrase lyase motif (PF00463). The termination codon of SB43 and initiation codon of SB44 overlap each other in the sequence UAAUG. If an arbitrary insertion (e.g. base C immediately after base A of the start codon) could be introduced to join SB43 and SB44, a gene product (471 aa or 457 aa if the second AUG is taken as the initiation codon in SB43) with a comparable size to the P22 gp12 (459 aa) would be obtained. This region was carefully analysed and re–sequenced to ascertain the possibility of a single base sequencing error, which would allow merging of the two overlapping ORFs, however, repeated sequencing confirmed the original result. Nevertheless, in lambdoid phages, gene *O* is followed by gene *P*. Based on the fact that SB42 probably encodes a λ–like gpO protein, it is possible that both (SB43 and SB44) or at least one (SB43) gene product(s) would encode a gpP orthologue.

5.3.2.7 SB45 to SB47

SB45. This 390 bp ORF (55 %mol GC) encodes a 129 aa putative protein with a predicted mass of 14,383 Da and pI of 9.33. A potential RBS (GGAGG) was located 5 bp upstream of the putative initiation codon. BlastP search revealed 56%, 43%, 42% and 39%

sequence identity of the SB45 gene product to putative endodeoxyribonuclease proteins (Rus) in many bacteriophages including CP–933O, CP–933U, HK620 and CP–933P, respectively. In *E. coli*, homologous pairing and strand exchange are catalysed by the RecA protein, responsible for *in vivo* recombination (Kowalczykowski, *et al.*, 1994). RecA intermediates are processed by junction–specific DNA helicases such as RuvAB and RecG (Tsaneva and West, 1994). RuvAB is predominantly targeted to Holliday junctions where it drives the migration of the junction along the DNA by a reaction involving unwinding and rewinding, thereby rotating the bound arms through the RuvB rings (Mahdi, *et al.*, 1996). RusA and RuvC are endonucleases linked with the resolution of the Holliday junctions into duplex products. RusA resolves junctions by making a dual incision at specific sequences and is encoded within an *E. coli* phage 82 and a defective phage related to 82 (Mahdi, *et al.*, 1996). However, it has been shown that deletion of *rusA* from the chromosome has no effect on recombination or DNA repair (Mahdi, *et al.*, 1996).

SB46. This 891 bp ORF (51 %mol GC), encodes a 286 aa putative protein with a predicted mass and pI of 31,627 Da and 9.23, respectively. A potential RBS (GGAGA) was located 8 bp upstream of the putative initiation codon. Pfam search revealed one conserved motif identified as a bacterial regulatory protein of the GntR family (PF00392). BlastP analysis revealed 40%, 37% and 27% sequence identity to a KilA protein of phage P1, a phage related putative protein in *X. fastidiosa* (accession no. XF2294) and an unknown protein similar to the KilA protein of phage P1 in HK620, respectively.

SB47. This 762 bp ORF (54 %mol GC) did not have a potential RBS sequence upstream of the putative start codon. However, a second potential AUG initiation codon, 36 nt downstream of the initial AUG may serve as a potential initiation codon. If this second AUG is used as an initiation codon, then a GGGA 9 bp upstream of this second putative initiation codon would serve as a RBS. BlastP analysis revealed 97% sequence identity to a putative cytoplasmic protein in *S.* Typhimurium LT2 strain (accession no.

NP_461183). Furthermore, 43% and 42% identity was shown to unknown proteins in *E. coli* O157:H7 (BAB35618) as well as the phages CP–933O and CP–933N, respectively. BlastN showed 99% sequence identity to the *sspH2* gene of *S.* Typhimurium. This region however, covered the last 100 bp of SB46 and extended to 704 bp of the 742 bp of SB47. Furthermore, comparison of this putative region with the *S.* Typhimurium *sspH2* gene revealed an excellent sequence identity from position 2,871 of the *S.* Typhimurium *sspH2* gene to its 3' end (Figure 5.15). This leucine–rich region, is associated with the SPI–1 and SPI–2 type III secretion systems. However, it has been shown that the N–terminus of *sspH2* gene product (which is missing in the ST64B encoded *sspH2*–like gene product) encodes the secretion/translocation signal (Miao, *et al.*, 1999).

SB49. This 762 bp ORF (54 %mol GC) encodes a 207 aa putative protein with a predicted mass of 24,093 Da and pI of 9.3. A potential RBS (UAAGGA) was located 1 bp upstream of the putative initiation codon. The inferred gene product encoded by SB49 did not show any significant sequence similarity to proteins in the database.

SB50. This 360 bp ORF (45 %mol GC) has a potential RBS (AGGU) sub–optimally located 21 bp upstream of the initiation codon and is positioned upstream of the lysis genes, suggesting that it would encode a late gene regulator analogous to Q_{λ} or $gp23_{P22}$. However, both BlastP and BlastN failed to identify any related sequences in the NCBI database. A search for the CII–binding site ($TTGCN_6TTGC$) within or close to SB50 was unsuccessful. However, as the probable ST64B gene *cII* seems to be distinct from the cII_{λ} and $c1_{P22}$, the binding sites may also be significantly different. It is therefore possible that SB50 encodes an ST64B Q–like protein, distinct from its orthologues in P22 and λ and may be involved in the anti–termination of its late genes. The gene product encoded by this ORF is 119 aa with a predicted mass of 13,952 Da and pI of 9.57.

5.3.2.8 The genes involved in lysis

SB51 (Holin). This 312 bp (48 %mol GC) encodes a putative protein of 103 aa with a predicted mass and pI of 12,023 Da and 6.26, respectively. A potential RBS (GAGG) was located 3 bp upstream of the putative initiation codon. The inferred gene product encoded by SB51 revealed a 100% sequence identity (over the first 84 aa of a total of 103 aa) to a protein in *S.* Typhimurium prophage Fels–1 (Figure 5.16). Furthermore, 36% sequence identity was shown to an unknown 12.2 kDa protein in *X. nematophila* (accession no. CAB58444). Additionally, the putative holin of ST64B shared 36% sequence identity with $gp13_{P22}$ from position 29 – 58 aa. Weak similarities of SB43 putative gene product with holins of other lambdoid phages, HK97 and HK620 was also observed. Moreover, the conserved dual start motif feature observed in most lambdoid phage holin proteins, was also observed in the putative holin of ST64B as well as in the putative protein in Fels–1, with initiation codons at Met–1 and Met–3. This motif seems to be important in the timing of lysis. TMHMM analysis revealed 1 transmembrane domain. The 100% sequence identity of the ST64B putative holin protein with the putative protein in Fels–1 suggests that the putative protein in Fels–1 phage is probably a holin.

SB52 (Endolysin). This 615 bp ORF (54 mol% GC) encodes a putative protein of 204 aa with a predicted mass of 22,350 Da and pI of 9.19. A weak potential RBS (AGGU) was located 7 bp upstream of the putative initiation codon. Pfam search revealed one motif of the chitinase class I family (Glyco_hydro_19) (PF00182). The product specified by SB52 is most probably involved in lysis but has almost no similarity with both P22 lysozyme and λ transglycosylase, suggesting that ST64B probably uses an endolysin with a different activity. BlastP analysis revealed that the putative endolysin of ST64B shared 93% sequence identity with a protein related to the *P. aeruginosa* lytic enzyme found in Fels–1. Furthermore, the putative endolysin of ST64B shared 41% and 34% sequence identity with a putative glycohydrolase lytic enzyme in *P. aeruginosa* (accession no.

BAA83137) and *Deinococcus radiodurans* (accession no. AAF12239), respectively. Both the ST64B and Fels–1 ORFs encode inferred gene products of the same size (204 aa).

SB53 (Rz). This 486 bp ORF (53 %mol GC) encodes a putative protein of 161 aa with a predicted mass of 16,557 Da and pI of 9.27. A potential RBS (GGUG) was located 7 bp upstream of the initiation codon. The inferred gene product encoded by SB53 exhibited 93% sequence identity to an unknown Fels–1 protein. Additionally, the inferred SB43 gene product exhibited 82%, 67%, 65%, 64% sequence identity to $gp15_{ES18}$, Rz_{21} and Rz_{λ}, Rz_{Rac} and Rz_{H-19B}, respectively. The carboxy termini of these phage proteins exhibited a high level of identity. However, Rz_{H-19B} has a carboxy terminal extension beyond the region of identity (Figure 5.17). Furthermore, Pfam revealed an ATP synthetase B/B' CF (0) motif (PF00430). The putative *Rz* gene of ST64B signifies the end of the lysis genes and appears to be conserved within the lambdoid family.

SB56. This 351 bp ORF (53%) with a potential RBS (GGGAGU) located 5 bp upstream of the putative initiation codon would encode a putative 119 aa protein with a predicted mass and pI of 13,952 Da and 9.57, respectively. Pfam search revealed a conserved motif in the family of HNH endonuclease (PF01844). BlastP search revealed 60% sequence identity to a 13.3 kDa hypothetical provirus protein in *Xenorhabdus nematophila* (accession no. CAB58450).

5.3.3 Identification of the integration site of ST64B

As a prophage, the ST64B genome is probably integrated within the host *S.* Typhimurium DT 64 chromosome between a phage attachment site (*attP*) and a host bacterial attachment site (*attB*) thereby generating two additional sites at the right (*attR*) and left (*attL*) junctions. In order to map the phage–host junctions, genomic DNA extracted from ST64B and *S.* Typhimurium DT 64 were digested with several restriction enzymes (*Bam*HI, *Eco*RI, *Hin*dIII, *Cla*I, *Bgl*II, *Acc*I, *Apa*I, *Spe*I, *Pst*I and *Kpn*I) and

subjected to Southern blot analysis. Two oligonucleotide primers, F1 and F2, were designed at positions 20,268 and 20,724 (5' –GGCAGAGTAATAGCTGGAGTTGC –3' and 5' – GTTATCGGCTAACCGGAAGGG – 3'). These were used as forward primers. A reverse primer, R1, was designed at position 21,215 (5' –CCTGGACAGGAAGCGTTGAGC –3'), 33 nt downstream of the *int* gene. Using F1 and R1 as well as F2 and R1 primers, a fragment from each set of primers was amplified from the phage genome but could not be amplified from the bacterial genome. This supported the idea that the amplified fragments carried *attP* which, is split by integration of the phage genome into the bacterial genome and thus could not be amplified from the genomic DNA of the lysogen.

The amplified fragment using degenerate primers F2 and R1 was digoxigenin–labelled and used as a probe in Southern analysis of restriction endonuclease digested phage and lysogen genomes. Two bands could be seen in the bacterial genome restricted with *Eco*RI, *Hin*dIII, *Cla*I and *Acc*I (Figure 5.18). These bands were thought to carry the *attL* and *attR* junction sites. The *Hin*dIII fragments were chosen because of their smaller size (approximately 9.4 and 6.1 kb) as compared to the *Eco*RI fragments (approximately 12.7 and 6.5 kb). The two *Hin*dIII fragments were gel purified and cloned into two low copy number plasmids (pGB2 and pCL1920) (Table 2.4) and electroporated into DH5α cells (Table 2.2). Colony blot hybridisation was used for the selection of transformants using the method of (Paton, *et al.*, 1992) as explained in chapter 2. The smaller *Hin*dIII fragment (~6.1 kb) was successfully cloned into both pGB2 and pCL1920 (Table 2.4). However, the larger fragment (~9.4 kb) probably containing the other junction could not be cloned. Sequencing of the ends of the pGB2 insert was achieved by dye primer sequencing of small scale DNA using the M13 forward primer. The remaining sequence was determined by primer walking with dideoxy chain–termination, using CsCl purified plasmid DNA as described in chapter 2. Analysis of the sequence revealed the phage/bacterial junction. The junction was found 162 bp downstream from gene *int*.

ST64B therefore integrates within the tRNA gene *serU*, immediately upstream from *umuCD* operon. Comparison of the nucleotide sequences of ST64B, the ST64B *attP/attB* clones amplified from *S.* Typhimurium DT 64 and the *S.* Typhimurium LT2 (accession number AE008788) revealed a sequence shared by all three genomes (Figure 5.19), with the *attB/attP* clone showing perfect homology with the ST64B left of the junction and almost perfect homology with both the *S.* Typhimurium DT 64 and LT2 sequence. LT2 however has a C deletion at position 269 at the right of the junction.

5.3.4 Potential CII–binding sites and integration host factors

A search for the CII–binding site (TTGC(N_6)TTGC) motif in ST64B nucleotide sequence revealed 3 sites at positions 28,623 – 28,636 (**TTGC**TTTTGC**TTGC**), 28,790 – 28,803 (**TTGC**GCAAGT**TTGC**), and 39,152 – 39,165 (**TTGC**TTCACG**TTGC**). Examination of the genomic sequence upstream of the putative *int* and *xis* genes did not reveal any CII–binding site motif. Attempts to identify CII–binding sites for promoters P_{RE} and P_{aQ} were unsuccessful probably because the ST64B putative CII protein may be distinct from its λ and P22 orthologues. Three CII–binding site motifs were located within the putative *cI* gene at position 28,623 – 28,636, upstream of the *cI* gene at position 28,790 – 28,803 and within the putative replication gene *O* at position 31,197 – 31,210.

Large numbers of IHF–binding sites were discovered with the on–line program MacTargsearch (Goodrich, *et al.*, 1990) but only five sites with scores of greater than 48 are presented in Table 5.2. These included the sequence AGAATTAC**ATTCA**CACA**ATTA**TGGCGAAAATTT that has a score of 55.7 and is located within the *int* gene.

5.3.5 Potential rho–independent terminators and promoters

A total of 5 potential rho–independent terminators were identified using the search algorithm of Brendel and colleagues (Brendel, *et al.*, 1986, Brendel and Trifonov, 1984). A list is presented in Table 5.2 and includes the rho–independent terminator identified at position 28,049 – 28,010, transcribed in the leftward direction, (AAAGCGTCGACAAG**GCCACCATT**A**TGGTGGC**TTTTTTTTT). A search for promoters using Martin Reese's Promoter prediction program (http://www.fruitfly.org/seq_tools/promoter.html) revealed several promoters including a probable P_L at positions 27,989 – 27,941 downstream of putative *cI* gene and transcribed in the leftward direction. A probable P_{RM} promoter was located at position 28,855 – 28,810 upstream of the *cI* gene and transcribed in the leftward direction. A probable P_R promoter was located at position 28,794 – 28,839, immediately upstream of the *cro* gene and transcribed rightwardly. Furthermore, a probable P_{Late} promoter was identified at positions 36,838 – 36,883 148 bp upstream of gene *13*. Other promoters identified are listed in Table 5.3.

5.4 Discussion

The genomic architecture of ST64B resembles that of the lambdoid phages (Figure 5.2). ST64B and ST64T, which are co–lysogens of *S.* Typhimurium DT 64 strain, did not show a strong relationship, except for their general lambdoid genomic architecture. Most predicted ST64B gene products showed low sequence identity to proteins in the databases. A similar scenario was observed with the gene products of *Streptomyces* temperate phage ϕC31 (Smith, *et al.*, 1999).

5.4.1 Genes involved in packaging and morphogenesis

DNA replication in many viruses including the lambdoid phages, involves the synthesis of concatamers, which are used as substrates for the packaging enzymes (Gilakjan and Kropinski, 1999). Packaging of λ DNA involves cutting of the DNA from a concatemer and translocating the DNA into a prohead. The site containing DNA packaging signal, *cos*, consists of three segments, *cosB*, *cosN* and *cosQ* (Cue and Feiss, 1993). The λ terminase binds at *cosB* sites and introduces staggered ends at *cosN,* generating the 12– base cohesive ends of virion DNA (Cue and Feiss, 1997). The *cosQ* site is required for termination of packaging (Cue and Feiss, 1993). In ST64B, a putative terminase small subunit was identified upstream of the putative terminase large subunit by map position and sequence identity to other phage terminase small subunits. Interestingly, it had sequence identity to terminases of the Gram positive bacterial phages *Lactococcus* BK5–T and *Streptococcus* Sfi21, which recognise the 3' *cos* sites for DNA packaging (Desiere, *et al.*, 1999). However, when ST64B was digested with *Eco*RI, *Hin*dIII and *Sma*I no under–represented band(s) were found. It is likely that ST64B packages its DNA similar to the 3' *cos* recognising phages. Further studies are required to confirm this.

The putative terminase large subunit which forms the staggered cuts during DNA packaging, showed very high sequence identity to a hypothetical YMFN protein in *E. coli* as well as to the large terminase subunit of *P. aeruginosa* phage D3. Furthermore, the large terminase subunit of phage D3 was shown to be closely related to the product of *ymfN* gene (Gilakjan and Kropinski, 1999), which also revealed 93% sequence identity to putative large terminase subunit protein of ST64B. Interestingly, the ST64B large terminase subunit showed considerable sequence identity to terminases of phages from Gram positive bacteria including, *Staphylococcus*, *Clostridium*, *Bacillus*, and *Lactococcus.* Furthermore, sequence identity was also shown to a protein found in the purple non–sulphur photosynthetic bacterium, *Rhodobacter capsulatus*. Phylogenetic studies on large terminase subunits of phages from diverse bacterial genera showed that phage

terminases, like the integrases, are a diverse group of proteins (Gilakjan and Kropinski, 1999). The similarity of the putative terminase large subunit to those of the *cos*–recognising phages supports the hypothesis that the ST64B genome has the 3' extended termini analogous to those found in phage D3 and phages that largely infect Gram positive bacteria even though these *cos* sites were not found during sequencing. Further studies are required to confirm the nature of the DNA packaging process that occurs in ST64B.

The mature head shell of most dsDNA phages consists of several copies of a major protein subunit plus a small number of minor subunits (Duda, *et al.*, 1995). In many phages, shell assembly is thought to compose of both the major shell subunit and the scaffolding protein which is thought to promote correct assembly of the shell subunit (Duda, *et al.*, 1995, Prevelige, *et al.*, 1988). Amongst the minor capsid proteins, portal protein seems to be the only universally conserved element among the known phages (cited in Duda, *et al.*, 1995). Unlike ST64T which is almost identical to P22 on the basis of its capsid assembly, ST64B capsid genes resemble those of HK97, HK022 and D3. Coliphage HK97, possesses a unique mechanism of head assembly, which does not involve a scaffolding protein. The first shell structure, prohead I, is composed of 420 copies of the 42 kDa of major head protein and 12 copies of the portal protein (Duda, *et al.*, 1995). Almost one quarter of the major head protein subunit mass is lost leaving prohead II with a mass of 31 kDa (Duda, *et al.*, 1995). Shell expansion of prohead II due to the conformational change in the subunits, leads to an angular structure head I, and ultimately the mature phage head (head II) comprising 31 kDa subunits, with the autocatalytic formation of covalent bonds between subunits (Duda, *et al.*, 1995). A number of phages belonging to the *Siphoviridae* family, seem to have this "chain mail" feature (Gilakjan and Kropinski, 1999). For example, *Mycobacteriophage* L5 capsid undergoes conformational changes leading to cross–linking of its subunits (Hatfull and Sarkis, 1993). The arrangement of the head genes of both *Pseudomonas* D3 and *Streptomyces* ϕC31 is similar

to that of HK97 with portal, protease and major capsid genes arranged without a scaffold (Gilakjan and Kropinski, 1999, Smith, *et al.*, 1999). Furthermore, the capsid proteins of D3 display an unusual electrophoretic pattern, reminiscent of that of HK97 in that they lack a 43 kDa major capsid band together with high molecular weight proteins including protein which failed to enter the gel matrix (Gilakjan and Kropinski, 1999).

The data presented here suggests that ST64B probably has a similar capsid assembly process as HK97 and related phages. The arrangement of the head genes of ST64B is similar to that of HK97, D3 and ϕC31 with portal, protease and major head genes arranged without a scaffold. Both putative portal and protease proteins from ST64B showed significant sequence identity with analogous proteins in HK97, HK022 and ϕC31. The putative major capsid protein of ST64B showed sequence similarity to many proteins from other bacteriophages of the Gram positive bacteria. However, 24% identity was found with major head subunit precursor of both HK97 and HK022 phages. Studies on the head structure of HK97 have shown that the head protein subunits are cross–linked through the side chains of Lys169 and Asn356 and the bridges hexamers and pentamers, are found in a complex catenated structure referred to as "protein chain mail" (Duda, 1998). A very thin but stable capsid structure is formed by these conformational changes.

Alignment of the ST64B major head protein with that of HK97 using Align algorithm (Person, *et al.*, 1997) did not show conservation of Lys169 and Asn356 residues in the ST64B major head protein, even though, similar residues were found at Arg171 (Arg is similar to Asn) and Lys372 (Lys is similar to Asn). SDS–PAGE was used to estimate the size of the major head protein of ST64B, however, a band corresponding to the predicted 44 kDa putative capsid protein could not be found. Nevertheless, a band with a lower molecular weight of approximately 27 – 30 kDa was observed (Figure 5.6). It is possible that the larger pre–protein is processed prior to maturation as the major head protein. In contrast to D3 and HK97 phage proteins where high molecular weight material did not enter the gel, no similar material was shown to be trapped within the wells of the

SDS–PAGE gel when ST64B proteins were electrophoresed. This may be explained by the absence of conserved residues Lys169, Asn356 found in HK97 and Lys178, Asn363 found in D3 that form cross–links, with the resulting covalent modification which can be observed by SDS–PAGE gel as the appearance of very high molecular weight bands (Duda, 1998, Gilakjan and Kropinski, 1999).

The putative tail proteins of ST64B were shown to be similar by sequence to those of phage Mu and to a lesser extent to those of P2. In addition, other putative tail gene products encoded by SB24 and SB25 showed high sequence identity to tail proteins of *S.* Typhimurium prophages Fels–1, Fels–2 and Gifsy–1, in addition to those of phage Mu. The tail fiber proteins of Mu are similar to those of P1 (Chow and Bukhari, 1976, Iida, 1984) and possibly to those of P2 (Haggård-Ljungquist, *et al.*, 1992) although both P1 and P2 are unrelated to Mu. This is supported by the fact that the receptors of Mu, P1 and P2 are located in the core structure of the LPS on the outer surface of the bacterial cell (cited in Haggård-Ljungquist, *et al.*, 1992). Furthermore, bacterial strains resistant to P2 are usually also resistant to P1 and Mu (Grundy and Howe, 1984). The gene products encoded by SB13 and SB14 shared 37% and 24% sequence identity with the tail sheath protein (gpL) and tail tube protein (gpM) of Mu, respectively. Other ORFs encoding the inferred tail gene products include SB16 (which potentially encodes a gene product with sequence similarity to phage related proteins gpN of Mu and T of P2) and SB18 (which potentially encodes a tail protein similar to gpP of Mu). However, BlastN revealed a 3' end of the *orgA* gene, an oxygen–regulated gene involved in the entry process of *Salmonella* into eukaryotic cells, at position 10,579 to 10,896 which is within SB16 (Jones and Falkow, 1994). Two additional putative tail genes (SB24 and SB25) revealed *sopE* related sequences at positions 17,222 – 17,294 and 18,480 – 19,049, respectively. However, when the inferred gene products of these ORFs were examined, both the tail fiber and *sopE* sequence were in the same reading frame. Nevertheless, it has been shown in phage Mu that all amber mutants defective in any of the tail genes (*K, L, M, Y, N, P, Q, V, W* and *R*)

produced normal heads with no visible tail structures (Grundy and Howe, 1985). Similar results were obtained in λ where none of the tail–defective mutants, other than mutants in *Z* or *U*, produced tail–related structures detectable in crude lysates by electron microscopy (cited in Grundy and Howe, 1985). Both the U and Z gene products are thought to act late in λ tail assembly to produce tails competent for head attachment (Katsura and Kühl, 1974).

The phage Mu particle consists of an icosahedral head with a contractile tail. The six tail fibers are attached to the tail with a knob–like structure at the junction of the head and tail (cited in Grundy and Howe, 1985). When phage lysate containing both ST64B and ST64T was examined by Transmission Electron Microscopy (TEM), icosahedral heads with short tails were observed. Additionally, when the CsCl gradient purified ST64B lysate was examined, the majority of phages showed icosahedral heads with no visible tails, with the presence of a few phages with short tails. Because ST64B and ST64T are carried by the same *S.* Typhimurium DT 64 strain and ST64B could not be propagated on any strain tested, it was not possible to test a pure lysate of ST64B. This may explain why a few phages with short tails that were observed under TEM, could be ST64T contaminants. Furthermore, unlike ST64T which has one putative tail gene *9*, ST64B has many putative tail genes, suggesting that it should have a longer tail than that of P22 and ST64T (if the tail genes are expressed). Although ST64B could be isolated from the *S.* Typhimurium DT 64 isolate, it failed to plaque on diverse *Salmonella* and *E. coli* strains. Further studies on ST64B putative tail genes are required to confirm whether they are expressed.

5.4.2 Integration and excision

In 1962, Campbell proposed that the intracellular phage DNA circularises after infection and a region of the genome aligns with a homologous region on the bacterial

genome. The phage DNA is then inserted into the chromosome by a site–specific recombination process (Smith and Levine, 1967). Site–specific recombination in lambdoid phages is catalysed by an integrase gene *int*. ST64B integrase showed sequence similarity to several lambdoid phage integrases including that of λ. The ST64B genome integrates within *serU,* which encodes a tRNA for serine. Integration of phages within, or close to tRNAs is not uncommon. In addition, many pathogenicity islands insert at sites close to tRNAs. For example, the pathogenicity islands of *E. coli*, PAI–I and LEE (enteropathogenic locus for enterocyte effacement) and a retronphage ϕ73 insert downstream from the *selC* gene, which encodes a tRNA for selenocysteine (Blum, *et al.*, 1994, McDaniel, *et al.*, 1995, Sun, *et al.*, 1991) whereas PAI–II and phage P4 insert downstream *leuX*, encoding a tRNA for leucine (Blum, *et al.*, 1994, Pierson and Kahn, 1987). Although the Int gene product of ST64B showed homology with other Int proteins in the database, no homologues were found for the putative excisionase gene product. However, the map position and size of SB29 suggested that its gene product may be a Xis protein responsible for excision of the ST64B genome during induction of the phage.

5.4.3 Genes involved in homologous recombination

The region in P22 upstream of the *eai* gene encompasses three genes (*abc2*, *abc1* and *erf*) involved in homologous recombination. Similarly, λ has genes *exo*, *gam* and *bet* that are also involved in homologous recombination (Poteete, 1988). None of the ORFs in ST64B shared any sequence similarity with any of these genes. It is possible that SB33 may encode a putative Erf product based on size and map position, however, when compared to both the P22 Erf and the λ Bet proteins, no significant sequence similarity was observed.

5.4.4 Regulation of gene expression by the putative ST64B N protein (λ N and P22 gp24 orthologue)

In many lambdoid phages, the P_L promoter regulates the expression of genes by the phage early in infection. A small protein (78 aa to 127 aa) called N in λ and gp24 in P22, regulates the switch from immediate–early to delayed–early gene expression (Franklin, 1985b). In addition, anti–termination is dependent on several host proteins including RNA polymerase, NusA, –B, –E, –G, Rho, ribosomal protein S10 as well as *cis*–acting sites on the RNA (Kropinski, 2000). Unlike ST64T where an early gene regulatory protein (gp24) was easily identified by sequence similarity and the identification of the *nut* sites associated with it, attempts to identify N–like proteins using BlastP or searches for conserved motifs failed to identify an N orthologue in ST64B. A candidate for this gene product could be encoded by SB37 based upon map position, size (123 aa) which is between 78 and 127 aa (Franklin, 1985b) and the direction of transcription. Furthermore, a potential P_L promoter

(CGCAATT**TTACA**CTTGCGCAATGTGCAAT**TTAAAT**GTAAT**T**TGCACTCAT)
−35 −10 +1

was identified downstream of putative repressor gene, *cI*. A rho–independent terminator was also identified as explained in section 5.3.5. However, no *nut* sites were detected in this area, which is not surprising since the putative *N* sequence of ST64B is distinct from the N_λ or 24_{P22} sequences, suggesting that the *nut* sites may also be distinct.

5.4.5 Immunity genes

A putative repressor protein, CI of ST64B was identified by its similarity to repressors of many other bacteriophages and motifs found within these repressors, however, the inferred Cro product was identified by size, map position as well as a predicted helix–turn–helix motif. A CII–binding site (**TTGC**TTTTGC**TTGC**) was found

at position 28,623 – 28,636 within the *cI* (*SB38*) repressor gene. Another CII–binding site (**TTGC**GCAAGT**TTGC**) was located at position 28,790 – 28,803 upstream of the *cI* (*SB38*) repressor gene. However, the last of these three genes predicted to be involved in ST64B immunity, the product of the *cII* gene (λ terminology) or *c1* (P22 terminology) could not be identified. A small gene product (93 aa) lies downstream of the putative *cro*–like gene. This ORF is the same size as the putative *c1*–like gene of ST64T, which itself showed high similarity with the *c1* gene of phage L. Furthermore, SB40 maps at the same position as other lambdoid *cII* (P22 *c1* orthologue) genes. The absence of the CII–binding sites within or close to SB40 probably indicates that this gene is distinct from the λ CII and P22 C1 proteins. A putative ST64B CIII gene product is probably encoded by SB34, which is located downstream of this region based upon map position and size. This region encompassing the putative *cI*, *cro* and *cII*–like genes would be the immunity C (*immC*) region of ST64B.

Downstream of SB40, lies ORF SB41 which encodes a gene product with 58% sequence identity to ORF199 of a satellite bacteriophage P4. The latter ORF has been thought to encompass the whole immunity region of P4 (Ghisotti, *et al.*, 1992). The similarity of the SB41 gene product with P4 ORF199 gene product could suggest that it plays a role in the immunity of ST64B. However, a 67 bp insert, homologous to IS*911* was found within this ORF. A similar scenario was observed in *S. flexneri* where a 1,586 bp region containing a 91 bp sequence, nearly identical to the *pap* operon sequence of *E. coli* J96 was followed by a partial sequence of IS*911* (61 bases) (Faubladier and Bouche, 1994). It was suggested that the presence of this fragment of IS*911* would inactivate the upstream regions due to illegitimate excision of the IS*911* with adjacent phage sequences (Faubladier and Bouche, 1994).

5.4.6 Replication genes

The replication genes of λ and P22 are found downstream of the immunity genes. Initiation of replication occurs at the formation of a nucleoprotein complex by (O_λ or $gp18_{P22}$) at a specific site of the DNA (*ori*) and replication of the DNA proceeds bi–directionally. An 11 bp direct repeat (**GAACCAGTCAAaGAACCAGTCAA**) with one base as a spacer, was detected in ST64B at position 391 – 413 of SB42. A similar scenario is observed in the P22 gene *18*, where a 13 bp direct repeat (**ACAAAAGACAATtACAAAAGACAATA**) with one base spacer is found within gene *18*, at position 439 – 469. This repeat is considered as the origin of replication and also marks the 3' end of the *ori*–binding domain of gp18 which is highly conserved in P22, ES18 and λ. Similarly, shorter direct repeats with longer spacers are also found in the *O* genes of both λ and ES18 (**ACAAAAGA**cactatt**ACAAAAGA** and **CAAAAGA**cactatt**CAAAGA**, respectively. It is likely that SB42 encodes a putative *ori* of ST64B, based upon these observations.

Although the SB43 inferred gene product did not reveal any sequence similarity to any helicase proteins in the database, an ATP/GTP–binding site motif with a P loop was found in the SB43 gene product. DNA and RNA helicases comprise members of the ATP/GTP–binding family of proteins with a P–loop (Gorbalenya, *et al.*, 1989). As already mentioned in chapter 4, the gene product of 12_{P22} provides the helicase function, which results in P22 being independent of the helicase DnaB, in the host. The size of SB43 was, however, small (525 bp) when compared to that of 12_{P22} (1,377 bp). The initiation codon of SB44 (894 bp) overlaps the termination codon of SB43 and if the two ORFs were combined, an ORF of a comparable size to P22 gene *12* would be produced. Alternatively, the two ORFs may encode two small interacting proteins, which together provide the function required. Another supporting factor (although not completely convincing) in favour of the interpretation that SB42 and SB43+SB44 are analogous to the replication

genes *18* and *12* in P22 or *O* and *P* in λ, is that all three reading frames overlap each other, and would thus be translationally coupled.

It is possible that the gene product encoded by SB45 is involved in recombination as a resolvase. This prediction is based on its similarity to the *rusA* gene of *E. coli* cryptic phages (CP) as well as *E. coli* phage HK620. The presence of the *rusA* gene in phage 82 as well as in other lambdoid phages provides evidence that the lambdoid phages encode their own recombination function (Mahdi, *et al.*, 1996). On the basis of conservation of genome organisation in the lambdoid phages, it was predicted that the Rap protein of λ may also be a resolvase (Mahdi, *et al.*, 1996). Moreover, Rap was linked with recombination *in vivo* (Stahl, *et al.*, 1995).

The lytic or lysogenic switch in bacteriophage P1 is governed by a tripartite immunity system encompassing *immC*, *immI* and *immT* (Hansen, 1989, Heinrich, *et al.*, 1995, Heisig, *et al.*, 1989). The *immC* and *immI* regions are similar to those of P22. Furthermore, P1 consists of three replicons, R, L and P1dR. The principal replicon L, used for replication during the lytic cycle is under control of the P1 *c1* repressor. The region around the L–replicon contains the *c4* and *ant* genes (*immI* region) and a *kilA* gene which was shown to be non–essential for both replication and lytic development (Hansen, 1989). However, expression of *kilA* is lethal to *E. coli* (Hansen, 1989). Two *kilA*::IS*50* insertion mutants were isolated and characterised as defective in lytic replication. The authors emphasised that insertions in *kilA* prevented the accumulation of the lethal KilA product (Hansen, 1989). SB48 gene product showed 40% sequence identity to KilA protein in P1. This ORF is situated between the replication and lysis genes of ST64B.

The carriage of an 804 bp long region, that showed 99% sequence identity to the 3' end of *S.* Typhimurium *sspH2* gene, which covered the last 100 bp of SB48 and extended into SB49 (covering 704 of the 762 bp SB49 putative gene) could probably account for the absence of a potential RBS upstream of SB49.

5.4.7 Late gene antitermination and the lysis genes

The putative ST64B Q protein (probably encoded by SB50) did not have any orthologues in the database and did not have the CII–binding sites recognised by λ P_{aQ} and P22 P_{a23} probably because the putative CII protein is also distinct from the well known λ and P22 orthologues. A promoter region was located 148 bp upstream of the lysis genes (Table 5.3). All three putative lysis genes of ST64B had strong sequence identity 93 – 100%) with the *S.* Typhimurium Fels–1 proteins which may also be involved in lysis of the bacterial cell.

The small membrane protein, holin, causes a non–specific lesion in the cytoplasmic membrane, allowing the endolysin to gain access to the peptidoglycan (Bläsi and Young, 1996). A dual–start motif is characteristic of the holins consisting of the two Met initiation codons separated by one or two codons, with at least one specifying Arg or Lys (Bläsi and Young, 1996). Two gene products are observed with the shorter product being the lysis–effector (the active holin) and the longer product being an inhibitor of holin (Bläsi, *et al.*, 1990, Bläsi, *et al.*, 1989). The SB51 gene product is 103 aa long and has 100% sequence identity with a similar putative protein in the Fels–1 prophage. However, the Fels–1 prophage sequence was detected during the recent sequencing of the *S.* Typhimurium LT2 genome (accession no. AE008788) and no putative function was assigned for this particular gene. Furthermore, weak similarity was also shown with putative proteins of prophages Gifsy–1 and Gifsy–2. The dual–start motif feature observed in both ST64B and Fels–1 putative proteins at the first and third positions, the map position of SB51 in ST64B and its weak similarity to a region of P22 holin, strongly suggest that both the ST64B and also the Fels–1 gene products are holins.

During the lytic cycle of most bacteriophages, a phage–encoded peptidoglycan–degrading activity is elaborated (Young and Bläsi, 1995). This role is fulfilled by at least four entirely distinct types of enzymes (lysozyme, transglycosylase,

amidase and endopeptidase) which were given a generic name 'endolysin' (Young, 1992). These enzymes depend on the holin to cause a lethal lesion in the cytoplasmic membrane thereby allowing access of endolysin to the peptidoglycan in order to lyse the cell and end the productive cycle (cited in Bläsi and Young 1996). The putative endolysin of ST64B encoded by SB52 shared 93% amino acid identity with a putative chitinase–like protein of Fels–1 and a putative bacteriophage encoded protein in *S.* Typhi. Furthermore, a lower sequence identity (41%) was shown with a glycohydrolase of *P. aeruginosa*. Lytic enzymes of bacteriophages hydrolyse bacterial peptidoglycan consisting of N–acetyl–D–glucosamine and N–acetyl–D–muramic acid whereas chitinases hydrolyse homopolyssacharide of N–acetyl–D–glucosamine (Nakayama, *et al.*, 2000). It has been shown that the R–type pyocin of *P. aeruginosa* is related to phage P2 and that the F–type pyocin is related to phage λ (Nakayama, *et al.*, 2000). Furthermore, the authors identified a lysis gene cassette (a holin and a lytic enzyme designated a glycohydrolase, belonging to the chitinase family) similar to those of bacteriophages. Similar to the putative holin of ST64B, the putative ST64B endolysin is very similar to the Fels–1 putative chitinase–like protein, which is probably an endolysin. The last of the putative ST64B lytic proteins, Rz, encoded by SB53 is related to the Fels–1 unknown protein and to many other actual and putative Rz (λ terminology) and gp15 (P22 terminology) proteins found in the lambdoid phages.

5.4.8 Fragments of genes associated with virulence genes

Interestingly, ST64B carries fragments of genes involved in the virulence of *Salmonella*. Two of the genes, *sopE* and *sspH2* (*Salmonella* secreted protein H2) are associated with the SPI–1 and SPI–2 virulence associated type III secretion systems (Miao, *et al.*, 1999, Mirold, *et al.*, 1999). Virulence associated type III secretion systems permit Gram negative pathogens to translocate effector proteins directly into the host cytoplasm.

(Hueck, 1998). A *sopE*–encoding phage, SopEϕ was isolated (Mirold, *et al.*, 1999). It was discovered that *sopE* gene is flanked by sequences resembling tail and tail fiber genes of P2–like phages (Hardt, *et al.*, 1998b). Indeed, fragments of *sopE* genes identified in ST64B were found within the putative tail and tail fiber genes. The third virulence associated gene (*orgA*) fragment found in ST64B was also found within a tail gene. However, the 3' end of the *sspH2* gene was not found within a tail gene. Although the *sspH2* and the *orgA* genes found in ST64B did not have the core sequences for virulence and the *sopE* gene fragments were small, their presence in this phage supports the notion of horizontal gene transfer mediated by temperate phages. For example, effector proteins that evolved from other bacterial pathogens may become integrated into the genome of a temperate phage and integration of this phage into the genome of a bacterium, may provide these proteins to any susceptible pathogen bearing the type III secretion system (Mirold, *et al.*, 1999). Moreover, both SPI–1 and SPI–2 are approximately 40 kb, a size which can be transduced by generalised transducing phages (Mills, *et al.*, 1995, Shea, *et al.*, 1996) and ST64B integrates in the *tRNA*, a target of many pathogenicity islands for integration.

5.4.9 Conclusion

Several expected putative genes in ST64B could not be identified by sequence similarity to analogous genes in the database. For example, putative genes *N*, *cIII*, *cII* and *Q* (λ terminology) could not be easily identified by sequence similarity, although possible candidates could be inferred based mostly on map position, size of the ORF as well as the direction of transcription. One would think that ST64B may depend on ST64T gene products for its other functions, with ST64T acting as a helper phage. For example, the putative $gp23_{ST64T}$ would supply the function of antitermination in order to express the ST64B late genes. However, an ST64B–like phage isolated from *S.* Typhimurium DTs 9 and 135 (identified by hybridisation and restriction analyses), was excised from the

lysogens which do not carry ST64T. Although the phages from DTs 9 and 135 were not sequenced, an assumption can be made that the sequence of the late genes would be closely related to those of ST64B. Moreover, ST64B *Sma*I restricted genome and Southern hybridisation analysis showed a similar pattern of banding to the *Sma*I restricted genomes of phages from both *S.* Typhimurium DT 9 and 135.

The genome of ST64B is clearly a mosaic, composed of putative genes with similarities to genes from phages of diverse bacterial groups, including genes from bacteria infecting nematodes. Unlike ST64T, ST64B is a λ–like phage and does not share extensive sequence identity with P22. Although the capsid genes showed similarity to those of the lambdoid phages HK97, HK022 and D3, sequence similarity was also observed with capsid genes from diverse phages. The putative tail genes however, showed similarity with those of phage Mu with two putative tail genes showing similarity to the putative tail genes of *S.* Typhimurium phages Fels–2, Fels–1 and Gifsy–2. It is likely that ST64B packages its DNA similarly to phage D3 and phages infecting the Gram positive bacteria which possess the 3' extended termini, however, more work needs to be performed to confirm this. Carriage of putative genes related to bacterial chromosomal genes by ST64B, suggests this phage could have a role in the transfer of virulence determinants, although the ability of ST64B to mediate transduction could not be established due to lack of a suitable indicator strain for propagation.

Finally, it is interesting that ST64B integrates into tRNA *serU.* It has been proposed that the integration of numerous phages close to or into tRNA genes could have a regulatory effect on expression of various genes (Allison and Verma, 2000, Cheetham and Katz, 1995). Moreover, hybridisation studies have shown that ST64B is carried by most *S.* Typhimurium isolates tested. Comparison of ST64B and ST64T genomic sequences with emphasis on their evolution is discussed in chapter 6. Furthermore, since ST64B encodes virulence factor–like gene fragments, it is possible that the genome may represent a mechanism for evolution of the host bacterium as a pathogen. The mechanism and site of

integration in the bacterial genome is reminiscent of the pathogenicity islands of *Salmonella*.

Figure 5.1

Schematic representation of phage ST64B genome showing putative ORFs with *Eco*RI and *Sma*I restrictions sites. The ORFs are labelled SB1 to SB56. ‘ indicates that the ORF extends to the other side of the map, ’ indicates that the ORF is continuing from the other side of the map. The map is circular with the end of the map (40,149) joined to the start of the map. Phage λ nomenclature was used to designate a few ORFs (in brackets and italics) where possible.

- Packaging (Terminase subunits)
- Head
- Tail
- Integration and excision
- Immunity
- Replication
- Lysis
- Unknown or related to bacterial genes

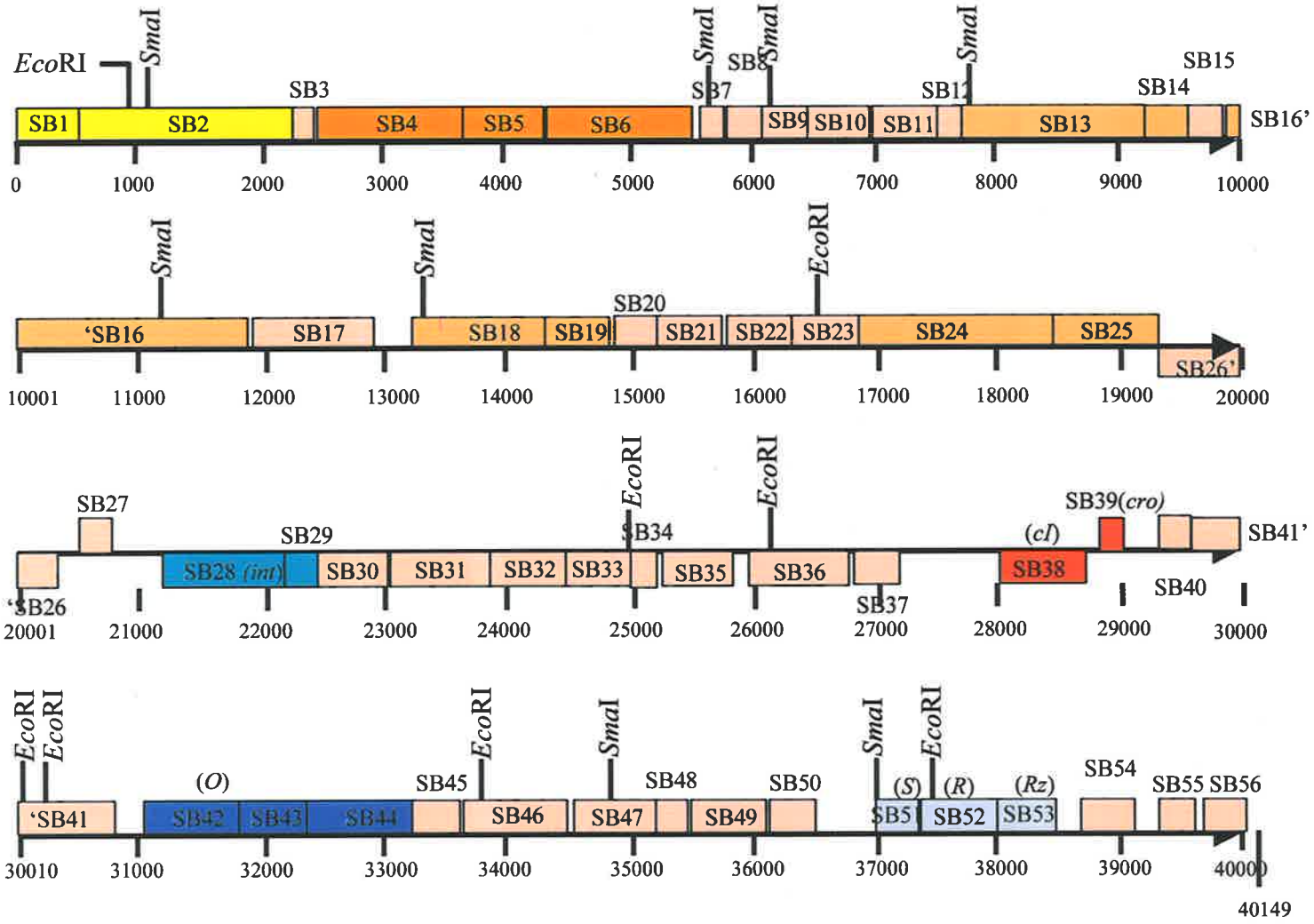
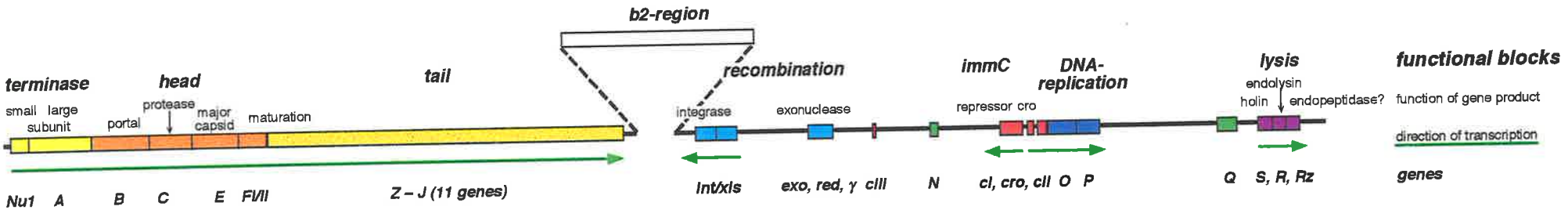

EcoRI
SmaI
SmaI
SmaI
SmaI
SB1
SB2
SB3
SB4
SB5
SB6
SB7
SB8
SB9
SB10
SB11
SB12
SB13
SB14
SB15
SB16'
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
SmaI
SmaI
EcoRI
'SB16
SB17
SB18
SB19
SB20
SB21
SB22
SB23
SB24
SB25
SB26'
10001
11000
12000
13000
14000
15000
16000
17000
18000
19000
20000
SB27
SB29
EcoRI
EcoRI
SB34
SB39(cro)
(cI)
SB41'
'SB26
SB28 (int)
SB30
SB31
SB32
SB33
SB35
SB36
SB37
SB38
SB40
20001
21000
22000
23000
24000
25000
26000
27000
28000
29000
30000
EcoRI
EcoRI
(O)
SB45
EcoRI
SmaI
SB48
SB50
SmaI
EcoRI
(S)
(R)
(Rz)
SB54
SB55 SB56
'SB41
SB42
SB43
SB44
SB46
SB47
SB49
SB51
SB52
SB53
30010
31000
32000
33000
34000
35000
36000
37000
38000
39000
40000
40149

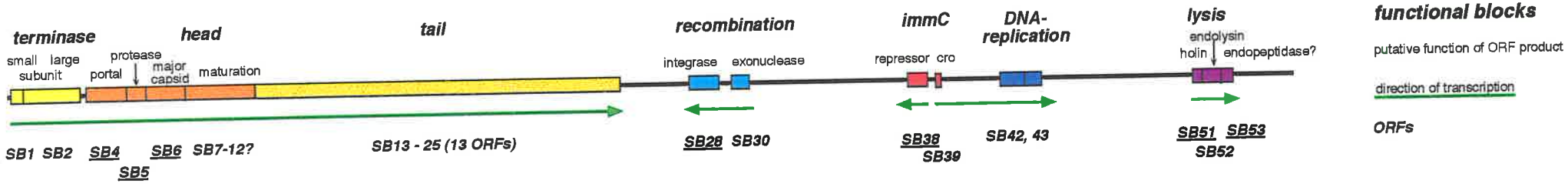
Figure 5.2

Comparison of the genetic maps of phages λ and ST64B. The maps are drawn to scale. The ORFs included in SB64B are those whose inferred gene products functions as the corresponding gene products of λ. The underlined ORFs show sufficient similarity with the corresponding genes of lambdoid phages and they are considered homologues, whereas the other ORFs whose function is deduced from similarities with gene products of distant phages or bacteria are not underlined and are considered as orthologues. The figure indicates high similarity between λ and ST64B ORFs positions, the sizes of most of the corresponding genes and their direction of transcription.

Comparison of the Genetic Maps of the Phages Lambda and ST64B



Phage Lambda



Phage ST64B

Figure 5.3

Comparison of the putative terminase large subunit protein (SB2) of ST64B with the putative terminase large subunit protein of *P. aeruginosa* phage D3 and the hypothetical YMFN protein (accession no. P975978) in *E. coli* using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

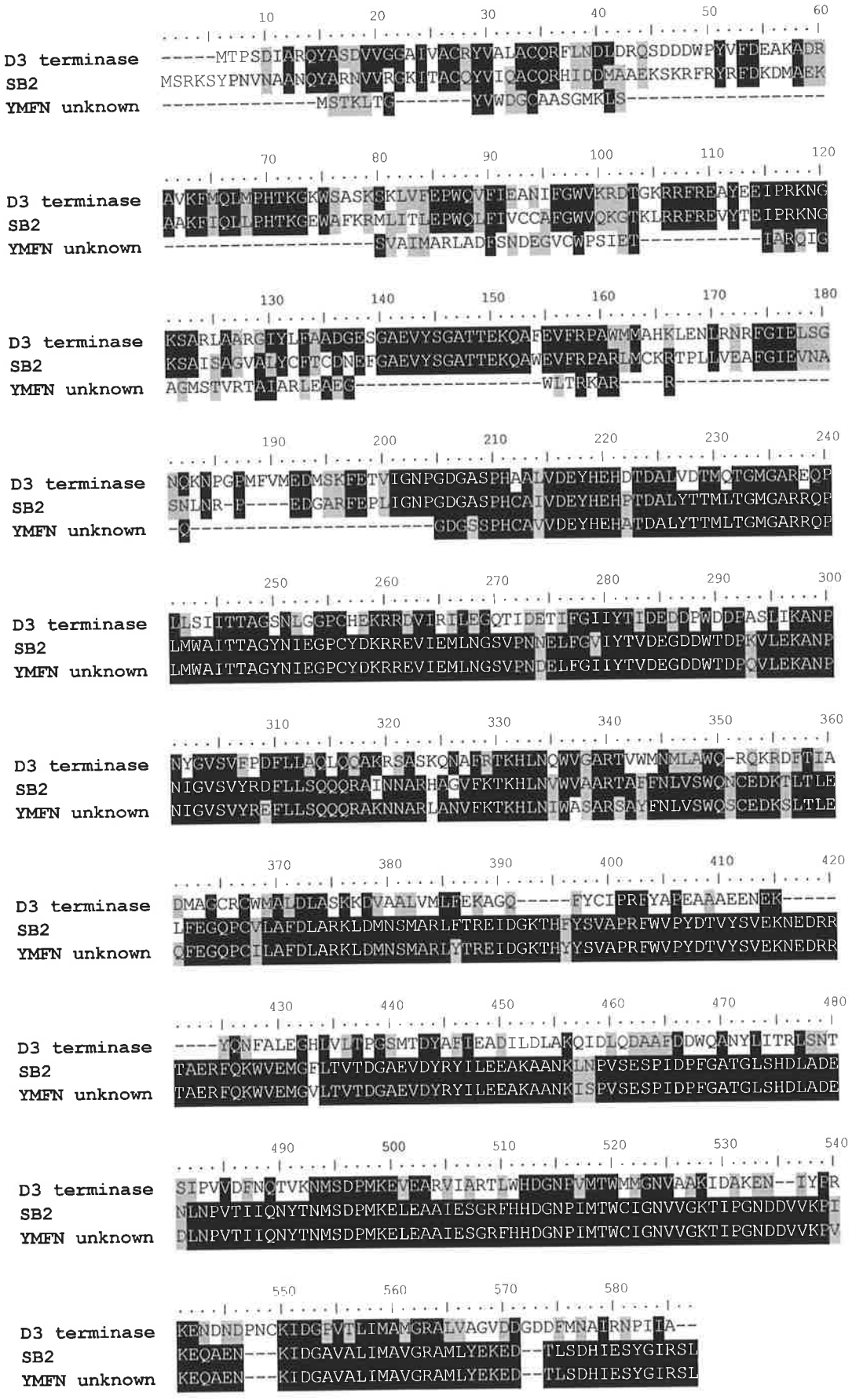

D3 terminase
SB2
YMFN unknown

Figure 5.4

Comparison of the amino acid sequence of ST64B (SB4 gene product) and *E. coli*_YMFO protein using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align–guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons while similar residues are indicated by periods.

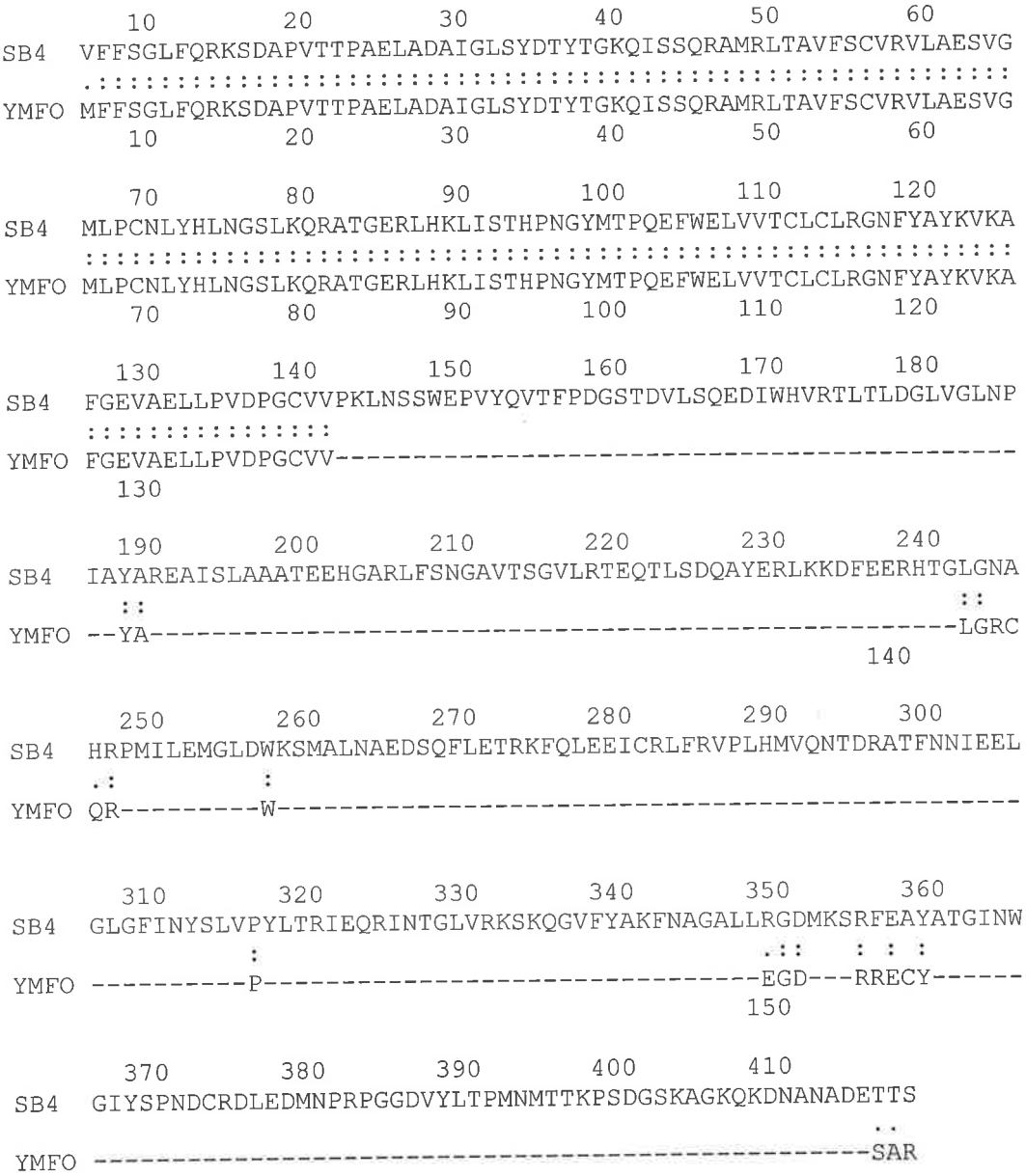

10 20 30 40 50 60
SB4 VFFSGLFQRKSDAPVTTPAELADAIGLSYDTYTGKQISSQRAMRLTAVFSCVRVLAESVG
YMFO MFFSGLFQRKSDAPVTTPAELADAIGLSYDTYTGKQISSQRAMRLTAVFSCVRVLAESVG
10 20 30 40 50 60
70 80 90 100 110 120
SB4 MLPCNLYHLNGSLKQRATGERLHKLISTHPNGYMTPQEFWELVVTCLCLRGNFYAYKVKA
YMFO MLPCNLYHLNGSLKQRATGERLHKLISTHPNGYMTPQEFWELVVTCLCLRGNFYAYKVKA
70 80 90 100 110 120
130 140 150 160 170 180
SB4 FGEVAELLPVDPGCVVPKLNSSWEPVYQVTFPDGSTDVLSQEDIWHVRTLTLDGLVGLNP
YMFO FGEVAELLPVDPGCVV----------
130
190 200 210 220 230 240
SB4 IAYAREAISLAAATEEHGARLFSNGAVTSGVLRTEQTLSDQAYERLKKDFEERHTGLGNA
YMFO --YA----------LGRC
140
250 260 270 280 290 300
SB4 HRPMILEMGLDWKSMALNAEDSQFLETRKFQLEEICRLFRVPLHMVQNTDRATFNNIEEL
YMFO QR---------W----------
310 320 330 340 350 360
SB4 GLGFINYSLVPYLTRIEQRINTGLVRKSKQGVFYAKFNAGALLRGDMKSRFEAYATGINW
YMFO ----------P----------EGD---RRECY------
150
370 380 390 400 410
SB4 GIYSPNDCRDLEDMNPRPGGDVYLTPMNMTTKPSDGSKAGKQKDNANADETTS
YMFO ----------SAR

Figure 5.5

Comparison of the putative major capsid protein of ST64B (SB6 gene product) with the phage major capsid proteins from phages in *Mesorhizobium loti* (accession no. NP_108602), *Agrobacterium tumefaciens* (accession no. AAK86763), *Caulobacter crescentus* (accession no. AAK86763) and *Streptomyces* phage ϕC31 using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

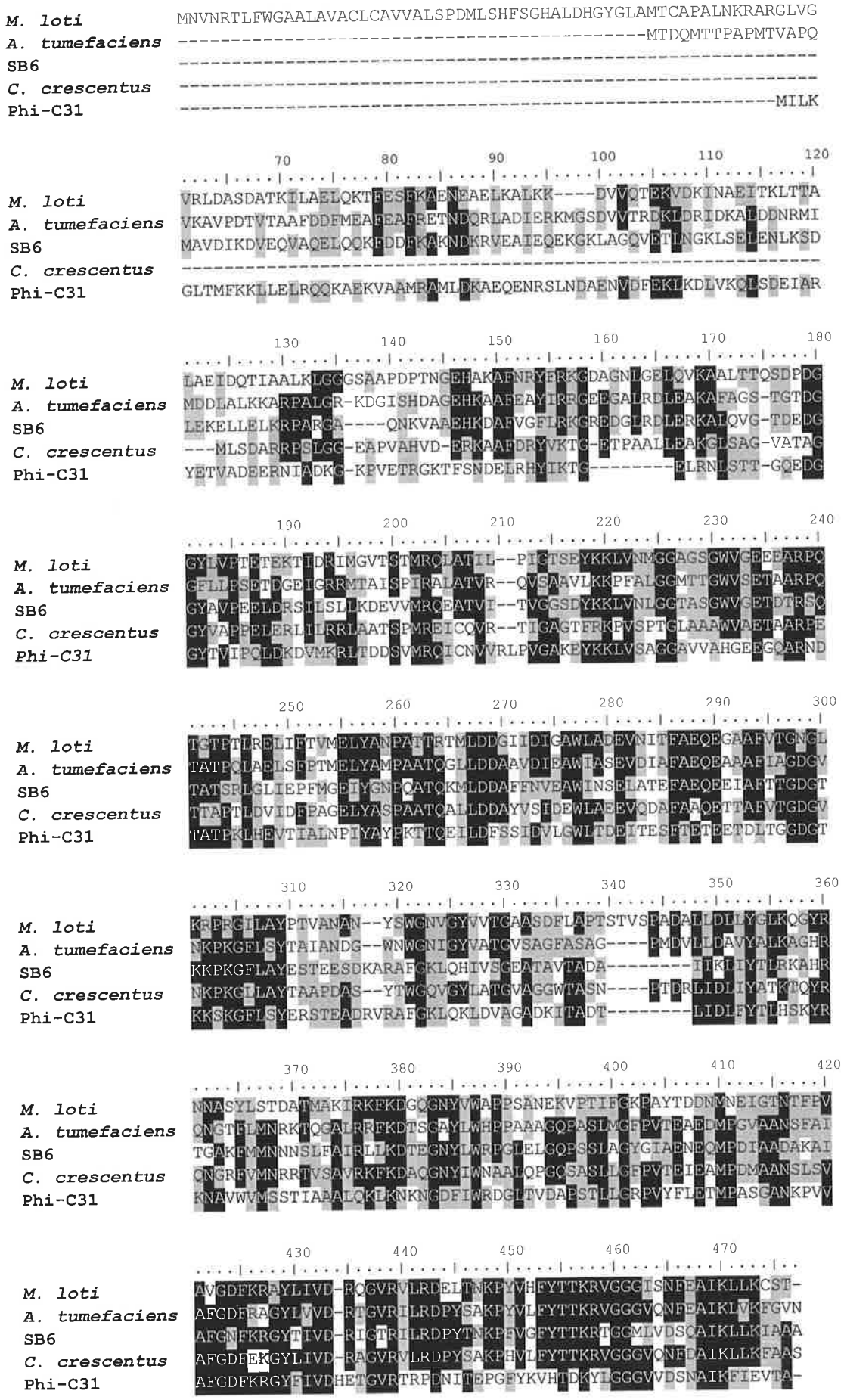

M. loti
A. tumefaciens
SB6
C. crescentus
Phi-C31
MNVNRTLFWGAALAVACLCAVVALSPDMLSHFSGHALDHGYGLAMTCAPALNKRARGLVG
MTDQMTTPAPMTVAPQ
MILK
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470

Figure 5.6

ST64B phage particle proteins separated on an SDS–PAGE gel with a separating gel of 12% acrylamide and 4% acrylamide stacker gel (Laemmli, 1970), to approximate the molecular size of the major head protein. A band at approximately 28 kDa was identified rather than the 44 kDa predicted by the sequence of the inferred protein encoded by SB6.

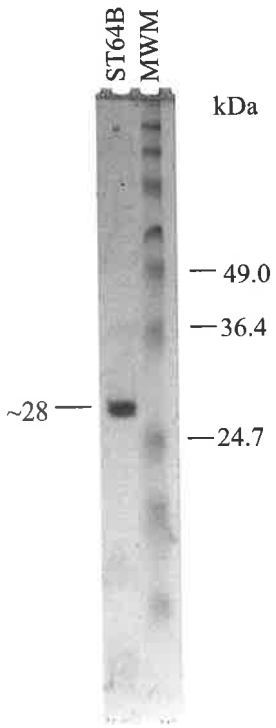

ST64B
MWM
kDa
49.0
36.4
~28
24.7

Figure 5.7

Comparison of the amino acid sequence of ST64B (SB22 gene product) and a hypothetical YMFP protein of the defective *E. coli* phage E14 using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align–guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons while similar residues are indicated by periods.

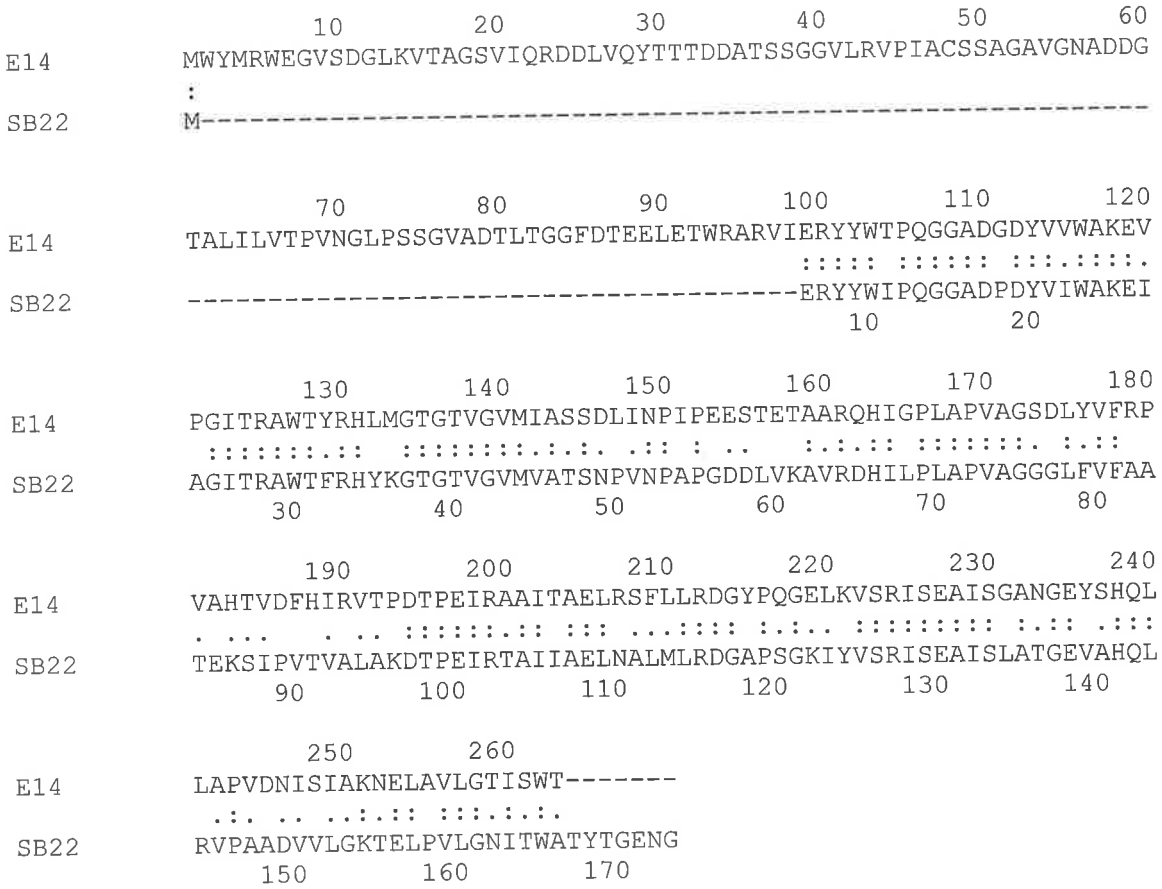

10 20 30 40 50 60
E14 MWYMRWEGVSDGLKVTAGSVIQRDDLVQYTTTDDATSSGGVLRVPIACSSAGAVGNADDG
SB22 M
70 80 90 100 110 120
E14 TALILVTPVNGLPSSGVADTLTGGFDTEELETWRARVIERYYWTPQGGADGDYVVWAKEV
SB22 ERYYWIPQGGADPDYVIWAKEI
10 20
130 140 150 160 170 180
E14 PGITRAWTYRHLMGTGTVGVMIASSDLINPIPEESTETAARQHIGPLAPVAGSDLYVFRP
SB22 AGITRAWTFRHYKGTGTVGVMVATSNPVNPAPGDDLVKAVRDHILPLAPVAGGGLFVFAA
30 40 50 60 70 80
190 200 210 220 230 240
E14 VAHTVDFHIRVTPDTPEIRAAITAELRSFLLRDGYPQGELKVSRISEAISGANGEYSHQL
SB22 TEKSIPVTVALAKDTPEIRTAIIAELNALMLRDGAPSGKIYVSRISEAISLATGEVAHQL
90 100 110 120 130 140
250 260
E14 LAPVDNISIAKNELAVLGTISWT-------
SB22 RVPAADVVLGKTELPVLGNITWATYTGENG
150 160 170

Figure 5.8

Comparison of the putative invertase pin protein of ST64B encoded by SB27 with the DNA invertase pin proteins of the *E. coli* cryptic prophage E14 and CP–933H using BioEdit version 4.8.10 (Hall, 1999). The SB27 gene product exhibits 81 to 83% amino acid identity over the terminal 73 aa of CP–933H and E14, respectively. Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

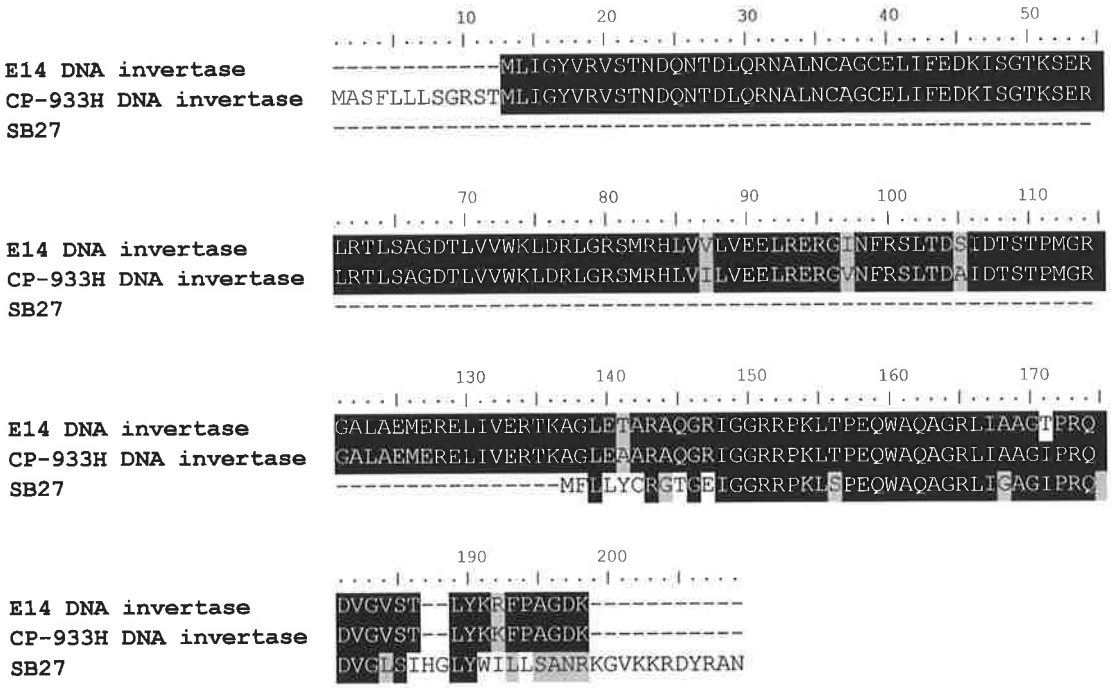
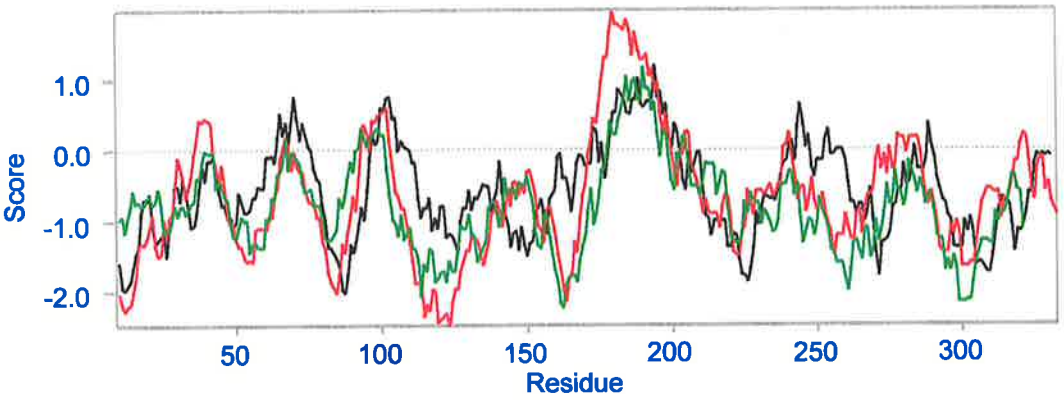

10
20
30
40
50
E14 DNA invertase
CP-933H DNA invertase
SB27
MLIGYVRVSTNDQNTDLQRNALNCAGCELIFEDKISGTKSER
MASFLLLSGRSTMLIGYVRVSTNDQNTDLQRNALNCAGCELIFEDKISGTKSER
70
80
90
100
110
E14 DNA invertase
CP-933H DNA invertase
SB27
LRTLSAGDTLVVWKLDRLGRSMRHLVVLVEELRERGINFRSLTDSIDTSTPMGR
LRTLSAGDTLVVWKLDRLGRSMRHLVILVEELRERGVNFRSLTDAIDTSTPMGR
130
140
150
160
170
E14 DNA invertase
CP-933H DNA invertase
SB27
GALAEMERELIVERTKAGLETARAQGRIGGRRPKLTPEQWAQAGRLIAAGTPRQ
GALAEMERELIVERTKAGLEAARAQGRIGGRRPKLTPEQWAQAGRLIAAGIPRQ
MFLLYCRGTGEIGGRRPKLSPEQWAQAGRLIGAGIPRQ
190
200
E14 DNA invertase
CP-933H DNA invertase
SB27
DVGVST--LYKRFPAGDK
DVGVST--LYKKFPAGDK
DVGLSIHGLYWILLSANRKGVKKRDYRAN

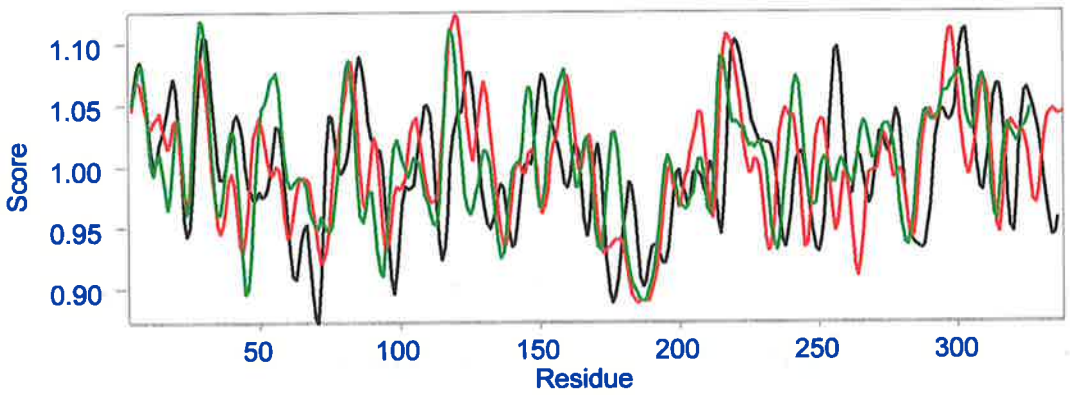
Figure 5.9

Hydropathy predictions of the putative Int protein compared to the putative Int proteins of CP–933M and CP–933U bacteriophages. The plots were produced by OMIGA 1.1. References are as follows: Kyte–Doolittle hydropathy (Kyte and Doolittle, 1982), Karplus protein flexibility (Karplus and Schultz, 1985), Parker antigenicity (Parker, *et al.*, 1986) and von Heijne hydrophilicity (von Heijne, 1981).

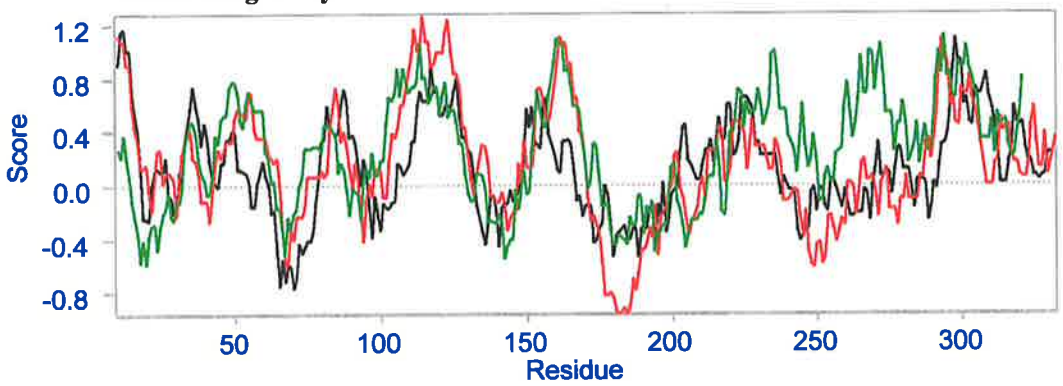
Kyte-Doolittle hydropathy



Karplus protein flexibility



Parker antigenicity



Von Heijne hydrophilicity

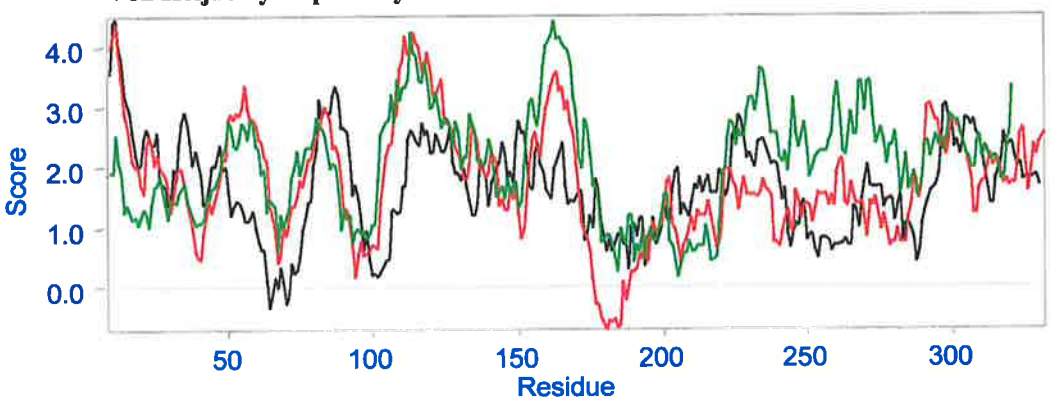


Figure 5.10

Comparison of putative excisionase (Xis) protein of ST64B encoded by SB29 with the putative and actual Xis proteins of phages SfV, HK022, L, ST64T and P22 using BioEdit version 4.8.10 (Hall, 1999). Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

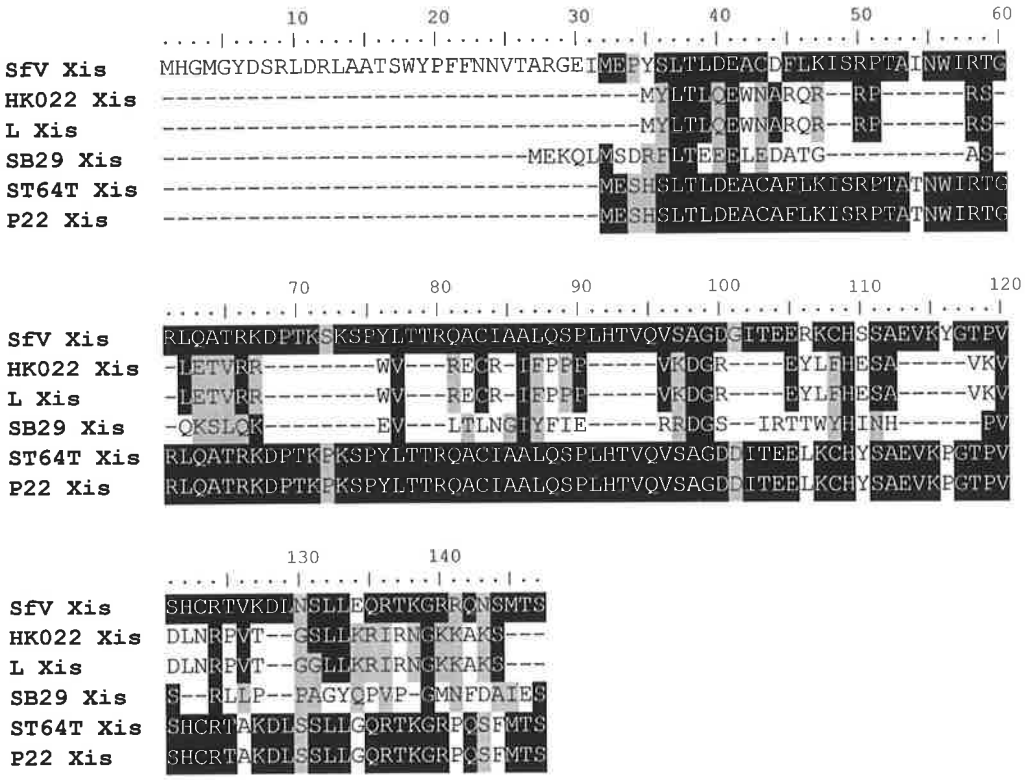

10
20
30
40
50
60
SfV Xis
HK022 Xis
L Xis
SB29 Xis
ST64T Xis
P22 Xis
70
80
90
100
110
120
SfV Xis
HK022 Xis
L Xis
SB29 Xis
ST64T Xis
P22 Xis
130
140
SfV Xis
HK022 Xis
L Xis
SB29 Xis
ST64T Xis
P22 Xis

Figure 5.11

Comparison of the putative exonuclease protein of ST64B encoded by SB30 with the exonuclease proteins of the cryptic phages (CP) of *E. coli*, CP–933P, CP–933U, CP–933M and CP–933O using BioEdit version 4.8.10 (Hall, 1999). The SB30 gene product aligns from residue 640 aa to the C–termini of the other phage exonucleases. Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

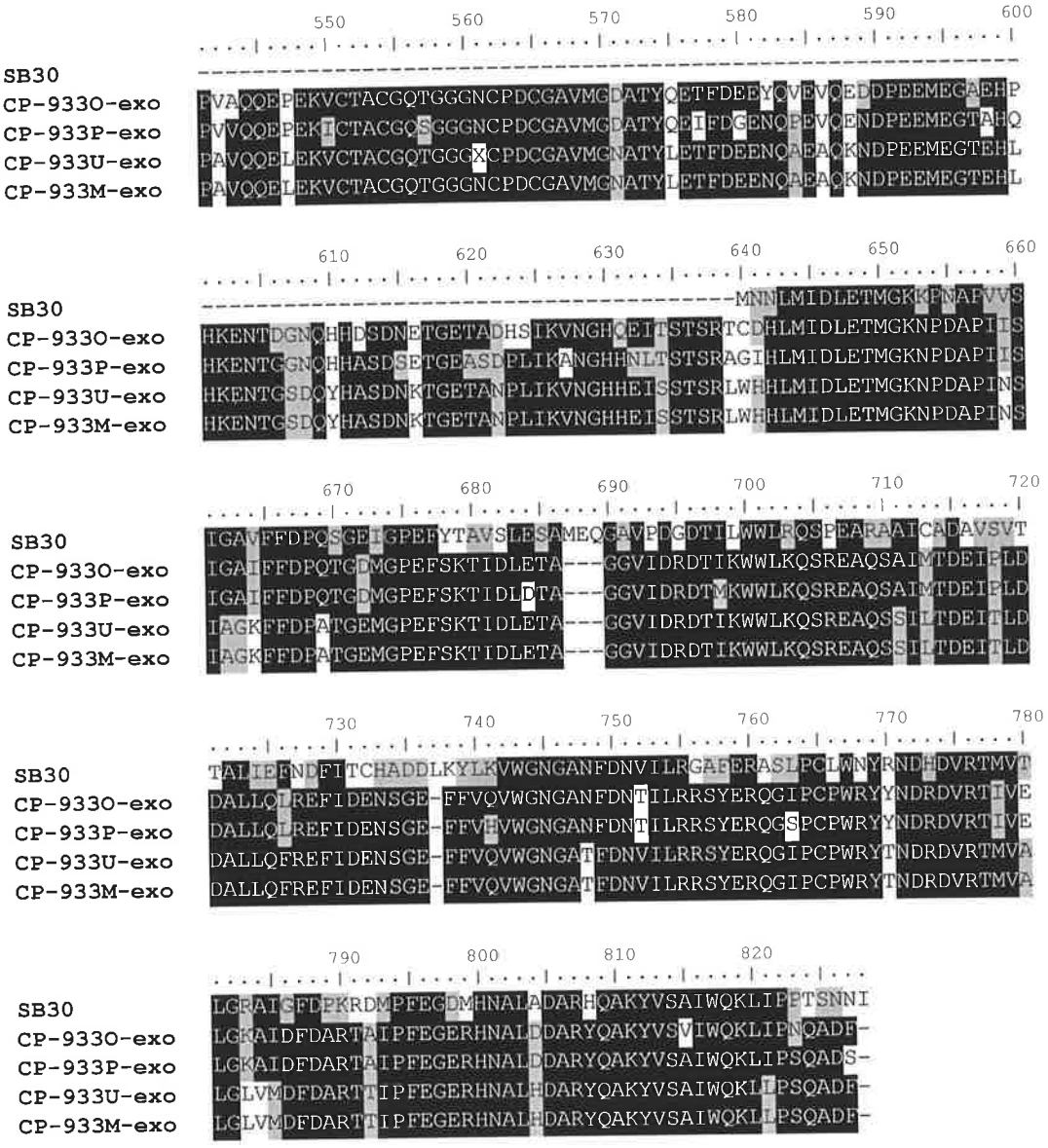

550
560
570
580
590
600
SB30
CP-933O-exo
CP-933P-exo
CP-933U-exo
CP-933M-exo
610
620
630
640
650
660
SB30
CP-933O-exo
CP-933P-exo
CP-933U-exo
CP-933M-exo
670
680
690
700
710
720
SB30
CP-933O-exo
CP-933P-exo
CP-933U-exo
CP-933M-exo
730
740
750
760
770
780
SB30
CP-933O-exo
CP-933P-exo
CP-933U-exo
CP-933M-exo
790
800
810
820
SB30
CP-933O-exo
CP-933P-exo
CP-933U-exo
CP-933M-exo

Figure 5.12

Comparison of the putative Eaa protein of ST64B encoded by SB31 with P22 Eaa protein using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align–guess.cgi) (Person, *et al.*, 1997). Identical residues are indicated by colons while similar residues are indicated by periods.

```
                10         20         30         40         50
SB31    MTTITKEWLQQTI-AEFENTRDDIPFGLDDDDAKILLVLKRALASLDAEPVRYLNKFSGT
        ::::::: ..  . . .::    .  : . . :.:      :::::::: ::::::::::
P22     MTTITKERIELFVKSPLEN---GLTRGEQMELARI------ALASLDAETVRYLNKFSGT
                10         20         30         40         50

      60         70         80         90
SB31    CVTLEQQSNAADDVAVYIPLYTAPPAP----VVPEEA------YSDDCP----------D
        ::::::: ::::::::::::::.:::.:    .  :.:      ..:  :
P22     CVTLEQQPNAADDVAVYIPLYAAPPVPERERIRREHAEWSDKTFGDVGPVGPLKHLSKEA
               60         70         80         90        100        110

    100          110                                                  120
SB31    LYA----SQPEAWAAG----WNA----------CRAAL--------------------L
        : :    :.:  ::      :.:           :: .                      :
P22     LEAAADPSDPLEWADMQFLLWDAQRRMGISDEFITRAMIEKLEINKTRQWPEPKDGEPRL
              120        130        140        150        160        170

                    130        140        150        160        170
SB31    HGAE------PVSQTYELPQTQFEQVADLYEMQFDDGRTCAFHTDGAKAAQWLLACDGNK
        :  :      :     ::: .: . ::::: . ...:.. ..  :  ::. :.   .:.
P22     HIKEQPESVVPEECPAELPYAQVKAVADLYALCWQSGEVVTYTPDPEKATIWINNYSGTC
              180        190        200        210        220        230

          180        190        200        210         220        230
SB31    VQEYVRLERYHEALIGNSPVIPDDWVMVPKKLTAENGAK--SLLSGEFL-ETTFISFPEC
        ::::.::: .::: :::::::  :.   ...  .. .:  ....:. : .  :..
P22     VQEYVKLERLQEALAGNSPVIPGGWISCSERMPDNDESKPIAIFTGKCLGQGMFVA----
              240        250        260        270        280

              240        250        260        270
SB31    LADEECESCDGSGRIKIEVPVRWTTIKAIWNKGVEHFRSSTATGDN
          . : .: .       :  .. :   :: :. .
P22     -------TYDDDGFFDY-----WEGMEII---GVSHWMQ-LPDPPL
               290        300        310
```

Figure 5.13

Comparison of the SB32 gene product with Ead_{P22} and $gp45_{N15}$ protein using BioEdit version 4.8.10 (Hall, 1999). The carboxy termini of these proteins showed considerable similarity. Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

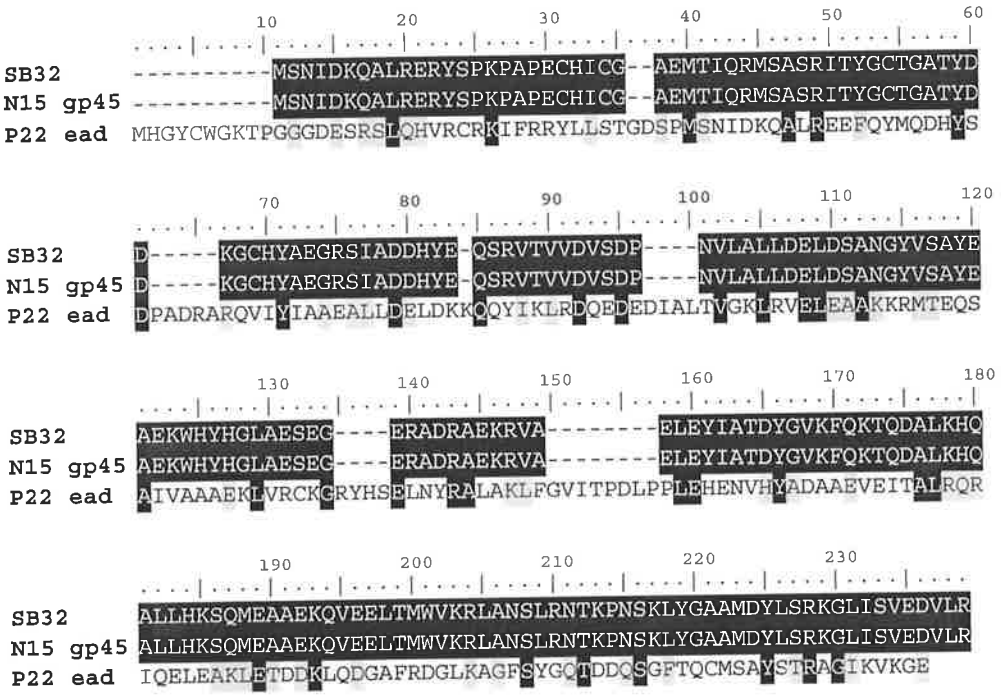

SB32
N15 gp45
P22 ead

Figure 5.14

Comparison of ST64B SB43 gene product with a hypothetical protein in *E. coli* using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align–guess.cgi) (Person, *et al.*, 1997) was used. A higher similarity is shown from the second methionine of SB43 gene product. Identical residues are indicated by colons while similar residues are indicated by periods.

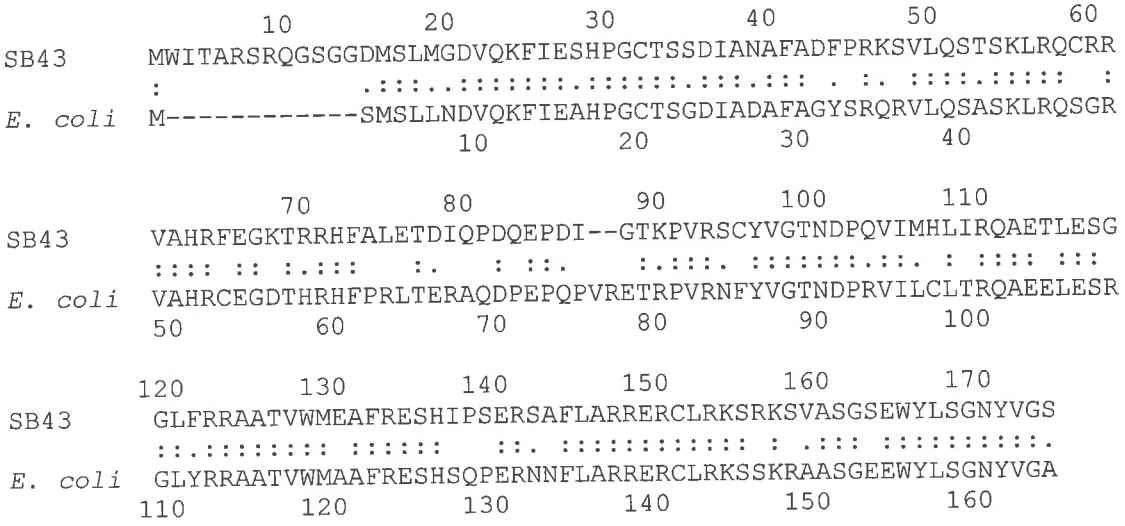
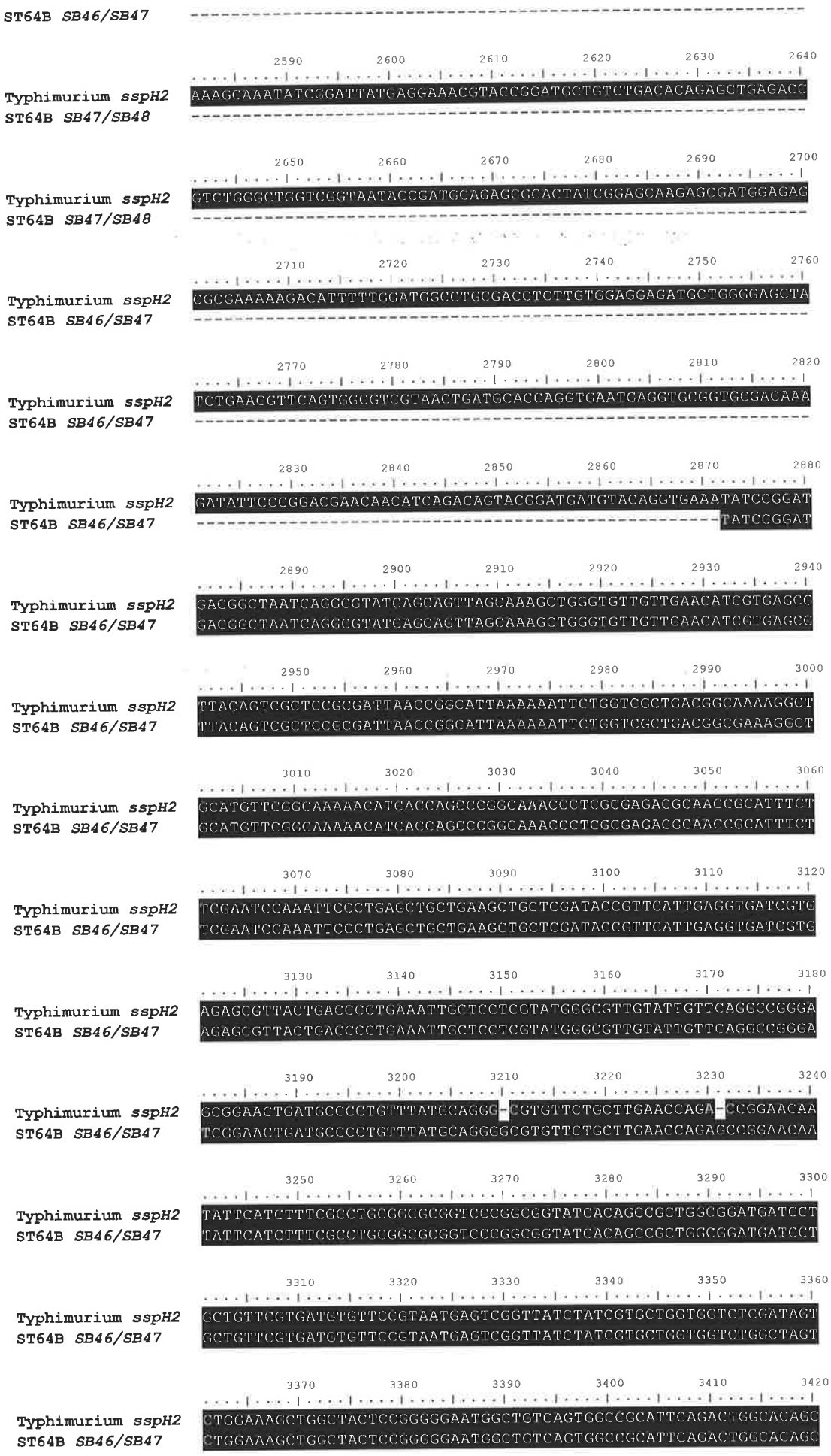

10 20 30 40 50 60
SB43 MWITARSRQGSGGDMSLMGDVQKFIESHPGCTSSDIANAFADFPRKSVLQSTSKLRQCRR
E. coli M----------SMSLLNDVQKFIEAHPGCTSGDIADAFAGYSRQRVLQSASKLRQSGR
10 20 30 40
70 80 90 100 110
SB43 VAHRFEGKTRRHFALETDIQPDQEPDI--GTKPVRSCYVGTNDPQVIMHLIRQAETLESG
E. coli VAHRCEGDTHRHFPRLTERAQDPEPQPVRETRPVRNFYVGTNDPRVILCLTRQAEELESR
50 60 70 80 90 100
120 130 140 150 160 170
SB43 GLFRRAATVWMEAFRESHIPSERSAFLARRERCLRKSRKSVASGSEWYLSGNYVGS
E. coli GLYRRAATVWMAAFRESHSQPERNNFLARRERCLRKSSKRAASGEEWYLSGNYVGA
110 120 130 140 150 160

Figure 5.15

Alignment of SB47 region found in ST64B with the *S.* Typhimurium *sspH2* gene. There is 99% sequence identity between the two sequences at the 3' end of the 3,815 bp long *sspH2* gene. The short region found in ST64B shows sequence identity from 2,871 to 3,815 of the *sspH2* gene. The alignment was generated using BioEdit version 4.8.10 (Hall, 1999).


ST64B SB46/SB47
2590 2600 2610 2620 2630 2640
Typhimurium sspH2
AAAGCAAATATCGGATTATGAGGAAACGTACCGGATGCTGTCTGACACAGAGCTGAGACC
ST64B SB47/SB48
2650 2660 2670 2680 2690 2700
Typhimurium sspH2
GTCTGGGCTGGTCGGTAATACCGATGCAGAGCGCACTATCGGAGCAAGAGCGATGGAGAG
ST64B SB47/SB48
2710 2720 2730 2740 2750 2760
Typhimurium sspH2
CGCGAAAAAGACATTTTTGGATGGCCTGCGACCTCTTGTGGAGGAGATGCTGGGGAGCTA
ST64B SB46/SB47
2770 2780 2790 2800 2810 2820
Typhimurium sspH2
TCTGAACGTTCAGTGGCGTCGTAACTGATGCACCAGGTGAATGAGGTGCGGTGCGACAAA
ST64B SB46/SB47
2830 2840 2850 2860 2870 2880
Typhimurium sspH2
GATATTCCCGGACGAACAACATCAGACAGTACGGATGATGTACAGGTGAAATATCCGGAT
ST64B SB46/SB47
TATCCGGAT
2890 2900 2910 2920 2930 2940
Typhimurium sspH2
GACGGCTAATCAGGCGTATCAGCAGTTAGCAAAGCTGGGTGTTGTTGAACATCGTGAGCG
ST64B SB46/SB47
GACGGCTAATCAGGCGTATCAGCAGTTAGCAAAGCTGGGTGTTGTTGAACATCGTGAGCG
2950 2960 2970 2980 2990 3000
Typhimurium sspH2
TTACAGTCGCTCCGCGATTAACCGGCATTAAAAAATTCTGGTCGCTGACGGCAAAAGGCT
ST64B SB46/SB47
TTACAGTCGCTCCGCGATTAACCGGCATTAAAAAATTCTGGTCGCTGACGGCGAAAGGCT
3010 3020 3030 3040 3050 3060
Typhimurium sspH2
GCATGTTCGGCAAAAACATCACCAGCCCGGCAAACCCTCGCGAGACGCAACCGCATTTCT
ST64B SB46/SB47
GCATGTTCGGCAAAAACATCACCAGCCCGGCAAACCCTCGCGAGACGCAACCGCATTTCT
3070 3080 3090 3100 3110 3120
Typhimurium sspH2
TCGAATCCAAATTCCCTGAGCTGCTGAAGCTGCTCGATACCGTTCATTGAGGTGATCGTC
ST64B SB46/SB47
TCGAATCCAAATTCCCTGAGCTGCTGAAGCTGCTCGATACCGTTCATTGAGGTGATCGTG
3130 3140 3150 3160 3170 3180
Typhimurium sspH2
AGAGCGTTACTGACCCCTGAAATTGCTCCTCGTATGGGCGTTGTATTGTTCAGGCCGGGA
ST64B SB46/SB47
AGAGCGTTACTGACCCCTGAAATTGCTCCTCGTATGGGCGTTGTATTGTTCAGGCCGGGA
3190 3200 3210 3220 3230 3240
Typhimurium sspH2
GCGGAACTGATGCCCCTGTTTATGCAGGG-CGTGTTCTGCTTGAACCAGA-CCGGAACAA
ST64B SB46/SB47
TCGGAACTGATGCCCCTGTTTATGCAGGGGCGTGTTCTGCTTGAACCAGAGCCGGAACAA
3250 3260 3270 3280 3290 3300
Typhimurium sspH2
TATTCATCTTTCGCCTGCGGCGCGGTCCCGGCGGTATCACAGCCGCTGGCGGATGATCCT
ST64B SB46/SB47
TATTCATCTTTCGCCTGCGGCGCGGTCCCGGCGGTATCACAGCCGCTGGCGGATGATCCT
3310 3320 3330 3340 3350 3360
Typhimurium sspH2
GCTGTTCGTGATGTGTTCCGTAATGAGTCGGTTATCTATCGTGCTGGTGGTCTCGATAGT
ST64B SB46/SB47
GCTGTTCGTGATGTGTTCCGTAATGAGTCGGTTATCTATCGTGCTGGTGGTCTGGCTAGT
3370 3380 3390 3400 3410 3420
Typhimurium sspH2
CTGGAAAGCTGGCTACTCCGGGGGAATGGCTGTCAGTGGCCGCATTCAGACTGGCACAGC
ST64B SB46/SB47
CTGGAAAGCTGGCTACTCCGGGGGAATGGCTGTCAGTGGCCGCATTCAGACTGGCACAGC

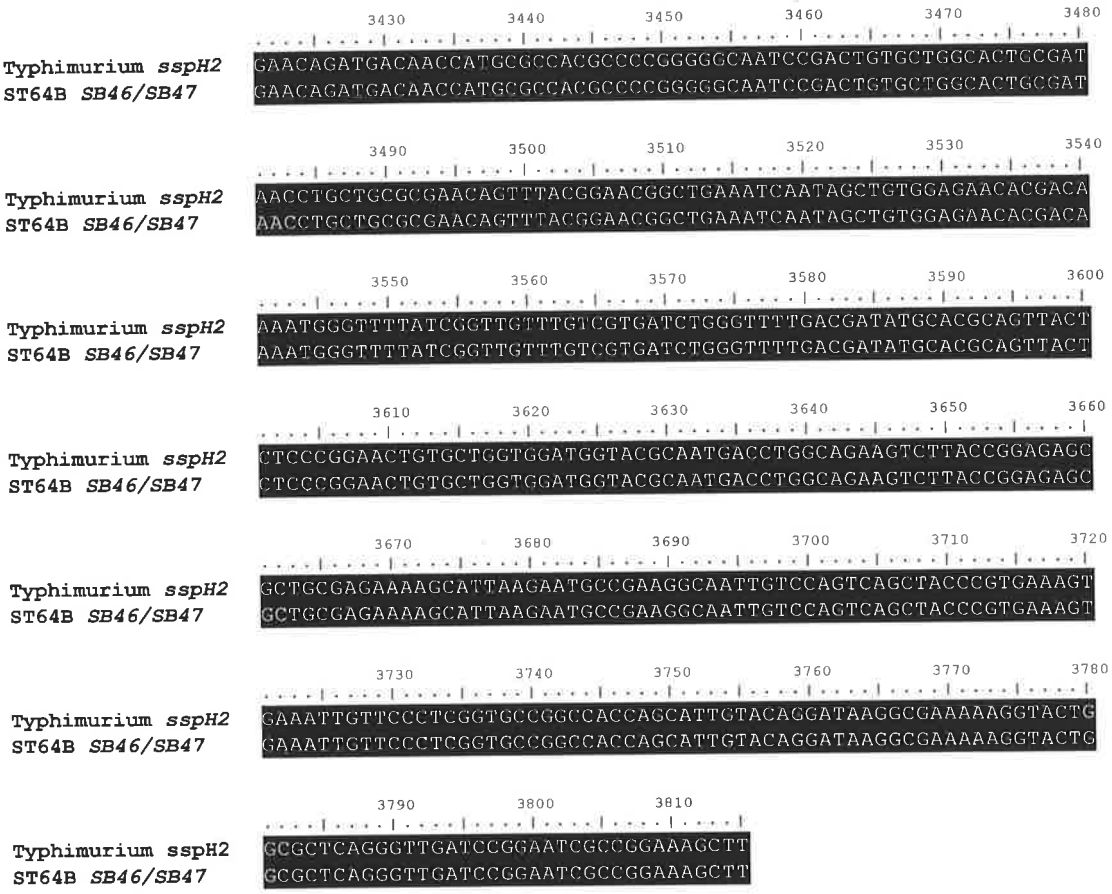

3430 3440 3450 3460 3470 3480
Typhimurium sspH2
ST64B SB46/SB47
GAACAGATGACAACCATGCGCCACGCCCCGGGGGCAATCCGACTGTGCTGGCACTGCGAT
GAACAGATGACAACCATGCGCCACGCCCCGGGGGCAATCCGACTGTGCTGGCACTGCGAT
3490 3500 3510 3520 3530 3540
Typhimurium sspH2
ST64B SB46/SB47
AACCTGCTGCGCGAACAGTTTACGGAACGGCTGAAATCAATAGCTGTGGAGAACACGACA
AACCTGCTGCGCGAACAGTTTACGGAACGGCTGAAATCAATAGCTGTGGAGAACACGACA
3550 3560 3570 3580 3590 3600
Typhimurium sspH2
ST64B SB46/SB47
AAATGGGTTTTATCGGTTGTTTGTCGTGATCTGGGTTTTGACGATATGCACGCAGTTACT
AAATGGGTTTTATCGGTTGTTTGTCGTGATCTGGGTTTTGACGATATGCACGCAGTTACT
3610 3620 3630 3640 3650 3660
Typhimurium sspH2
ST64B SB46/SB47
CTCCCGGAACTGTGCTGGTGGATGGTACGCAATGACCTGGCAGAAGTCTTACCGGAGAGC
CTCCCGGAACTGTGCTGGTGGATGGTACGCAATGACCTGGCAGAAGTCTTACCGGAGAGC
3670 3680 3690 3700 3710 3720
Typhimurium sspH2
ST64B SB46/SB47
GCTGCGAGAAAAGCATTAAGAATGCCGAAGGCAATTGTCCAGTCAGCTACCCGTGAAAGT
GCTGCGAGAAAAGCATTAAGAATGCCGAAGGCAATTGTCCAGTCAGCTACCCGTGAAAGT
3730 3740 3750 3760 3770 3780
Typhimurium sspH2
ST64B SB46/SB47
GAAATTGTTCCCTCGGTGCCGGCCACCAGCATTGTACAGGATAAGGCGAAAAAGGTACTG
GAAATTGTTCCCTCGGTGCCGGCCACCAGCATTGTACAGGATAAGGCGAAAAAGGTACTG
3790 3800 3810
Typhimurium sspH2
ST64B SB46/SB47
GCGCTCAGGGTTGATCCGGAATCGCCGGAAAGCTT
GCGCTCAGGGTTGATCCGGAATCGCCGGAAAGCTT

Figure 5.16

Comparison of SB51 inferred gene product with a putative protein of Fels–1 using Align at the Genestream network server IGH (http://www2.igh.cnrs.fr/bin/align–guess.cgi) (Person, *et al.*, 1997) was used. Identity of 100% over the first 84 amino acids is evident. Identical residues are indicated by colons while similar residues are indicated by periods.

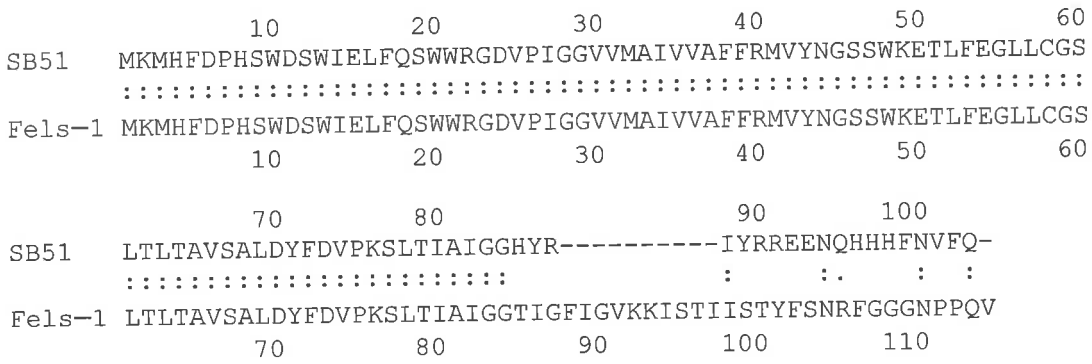

10 20 30 40 50 60
SB51 MKMHFDPHSWDSWIELFQSWWRGDVPIGGVVMAIVVAFFRMVYNGSSWKETLFEGLLCGS
Fels-1 MKMHFDPHSWDSWIELFQSWWRGDVPIGGVVMAIVVAFFRMVYNGSSWKETLFEGLLCGS
10 20 30 40 50 60
70 80 90 100
SB51 LTLTAVSALDYFDVPKSLTIAIGGHYR----------IYRREENQHHHFNVFQ-
Fels-1 LTLTAVSALDYFDVPKSLTIAIGGTIGFIGVKKISTIISTYFSNRFGGGNPPQV
70 80 90 100 110

Figure 5.17

Comparison of the putative Rz_{ST64B} encoded by SB53 with Rz_{λ}, Rz_{H-19B}, $gp15_{ES18}$, $gp15_{P22}$ and Fels–1 proteins using BioEdit version 4.8.10 (Hall, 1999). The carboxy termini of these proteins showed considerable of similarity. Identical amino acid residues and similar residues are indicated by black and grey areas, respectively.

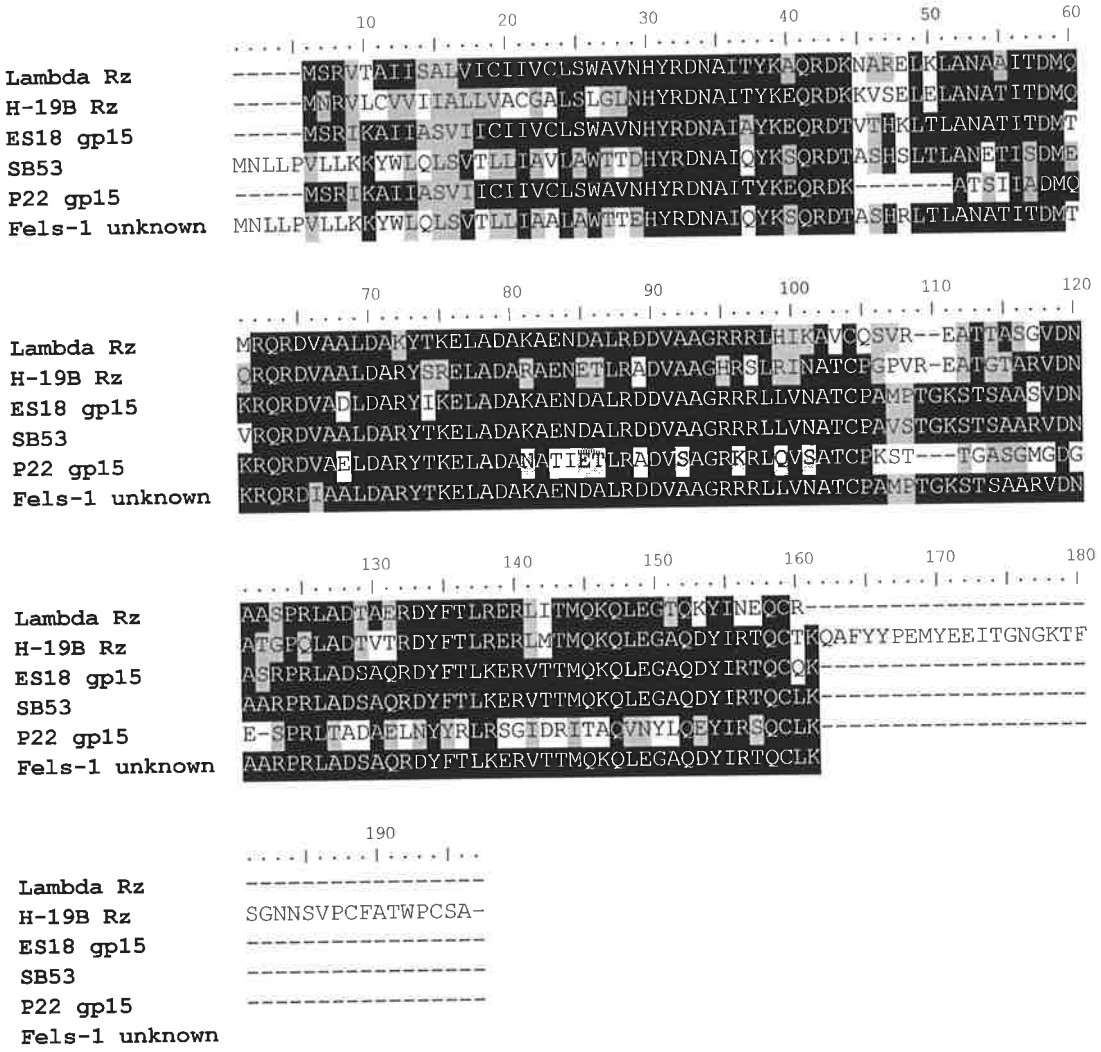

10
20
30
40
50
60
Lambda Rz
H-19B Rz
ES18 gp15
SB53
P22 gp15
Fels-1 unknown
70
80
90
100
110
120
Lambda Rz
H-19B Rz
ES18 gp15
SB53
P22 gp15
Fels-1 unknown
130
140
150
160
170
180
Lambda Rz
H-19B Rz
ES18 gp15
SB53
P22 gp15
Fels-1 unknown
190
Lambda Rz
H-19B Rz
ES18 gp15
SB53
P22 gp15
Fels-1 unknown

Figure 5.18

Southern hybridisation analysis of digested *S.* Typhimurium DT 64 and phage ST64B genomic DNA using a variety of different restriction enzymes. The amplified fragment using primers F2 (forward) and R1 (reverse primer), was labelled and used as a probe in Southern analysis of the restricted genomes of both phage ST64B and the ST64B lysogen. Lane 1 MWM: Hi–Lo mixed DNA markers (in kb), which are a combination of equal amounts of three DNA digests; λ/*Hin*dIII, λ/*Pvu*I and SPP–I/*Eco*RI resulting in a range of commercially available DNA fragments (Geneworks, Therbaton, SA, Australia) were used as molecular weight markers (in kb). Lanes 2 and 3: *Bam*HI restricted bacterial and phage DNA, respectively; lanes 4 and 5: *Eco*RI restricted bacterial and phage DNA; lanes 6 and 7: *Hin*dIII restricted bacterial and phage DNA; lanes 8 and 9: *Cla*I restricted bacterial and phage DNA; lanes 10 and 11: *Bgl*II restricted bacterial and phage DNA; lanes 12 and 13: *Acc*I restricted bacterial and phage DNA. Two bands signifying the two phage and bacterial junctions, *attR* and *attL*, were evident in lanes 4, 6, 8 and 12. **Hin*dIII fragments (lane 6) from the lysogen, *S.* Typhimurium DT 64 were used for further analysis.

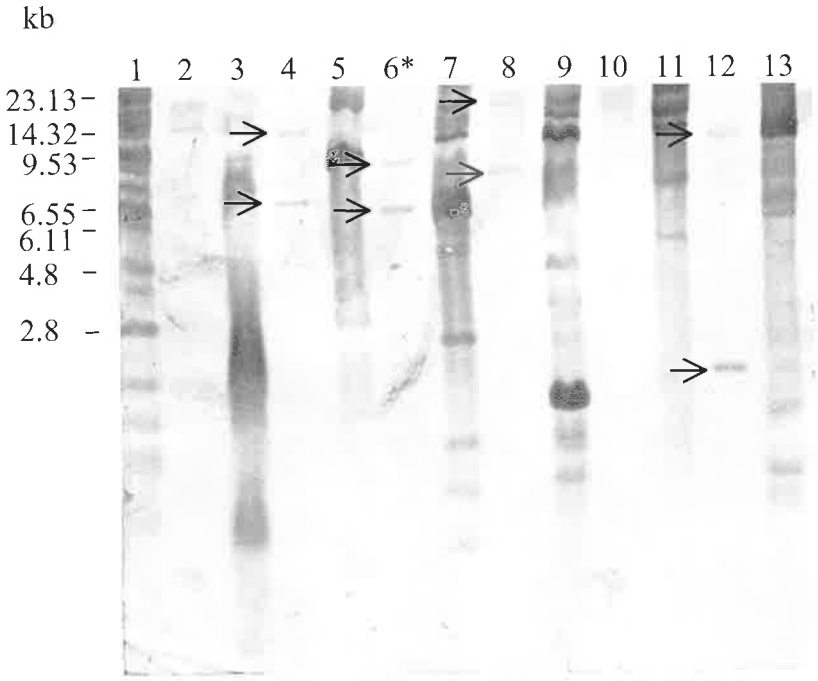

kb
1 2 3 4 5 6* 7 8 9 10 11 12 13
23.13
14.32
9.53
6.55
6.11
4.8
2.8

Figure 5.19

A graphical map showing the nucleotide sequences of ST64B, *attP/attB* sequence in the lysogen, *S.* Typhimurium DT 64 and compared to *S.* Typhimurium LT2 sequence (accession no. AE008788). The ST64B sequence is in yellow, the lysogen (*S.* Typhimurium DT 64) as well as the LT2 sequences are shown in blue. The junction of the three sequences is shown in green. Phage ST64B integrates within the *tRNA serU*, just upstream of the *umuDC* operon. The blue arrow pointing from right to left and designated tRNA *serU* indicates the region coding for tRNA *serU* and the direction of transcription. *S.* Typhimurium DT 64 and *S.* Typhimurium LT2 sequences, show perfect homology, except for a deletion (C) at position 269 in the *S.* Typhimurium LT2 strain.

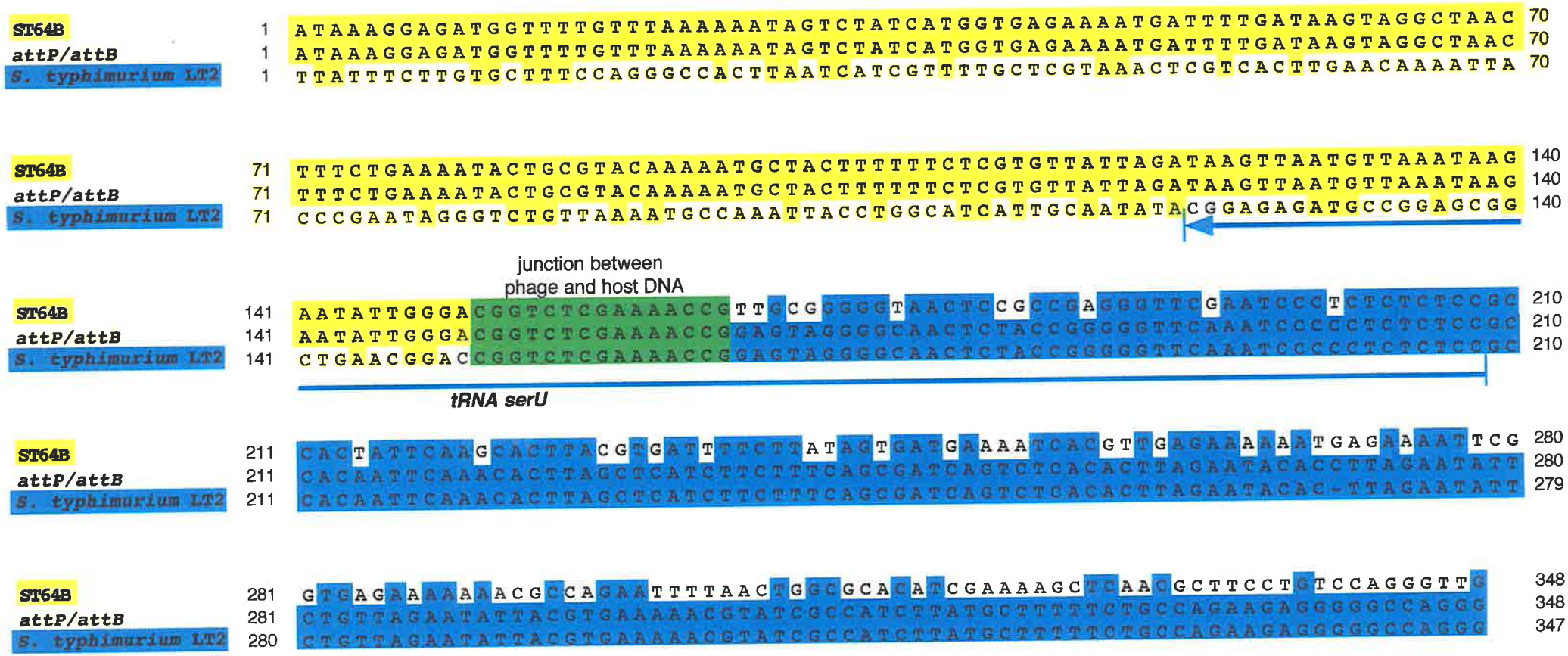

ST64B
attP/attB
S. typhimurium LT2
1 ATAAAGGAGATGGTTTTGTTTAAAAAATAGTCTATCATGGTGAGAAAATGATTTTGATAAGTAGGCTAAC 70
1 ATAAAGGAGATGGTTTTGTTTAAAAAATAGTCTATCATGGTGAGAAAATGATTTTGATAAGTAGGCTAAC 70
1 TTATTTCTTGTGCTTTCCAGGGCCACTTAATCATCGTTTTGCTCGTAAACTCGTCACTTGAACAAAATTA 70
71 TTTCTGAAAATACTGCGTACAAAAATGCTACTTTTTTTCTCGTGTTATTAGATAAGTTAATGTTAAATAAG 140
71 TTTCTGAAAATACTGCGTACAAAAATGCTACTTTTTTTCTCGTGTTATTAGATAAGTTAATGTTAAATAAG 140
71 CCCGAATAGGGTCTGTTAAAATGCCAAATTACCTGGCATCATTGCAATATACGGAGAGATGCCGGAGCGG 140
junction between
phage and host DNA
141 AATATTGGGACGGTCTCGAAAACCGTTGCGGGGGTAACTCCGCCGAGGGTTCGAATCCCTCTCTCTCCGC 210
141 AATATTGGGACGGTCTCGAAAACCGGAGTAGGGGCAACTCTACCGGGGGTTCAAATCCCCCTCTCTCCGC 210
141 CTGAACGGACCGGTCTCGAAAACCGGAGTAGGGGCAACTCTACCGGGGGTTCAAATCCCCCTCTCTCCGC 210
tRNA serU
211 CACTATTCAAGCACTTACGTGATTTTCTTATAGTGATGAAAATCACGTTGAGAAAAAATGAGAAAATTCG 280
211 CACAATTCAAACACTTAGCTCATCTTCTTTCAGCGATCAGTCTCACACTTAGAATACACCTTAGAATATT 280
211 CACAATTCAAACACTTAGCTCATCTTCTTTCAGCGATCAGTCTCACACTTAGAATACAC-TTAGAATATT 279
281 GTGAGAAAAAAACGCCAGAATTTTAACTGGCGCACATCGAAAAGCTCAACGCTTCCTGTCCAGGGTTG 348
281 CTGTTAGAATATTACGTGAAAAACGTATCGCCATCTTATGCTTTTTCTGCCAGAAGAGGGGGCCAGGG 348
280 CTGTTAGAATATTACGTGAAAAACGTATCGCCATCTTATGCTTTTTCTGCCAGAAGAGGGGGCCAGGG 347

Table 5.1 ST64B ORFs and putative function

ORF No.	Strand	Position (nt)[a]	Length (aa)	Putative function	Phage related sequences[b]	BlastP e value	% Identity
SB1	+	16 – 510	164	Terminase small subunit	GMSE-1; AAG50266	3e–04	36
					BK5–T; AAK56801	1e–08	34
					bIL286 *orf40*; NP_076674	3e–05	29
					Pi3 27; AAK05498		29
					PhiPV83; BAA97845	5e–04	25
					Phi–105; BAA36628	5e–37	28
					A2; CAB63682	6e–32	24
					bIL285 *orf41*; NP_076613	3e–27	23
						3e–24	
SB2	+	507 – 2240	577	Terminase large subunit	D3; AAD38954	e–134	46
					PhiSLT; AB045978	2e–54	30
					CP–933C; NP_287339	2e–52	30
					phiPV83; NP_061628	5e–37	25
					A2; CAB63682	3e–27	24
					bIL285; CAC97806	3e–24	26
					HK022 gp2; NP_037663	5e–21	26
					HK97 gp2; NP_037698	3e–21	26
SB3	+	2252 – 2434	60	Unknown			
SB4	+	2431 – 3675	414	Portal protein	HK97 gp3; AAA80202	3e–43	32
					HK022; AAF30355	8e–43	32
					Phi–C31 gp34; NP_047925	6e–41	31
					Phi-105; AAF27362	5e–40	27
SB5	+	3617 – 4303	228	Prohead protease	HK022; AAF30356	1e–16	35
					HK97	4e–16	35
					BFK20; CAB93910	9e–16	37

SB6	+	4318 – 5523	401	Major capsid protein precursor	CP–933C; AAG55944	1e–50	32
					phi–C31 gp36; NP_047927	2e–47 8e–14	25 25
					bIL285; NP_076613	4e–04	21
					BK5–T; NP_116499	6e–04	21
					DT1; NP_049396	0.013	24
					HK97; AAF31098	0.013	24
					HK022; AAF30354		
SB7	+	5574 – 5774	66	Unknown			
SB8	+	5777 – 6100	107	Unknown	CP–933C; AAG55948	4e–05	29
SB9	+	6097 – 6501	134	Unknown			
SB10	+	6467 – 6984	170	Unknown			
SB11	+	6982 – 7542	186	Unknown			
SB12	+	7546 – 7710	54	Unknown			
SB13	+	7700 – 9196	498	Tail sheath protein	Mu gpL; P44233	2e–73	37
SB14	+	9196 – 9552	118	Tail tube protein	Mu gpM; BAA19196	0.071	24
SB15	+	9540 – 9875	108	Unknown			
SB16	+	9960 – 11888	642	Phage related tail protein	P2 T; BAA36253	0.002	19
SB17	+	11922 – 12935	337	Unknown	Mu N; AAC23161	1e–18	32
SB18	+	13259 – 14317	352	Tail protein phage Mu	Mu P; NP_050648	2e–35	31
SB19	+	14317 – 14850	177	Base plate assembly	Mu gp45; AAF01123	3e–12	30
SB20	+	14855 – 15268	137	Unknown	Mu gp46; AAF01124	1e–08	37
SB21	+	15240 – 15785	181	Unknown			
SB22	+	15820 – 16341	173	Unknown	Prophage E14 YMFP; P75981	4e–48	57
					Mu gp47; AAF01125	1e–17	33
SB23	+	16344 – 16931	195	Unknown	SfV *orf1*; AAB72129	3e–27	54
					FluMu gp48; P44241		28

						3e–07	
SB24	+	16918 – 18480	520	Phage tail protein	Fels–2	1e–93	72
SB25	+	18480 – 19049	189	Tail fiber assembly protein	Fels–2; AAL21592	6e–63	63
					Fels–1; AAL19863	9e–11	54
					Gifsy–2; AAL19984	5e–11	54
					Mu gpU; AAF01128	6e–29	40
					P2 gpG; AAD03287	1e–27	38
					186 unnamed; AAC34165	2e–19	37
SB26	–	19334 – 20341	335	Unknown	CP–933K; AAG55140	3e–85	47
SB27	+	20554 – 20775	73	DNA invertase pin	CP–933H AAG54580	2e–14	83
					E14; pin AAA24391	6e–14	81
					Mu gin; AAF01129	2e–09	61
SB28 (*int*)	–	21217 – 22206	329	Integrase	CP–933U; AAG57039	7e–84	49
					CP–933M; AAG55457	6e–21	30
					HK022 *int*; S04990	3e–17	25
					HK97 *int*; NP_037720	7e–15	25
					H19B *int*; CAB38715	1e–14	25
					434 *int*; A45584	2e–15	25
					λ *int*; P03700	5e–14	25
SB29	–	22208 – 22450	80	Unknown			
SB30	–	22475 – 23044	189	Endodeoxyribonuclease (3'5' Exonuclease)	CP–933O ; AAG56113	1e–58	58
					CP–933P; AAK16992	1e–57	56
					CP–933M; AAG55458	4e–56	56
					CP–933U; AAG57038	4e–56	56

SB31	–	23048 – 23881	277	Unknown	H19J; CAB38712 933W ; AAD25483 VT2–Sa; BAA84359 P22 eaA; AAF75006	3e–33 1e–30 4e–30 4e–19	73 71 71 60
SB32	–	23878 – 24495	205	Unknown	N15 gp45; AAC19083 CP–933R; AAG56431 CP–933R; CAB38711	1e–09 1e–07 4e–07	68 61 60
SB33	–	24492 – 25007	171	Unknown			
SB34		25004 – 25234	76	Unknown			
SB35	–	25305 – 25844	179	Unknown			
SB36	–	25981 – 26808	275	Unknown			
SB37	–	26866 – 27237	123	Unknown			
SB38 (*cI*)	–	28053 – 28748	231	Regulatory protein	HK022 *cI*; CAA34222 D3 *cI*; NP_061565	4e–34 7e–07	42 26
SB39 (*cro*)	+	28846 – 29070	74	Unknown			
SB40	+	29372 – 29653	93	Unknown	E14 *ymfI*; P75976	1e–29	71
SB41	+	29650 – 30807	385	Unknown	P4 *orf199*; CAA80647 Phi–80 rha; A57258	6e–27 8e–21	58 52
SB42 (*O*)	+	31025 – 31843	272	Replication protein	CP–933M; AAG55470	4e–26	37
SB43	+	31803 – 32327	174	Unknown			
SB44	+	32327 – 33220	297	Unknown	CP–933V; AAG57233	2e–04	33
SB45	+	33217 – 33606	129	Cross over junction (Holliday junction nuclease Rus)	CP–933U; AAG57024 CP–933O; AAG56131 CP–933P; AAK16974 HK620; AAK28880	1e–17 1e–16 2e–15 2e–14	43% 56 39 42

SB46	+	33623 – 34483	286	Unknown	P1 KilA; KIBPP1 HK620; AAK28886	9e–10 2e–05	40 27
SB47	+	34527 – 35246	239	Unknown	CP–933O; AAG56170 CP–933N; AAG55888	8e–57 2e–55	43 42
SB48	+	35221 – 35481	86	Unknown	CP–933U; AAG57025 CP–933O; AAG56170	8e–35 2e–34	82 80
SB49	+	35492 – 36115	207	Unknown			
SB50	+	36146 – 36505	119	Unknown			
SB51 (*S*)	+	37031 – 37342	103	Lysis protein (holin)	Fels–1; NP_459883 P22 gp13; AAF75039	2e–41 1.3	100 36
SB52 (*R*)	+	37379 – 37993	204	Lysis	Fels–1; NP_459884	e–106	93%
SB53 (*Rz*)	+	37990 – 38475	161	Lysis	Fels–1 prophage; NP_459885 ES18 gp15; CAA47618 Gifsy–2 gp15; NP_460004 H–19B Rz; AAD17383 21 Rz; AAA32351 RAC Rz; AAC74444 λ Rz; AAA96599	93 2e–50 4e–79 3e–42 5e–41 2e–40 2e–40	5e–57 82 78 64 67 65 67
SB54	+	38687 – 39106	139	Unknown			
SB55	+	39326 – 39628	100	Unknown			
SB56	+	39689 – 40039	116	Unknown			

[a] Nucleotide coordinates corresponding to the first nucleotide of the initiation codon and the last nucleotide of the termination codon.
[b] Where similar sequences exist, the phage name precedes the gene designation (where possible) and the GenBank accession number.

Table 5.2 Potential CII–, IHF–binding sites and rho–independent terminators in ST64B genome

Location	Similarity score[a]	Sequence
Potential CII–binding sites		
28623 – 28636		**TTGC**TTTTGC**TTGC**
28790 – 28803		**TTGC**GCAAGT**TTGC**
31197 – 31210		**TTGC**GGCCC**TTGT**
28509 – 28522		**TTGC**TTCACG**TTGC**
Potential IHF–binding sites		
{22063 – 22096}	55.7	AGAATTAC**ATT**C**A**CACA**ATTA**TGGCGAAAATTT
24490 – 24522	53.7	ATCAGAAT**ATCAA**AAAT**TT**CAGACTTATTGACG
{25948 – 25196}	50.1	AAGAAACT**ATCA**GATATTTATTTAGAAAACATC
{26945 – 26977}	48.5	GCAGTATTT**TCA**G**A**AAG**TTA**GCCTACTTATCAA
38887 – 38920	53.3	GAAAGAAAT**TCAA**CCAC**TTA**AAGTAATAAAGGG
Lambda *cro/cII*	61.5	TGCATACAT**TCAAT**CAA**TTG**TTATCTA
Potential rho–independent terminators		
2324 – 2350		TTTATCCCCCGGCGGGTTTTGTTGTT
5433 – 5451		ACCGTTTGTCGGTTTTTAT
6110 – 6137		AGGCGCAGACCAGCGCCACATACCTTTT
28983 – 29001		GCGCCTGGCTCAATTTTAC
{28049 – 28010}		AAAGCGTCGACAAGGCCACCATTATGGTGGCTTTTTTTTT

Putative C1–binding sites (boldface) were determined by scanning the sequence for the consensus sequence (TTGCN6TTGY). IHF–binding sites were scanned using MacTargsearch (Goodrich *et al.*, 1990). [a]Similarity scores were included in the case of IHF–binding sites. Only those with scores of more than 48 were included. For comparison, a characterised IHF–binding site in phage λ was included and bases identical to the consensus sequence are highlighted in boldface. Rho–independent terminators were determined by scanning the sequence for a stem–loop structure (boldface and underlined) followed by a region rich in thymine residues (Brendel and Trifonov, 1986). The sites in brackets {} are those associated with the complementary strand.

Table 5.3 Potential promoters in phage ST64B genome

Location	Putative function	Sequence
5443 – 5488		GGTTT**TTATAC**CACCAAGCGCACCGGGGG**TATGCT**GGTCG**A**TTCACAGGC
7601 – 7646		TTTTAC**CTGAAG**GAGGTCGAAATGTTGAT**GAGAAT**AACTA**C**TGGCTGCGC
10317 – 10362		GATTTG**ATGCAG**TTTGCGAACGACGCAGT**GAAAAT**GGGTG**T**GGCGTTTGA
11160 – 11205		GTATCCA**TTGGCA**GCCTGTCCCGGGCTGT**CAAAAT**CCTGA**A**CAGTGTCAT
13097 – 13142		ACAGG**TTGAAC**GTACTGCATGCGAACGCC**GATGAT**GTTCG**G**CCCGCACTG
16924 – 16969		GAA**TTGATA**CGCCCACCGCGCAAAAA**GATAAA**TTTGGTC**A**GGGAAAAAA
18776 – 18821		TGATGA**TTGCCG**TGGCGATGGTACGTGGT**TATATC**AGGAC**G**GTAAAGTCG
20426 – 20469		GAAGATTT**TTGTAT**GAAGATTTGCTACCG**TATTTT**CGTGC**C**ATAGCTATC
[22948 – 22903]	*xis*/*int*	GTGGTGAAA**TTGGAC**CTGAGTTCTATACCGC**CGTTAG**CCT**T**GAAAGCGCA
[27989 – 27941]	P_L	CGCAATT**TTACAC**TTGCGCAATGTGCAAT**TTAAAT**GTAAT**T**TGCACTCAT
[28855 – 28810]	P_{RM}	CAATGGTGA**TTGCAT**AAAACCTCCTTCTTTTG**CGTATA**TC**G**CATGTTATT
28794 – 28839	P_R	GCAAGT**TTGCTG**TATGAAATAACATGCGA**TATACG**CAAAA**G**AAGGAGGTT
36034 – 36079		ATGGT**GTGTGA**ATCACTTTAATTGTGC**TGTATT**TCTCGCA**T**TCAGGGATG
36838 – 36883	P_{LATE}	ATGAT**TTGGAA**AAGGTCTGAGTA**ATTACT**CATTTTTACCC**C**GTTTAACGT

The ST64B sequence was scanned for probable promoter regions using Martin Reese's Promoter Prediction by Neutral Network program (http://www.fruitfly.org/seq_tools/promoter.html). Sequence in boldface resemble those of the −35 and −10 consensus sequence. For some putative promoters, functions are inferred.

Chapter 6

Comparison of phages ST64T and ST64B

6.1 Introduction

Bacterial viruses are ubiquitous in the environment. Newly developed techniques have made it possible to estimate the total numbers of phages in environmental samples (Fuhrman, 1999, Noble and Fuhrman, 1998, Sander and Schmieger, 2001). The high phage titers (>10^6/ml) found by such methods in soil and aquatic environments such as the sea (Miller, 1998), indicate that phages constitute a large fraction of the biomass (Tétart, *et al.*, 2001). Although phages are probably the most abundant biological entities on the planet (Hendrix, *et al.*, 1999), their diversity makes it difficult to assess the impact they make on the biosphere (Fuhrman, 1999).

The modular theory of phage evolution states that the product and the unit of phage evolution is not a given phage, but a family of interchangeable genetic elements (modules) which can be considered as a functional unit (Botstein, 1980). The dsDNA tailed bacteriophages which form a large and diverse group of viruses found in natural settings, vary considerably, with known genome sizes ranging from <19,000 to >160,000 bp (Casjens, *et al.*, 1992). The availability of phage genomic sequences, allows the calculation of relationships that have led to the suggestion that the worldwide population of today has a common ancestry (van Belkum, *et al.*, 2001). A recent hypothesis based on genomic sequencing data suggests all dsDNA tailed phage genomes are mosaics with access, by horizontal genetic exchange, to a large common pool, but in which access to this gene pool is not uniform for all phages (Hendrix, *et al.*, 1999). In contrast, comparative phage genomics and the evolution of *Siphoviridae* showed relatedness within the λ supergroup (λ–, ΨM2–, L5, Sfi21–, Sfi11–, φC31–, sk1– and TM4–like phages). This

relatedness suggested elements of vertical evolution in the capsid modules of the family, *Siphoviridae* (Brüssow and Desiere, 2001).

In this chapter, the genomic sequences of two members of the lambdoid family of phages, *S.* Typhimurium phages ST64T and ST64B are compared with those of *S.* Typhimurium phage P22 and *E. coli* phage λ, with an emphasis on evolution.

6.2. Results

6.2.1 Genome architecture

The genomes of both ST64T and ST64B appear to be similar with regard to organisation and direction of transcription (Figure 6.1). Both genomes resemble the two archetypes of the lambdoid phages, P22 and λ. Although λ and P22 appear to be distant in evolutionary terms from each other based upon the host range, morphology and the mode of DNA packaging, their genomic architecture is almost identical. Based on sequence comparisons of ST64T and ST64B with other phage sequences in the database, these two phages appear to be unrelated. For example, their target sites for packaging appear to be distinct, since the terminase of ST64T like that of P22, probably recognises the putative *pac* site whereas ST64B being similar to λ, probably starts DNA packaging at as yet undetected *cos* sites for the reasons discussed in chapter 5. The putative capsid proteins of ST64T and ST64B appear to be distinct, since ST64T has a putative scaffold whereas ST64B has no scaffold. ST64B has a putative protease whereas ST64T has no protease. Although ST64T resembles P22 in the head and tail genes, ST64B was shown to be a mosaic in this region of the genome. The putative portal protein of ST64B exhibited some similarity to a probable phage portal–like protein found in *Mesorhizobium loti*, and to the portal proteins of the *E. coli* phages HK97, HK022 and the *Streptomyces* phage ϕC31. The putative protease exhibited sequence similarity to proteases in HK97, HK022 and the phages of *Rhodobacter* and *Caulobacter*. Surprisingly, the putative major head protein

was most similar to the major head proteins of phages in *A. tumefaciens* (AAK86763), *M. loti* (NP_108602), *C. crescentus*; (AAK24747), *Streptomyces* ϕC31 and less similar to the head proteins of HK97 and HK022.

A similar scenario was observed in the examination of protease/head protein sequences of *Streptomyces* ϕC31 when compared to related phage proteins (Smith, *et al.*, 1999). The authors suggested that the inability to discern a more comprehensive relationship with a single phage genome is suggestive of horizontal gene transfer between phages. The genomes of ST64T and ST64B share very little sequence identity with each other and are not easily comparable. They appear to belong to two different "sub–groups" within the lambdoid family with ST64T belonging to a P22–like group and ST64B most probably belonging to a λ–like group.

6.2.2 Genomic mosaicism

Figure 6.2 presents the genetic map of ST64T with representation of the degree of sequence similarity to P22 presented graphically in a histogram. The map clearly shows mosaicism of segments which are almost 100% identical with the corresponding areas of P22 and segments which do not have any sequence identity with corresponding areas of P22. In most cases, between these segments of sequence identity and non-identity, there are sharp transitions. An interesting feature of this mosaicism is shown in the area of the immunity genes (Figure 6.2). The putative gene 24_{ST64T} is 96% identical to gene 24_{P22} and as such the *nut* sites of both phages are also almost identical. Figure 6.3 presents a comparison of the highly conserved short sequences in the genomes of phages P22, ST64T and PS100 (a prophage of *Salmonella*), flanking the immunity C region, which has very little sequence identity in the three phages. Such transitions identify sites of recombination in the ancestry of phages ST64T and P22. Indeed, Juhala, *et al.* (2000) have suggested that such a recombination event is representative of the non–homologous or illegitimate

mechanism because it is between two largely non–similar sequences. Moreover, there are not sufficient homologous regions on the phage genome to explain frequent module exchange. It is thus evident that these recombination events may be explained by the hypothesis that individual phages of the lambdoid family are mosaics that draw sequences from a common genetic pool. However, the putative gene *N* (SB37) in phage ST64B could not be identified by sequence similarity to any sequences in the database and as such, even the *nut* sites could not be identified. This indicates that ST64B and ST64T are completely unrelated in this region of their genomes, which suggests that their ancestry may be distinct.

6.2.2.1 Recombination between genes

Examination of the sites of sequence transitions or the putative sites of ancestral recombination indicated that the majority of them are located at gene boundaries. This is illustrated in Figure 6.4 in a comparison between ST64T and P22 at the end of gene *ninZ* and the start of gene *23* in both phages. The sequence shown in Figure 6.4 includes the putative *ninZ* of ST64T which shares 93% identity in nucleotide sequence with *ninZ* of P22. This sequence is followed by gene *23* in both ST64T and P22 that shows no sequence similarity between the two phages. The similarity between the shared *ninZ* genes persists until the first base after the termination codon, then the two sequences diverge completely. The putative gene *23* of ST64T was shown to be closely related to gene *Q* of phage 21 by sequence similarity (81% sequence identity).

A similar scenario was shown with ORF146 of λ and gene *61* of HK97 (Juhala, *et al.*, 2000). The two sequences shared 96% nucleotide sequence identity and the similarities persisted until the last base of the termination codon, after which the sequences diverged completely at the start of gene *62* in HK97 and ORF290 in λ. The same region was also examined in P22 and HK97 where the apparent recombination point was several codons upstream of the termination codon of the two genes (Juhala, *et al.*, 2000). Figure

6.4 is representative of the putative recombination sites seen throughout the genome of ST64T where the sites are sometimes at the gene boundaries or occur a short distance away from the end of the putative gene.

6.2.2.2. Recombination between coding regions

The recombination region of ST64T is very similar to that of P22. Figure 6.2 shows that putative genes *abc2*, *abc1* and the 3' end of *erf* in ST64T are >90% identical to analogous genes in P22. Furthermore, a comparison between the P22 and ST64T recombination genes with areas of similarity shaded, is presented in Figure 6.5. In the third recombination gene, *erf*, there is evidence of a recombination event within the coding region, with the resulting sequence transition occurring at codon 180 of the 235 inferred gene product in ST64T and at codon 150 of the 205 gene product in P22. The carboxy termini (approximately a quarter of the Erf protein) of Erf_{ST64T} and Erf_{P22} show high sequence identity (>90%). Previous studies have shown that Erf is a two–domain protein and that the two domains can be separated by protease treatment and assigned different functions. These domains can also be visualised separately by electron microscopy (Murphy, *et al.*, 1987a, Poteete, *et al.*, 1983). Interestingly, the domain boundary identified by these researchers corresponds with the location of the putative recombination site in ST64T where the putative smaller domain of the gene product Erf, is similar to the smaller domain of Erf in P22. A similar scenario was observed in HK97 when the recombination genes were compared to those of P22 (Juhala, *et al.*, 2000). In this case, the sequence transition occurred at approximately codon 150 of the 200 gene product with 75% of the amino terminal of the Erf_{HK97} and Erf_{P22} proteins predicted to be 89% identical (Juhala, *et al.*, 2000). Therefore, Erf_{HK97} is similar to the large domain of Erf_{P22} whereas Erf_{ST64T} is similar to the small domain of Erf_{P22}. It was suggested that the hybrid protein produced as a result of the recombination event did not disrupt the function of the protein

because recombination occurred at the boundary of the two domains which were earlier shown to be separately functional (Juhala, *et al.*, 2000).

6.2.2.3 Modular recombination of the lysis gene

A study of the lysis gene cassette of various P22–related lambdoid phages shows the mosaic structure of their genomes. The actions of various endolysins (e.g. P22 lysozyme or λ transglycosylase) lead to lysis of bacterial cells. The holin of λ and P22 are similar whereas their endolysins are distinct (Casjens, *et al.*, 1992). The lysis cassette of ST64T is almost identical (99–100% amino acid identity) to the lysis cassette of *S.* Typhimurium prophage PS3. Similarly, the lysis cassette of ST64B is 93 – 100% identical on a protein level to the putative cassette in *S.* Typhimurium prophage, Fels–1. A study on the lysis genes of 140 *Salmonella* P22–like phages, revealed one combination of the two lysis genes, a holin and an endolysin in PS3 (Prof. H. Schmieger, personal communication). This combination was only detected in phage PS3. Phage PS3 was harvested from a horse isolate of *S.* Typhimurium in Rhode Island, USA in 1987 (Beltran *et al.*, 1991). Interestingly, phage ST64T was harvested from a 1997 human *S.* Typhimurium DT 64 isolate from Australia and was shown to have a lysis cassette highly similar to that of PS3, including the highly variable gene *15* which shared 100% sequence identity with that of PS3 at the protein level. At the nucleotide level, approximately 3,000 bp region encompassing the entire lysis cassette of ST64T and PS3 showed only 6 mismatches. The endolysin of ST64T also showed weak overall sequence identity (26%) (with a short region of high similarity within the sequence) with an endolysin of phage A118 from *Listeria monocytogenes*. This endolysin has been shown to act as an L–alanine–D–glutamate peptidase (Loessner, *et al.*, 1995) (Figure 6.6).

Although the putative holin of phage ST64B showed a very high amino acid identity to an unknown protein of prophage Fels–1, it also showed a weak identity to P22

holin. The putative endolysin of ST64B shared sequence identity not only with the Fels–1 protein (100%) but also with the phage like transglycosylase protein of *P. aeruginosa* (41%) which was shown to be involved in lysis of the bacterial cell to release the pyocins (Nakayama, *et al.*, 2000). Although combinations of different modules of holins and endolysins can be found in different phages, the last of the lysis genes, *15*, appears to be conserved within the lambdoid phages (Figure 6.7).

6.2.3 Phylogenetic analysis

Maximum parsimony (MP) analysis was performed using PAUP* version 4.05b (Swofford, 1993). Different algorithms were used for analysis which provided different information contained within the data. Two putative gene products, terminase large subunit and major head protein were chosen for analysis. Protein sequences were used for analysis.

6.2.3.1 Terminase large subunit

A phylogenetically diverse range of bacteriophages use the *cos*–site mechanism for DNA packaging (Desiere, *et al.*, 2001, Gilakjan and Kropinski, 1999). Large subunit terminases have in the past been selected for phylogenetic analysis based upon sequence alignments (Desiere, *et al.*, 2001, Gilakjan and Kropinski, 1999). It has been suggested that the basic requirements of the enzymatic process has probably imposed the conservation of the large subunit terminase (Smith, *et al.*, 1999). The putative terminase large subunit from ST64B (encoded by SB2) when analysed by BlastP, yielded similarities to many actual and putative terminases. A limited number of proteins that crossed the e–value threshold of 3e–20 were chosen for analysis. For comparison of ST64T and ST64B terminases, both ST64T and P22 were included in the analysis even though they use a different mechanism for packaging of DNA (*pac* site recognition). An alignment of

the terminase amino acid sequence from phages was performed using ClustalW in BioEdit version 4.8.10 (Hall, 1999) (Figure 6.8). MP analysis was performed on the 705 amino acid consensus alignment with 69 characters excluded, using PAUP* version 4.05 (Swofford, 1993). Furthermore, 529 characters were parsimony informative. Both MP heuristic and exhaustive searches produced the same tree (Figure 6.9). The reliability of the trees obtained was tested by bootstrap analysis, which suggested that the terminase of ST64B is closely related to both *E. coli* YMFN hypothetical protein and the D3 terminase at a confidence level of 100% as shown in branch 2. Branches 1 and 3 were more distantly related to branch 2 at a bootstrap confidence level of 83% pseudo replicates. However, ST64T and P22 terminases were shown to be related to each other at a bootstrap confidence level of 100% but unrelated to the other groups. These results are supported by known facts in that P22 uses a headful mechanism for DNA packaging, recognising a *pac* site located within the terminase small subunit whereas D3 and other characterised phages in this tree bind to the 3' extended *cos* sites (Juhala, *et al.*, 2000) (in contrast to the 5' extended *cos* sites found in phage λ) for DNA packaging.

6.2.3.2 Major head protein

The SB6 encoded inferred gene product when analysed by BlastP yielded similarities to putative and actual major head proteins from different phages. A limited number of proteins that crossed the e–value threshold of 0.013 were selected for analysis. For comparison of ST64T and ST64B putative major head proteins, ST64T and P22 major head proteins were included in the analysis, although they follow a different head assembly mechanism. An alignment of the proteins was performed using ClustalW in BioEdit version 4.8.10 (Hall, 1999) (Figure 6.10). MP analysis was performed on the 509 amino acid consensus alignment with 39 characters excluded using PAUP* version 4.05 (Swofford, 1993). However, 422 characters were parsimony informative. A consensus of two equally parsimonious trees suggested that the putative major head protein of ST64B

was related to the major head protein of phages found in *A. tumefaciens*, *C. crescentus* and *M. loti* at a confidence level of 95% as shown in branch 1 of Figure 6.11. The branch 2 phages were shown to be related to the branch 1 phages at a confidence level of 100% pseudo replicates. However, ST64B was shown to be distantly related with the branch 3 phages D3, HK022 and HK97 at a confidence level of 90%. Similar to the terminase dendrogram, ST64T and P22 major head proteins were shown to be related to each other at a bootstrap confidence level of 100% but unrelated to the other groups. Indeed, BlastP search revealed that ST64B is more closely related to the major head protein of ϕC31 than to those of HK97 and HK022. The most parsimonious tree of the same topology requires a total of 100 changes. A partial sequence of the major head protein of phage Tu1a, a T4–type phage was employed as an outgroup.

Studies on HK97 head morphogenesis indicated that a major head protein band at 43 kDa could not be found, instead, a 31 kDa protein band could be found with the appearance of high molecular weight material which includes the protein that failed to enter the SDS PAGE gel matrix. This feature was also seen in phage D3 including the conserved residues that form cross–links within the head protein in order to link the proteins together in a kind of a chain mail arrangement (Gilakjan and Kropinski, 1999) (see chapter 5). Although ϕC31 follows the HK97 head assembly mechanism, the chain mail feature has not been observed (Smith, *et al.*, 1999). Furthermore, ST64B major head protein, similar to the characterised major head proteins of phages in branches 1, 2 and 3 was shown to be processed with the final molecular weight of approximately 28 kDa instead of 44 kDa. However, both the cross–links and the high molecular weight material including the protein which failed to enter the gel matrix in HK97 and D3, were not observed in ST64B. This may explain why ST64B is distantly related to HK97 on the basis of head assembly (Figure 6.11).

6.3 Discussion

Several authors have shown that there is a vigorous and ongoing horizontal exchange among the well studied lambdoid phages of *E. coli* and *Salmonella* (Campbell, 1988, Campbell, 1994, Casjens, *et al.*, 1992, Hendrix, *et al.*, 1999, Juhala, *et al.*, 2000). It is thought that the exchange probably occurs when two phage genomes are found in the same host as two co–infecting phages or as a single phage infecting a cell carrying another prophage (Hendrix, *et al.*, 1999). Although this has been evident in several phages, it is surprising that ST64T and ST64B which are prophages of the same *S.* Typhimurium DT 64 isolate, share very little sequence identity, with ST64B sharing sequence similarities with dsDNA phages of diverse origins, such as the non–lambdoid phages Mu, P1 and P2 as well as phages from *Lactococcus* and *Rhodobacter*. The proposition of Hendrix and coworkers that "all dsDNA phage genomes are mosaics with access, by horizontal exchange, to a large common genetic pool but in which access to the pool is not uniform for all phages" is strongly supported by the analyses of the genomes of phages ST64T and ST64B which appear to have undergone genetic exchange at different levels.

Evidently, ST64T has undergone recombinational exchange of gene modules or cassettes during its evolution. The high level of sequence identity between ST64T and P22 and the sharp transitions in ST64T genome leading to putative genes with little or no sequence similarity between these two phages, probably imply that horizontal gene exchange between ST64T and P22 occurred in the very recent evolutionary past. This has also been suggested for HK97 and P22 (Hendrix, *et al.*, 1999). Examples of mosaic boundaries in ST64T genome were found at the end or near the end of a putative gene. A similar boundary was shown within a coding region where it coincided with the inferred protein functional domain. Generally, when a mosaic boundary falls within a coding region, it often coincides with the functional domain boundary in the protein (Juhala, *et al.*, 2000). Juhala, *et al.* (2000) have suggested that recombination can occur anywhere in the genomes of phages and it is possible that such recombination may not disrupt the function

of the recombinant genome. This was the case with recombination sites seen in the genome of ST64T, which in all likelihood did not disrupt the function of the recombinant genes.

The theory of modular evolution states that modules may be genes, group of genes or only domains of genes. For example, the entire block of genes involved in morphogenesis (head and packaging genes) represents one module. However, the three products of the lysis gene cassette, may operate independently from each other, with each gene representing a module (Casjens, *et al.*, 1992). The lysis genes in both ST64T and ST64B have shown to be similar to the lysis genes in other *Salmonella* lambdoid phages (PS3 and Fels–1, respectively) with two potential AUG start codons found in their holin genes.

Comparison of the major head proteins and terminases of ST64T and ST64B have confirmed that these two phages are distant relatives in this region of their genomes. It is surprising that the putative large and small subunit terminases of ST64B are closely related to terminases of phages infecting the Gram positive bacteria that are known to have 3' *cos* extended termini, unlike the 5' *cos* extended termini found in phage λ. The 3' *cos* extended termini have also been shown in *P. aeruginosa* phage D3 (Sharp, *et al.*, 1996). It has been shown that the hypothetical protein encoded by gene *ymfN* (YMFN protein shares 93% sequence identity with putative terminase large subunit of ST64B) lies centrally in the *intE* (integrase) to *pin* (invertase) gene cluster of *E. coli* (Gilakjan and Kropinski, 1999). It is also documented that both of these genes are remnants of the cryptic lambdoid phage E14 (cited in Gilakjan and Kropinski, 1999). Similarly to D3, it may be hypothesised that the progenitor of ST64B was a coliphage related to the cryptic phage E14 or recombination occurred between the ST64B and the E14 progenitor (Gilakjan and Kropinski, 1999). This may be supported by the fact that ST64B unlike ST64T, shares much of sequence identity with genes of unknown function in *E. coli* (see Table 6.1), which may be remnants of

cryptic phages. Based on sequence identity, it is likely that ST64B like D3, possesses the 3' extended termini characteristic of phages that largely infect Gram positive bacteria.

Dramatic evidence for wide–spread genetic exchange among diverse phage groups comes from the discovery that the tail fiber genes of most of the *E. coli* phages in the databases match each other in a complex manner (Casjens, *et al.*, 1992). These studies revealed that segments of the two tail fiber genes shared sequence similarity with phages as genetically diverse as λ, P2, P1, T4, T7, Mu and two defective *E. coli* prophages (Casjens, *et al.*, 1992). The authors emphasised that these segments of tail fiber genes were found in different combinations in each phage as if the phages built their tail genes from a "menu" of available tail segments. A similar scenario was observed with the tail fiber genes of ST64B where 4 putative genes (SB13, SB14, SB17 and SB18) had sequence identity to tail genes of phage Mu, one (SB16) had sequence identity to tail genes of P2, SB24 had identity to tail genes of Fels–2 and SB24 had sequence identity to Fels–1, Fels–2, Gifsy–2, Mu and P2 tail genes. Such features imply horizontal gene exchange in the very recent evolutionary past (Casjens, *et al.*, 1992, Hendrix, *et al.*, 1999).

In conclusion, both ST64T and ST64B are mosaics and although they share very little sequence identity, they both appear to belong to the lambdoid family.

Figure 6.1

A schematic representation of the maps of ST64B and ST64T showing their genomic architecture. The sizes of the maps are approximated. Different colours represent groups of genes involved in a particular function. The P22 and λ nomenclature were used for designating the putative genes in ST64T and ST64B, respectively.

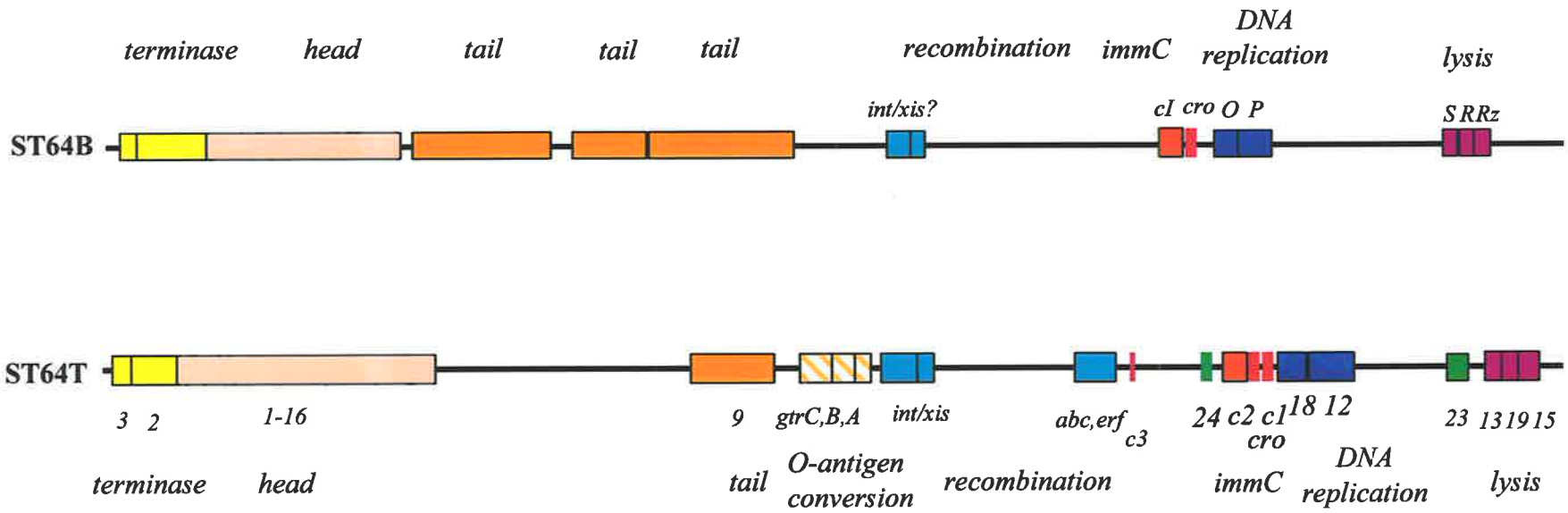
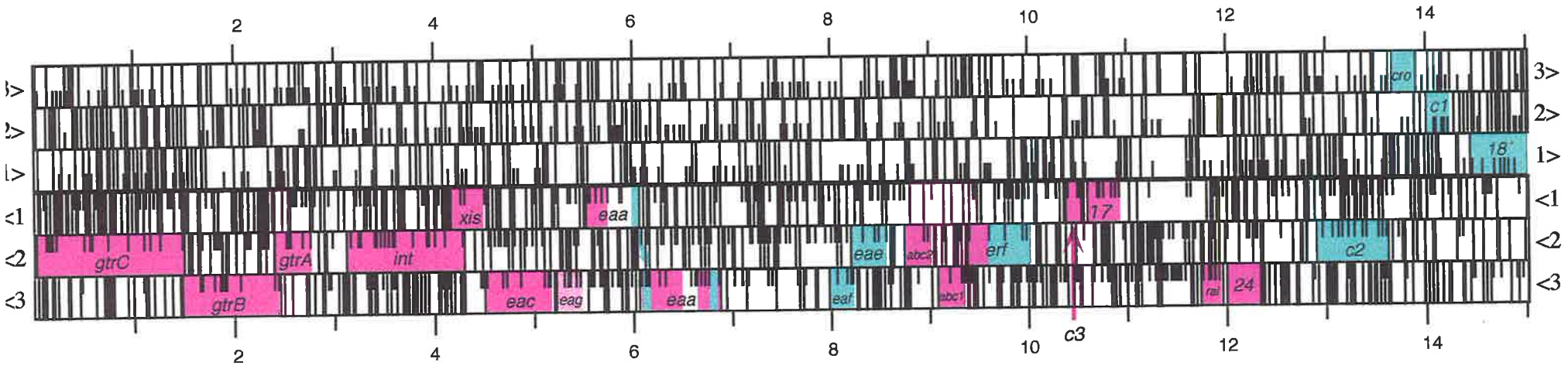
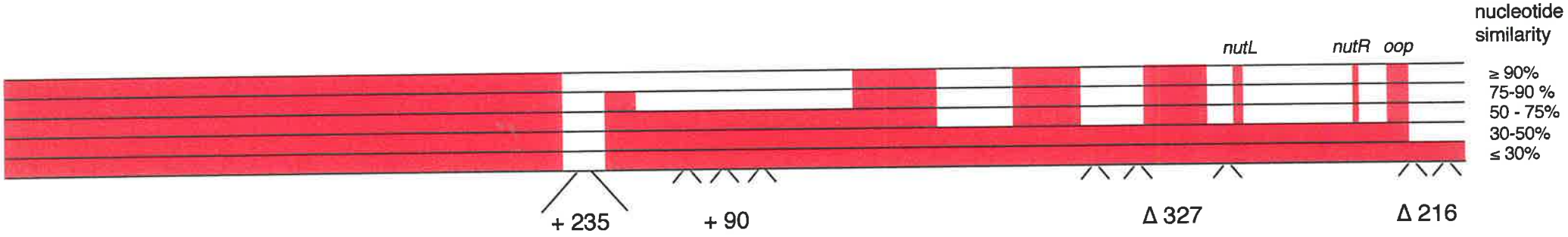

terminase
head
tail
tail
tail
recombination
immC
DNA replication
lysis
int/xis?
cI cro O P
S RRz
ST64B
ST64T
3 2
1-16
9
gtrC,B,A
int/xis
abc,erf
c3
24 c2 c1
18 12
23 13 19 15
cro
terminase
head
tail
O-antigen conversion
recombination
immC
DNA replication
lysis

Figure 6.2

ST64T map showing sequence similarity to P22. The coloured histograms show the degree and locations of sequence similarity between ST64T and P22. Numbers below the histograms preceded by the symbol Δ or + indicate positions of the deficiency or surplus, respectively, of the indicated number of base–pairs in ST64T relative to the P22 sequence.

ST64T: 1 - 15 kb

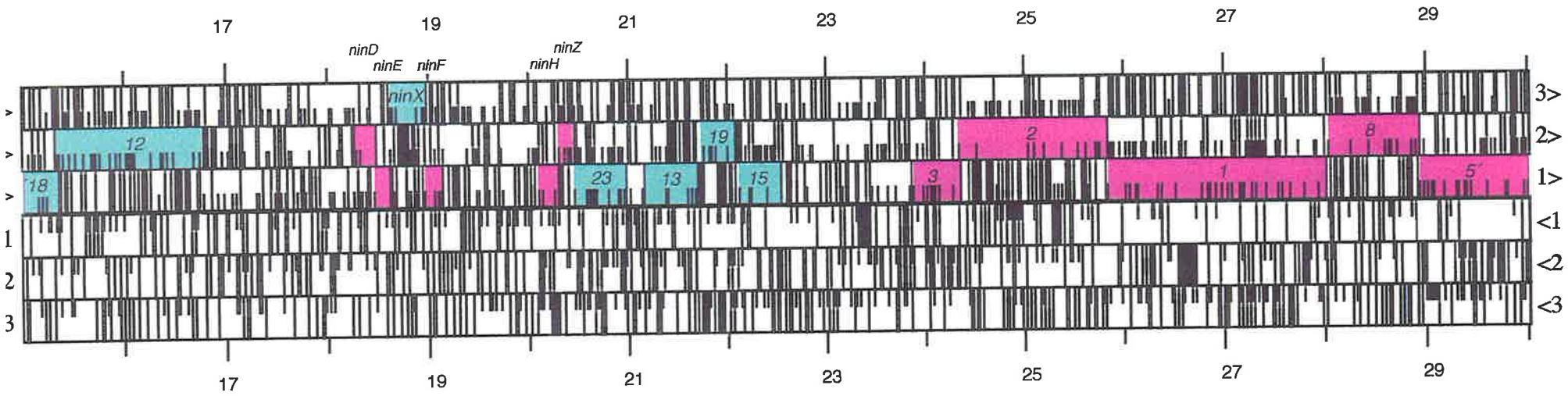


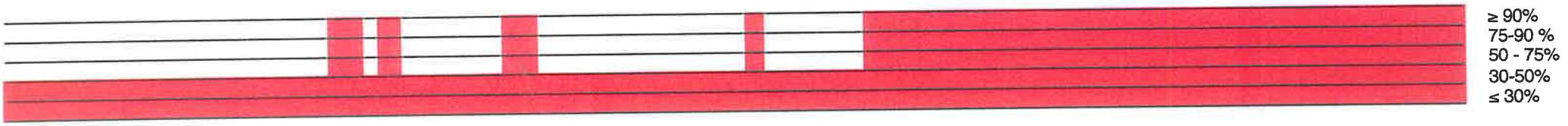


> 85% similarity with P22

gene assigned according to position and size, weak or no similarity with P22

ST64T: 15 - 30 kb





ST64T: 30 kb - End

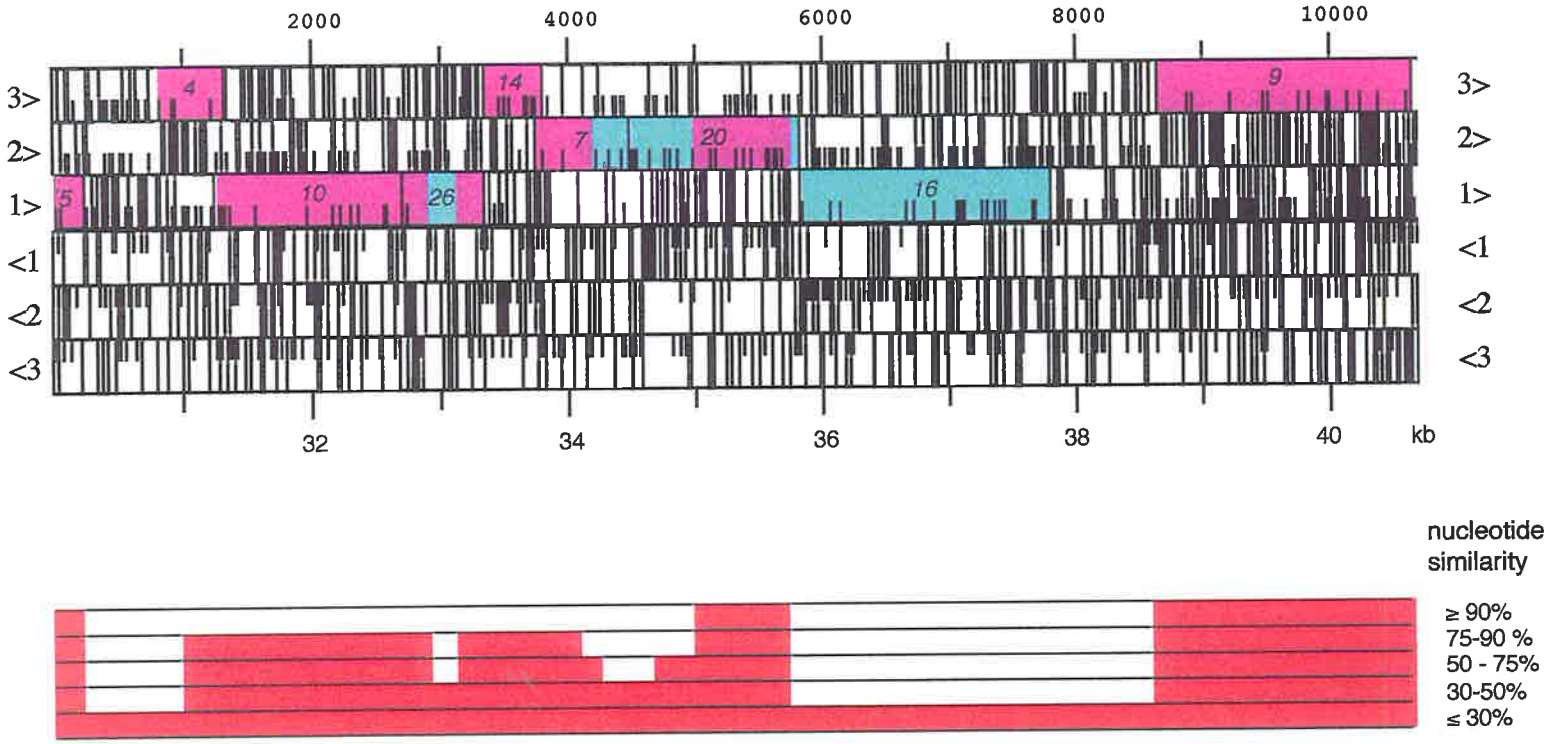


Figure 6.3

Sequence similarity of the immunity region of phages P22, ST64T and PS100. The figure shows the conserved *nut* sites, *nutL* and *nutR* in the three phages encompassing the immunity genes *c2* and *cro* which are distinct in P22, ST64T and PS100. Such highly conserved sequences allow module exchange by homologous recombination. The PS100 sequence was kindly provided by Prof. H. Schmieger.

The *immunity C* modules comprising the genes *c2* and *cro* are flanked by short, highly conserved sequences allowing module exchange by homologous recombination

Figure 6.4

Comparison of genes *ninZ* and *23* of ST64T and P22. The figure shows high sequence identity of *ninZ* and a sharp transition at the end of *ninZ*.

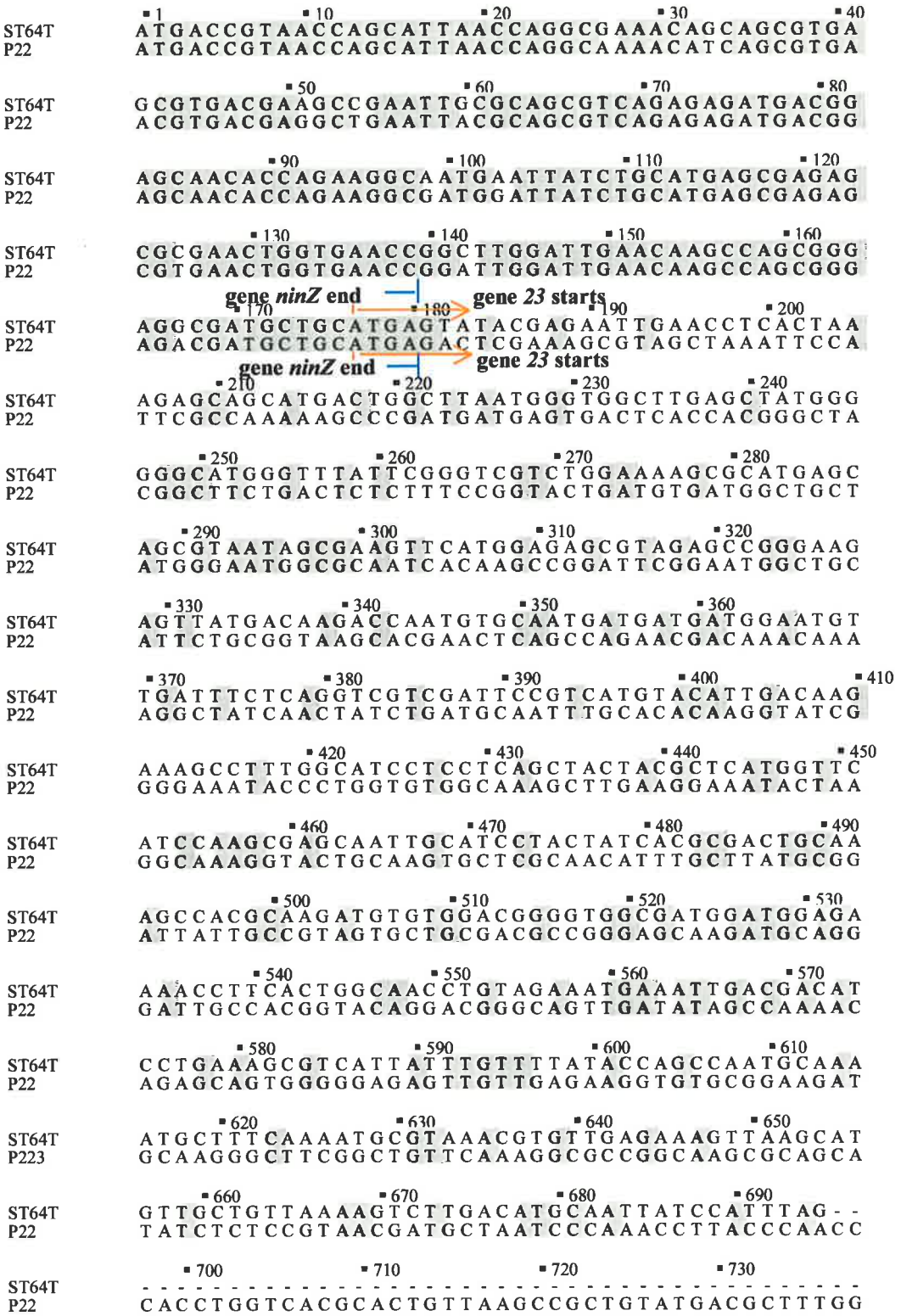

1 10 20 30 40
ST64T ATGACCGTAACCAGCATTAACCAGGCGAAACAGCAGCGTGA
P22 ATGACCGTAACCAGCATTAACCAGGCAAAACATCAGCGTGA
50 60 70 80
ST64T GCGTGACGAAGCCGAATTGCGCAGCGTCAGAGAGATGACGG
P22 ACGTGACGAGGCTGAATTACGCAGCGTCAGAGAGATGACGG
90 100 110 120
ST64T AGCAACACCAGAAGGCAATGAATTATCTGCATGAGCGAGAG
P22 AGCAACACCAGAAGGCGATGGATTATCTGCATGAGCGAGAG
130 140 150 160
ST64T CGCGAACTGGTGAACCGGCTTGGATTGAACAAGCCAGCGGG
P22 CGTGAACTGGTGAACCGGATTGGATTGAACAAGCCAGCGGG
gene ninZ end
gene 23 starts
170 180 190 200
ST64T AGGCGATGCTGCATGAGTATACGAGAATTGAACCTCACTAA
P22 AGACGATGCTGCATGAGACTCGAAAGCGTAGCTAAATTCCA
gene ninZ end
gene 23 starts
210 220 230 240
ST64T AGAGCAGCATGACTGGCTTAATGGGTGGCTTGAGCTATGGG
P22 TTCGCCAAAAAGCCCGATGATGAGTGACTCACCACGGGCTA
250 260 270 280
ST64T GGGCATGGGTTTATTCGGGTCGTCTGGAAAAGCGCATGAGC
P22 CGGCTTCTGACTCTCTTTCCGGTACTGATGTGATGGCTGCT
290 300 310 320
ST64T AGCGTAATAGCGAAGTTCATGGAGAGCGTAGAGCCGGGAAG
P22 ATGGGAATGGCGCAATCACAAGCCGGATTCGGAATGGCTGC
330 340 350 360
ST64T AGTTATGACAAGACCAATGTGCAATGATGATGATGGAATGT
P22 ATTCTGCGGTAAGCACGAACTCAGCCAGAACGACAAACAAA
370 380 390 400 410
ST64T TGATTTCTCAGGTCGTCGATTCCGTCATGTACATTGACAAG
P22 AGGCTATCAACTATCTGATGCAATTTGCACACAAGGTATCG
420 430 440 450
ST64T AAAGCCTTTGGCATCCTCCTCAGCTACTACGCTCATGGTTC
P22 GGGAAATACCCTGGTGTGGCAAAGCTTGAAGGAAATACTAA
460 470 480 490
ST64T ATCCAAGCGAGCAATTGCATCCTACTATCACGCGACTGCAA
P22 GGCAAAGGTACTGCAAGTGCTCGCAACATTTGCTTATGCGG
500 510 520 530
ST64T AGCCACGCAAGATGTGTGGACGGGGTGGCGATGGATGGAGA
P22 ATTATTGCCGTAGTGCTGCGACGCCGGGAGCAAGATGCAGG
540 550 560 570
ST64T AAACCTTCACTGGCAACCTGTAGAAATGAAATTGACGACAT
P22 GATTGCCACGGTACAGGACGGGCAGTTGATATAGCCAAAAC
580 590 600 610
ST64T CCTGAAAGCGTCATTATTTGTTTTATACCAGCCAATGCAAA
P22 AGAGCAGTGGGGGGAGAGTTGTTGAGAAGGTGTGCGGAAGAT
620 630 640 650
ST64T ATGCTTTCAAAATGCGTAAACGTGTTGAGAAAGTTAAGCAT
P223 GCAAGGGCTTCGGCTGTTCAAAGGCGCCGGCAAGCGCAGCA
660 670 680 690
ST64T GTTGCTGTTAAAAGTCTTGACATGCAATTATCCATTTAG--
P22 TATCTCTCCGTAACGATGCTAATCCCAAACCTTACCCAACC
700 710 720 730
ST64T ----------
P22 CACCTGGTCACGCACTGTTAAGCCGCTGTATGACGCTTTGG

Figure 6.5

Illustration of sequence similarities showing the genes responsible for homologous recombination in P22 and the corresponding putative genes of ST64T. The shaded areas show the high sequence similarity between the two phages and the sharp transition of no similarity in gene *erf* (indicated by the un–filled areas).

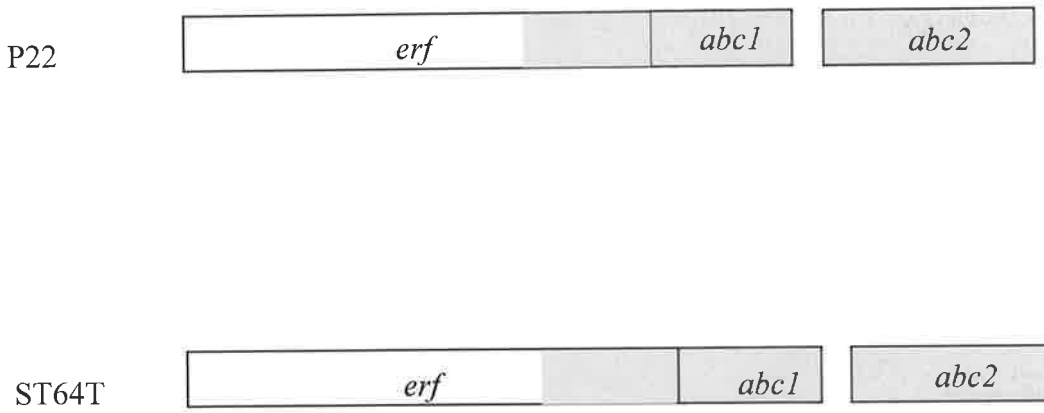

P22
erf
abc1
abc2
ST64T
erf
abc1
abc2

Figure 6.6

A diagram showing the different action of several endolysins of bacteriophages on the bacterial cell resulting in liberation of bacteriophages. The putative endolysins of ST64T and PS3 showed some sequence identity with the endolysin of phage A118 of *L. monocytogenes* which acts as a L–alanine–D–glutamate peptidase (Loessner, *et al.*, 2000). The point of action of different endolysins are indicated in the arrowed boxes. The figure was adapted from Young (1992).

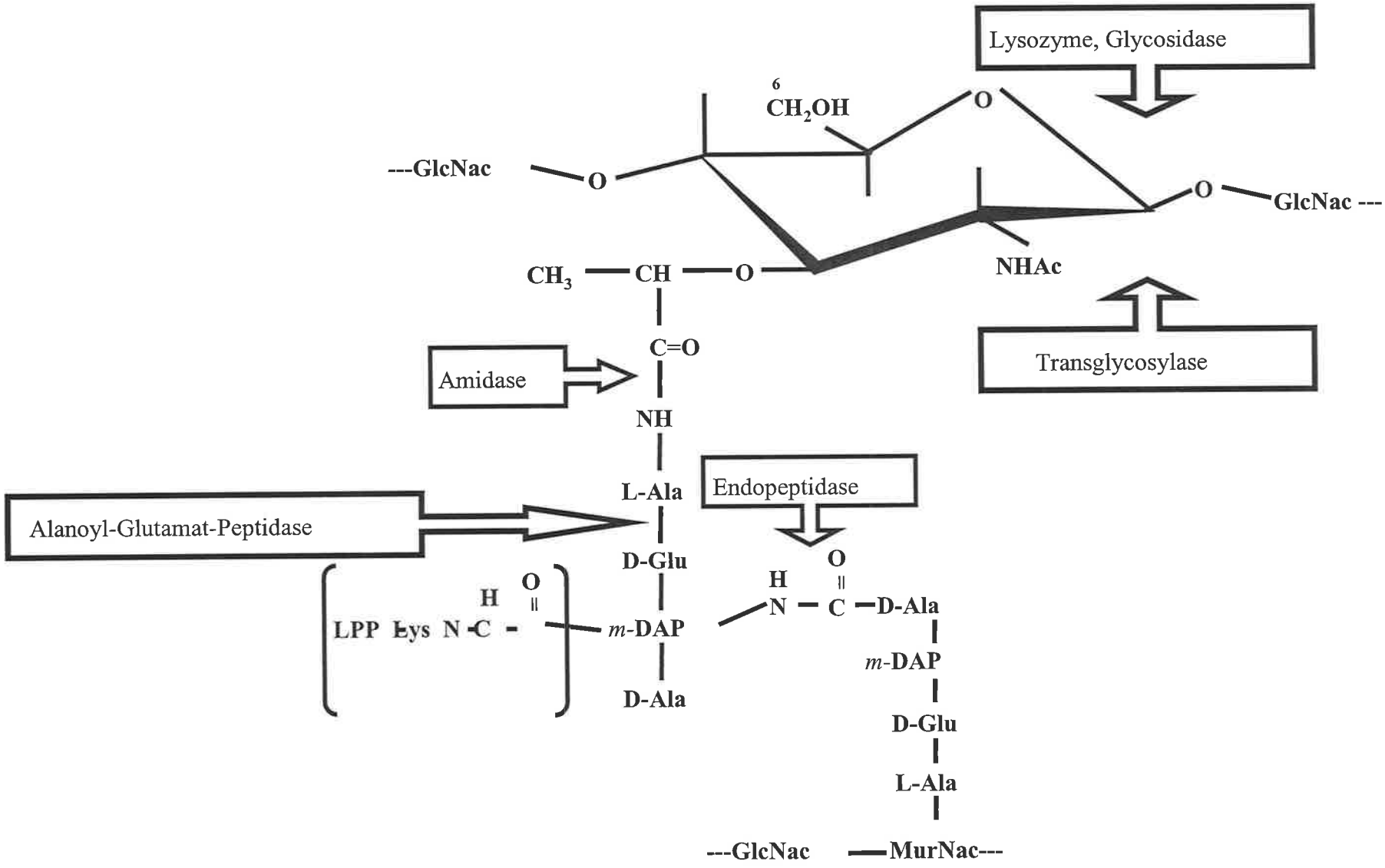

Lysozyme, Glycosidase
6
CH2OH
O
---GlcNac
O
O
GlcNac ---
NHAc
CH3
CH
O
C=O
Amidase
Transglycosylase
NH
L-Ala
Endopeptidase
Alanoyl-Glutamat-Peptidase
D-Glu
O
H
LPP Lys N C
m-DAP
H
O
N
C
D-Ala
m-DAP
D-Ala
D-Glu
L-Ala
---GlcNac
MurNac---

Figure 6.7

A schematic diagram of the lysis cassettes of different lambdoid phages, indicating the different combinations of holin and endolysin modules used by different phages (Casjens, *et al.*, 1992), with the last of the lysis genes (P22 *15* or λ *Rz*) showing conservation in all phages shown. Different colours are used to indicate the relationship of lysis modules used by different phages. The figure also suggests that both ST64T and ST64B (shown in bold) probably use different combinations of the holin and endolysin modules.

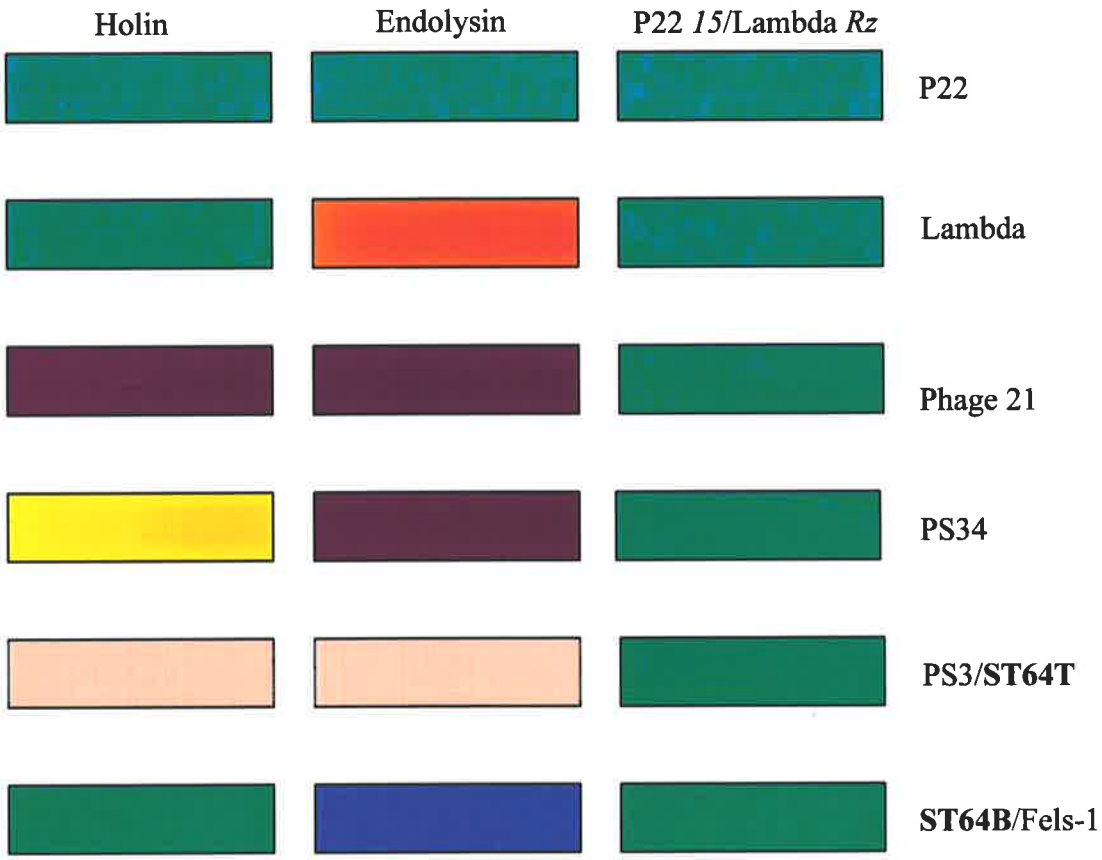
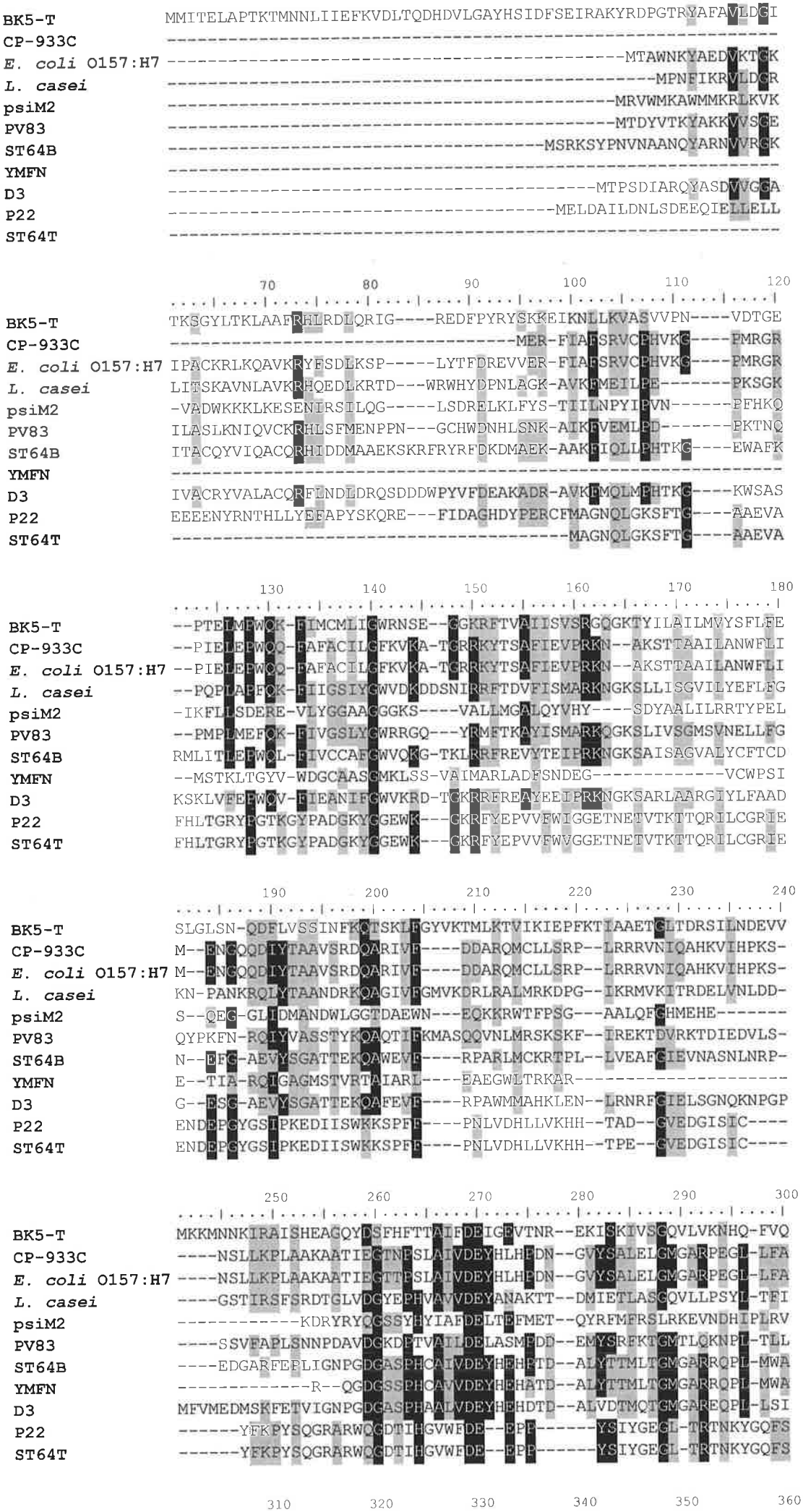
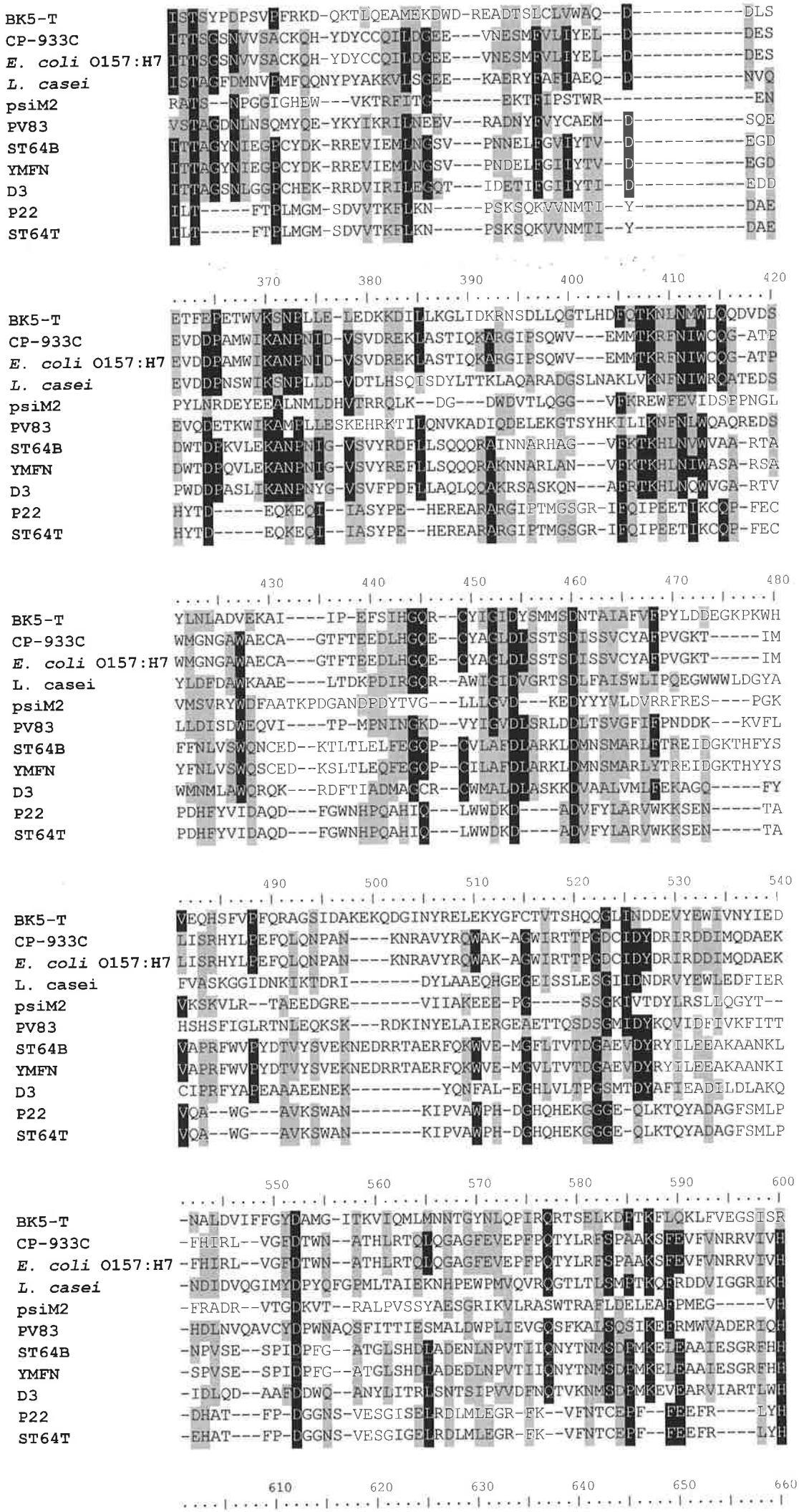

Holin
Endolysin
P22 *15*/Lambda *Rz*
P22
Lambda
Phage 21
PS34
PS3/**ST64T**
ST64B/Fels-1

Figure 6.8

Comparison of the putative and actual large subunit terminases from phages and prophages of diverse bacteria using clustalW at BioEdit version 4.8.10 (Hall, 1999). Identical and similar residues are indicated by black and grey areas, respectively. Dashes indicate absence of amino acid residues in that region of the protein sequence. The accession numbers for the proteins showed in this alignment are AAK56802 (*Lactococcus lactis* phage BK5–T), NP_287339 (*E. coli* prophage CP–933C), NP_309625 (phage–like terminase in *E. coli* O157:H7), CAB63682 (*Lactococcus casei* phage A2), NP_061628 (*Staphylococcus aureus* phage PV83), P75978 (*E. coli* hypothetical YMFN protein), AAD38954 (*Pseudomonas aeruginosa* phage D3); AAF75044 (*Salmonella* phage P22) and AF65411(*Methanobacterium* phage psiM2).


BK5-T
CP-933C
E. coli O157:H7
L. casei
psiM2
PV83
ST64B
YMFN
D3
P22
ST64T
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360


BK5-T
CP-933C
E. coli O157:H7
L. casei
psiM2
PV83
ST64B
YMFN
D3
P22
ST64T
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660

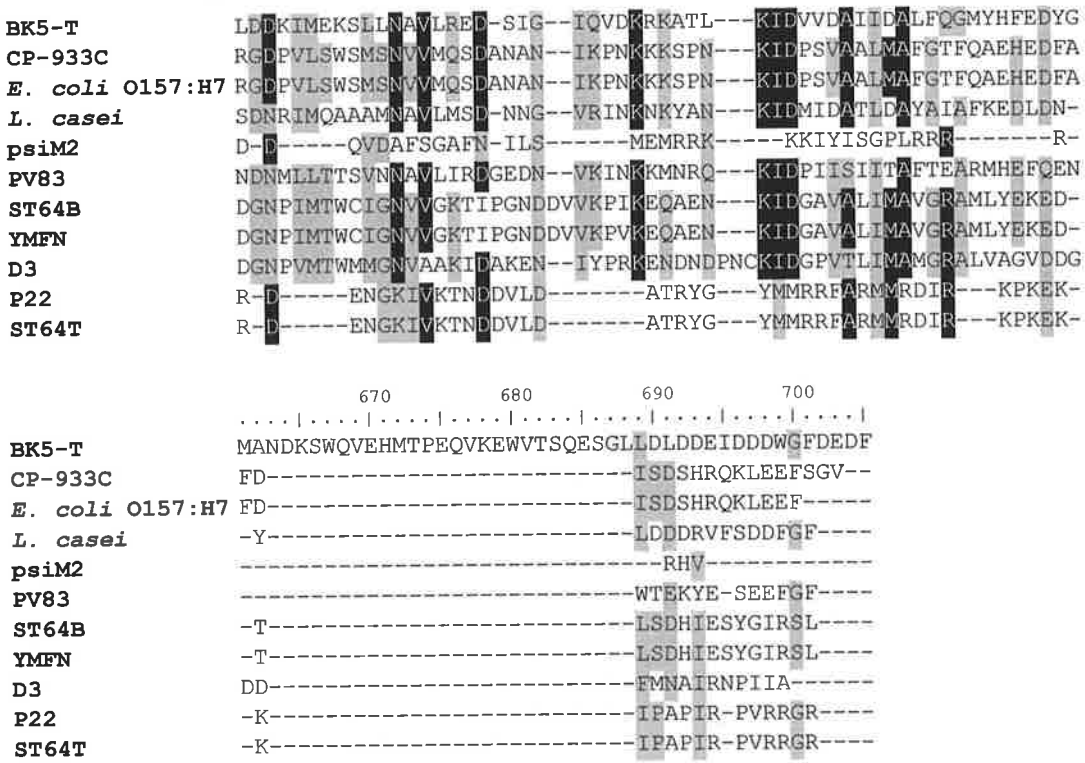

BK5-T
CP-933C
E. coli O157:H7
L. casei
psiM2
PV83
ST64B
YMFN
D3
P22
ST64T
670
680
690
700

Figure 6.9

A phylogenetic dendrogram of bacteriophage putative and actual large subunit terminases based upon the maximum parsimony analysis (MP). The sequence alignment used to create the dendrogram is shown in Figure 6.8. The putative *Methanobacterium* phage psiM2 (accession no. AF65411) was used to root the tree. The accession numbers for the proteins used to create this dendrogram are as shown in Figure 6.8. Branch numbers are shown in boldface. Bootstrap values are indicated on branches. The results indicated that the terminase large subunit of ST64B is closely related to the unknown protein YMFN in *E. coli* and to the terminase large subunit of *P. aeruginosa* phage D3. However, the dendrogram clearly indicates that ST64B is not related to ST64T with regard to the terminase large subunit.

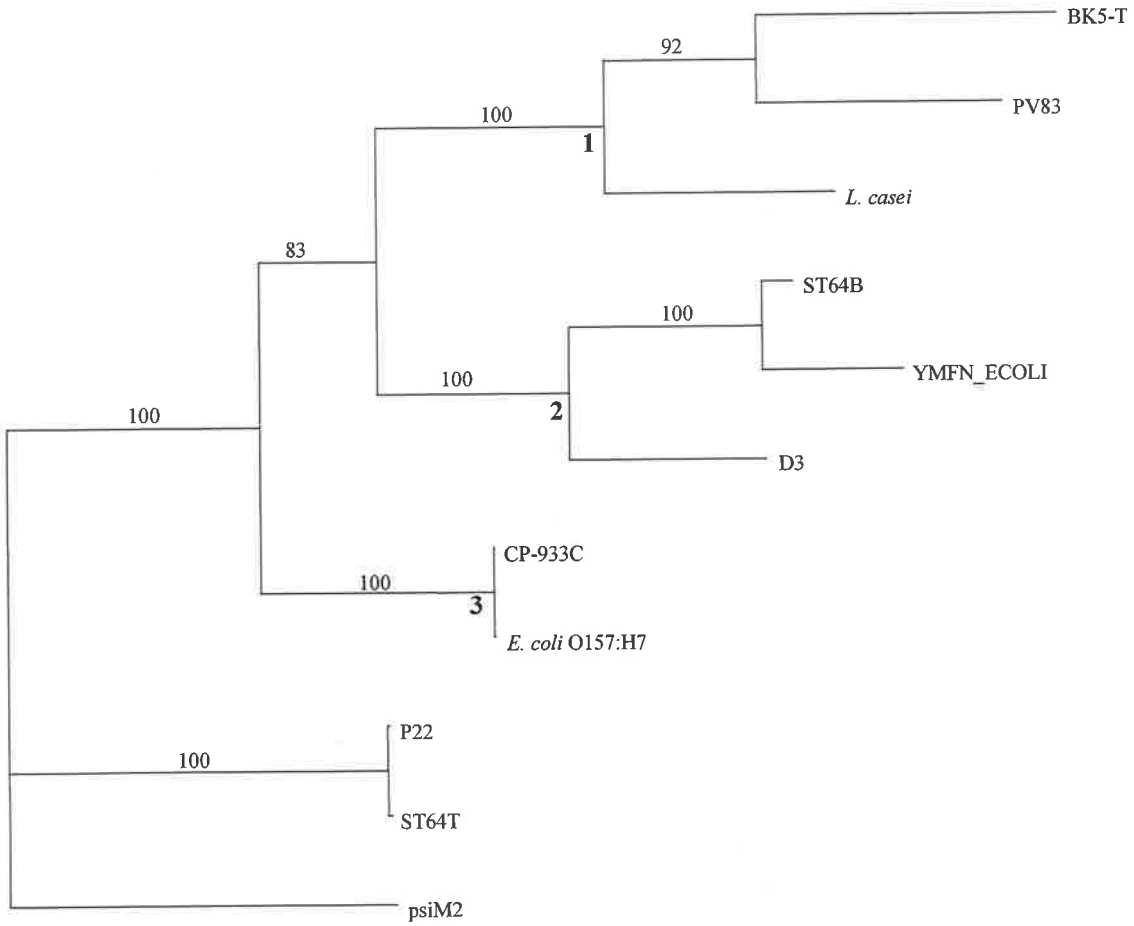
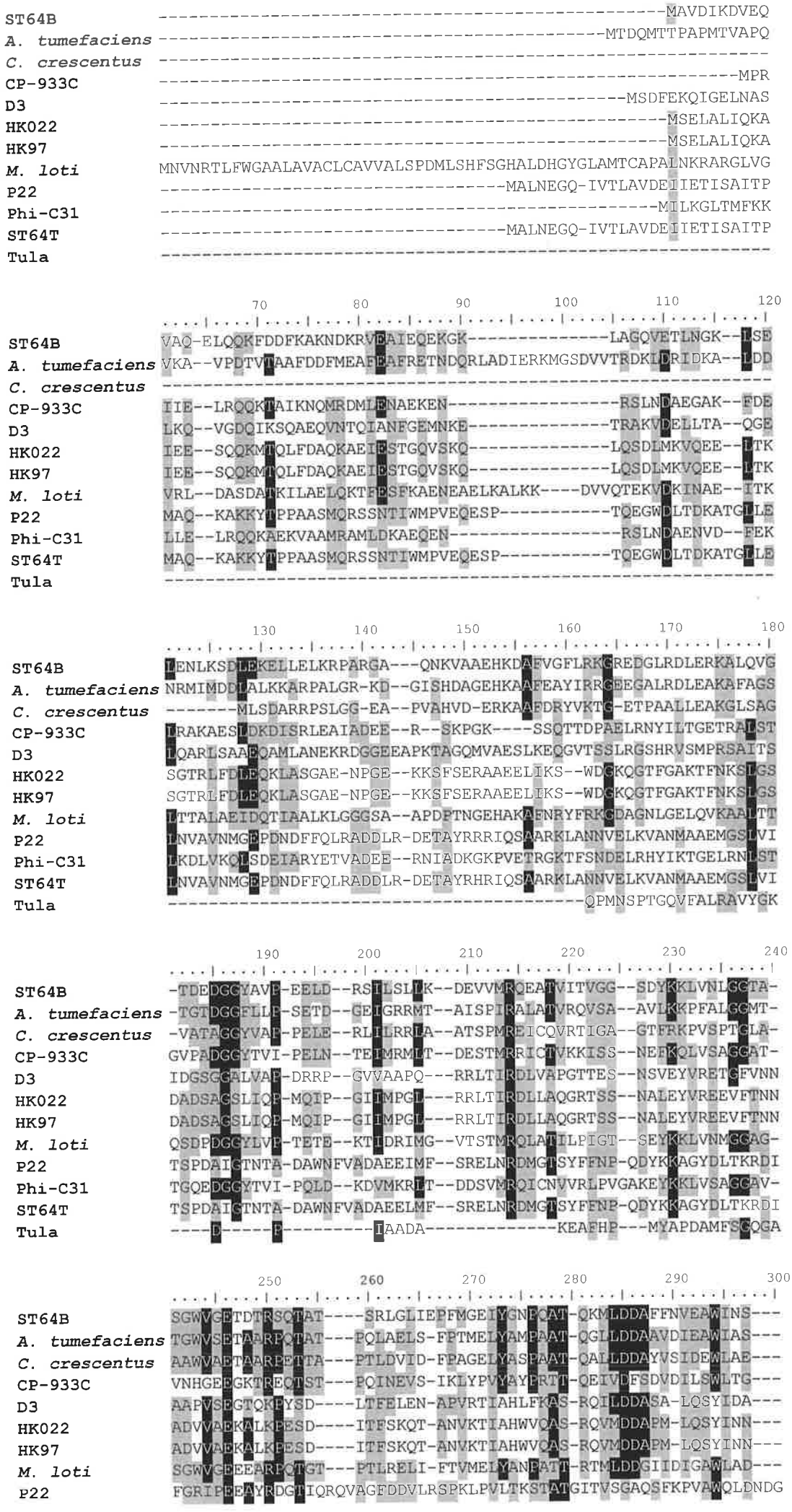

BK5-T
92
PV83
100
1
L. casei
83
ST64B
100
YMFN_ECOLI
100
2
100
D3
CP-933C
100
3
E. coli O157:H7
P22
100
ST64T
psiM2

Figure 6.10

Comparison of the putative and actual major head proteins from phages and prophages of diverse bacteria using ClustalW at BioEdit version 4.8.10 (Hall, 1999). The accession numbers of proteins used in this alignment are AAK86763 (*A. tumefaciens* phage–like protein), AAK24747 (*C. crescentus* phage–like protein), AAG55944 (*E. coli* phage CP–933C), AAD38957 (*P. aeruginosa* phage D3), AAF30354 (*E. coli* phage HK022), AAF31098 (E. coli phage HK97), NP_108602 (*M. loti* phage–like protein), NP_047927 (*Streptomyces* ϕC31), AAF75047 (*Salmonella* phage P22) and AF221994 (*E. coli* phage Tula). Identical and similar residues are indicated by black and grey areas, respectively. Dashes indicate absence of amino acid residues in that region of the protein sequence.


ST64B
A. tumefaciens
C. crescentus
CP-933C
D3
HK022
HK97
M. loti
P22
Phi-C31
ST64T
Tula
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300

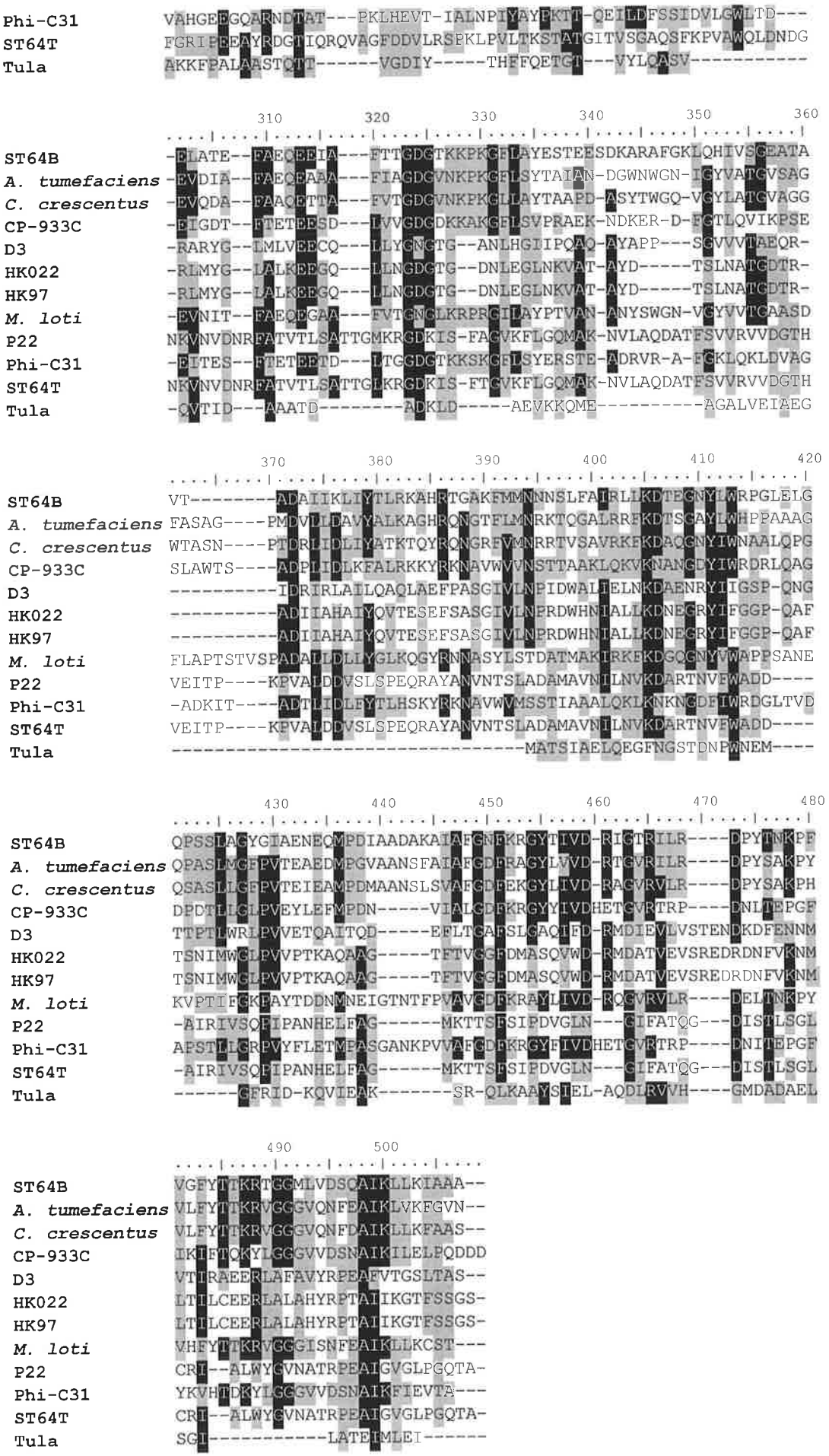

Phi-C31
ST64T
Tula
310
320
330
340
350
360
ST64B
A. tumefaciens
C. crescentus
CP-933C
D3
HK022
HK97
M. loti
P22
Phi-C31
ST64T
Tula
370
380
390
400
410
420
ST64B
A. tumefaciens
C. crescentus
CP-933C
D3
HK022
HK97
M. loti
P22
Phi-C31
ST64T
Tula
430
440
450
460
470
480
ST64B
A. tumefaciens
C. crescentus
CP-933C
D3
HK022
HK97
M. loti
P22
Phi-C31
ST64T
Tula
490
500
ST64B
A. tumefaciens
C. crescentus
CP-933C
D3
HK022
HK97
M. loti
P22
Phi-C31
ST64T
Tula

Figure 6.11

A phylogenetic dendrogram of bacteriophage putative and actual major head proteins of phages and prophages based upon the maximum parsimony analysis (MP). The sequence alignment used to create the dendrogram is shown in Figure 6.10. Phage Tula (accession no. AF221994), a T4–type coliphage was used to root the tree. Branch numbers are shown in boldface. Bootstrap values are indicated on branches. Bacterial names are given for phage related proteins isolated from these species. The accession numbers for the proteins used to create this tree are as shown in Figure 6.10. The results indicated that the putative major head protein of ST64B is closely related to those of the prophages in *A. tumefaciens, C. crescentus* and *M. loti* at a bootstrap confidence level of 95% pseudo replicates and distantly related to HK97, HK022 and D3. However, the dendrogram clearly showed that ST64B is not related to ST64T with regard to the major head protein.

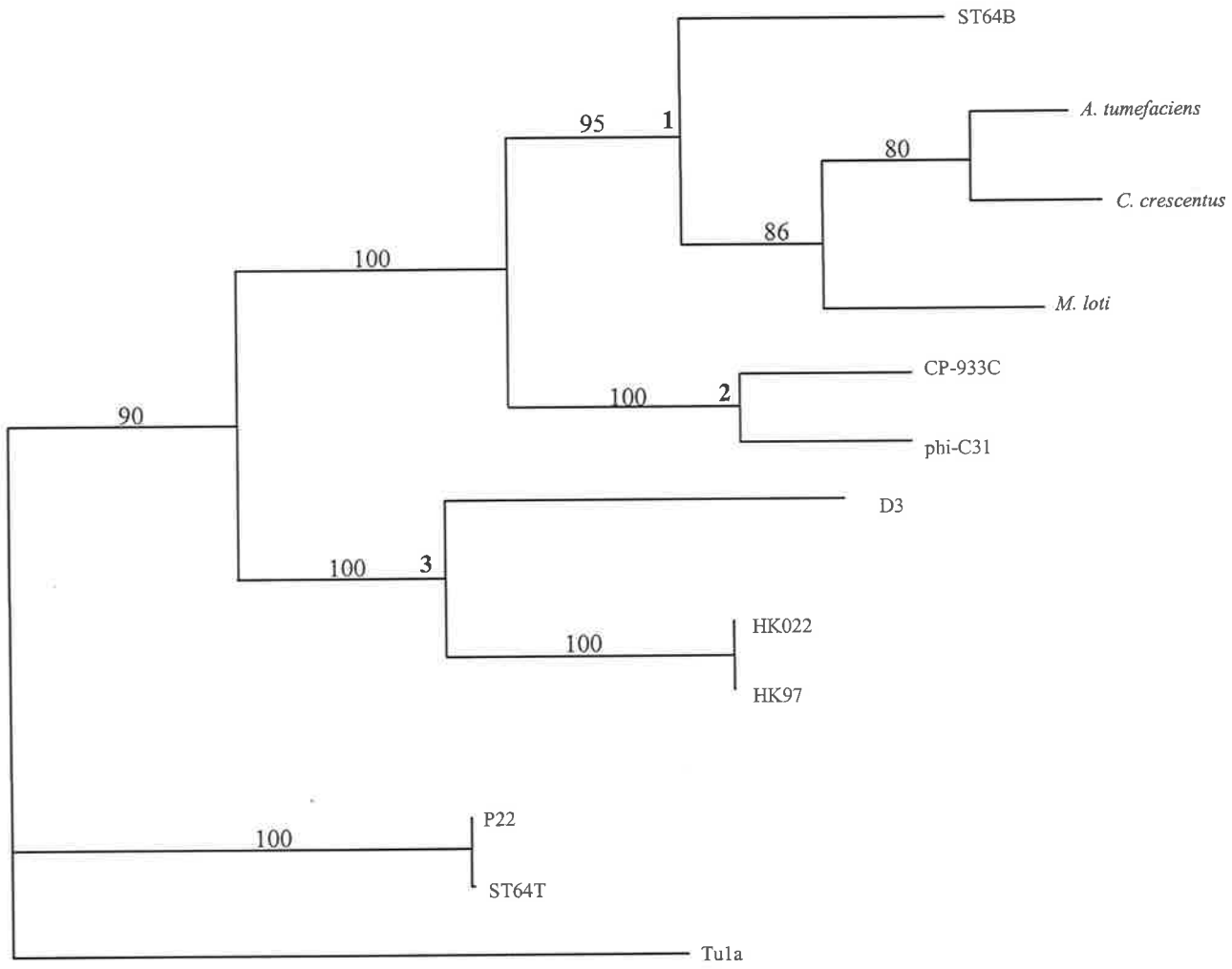

ST64B
A. tumefaciens
C. crescentus
M. loti
CP-933C
phi-C31
D3
HK022
HK97
P22
ST64T
Tula
95
1
80
86
100
100
2
90
100
3
100
100

Table 6.1 ST64B putative genes with similarity to *Escherichia coli* genes

ORF no.	Strand	Position (nt)	Putative function	Related sequences (BlastP)	BlastP e–value	% identity
SB2	+	507 – 2240	Terminase large subunit	*E. coli* YMFN; P75978	0.0	93
SB3	+	2252 – 2434	Unknown	*E. coli* YMFR; P75979	3e–13	96
SB22	+	15820 – 16341	Unknown	*E. coli* HI1520; BAA35978	3e–48	57
SB23	+	16344 – 16931	Unknown	*E. coli* 21.6 kd	4e–59 3e–27 3e–07	59 54 28
SB26	-	19334 – 20341	Unknown	*S.* Typhimurium LT2; AAF33527 *E. coli* 0157:H7 Z4328; AAG58112	e–100 1e–93	58 56
SB35	-	25305 – 25844	Unknown	*E. coli* b2361; F65009 *E. coli* 20.3 kDa; P76514	2e–60 5e–60	64 64
SB36	-	25981 – 26808	Unknown	*E. coli* 30.4 kDa INTC-DSDC; P76513	5e–74	53
SB37	-	26866 – 27237	Unknown	*E. coli* 16.5 kDa INTC-DSDC; P76512	4e–25	50
SB42	+	31025 – 31843	Unknown	*E. coli*; BAB34496 *E. coli*; BAA16223	5e–26 5e–05	37 83
SB43	+	31803 – 32327	Unknown	*E. coli* 18.8 kDa b2357; P76510 E. coli 0157:H7; BAB34011	5e–51 0.093	69 41
SB44	+	32327 – 33220	Unknown/ DNA methyltransferase	*E.coli* b2356; A65009 *E. coli* 10.3 kDa; P76509	1e-35 3e–35	75 75
SB47	+	34527 – 35246	Unknown	*E. coli* 0157:H7; BAB35618	4e–57	43
SB48	+	35221 – 35481	Unknown	*E. coli* b1560; C64911	2e–39	91

Chapter 7

Summary Discussion

Since 1960, *S.* Typhimurium has been the most common serovar isolated from humans in Australia (Australian *Salmonella* Reference Laboratory, Institute of Medical and Veterinary Science, Adelaide, Annual Reports, 1960 – 2000). Each year, about 6000 locally acquired *S.* Typhimurium cases are notified (Jeoffreys, *et al.*, 2001). At the time of writing this thesis, the most common *S.* Typhimurium definitive phage types causing infection in humans and chickens included 9, 64 and 135. Induction of temperate phages from DT 64 with mitomycin C followed by caesium chloride gradient purification, resulted in separation of two phages, ST64T from ST64B. However, only one phage band was evident when phages from DTs 9 and 135 were similarly induced and separated on a CsCl gradient. Restriction endonuclease digestion of these phage genomes with *Sma*I, resulted in a similar banding pattern to ST64B.

Lysogenisation of different *Salmonella* isolates with phage ST64T resulted in phage type conversion. *S.* Typhimurium DTs 9, 135 and 41 were converted to DTs 64, 16 and 29, respectively. *S.* Heidelberg PT 1 was converted to PT 4 and *S.* Enteritidis PT 26 was converted to PT 6a. These convertants were confirmed by AFLP analysis, which was capable of distinguishing the parental strains from the phage type convertants. It has been postulated that in some cases phage type conversion is the result of recombination of the incoming phage with the pre–existing prophages. Moreover, it has been shown that most of the *Salmonella* isolates are lysogenic for at least one prophage (Schicklmaier, *et al.*, 1998). It is documented that acquisition of new base plate modules may lead to a different host range for the recombinant phage. In addition, exchange of immunity regions between the infecting phage and the resident prophage within the strain to be tested, may also influence phage type (Schmieger, 1999). These results, in addition to the published phage type conversion studies (Baggesen, *et al.*, 1997, Chart, *et al.*, 1989, Harvey, *et al.*, 1993,

Hickman-Brenner, *et al.*, 1991, Rankin and Platt, 1995), raise the question of the stability of phage types in natural settings and the possibility of phage type conversion occurring during an outbreak scenario. It is important that phage typing should be complemented by newer molecular techniques such as PFGE and AFLP.

It was determined that the ST64B genome was widespread in most *S.* Typhimurium definitive phage types tested whereas the ST64T genome was limited to DTs 64 and 29. This interesting observation led to the sequencing of both the ST64T and ST64B genomes to facilitate analysis of their genomes for comparative studies with each other and with other phage genomes. The presence of ST64B in many isolates was of potential significance since recently, it was shown that phages Gifsy–1, Gifsy–2 and sopEϕ encode genes that potentially contribute to the virulence of *S.* Typhimurium (Figueroa-Bossi and Bossi, 1999, Mirold, *et al.*, 1999). Although, small, incomplete fragments of virulence associated genes were found in the genome of ST64B, it was unlikely that any functional virulence factors were produced. However, it is envisaged that the ST64B lysogenic hosts could be infected by other phages or defective phages carrying incomplete virulence factors that could lead to marker rescue and thereby reconstitute functional virulence genes due to recombination.

Comparative analysis of ST64T and P22 revealed striking similarities. Morphologically, the two phages are very similar and belong to the family, *Podoviridae*. A member of this family, P22, is capable of generalised transduction as first shown by Zinder and Lederberg, (1952). Similarly, in this study, ST64T was also shown to be capable of transduction of widely separated markers on the *S.* Typhimurium chromosome. Furthermore, the genomic architecture of ST64T is similar to that of P22 with 40% of ST64T putative genes exhibiting >90% sequence identity to P22 genes. The high similarity of ST64T and P22 tail spike gene *9* and the genes involved in O–antigen conversion (*gtrABC*), together with the presence of the putative promoter upstream of gene *gtrA*, suggest strongly that ST64T is also capable of mediating serotype conversion. Both

sequence data and biological tests showed that ST64T is heteroimmune to P22. The immunity region of ST64T is similar to that of phage L. However, unlike phage L, ST64T encodes a putative gene *mnt* at a similar map position and orientation as found in P22. Nevertheless, genes *ant* and *arc* were not found in that region of the ST64T map. It is possibile that imprecise excision could have occurred by illegitimate recombination within the *immI* region of a progenitor phage, leading to incorporation of the repressor gene, *mnt* in ST64T. It is also a possibility that in the past, the ST64T genome encoded the complete *immI* region but subsequently *ant* and *arc* genes were deleted. The presence of the *mnt* gene in ST64T may account for exclusion of heteroimmune phage P22 from a ST64T lysogen.

Although P22 has one *eaa* gene, two putative genes *eaa* (*eaa1* and *eaa2*) were observed in ST64T. Both of these putative genes had similarities to the *eaa* gene of P22. This is probably the result of an insertion in the *eaa* gene of the ST64T genome, leading to the two putative genes, both of which have good ribosomal binding sites upstream of their predicted start codons. It is speculative whether these two genes are expressed separately in ST64T. Nevertheless, since the function of the *eaa* gene in P22 is unknown, the observation of two putative genes in ST64T in contrast to one gene in P22 may not affect the ST64T phage growth. In summary, ST64T was shown to be a genetic mosaic with different modules or genes from diverse phages. As is the case for P22 (Vander Byl and Kropinski, 2000), ST64T has shown a significant degree of protein similarity with proteins of different bacteriophage families including *Podoviridae* (*AP*SE–1, L, 933W, H–19B, LP7, PS3, PS34 ES18 and SfV), *Siphoviridae* (λ, 21, HK022, 434 and VT2–Sa), *Myoviridae* (SfII) and *Inoviridae* (SfX).

The genomic architecture of ST64B resembles that of phage λ, with many putative ORFs encoding the inferred tail gene products with sequence similarity to the tail genes of phage Mu. Electron microscopy however showed that ST64B particles appeared to have only short tails or lacked tails. By contrast, phages λ and Mu are known to have long tails.

As explained in chapter 5, it is possible that the phages observed with short tails were ST64T contaminants, since both ST64T and ST64B were isolated from a *S.* Typhimurium DT 64 strain, and although well–separated when run on CsCl gradient, contamination of the two phage particles could not be ruled out. Since ST64T could propagate on many strains, it was possible to obtain pure particles for transduction, immunity and phage type conversion studies. However, ST64B could not be produced without possible ST64T contamination.

It is likely that the putative tail genes of ST64B may be non–functional, based upon several lines of evidence: (i) ST64B could not be propagated on any strain tested, (ii) many putative tail genes were observed in the genome of ST64B unlike a single tail gene observed in ST64T, (iii) morphologically, no long tails could be observed using TEM and those particles with short tails are likely to be ST64T contaminants. It is possible based upon these observations that ST64B is incapable of infection and may require helper phages to propagate. However, more studies are required to confirm this proposition.

This study has shown that the large terminase subunit of ST64B is unrelated to those of P22 and ST64T. The small and large terminase subunits of ST64B were shown to be similar on the protein level to the small and large subunits of the 3' *cos*–binding site phages. More studies are required to confirm whether ST64B has the 3' *cos* site termini, since no *cos* sites or evidence of under-represented band(s) were discovered during sequencing of the ST64B genome.

The size of the genomes of ST64B (40,149 bp) and ST64T (40,679 bp) correspond well with the reported sizes of the siphoviral and podoviral genomes (Ackermann and DuBow, 1995). In addition, the lambdoid phage group has been shown to have dsDNA chromosomes in the range of 40,000 – 50,000 bp (Casjens, *et al.*, 1992). Similar to λ and P22 (Casjens, *et al.*, 1992), the ST64B and ST64T putative head and tail genes have no recognisable sequence similarity, with the number of tail genes in both phages being completely distinct.

Using phage genome comparative studies, Hendrix and coworkers observed that phages and prophages associated with phylogenetically distant hosts shared a low level of sequence identity, which argues against direct horizontal genetic exchange in recent evolutionary time (Hendrix, *et al.*, 1999). Furthermore, for a phage sequence to travel to a phylogenetically distant host requires a long journey of many steps through a phylogenetic space (Hendrix, *et al.*, 1999). This is probably true for ST64B, which shared low sequence identity with phages from phylogenetically distant bacteria. Both ST64T and ST64B appear to have had access to a large common gene pool, with ST64B probably having genetic exchange beyond the confines of a local neighbourhood.

Conclusion

The discovery, isolation and characterisation of the two *S.* Typhimurium phages ST64T and ST64B is a contribution to the existing data of the dsDNA phage genomes. Inclusion of their genome sequences into the database will enhance the study and illustration of population genetics, evolution and epidemiology of the dsDNA phage genomes. Furthermore, the discovery that ST64T, which is capable of mediating generalised transduction, is carried by *S.* Typhimurium DT 64, a common human isolate, suggests that ST64T may be a vehicle for the transfer of drug resistance as well as virulence genes. Finally, ST64T has been shown to be a fully functional phage, capable of phage type conversion, generalised transduction and O–antigen serotype conversion. It is possible that ST64B lacks a functional tail and is therefore defective, however, more studies are required to confirm this hypothesis.

Bibliography

Aarts, H. J., L. A. van Lith, and J. Keijer. 1998. High-resolution genotyping of *Salmonella* strains by AFLP-fingerprinting. Lett Appl Microbiol. **26:**131-135.

Acheson, D. W., J. Reidl, X. Zhang, G. T. Keusch, J. J. Mekalanos, and M. K. Waldor. 1998. In vivo transduction with shiga toxin 1-encoding phage. Infect Immun. **66:**4496-4498.

Ackermann, H. W. 1998. Tailed bacteriophages: the order caudovirales. Adv Virus Res. **51:**135-201.

Ackermann, H.-W. 1969. Bactériophages-propriétés et premières étapes d'une classification. Pathol Biol. **17:**1003-1024.

Ackermann, H. W., and M. S. DuBow. 1987a. Viruses of Prokaryotes, vol. Volume I. CRC Press, Inc, Florida.

Ackermann, H. W., and M. S. DuBow. 1995. Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag.

Ackermann, H. W., and A. Eisenstark. 1974. The present state of phage taxonomy. Intervirology. **3:**201-219.

Ackermann, H. W., and M. Gersham. 1992. Morphology of phages of a general *Salmonella* typing set. Res Virol. **143:**303-310.

Adams, J. M., and S. E. Luria. 1958. Transduction by bacteriophage P1: abnormal phage function of the transducing particles. Proc Natl Acad Sci. **44:**621, 637.

Adams, M. B., A. R. Brown, and S. Casjens. 1985. Bacteriophage P22 tail protein gene expression. J Virol. **53:**180-184.

Adams, M. H. 1959. Bacteriophages. Interscience, New York.

Adhikari, P., G. Allison, B. Whittle, and N. K. Verma. 1999. Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the *Shigella flexneri* chromosome. J Bacteriol. **181:**4711-4718.

Admiraal, G., and J. E. Mellema. 1976. The structure of the contractile sheath of bacteriophage Mu. J Ultrastruct Res. **56:**48-64.

Allison, G. E., and N. K. Verma. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. Trends Microbiol. **8:**17-23.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol. **215:**403-410.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database serach programs. Nucleic Acids Res. **25:**3389-4022.

Amann, R., W. Ludwig, and K. H. Schleifer. 1988. Beta-subunit of ATP-synthase: a useful marker for studying the phylogenetic relationship of eubacteria. J Gen Microbiol. **134:**2815-2821.

Anderson, E. S. 1959. Use of phages in epidemiological studies, pp. 395-420, Bacteriophages. Interscience Publishers, New York.

Anderson, E. S., and A. Felix. 1953. The Vi-type determining phages carried by *Salmonella typhi*. J Gen Microbiol. **9:**65.

Anderson, E. S., L. R. Ward, M. J. Saxe, and J. D. de Sa. 1977. Bacteriophage-typing designations of *Salmonela typhimurium*. J Hyg (Lond). **78:**297-300.

Anderson, E. S., and R. E. E. O. Williams. 1956. Bacteriophage typing of enteric pathogens and staphylococci and its use in epidemiology. J Clin Pathol. **9:**94-127.

Andrews, R., J. Feldheim, R. Givney, J. Carman, C. Murray, M. Beers, J. Lanser, M. Nguyen, S. Cameron, and R. Hall. 1997. Concurrent outbreaks of *Salmonella* Typhimurium in South Australia. Commun Dis Intell. **21:**61-62.

Angulo, F. J. 1998. *Salmonella enteritidis* infections in the United States. J AM Vet Med Assoc. **12:**1729-1731.

Anilionis, A., and A. Riley. 1980. Conservation and variation of nucleotide sequences within related bacterial genomes: *Escherichia coli* strains. J Bacteriol. **143:**355-365.

Arbeit, R. D., M. Arthur, R. Dunn, C. Kim, R. K. Selander, and R. Goldstein. 1990. Resolution of recent evolutionary divergence among *Escherichia coli* from related lineages: the application of pulsed-field electrophoresis to molecular epidemiology. J Infect Dis. **161:**230-235.

Backhaus, H., and J. B. Petri. 1984. Sequence analysis of a region from the early right operon in phage P22 including the replication genes *18* and *12*. Gene. **32:**289-303.

Baggesen, D. L., H. C. Wegener, and M. Madsen. 1997. Correlation of conversion of *Salmonella enterica* serovar enteritidis phage type 1, 4, or 6 to phage type 7 with loss of lipopolysaccharide. J Clin Microbiol. **35:**330-333.

Banic, S. 1959. Transduction to penicillin and chloramphenicol resistance in *Salmonella typhimurium*. Genetics. **44:**449-455.

Barksdale, L., and S. B. Arden. 1974. Persisting bacteriophage infections, lysogeny, and phage conversions. Ann Rev Microbiol. **28:**265-299.

Barnhart, B. J., S. H. Cox, and J. H. Jett. 1976. Prophage induction and inactivation by UV light. J Virol. **18:**950-955.

Bateman, A., E. Birney, R. Durbin, S. R. Eddy, R. D. Finn, and E. L. L. Sonnhammer. 1999. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res. **27:**260-262.

Becker, B., N. de la Fuente, M. Gassel, D. Gunther, P. Tavares, R. Lurz, T. A. Trautner, and J. C. Alonso. 1997. Head morphogenesis genes of the *Bacillus subtilis* bacteriophage SPP1. J Mol Biol. **268:**822-839.

Beltran, P., S. A. Plock, N. H. Smith, T. S. Whittam, D. C. Old, and R. K. Selander. 1991. Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. J Gen Microbiol. **137:**601-606.

Bennet, I. L., and E. W. Hook. 1959. Infectious diseases (some aspects of salmonellosis). Ann Rev Med. **10:**1.

Berenstein, D. 1986. Prophage induction by ultra violet light in *Acinetobacter calcoaceti.* J Gen Microbiol. **132:**2633-2636.

Bergan, T., and T. Midtvedt. 1975. Epidemiological markers for *Pseudomonas aeruginosa*. Acta Pathol Microbiol Scand [B]. **83:**1-9.

Bertani, G. 1953. Infections bactériophagiques secondaires des bactéries lysogènes. Ann Inst Pasteur (Paris). **84:**409, 447, 468.

Bertani, L. E., and G. Bertani. 1971. Genetics of P2 and related phages. Adv Genetics. **16:**199-237.

Betley, M. J., and J. J. Mekalanos. 1988. Nucleotide sequence of the type A staphylococcal enterotoxin gene. J Bacteriol. **170:**34-41.

Beutin, L., E. Strauch, and I. Fischer. 1999. Isolation of *Shigella sonnei* lysogenic for a bacteriophage encoding gene for production of shiga toxin. Lancet. **353:**1498.

Bezdek, M., and P. Amati. 1967. Properties of P22 and a related *Salmonella typhimurium* phage. Virology. **31:**272-278.

Binkin, N., Scuderi, G., Novaco, F., Giovanardi, G. L., Paganelli, G., Ferrari, G., Cappelli, O., Ravaglia, L., Zillioli, F., Amadei, V., Magliani, W., Viani, I., Ricco, D., Borrini, B., Magri, M., Alessandrini, A., Bursi,G., Barigazzi, G., Fantasia, M., Feletici, E. and Salmaso, S. 1993. Egg-related *Salmonella enteritidis*, Italy, 1991. Epidemiol Infect. **110:**227-237.

Black, K., W. J. Buikema, and R. Haselkorn. 1995. The *hglK* gene is required for localization of heterocyst-specific glycolipids in the cyanobacterium *Anabaena sp.* strain PCC 7120. J Bacteriol. **177:**6440-6448.

Blahova, J., M. Hupkova, M. Babalova, V. Krcmery, and V. Schafer. 1993. Transduction of resistance to Imipenem, Aztreonam and Ceftazidime in nosocomial strains of Pseudomonas aeruginosa by wild-type phages. Acta Virol. **37:**429-436.

Blahova, J., K. Kralikova, V. Krcmery, and P. Jezek. 2000. Low-Frequency transduction of imipenem resistance and high-frequency transduction of ceftazidime and aztreonam resistance by the bacteriophage AP-151 isolated from a *Pseudomonas aeruginosa* strain. J Chemother. **12:**482-486.

Blahova, J., K. Kralikova, V. Krcmery, Sr., A. Mikovicova, and N. Bartonikova. 1998a. Two high-frequency-transduction phage isolates from lysogenic strains of *Pseudomonas aeruginosa* transducing antibiotic resistance. Acta Virol. **42:**175-179.

Blahova, J., M. Lesicka-Hupkova, K. Kralikova, V. Krcmery, Sr., T. Krcmeryova, and K. Kubonova. 1998b. Further occurrence of extended-spectrum beta-lactamase-producing *Salmonella enteritidis*. J Chemother. **10:**291-294.

Bläsi, U., C.-Y. Chang, M. T. Zagotta, K. Nam, and R. Young. 1990. The lethal lambda *S* gene encodes its own inhibitor. EMBO J. **9:**981-989.

Bläsi, U., K. Nam, D. Hartz, L. Gold, and R. Young. 1989. Dual translational initiation sites control function of the lambda *S* gene. EMBO J. **8:**3501-3510.

Bläsi, U., and R. Young. 1996. Two beginnings for a single purpose: the dual-start holins in the regulation of phage lysis. Mol Microbiol. **21:**675-682.

Blears, M. J., N. J. Pokorny, R. A. Carreno, S. Chen, S. A. De Grandis, H. Lee, and J. T. Trevors. 2000. DNA fingerprinting of *Cryptosporidium parvum* isolates using amplified fragment length polymorphism (AFLP). J Parasitol. **86:**838-841.

Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA- specific

loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect Immun. **62:**606-614.

Bopp, C. A., F. W. Brenner, J. G. Wells, and N. A. Stockbine. 1999. *Escherichia*, *Shigella* and *Salmonella*, pp. 459 - 474. *In* P. R. Murray (ed.), Manual of Clinical Microbiology. 7th ed. ASM press. Washington, D. C.

Botstein, D. 1980. A theory of modular evolution for bacteriophages. Ann NY Acad Sci. **354:** 484-491.

Botstein, D., R. K. Chan, and C. H. Waddell. 1972. Genetics of bacteriophage P22. II. Gene order and gene function. Virology. **49:**268-282.

Botstein, D., and I. Herskowitz. 1974. Properties of hybrids between *Salmonella* phage P22 and coliphage lambda. Nature. **251:**584-589.

Botstein, D., and M. J. Matz. 1970. A recombination function essential to the growth of bacteriophage P22. J Mol Biol. **54:**417-440.

Botstein, D., C. H. Waddell, and J. King. 1973. Mechanism of head assembly and DNA encapsulation in *Salmonella* phage P22. J Mol Biol. **80:**669-695.

Bradley, D. E. 1967. Ultrastructure of bacteriophages and bacteriocins. Bacteriol Rev. **31:**230-314.

Brendel, V., G. H. Hamm, and E. N. Trifonov. 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J Biomol Struct Dyn. **3:**705-723.

Brendel, V., and E. N. Trifonov. 1984. A computer algorithm for testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411-4427.

Brenner, F. W., R. G. Villar, J. Angulo, R. Tauxe, and B. Swaminathan. 2000. *Salmonella* nomenclature. J Clin Microbiol. **2000:**2465-2467.

Bronson, M. J., and M. Levine. 1971. Virulent mutants of bacteriophage P22. I. J Virol. **7:**559-568.

Bronson, M. J., and M. Levine. 1972. Virulent mutants of phage P22. II. Virology. **47:**644-655.

Brown, D. J., D. L. Baggesen, D. J. Platt, and J. E. Olsen. 1999. Phage type conversion in *Salmonella enterica* serotype Enteritidis caused by the introduction of a resistance plasmid of incompatibility group X (IncX). Epidemiol Infect. **122:**19-22.

Brown, T. A. 1991. Essential molecular biology: A practical approach, vol. I. IRL Press, Oxford.

Brüssow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of Siphoviridae: insights from dairy phages. Mol Microbiol. **39:**213-223.

Burrows, W. 1959. Textbook of microbiology, 7 ed. The W. B. Saunders Co., Philadelphia, Pa.

Callow, B. R. 1959. A new phage-typing scheme for *Salmonella typhimurium*. J. Hygiene. **57:**346-359.

Campbell, A. 1962. Episomes. Adv Genetics. **11:**110-145.

Campbell, A. 1988. Phage evolution and speciation, pp. 1-14, The Bacteriophages, ed. Plenum, New York.

Campbell, A. 1994. Comparative molecular biology of lambdoid phages. Annu Rev Microbiol. **48:**193-222.

Campbell, A., S. J. Schneider, and B. Song. 1992. Lambdoid phages as elements of bacterial genomes (integrase/phage21/*Escherichia coli* K-12/icd gene). Genetica. **86:**259-267.

Casjens, S., and M. B. Adams. 1985. Posttranscriptional modulation of bacteriophage P22 scaffolding protein gene expression. J Virol. **53:**185-191.

Casjens, S., K. Eppler, R. Parr, and A. R. Poteete. 1989. Nucleotide sequence of the bacteriophage P22 gene 19 to 3 region: identification of a new gene required for lysis. Virology. **171:**588-598.

Casjens, S., G. F. Hatful, and R. Hendrix. 1992. Evolution of dsDNA tailed bacteriophage genomes, pp. 383-397, Seminars in virology. Academic press, London.

Casjens, S., and R. Hendrix. 1988. Control mechanisms in dsDNA bacteriophage assembly, pp. 15-90, The Bacteriophages, vol. I. Plenum Press, New York.

Casman, E. P. 1965. *Staphylococcus* enterotoxin. Ann New York Acad Sci. **128:**124-131.

Chambers, R. M., P. McAdam, D. S. J. D. H, L. R. Ward, and B. Rowe. 1987. A phage typing scheme for *Salmonella virchow*. FEMS Microbiol Lett. **40:**155-157.

Chan, R. K., and D. Botstein. 1976. Specialized transduction by bacteriophage P22 in *Salmonella typhimurium*: genetic and physical structure of the transducing genomes and the prophage attachment site. Genetics. **83:**433-458.

Chan, R. K., D. Botstein, T. Watanabe, and Y. Ogata. 1972. Specialized transduction of tetracycline resistance by7 phage P22 in *Salmonella typhimurium*. II. Properties of a high frequency transducing lysate. Virology. **50:**883-898.

Chart, H., B. Row, E. J. Threlfall, and L. R. Ward. 1989. Conversion of *Salmonella enteritidis* phage type 4 to phage type 7 involves loss of lipopolysaccharide with concomitant loss of virulence. FEMS Microbiol Lett. **51:**37-40.

Cheetham, B. F., and M. E. Katz. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol Microbiol. **18:**201-208.

Chow, L. T., and A. I. Bukhari. 1976. The invertible DNA segments of coliphage Mu and P1 are identical. Virology. **74:**242-248.

Churchward, G., D. Belin, and Y. Nagamine. 1984. A pSC101-derived plasmid which shows no sequence homology to otehr commonly used cloning vectors. Gene. **31:**165-171.

Cianciotto, N. P., J. M. Bangsborg, B. I. Eisensteain, and N. C. Engleberg. 1990. Identification of mip-like gens in the genus *Legionella*. Infect Immun. **58:**2912-2918.

Clark, C. A., J. Beltrame, and P. A. Manning. 1991. The oac gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. Gene. **107:**43-52.

Coetzee, J. N. 1987. Phage ecology, pp. 45-85. Wiley, New York.

Coetzee, J. N., and T. G. Sacks. 1960. Intrastrain transduction in *Proteus mirabilis*. Nature. **185:**621.

Cohen, M. L. and R. V. Tauxe. 1986. Drug resistant *Salmonella* in the United States: An epidemiological perspective. Science. **234:**964-969.

Coleman, D. C., J. P. Arbuthnott, H. M. Pomeroy, and T. H. Birkbeck. 1986. Cloning and expression *in Escherichia coli* and *Staphylococcus aureus* of the beta-lysin determinant from *Staphylococcus aureus*: evidence that bacteriophage conversion of beta-lysin activity is caused by insertional inactivation of the beta-lysin determinant. Microb Pathog. **1:**549-564.

Coleman, D. C., D. J. Sullivan, R. J. Russell, J. P. Arbuthnott, B. F. Carey, and H. M. Pomeroy. 1989. *Staphylococcus aureus* bacteriophages mediating the simultaneous lysogenic conversion of beta-lysin, staphylokinase and enterotoxin A: molecular mechanism of triple conversion. J Gen Microbiol. **135:**1679-1697.

Collazo, C. M., and J. E. Galan. 1997. The invasion-associated type III system of *Salmonella typhimurium* directs the translocation of Sip proteins into the host cell. Mol Microbiol. **24:**747-756.

Collins, M. T. 1986. *Legionella* infections in animals. Isr J Med Sci. **22:**662-673.

Colson, C., and A. van Pel. 1974. DNA restriction and modification systems in *Salmonella*. I. SA and SB, two *Salmonella typhimurium* systems determined by genes with a chromosomal location comparable to that of *Escherichia coli hsd* genes. Mol Gen Genet. **129:**325-327.

Cowden, J. M., D. Chisholm, and M. e. a. O'Mahony. 1989a. Two outbreaks of *Salmonella enteritidis* phage type 4 infection associated with the consumtpion of fresh shell-egg products. Epidemiol Infect. **103:**47-52.

Cowden, J. M., D. Lynch, and C. A. e. a. Joseph . 1989b. Case-control study of infections with *Salmonella enteritidis* phage type 4 in England. Br Me J. **299:**771-773.

Craig, N. L. 1988. The mechanism of conservative site-specific recombination. Annu Rev Genet. **22:**77-106.

Cruickshank, R., J. P. Duguid, B. P. Marmion, R. H. A. Swain. 1975. Medical Microbiology. 12th ed. W. B. Saunders Co., London.

Cue, D., and M. Feiss. 1993. A site required for termination of packaging of the phage lambda chromosome. Proc Natl Acad Sci. **90:**9290-9294.

Cue, D., and M. Feiss. 1997. Genetic evidence that recognition of *cosQ*, the signal for termination of phage lambda DNA packaging, depends on the extent of head filling. Genetics. **147:**7-17.

D'Aoust, J.-V. 1989. *Salmonella*. In Foodborne Bacterial Pathogens, ed. Doyle M. P. Marcel Dekker Inc., New York.

Darwin, K. H., and V. L. Miller. 1999. Molecular basis of the interaction of *Salmonella* with the intestinal mucosa. Clin Microbiol Rev. **12:**405-428.

Davison, J. 1999. Genetic exchange between bacteria in the environment. Plasmid. **42:**73-91.

De Groote, M. A., U. A. Ochsner, M. U. Shiloh, C. Nathan, J. M. McCord, M. C. Dinauer, S. J. Libby, A. Vazquez-Torres, Y. Xu, and F. C. Fang. 1997. Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. Proc Natl Acad Sci U S A. **94:**13997-40001.

De Sa, J. D. H., L. R. Ward, and B. Rowe. 1980. A scheme for the phage typing of *Salmonella hadar*. FEMS Microbiol Lett. **9:**175-177.

DeMarini, D. M., H. G. Brooks, and D. G. Parkes. 1990. Induction of prophage lambda by chlorophenols. Environ Mol Mutagen. **15:**1-9.

DeMarini, D. M., and B. K. Lawrence. 1992. Prophage induction by DNA topoisomerase II poisons and reactive oxygen species: role of DNA breaks. Mutat Res. **267:**1-17.

Demerec, M., and U. Fanu. 1945. Bacteriophage-resistant mutants in *Escherichia coli.* Genetics. **30:**119-136.

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

Desiere, F., C. Mahanivong, A. J. Hillier, P. S. Chandry, B. E. Davidson, and H. Brüssow. 2001. Comparitive genomics of Lactococcal phages: Insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. Virology. **283:**240-252.

Dobrindt, U., and J. Reidl. 2000. Pathogenicity islands and phage conversion: evolutionary aspects of bacterial pathogenesis [In Process Citation]. Int J Med Microbiol. **290:**519-527.

Dower, W. J., J. F. Miller, and C. W. Ragsdale. 1988. High efficiency of transformation of *Escherichia coli* by high voltage electroporation. Nucleic Acids Res. **16:**6127-6145.

Duda, R. L. 1998. Protein chainmail: catenated protein in viral capsids. Cell. **94:**55-60.

Duda, R. L., K. Martincic, and R. W. Hendrix. 1995. Genetic basis of bacteriophage HK97 prohead assembly. J Mol Biol. **247:**636-647.

Duim, B., E. Kruijt, W. J. van Leeuwen, F. van Zijderveld, and J. Wagenaar. 1997. Application of the DNA fingerprinting method AFLP to differentiate strains of *Salmonella*

enteritidis and *Salmonella dublin*, pp. 81-86, Proceedings of the molecular epidemiology of *Campylobacter* and *Salmonella*.

Ebel-Tsipis, J., D. Botstein, and G. E. Fox. 1972. Generalised transduction by bacteriophages P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. J Mol Biol. **71**.

Eddy, S. R., and R. Durbin. 1994. RNA sequence analysis using covariance models. Nucleic Acids Res. **22:**2079-2088.

Edelman, R., and M. M. Levine. 1986. Summary of an international workshop on typhoid fever. Rev Infect Dis. **8:**329-349.

Eiserling, F. A., and E. Boy de la Tour. 1965. Capsomeres and other structures observed on some bacteriophages. Pathol Microbiol. **28:**175-180.

Ellis, E. L., and M. Delbrück. 1939. The growth of bacteriophage. J Gen Physiol. **22:**365.

El-Mabrouk, N., and F. Lisacek. 1996. Very fast identification of RNA motifs in genomic DNA. Application to tRNA search in the yeast genome. J Mol Biol. **264:**46-55.

Espion, D., K. Kaiser, and C. Dambly-Chaudiere. 1983. A third defective lambdoid prophage of *Escherichia coli* K-12 defined by the λ derivative, λ qin111. J Mol Biol. **170:**611.

Ewing, W. H. 1986. Edwards and Ewing's identification of Enterobacteriaceae, 4th ed. Elsvier Science Publishing Co. Inc., New York, N.Y.

Ewing, W. H. 1972. The nomenclature of *Salmonella*, its usage, and definitions for the three species. Can J Microbiolo. **18:**1629-1637.

Farber, J. M. 1996. An introduction to the hows and whys of molecular typing. J Food Protect. **59:**1091-1101.

Farkas-Himsley, H., J. Jessop, and P. Corey. 1977. Development and standardization of cytotoxic micro-assay for detection of enterotoxins: survey of enterotoxins from *Escherichia coli* of infant origin. Microbios. **18:**195-212.

Farmer III, J. J. 1999. Enterobacteriaceae, pp. 442-458, Manual of clinical microbiology, 7th ed. ASM Press, Washington, D.C.

Farmer III, J.J., B. R. Davis, F. W. Hickmann-Brenner, A. McWhorter, G. Huntle, P. Carter, M. A. Asbury, C. Riddle, H. G. Wathern-Grady, C. Ellias, G. R. Fanning, A. G. Steigerwalt, C. M. O'Hara, G. K. Morris, P. B. Smith, and D. J. Brenner, 1985. Biochemical identification of new species and biogroups of Enterobactericeae isolated from clinical specimens. J Clin Microbiol. 21: 46-76.

Farmer III, J. J. I., McWhorter, A. C. and Brenner, D.J. 1984. The *Salmonella*-Arizona group of Enterobacteriaceae: nomenclature, classification, and reporting. Clin Microbiol Newsletter. **6:**63-66.

Farrant, J. L., A. Sansone, J. R. Canvin, M. J. Pallen, P. R. Langford, T. S. Wallis, G. Dougan, and J. S. Kroll. 1997. Bacterial copper- and zinc-cofactored superoxide dismutase contributes to the pathogenesis of systemic salmonellosis. Mol Microbiol. **25:**785-796.

Faubladier, M., and J. P. Bouche. 1994. Division inhibition gene *dicF* of *Escherichia coli* reveals a widespread group of prophage sequences in bacterial genomes. J Bacteriol. **176:**1150-1156.

Feiss, M., R. A. Fischer, M. A. Crayton, and C. Egner. 1977. Packaging of the bacteriophage λ chromosome: effect of chromosome length. Virology. **77:**281-293.

Felix, A., and B. R. Callow. 1943. Typing of paratyphoid B bacilli by means of Vi bacteriophage. Br Med J. **ii:**127.

Fiandt, M., Z. Hradecna, H. A. Lozeron, and W. Szybalski. 1971. Electron micrographic mapping of deletions, inversions, and homologies in the DNAs of coliphages lambda and phi-80, pp. 329-354, The Bacteriophage Lambda. Cold Spring Harbor Laboratory, New York.

Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, F. Molemans, A. Raeymaekers, A. Van den Berghe, G. Volckaert, and M. Ysebaert. 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature. **260:**500-507.

Figueroa-Bossi, N., and L. Bossi. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. Eur J Epidemiol. **15:**395-401.

Finney, M. 1993. Pulsed-field gel electrophoresis, pp. 2.5.9-2.5.17. Current protocols in molecular biology, vol. 1. Greene-Wiley, New York.

Folkertsma, R. T., J. N. A. M. Rouppe van der Voort, and K. E. de Groot. 1996. Gene pool similarities of potato nematode populations assesed by AFLP analysis. Mol plant-Microbe Interact. **9:**47-54.

Fonden, R., U. Fitger, and L. Petterson. 1976. *Salmonella* bacteria in double cream. Nord Vet Med. **28:**385-391.

Franklin, N. C. 1985a. Conservation of genome form but not sequence in the transcription antitermination determinants of bacteriophages lambda, phi-21 and P22. J Mol Biol. **181:**75-84.

Franklin, N. C. 1985b. "N" transcription antitermination proteins of bacteriophage lambda, phi-21 and P22. J Mol Biol. **181:**85-91.

Franklin, N. C. 1992. The carboxy-terminal 14 amino acids of phage lambda N protein are indispensable for transcription termination. J Bacteriol. **174:**8144-8147.

Frech, G., M. Weide-Botjes, E. Nussbeck, W. Rabsch, and S. Schwarz. 1998. Molecular characterization of *Salmonella enterica* subsp. *enterica* serovar Typhimurium DT009 isolates: differentiation of the live vaccine strain Zoosaloral from field isolates. FEMS Microbiol Lett. **167:**263-269.

Friedman, D. I., and M. Gottesman. 1983. Lytic mode of lambda development, pp. 21-51, Lambda II. Cold Spring Harbor Laboratory Press, New York.

Frost, J. A., L. R. Ward, and B. Rowe. 1989. Acquisition of a drug resistance plasmid converts *Salmonella enteritidis* phage type 4 to phage type 24. Epidemiol Infect. **103:**243-248.

Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. Nature. **399:**541-548.

Galan, J. E. 1996. Molecular genetic bases of *Salmonella* entry into host cells. Mol Microbiol. **20:**263-272.

Galyov, E. E., M. W. Wood, R. Rosqvist, P. B. Mullan, P. R. Watson, S. Hedges, and T. S. Wallis. 1997. A secreted effector protein of *Salmonella dublin* is translocated into eukaryotic cells and mediates inflammation and fluid secretion in infected ileal mucosa. Mol Microbiol. **25:**903-912.

Garaizar, J., N. Lopez-Molina, I. Laconcha, D. Lau Baggesen, A. Rementeria, A. Vivanco, A. Audicana, and I. Perales. 2000. Suitability of PCR fingerprinting, infrequent-restriction-site PCR, and pulsed-field gel electrophoresis, combined with computerized gel analysis, in library typing of *Salmonella enterica* serovar Enteritidis. Appl Environ Microbiol. **66:**5273-5281.

Garger, S. J., O. M. Griffith, and L. K. Grill. 1983. Rapid purification of plasmid DNA by a single centrifugation two step caesium chloride-ethidium bromide gradient. Biochem Biophys Res Commun. **117:**835-842.

Garzon, A., D. A. Cano, and J. Casadesus. 1995. Role of Erf recombinase in P22-mediated plasmid transduction. Genetics. **140:**427-434.

Gemski, P., A. D. O'Brien, and J. A. Wohlhieter. 1978. Cellular release of heat-labile enterotoxin of *Escherichia coli* by bacteriophage induction. Infect Immun. **19:**1076-1082.

Gendrel, D., Kombila, M. Beaudoin-Leblevec, G. and Richard-Lenoble, D. 1994. Nontyphoidal *Salmonella* septicemia in Gabonese children infected with *Schistosoma intercalatum*. Clin Infect Dis. **18:**103-105.

Gerner-Smidt, P., L. M. Graves, S. Hunter, and B. Swaminathan. 1998. Computerized analysis of restriction fragment length polymorphism patterns: comparative evaluation of two commercial software packages. J Clin Microbiol. **36:**1318-1323.

Ghisotti, D., R. Chiaramonte, F. Forti, S. Zangrossi, G. Sironi, and G. Deho. 1992. Genetic analysis of the immunity region of phage-plasmid P4. Mol Microbiol. **6:**3405-3413.

Gilakjan, Z. A., and A. M. Kropinski. 1999. Cloning and analysis of the capsid morphogenesis genes of *Pseudomonas aeruginosa* bacteriophage D3: another example of protein chain mail? J Bacteriol. **181:**7221-7227.

Gille, J. J., N. M. Wientjies, M. V. Lafleur, H. Joenje, and J. Retel. 1996. Biological consequences of DNA damage introduced in bacteriophage MS2 DNA by hydrogen-mediated free radical reactions. Carcinogenesis. **17:**5-11.

Ginocchio, C. C., S. B. Olmsted, C. L. Wells, and J. E. Galan. 1994. Contact with epithelial cells induces the formation of surface appendages on *Salmonella typhimurium*. Cell. **76:**717-724.

Giphart-Gassler, M., C. Wijffelman, and J. Reeve. 1981. Structural polypeptides and products of late genes of bacteriophage Mu: Characterization and functional aspects. J Mol Biol. **145:**139-163.

Goodrich, J. A., M. L. Schwartz, and W. R. McClure. 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. **18:**4993-5000.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov. 1989. Two related superfamilies of putative helicase involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:**4713 - 4730.

Gough, M., and M. Levine. 1968. The circulairty of the phage P22 linkage map. Genetics. **58:**161-169.

Griffin, P. M. 1995. *Escherichia coli* O157:H7 and other enterohemorrhagic *Escherichia coli*, pp. 739-761. *In* M. J. Blaser, P. D. Smith, J. I. Ravdin, H. B. Greenberg, and R. L. Guerrant (eds.), Infections of the gastrointestinal tract. Raven press, New York.

Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. Cell. **87:**791-794.

Grundy, F. J., and M. M. Howe. 1984. Involvement of the invertible G segment in bacteriophage Mu tail fiber biosynthesis. Virology. **134:**296-317.

Grundy, F. J., and M. M. Howe. 1985. Morphogenetic structures present in lysates of amber mutants of bacteriophage Mu. Virology. **143:**485-504.

Guan, G. 1999. Functional analysis of the O antigen glucosylation gene cluster of *Shigella flexneri* bacteriophage SfX. Microbiology. **145:**1263-1273.

Guan, S., and N. K. Verma. 1998. Serotype conversion of a *Shigella flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. FEMS Microbiol Lett. **166:**79-87.

Hacker, J., and J. B. Kaper. 1999. The concept of pathogenicity islands, pp. 1-11, Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, D. C.

Haggård-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J Bacteriol. **174:**1462-1477.

Hakansson, U., R. Eitrem, B. Low, and S. Winblad. 1976. HLA-antigen b27 in cases of joint affections in an outbreak of salmonellosis. Scan J Infect Dis. **8:**245-248.

Hall, T. A. 1999. BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. **41:**95-98.

Hanahan, D. 1985. Bacterial transformation, pp. 109-135. *In* D. M. Glover (ed.), DNA cloning. IRL Press, Oxford, England.

Hansen, E. B. 1989. Structure and regulation of the lytic replicon of phage P1. J Mol Biol. **207:**135-149.

Hardt, W. D., L. M. Chen, K. E. Schuebel, X. R. Bustelo, and J. E. Galan. 1998a. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. Cell. **93:**815-826.

Hardt, W.-D., H. Urlaub, and J. E. Galan. 1998b. A substrate of the centisome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage. Proc Natl Acad Sci. **95:**2574-2579.

Harrington, C. S., J. A. Lanser, P. A. Manning, and C. J. Murray. 1991. Epidemiology of *Salmonella sofia* in Australia. Appl Environ Microbiol. **57:**223-227.

Harvey, D., C. Harrington, M. W. Heuzenroeder, and C. Murray. 1993. Lysogenic phage in *Salmonella enterica* serovar Heidelberg (*Salmonella* Heidelberg): implications for organism tracing. FEMS Microbiol Lett. **108:**291-295

Hatfull, G. F., and J. Sarkis. 1993. DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol Microbiol. **7:**395-405.

Hayden, M., M. B. Adams, and S. Casjens. 1985. Bacteriophage L: Chromosome physical map and structural proteins. Virology. **147:**431-440.

Hayes, W. 1964a. The physiology of bacteriophage infection, The genetics of bacteria and their viruses, 2 ed. Blackwell Scientific Publications, Oxford and Edinburg.

Hayes, W. 1964b. Transduction, pp. 620-649, The genetics of bacteria and their viruses, 2nd ed. Blackwell Scientific Publications, Oxford and Edinburg.

Head, S. C., M. Petric, S. Richardson, M. Roscoe, and M. A. Karmali. 1988. Purification and characterization of verocytotoxin 2. FEMS Microbiol Lett. **51:**211-216.

Heinrich, J., M. Velleman, and H. Schuster. 1995. The tripartite immunity system of phages P1 and P7. FEMS Microbiol Rev. **17:**121-126.

Heisig, A., H.-D. Riedel, B. Dobrinski, R. Lurz, and H. Schuster. 1989. Organization of the immunity system *immI* of bacteriophage P1 and synthesis of the P1 antirepressor. J Mol Biol. **209:**525-538.

Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Vaccine. **17:**770-778.

Hershey, A. D., and E. Burgi. 1965. Complementary structure of interacting sites at the ends of lambda DNA molecules. Proc Natl Acad Sci. **53:**325-328.

Hershey, A. D., and M. Chase. 1952. Independent functions of viral protein and nucleic acids in the growth of bacteriophage. J Gen Physiol. **36:**423.

Hickman-Brenner, F. W., A. D. Stubbs, and J. J. Farmer. 1991. Phage typing of *Salmonella enteritidis* in the United States. J Clin Microbiol. **29:**2817-2823.

Highton, P. J., Y. Chang, W. R. J. Macotte, and C. A. Schnaitman. 1985. Evidence that the outer membrane protein gene *ompC* of *Escherichia coli* lies within the defective qsr' prophage. J Bacteriol. **162:**256-262.

Highton, P. J., Y. Chang, and R. J. Myers. 1990. Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages. Mol Microbiol. **4:**1329-1340.

Ho, Y. S., D. Pfarr, J. Strickler, and M. Rosenberg. 1992. Characterization of the transcription activator protein C1 of bacteriophage P22. J Biol Chem. **267:**14388-14397.

Hofer, B., M. Ruge, and B. Dreiseikelmann. 1995. The superinfection exclusion gene (*sieA*) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. J Bacteriol. **177:**3080-3086.

Hohn, B. 1983. DNA sequences necessary for packaging of bacteriophage lambda DNA. Proc Natl Acad Sci. **80:**7456-7460.

Krieg, N. R., and J. G. Holt, (eds). 1984, 1986, 1989. *In* J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1-4. Williams and Wilkins, Baltimore.

Hook, E. W. 1990. *Salmonella* species (including typhoid fever), pp. 1700-1716. *In* D. Mandell, Bennet (ed.), Principles and practice of infectious diseases, 3rd ed. Churchill Livingstone, USA.

Hoppe, I., and J. Roth. 1974. Specialized transducing phages derived from *Salmonella* phage P22. Genetics. **76:**633-654.

Huan, P. T., B. L. Whittle, D. A. Bastin, A. A. Lindberg, and N. K. Verma. 1997. *Shigella flexneri* type-specific antigen V: cloning, sequencing and characterization of the glucosyl transferase gene of temperate bacteriophage SfV. Gene. **195:**207-216.

Huang, A., J. Friesen, and J. L. Brunton. 1987. Characterization of a bacteriophage that carries the genes for production of shiga-like toxin 1 in *Escherichia coli*. J Bacteriol. **169:**4308-4312.

Hubert, B., P. Dehaumont, G. Lelard, P. A. D. Grimont, and P. Bouvet. 1991. Les infections a *Salmonella enteritidis*: Situation en 1990. Bull Epidemiol Hebdomaidaire. **25:**103-105.

Hueck, C. J. 1998. Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev. **62:**379-433.

Hunter, P. R. 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol. **28:**1903-1905.

Huys, G., R. Coopman, P. Janssen, and K. Kersters. 1996. High-resolution genotypic analysis of the genus *Aeromonas* by AFLP fingerprinting. Int J Syst Bacteriol. **46:**572-580.

Iida, S. 1984. Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome. Virology. **134:**421-434.

Isaacson, R. E., and H. W. Moon. 1975. Induction of heat labile enterotoxin synthesis in enterotoxigenic *Escherichia coli* by mitomycin C. Infect Immun. **12:**1271-1275.

Iseki, S., and K. Kashiwagi. 1955. Induction of somatic antigen 1 by bacteriophage in *Salmonella* group B. Proc Jpn Acad. **31:**558-564.

Iwashita, S., and S. Kanegasaki. 1973. Smooth specific phage adsorption: Endorrhamnosidase activity of tail parts of P22. Biochem Biophys Res Commun. **55:**403.

Janssen, P., R. Coopman, G. Huys, J. Swings, M. Bleeker, P. Vos, M. Zabeau, and K. Kersters. 1996. Evaluation of the DNA fingerprinting method AFLP as an new tool in bacterial taxonomy. Microbiology. **142:**1881-1893.

Janssen, P., and L. Dijkshoorn. 1996. High resolution DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. FEMS Microbiol Lett. **142:**1881-1893.

Jeoffreys, N. J., G. S. James, R. Chiew, and G. L. Gilbert. 2001. Practical evaluation of molecular subtyping and phage typing in outbreaks of infection due to *Salmonella enterica* serotype Typhimurium. Pathology. **33:**66-72.

Jessop, A. P. 1972. A specialised transducing phage of P22 for which the ability to form plagues is associated with transduction of the proAB region. Mol Gen Genet. **114:**214-222.

Jessop, A. P. 1976. Specialized transducing phages derived from phage P22 that carry the proAB region of the host, *Salmonella typhimurium*: genetic evidence for their structure and mode of transduction. Genetics. **83:**459-475.

Johnson, L. P., M. A. Tomai, and P. M. Schlievert. 1986. Bacteriophage involvement in a group A streptococcal pyrogenic exotoxin A production. J Bacteriol. **166:**623-627.

Jones, B. D., and S. Falkow. 1994. Identification and characterization of a *Salmonella typhimurium* oxygen- regulated gene required for bacterial internalization. Infect Immun. **62:**3745-3752.

Jones, M. A., M. W. Wood, P. B. Mullan, C. B. Watson, T. S. Wallis, and E. E. Galyov. 1998. Secreted effector proteins of *Salmonella dublin* act in concert to induce enteritis. Infect Immun. **66:**5799-5804.

Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull, and R. W. Hendrix. 2000. Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol. **299:**27-51.

Kaiser, K., and N. E. Murray. 1979. Physical characterisation of the "Rac prophage" in *E. coli* K-12. Mol Gen Genet. **175:**159-161.

Kalinin, V. L., and L. V. Kuznetsova. 1995. Mutagenic effect of oxidizing agents on the thermally-induced prophage lambda c1857: effect of a system of oxidative stress. Genetika. **31:**784-787.

Karch, H., H. Schmidt, C. Janetzki-Mittmann, J. Scheef, and M. Kröger. 1999. Shiga toxins even when different are encoded at identical positions in the genomes of related temperate bacteriophages. Mol Gen Genet. **262:** 600-607.

Kariuki, S., C. Gilks, J. Kimari, J. Muyodi, P. Waiyaki and C. A. Hart. 1999. Analysis of *Salmonella enterica* serotype Typhimurium by phage typing, antimicrobial susceptibility and pulse field gel electrophoresis. J Med Microbiol. **48:**1037-1042.

Karlovsky, P., J. Soska, and J. Reich. 1984. Physical map of the bacteriophage L (*Salmonella typhimurium*). FEMS Microbiol Lett. **25:**117-120.

Karplus, P. A., and G. E. Schultz. 1985. Prediction of chain flexibility in proteins. A tool for the selection of peptide antigens. Naturwissenschaften. **72:**212-213.

Katsura, I., and P. W. Kühl. 1974. A regulator protein for the length determination of bacteriophage lambda tail. J Supramol Structure. **2:**239-253.

Kauffmann, F. 1954. Enterobacteriaceae, 2nd ed. Munksgaard, Copenhagen.

Khakhria, R., Duck, D. and Lior, H. 1991. Distribution of *Salmonella enteritidis* phage types in Canada. Epidemiol Infect. **106:**25-32.

Kimmitt, P. T., C. R. Harwood, and M. R. Barer. 2000. Toxin gene expression by shiga toxin-producing *Escherichia coli*: the role of antibiotics and the bacterial SOS response. Emerg Infect Dis. **6:**458-465.

Kitamura, J., and K. Mise. 1970. A new generalized transducing phage in *Salmonella.* Jpn J Med Sci Biol. **23:**99-102.

Kleckner, N., R. K. Chan, B.-K. Tye, and D. Botstein. 1975. Mutagenesis by insertion of a drug-resistance element carrying an inverted repetition. J Mol Biol. **97:**561-575.

Kleparnik, K., Z. Mala, Z. Havac, M. Blazkova, L. Holla, and P. Bocek. 1998. Fast detection of a (CA)18 microsatellite repeat in the IgE receptor gene by capillary electrophoresis with laser-induced fluorescence detection. Electrophoresis. **19:**249-255.

Knapp, S., J. Hacker, T. Jarchau, and W. Goebel. 1986. Large unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* O6 strain 536. J Bacteriol. **168:**22-30.

Knothe, H., G. Lebek, V. Krcmery, Z. Seginkova, J. Cervenka, M. Antal, and S. Mitsuhashi. 1981. Transduction of amikacin, gentamicin and tobramycin resistance in *Pseudomonas aeruginosa* with phage F 116 and AP 19, a new wildtype phage. Zentralbl Bakteriol Mikrobiol Hyg [A]. **250:**506-510.

Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer. 1994. Biochemistry of homologous recombination in *Escherichia coli.* Microbiol Rev. **58:**401-465.

Kroger, M., and G. Hobom. 1982. A chain of interlinked genes in the *ninR* region of bacteriophage lambda. Gene. **20:**25-38.

Kropinski, A. M. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. J Bacteriol. **182:**6066-6074.

Kuo, T. T., and B. A. D. Stocker. 1970. ES18, a general transducing phage for smooth and nonsmooth *Salmonella typhimurium*. Virology. **42:**621-632.

Kuzio, J., and A. M. Kropinski. 1983. O-antigen conversion in *Pseudomonas aeruginosa* PAO1 by bacteriophage D3. J Bacteriol. **155:**203-212.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol. **157:**105-132.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. **227:**680-685.

Laird, W., and N. Groman. 1976. Orientation of the *tox* gene in the prophage of corynebacteriophage beta. J Virol. **19:**228-231.

Lazar, S., and M. K. Waldor. 1998. ToxR-independent expression of cholera toxin from the replicative form of CTXϕ. Infect Immun. **66:**394-397.

Le, H., D. C. Fung, B. Yu, and R. J. Trent. 1998. Capillary electrophoresis: new technology for DNA diagnostics. Pathology. **30:**304-308.

Le Minor, L. 1984. Genus III. *Salmonella* Lignières 1900, 389AL, pp. 427-458, Bergey's manual of systematic bacteriology, vol. I. The Williams and Wilkins Co., Baltimore.

Le Minor, L. 1988. Typing of *Salmonella* species. Euro J Clin Microbiol and Infect Dis. **7:**214-218.

Le Minor, L., and M. Y. Popoff. 1987. Request for an opinion. Designation of *Salmonella enterica* sp. nov. , nom. rev., as the type and only species of the genus *Salmonella*. Int J Sys Bacteriol. **37:**465-468.

Lee, C. A. 1996. Pathogenicity islands and the evolution of bacterial pathogens. Infect Agent and Dis. **5:**1-7.

Lee, C. Y., and J. J. Iandolo. 1985. Mechanism of bacteriophage conversion of lipase activity in *Staphylococcus aureus*. J Bacteriol. **164:**288-293.

Lennox, E. S. 1955. Transduction of linked genetic characters of the host by bacteriophage P1. Virology. **1:**384, 621-646.

Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy. 1985. The phi-80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. J Biol Chem. **260:**4468-4477.

Lerner, C. G., and M. Inouye. 1990. Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability. Nucleic Acids Res. **18:**4631.

Letellier, A., S. Messier, J. Pare, J. Menard, and S. Quessy. 1999. Distribution of *Salmonella* in swine herds in Quebec. Vet Microbiol. **67:**299-306.

Liebana, E., L. Garcia-Migura, M. F. Breslin, R. H. Davies, and M. J. Woodward. 2001. Diversity of Strains of *Salmonella enterica* serotype Enteritidis from English Poultry Farms Assessed by Multiple Genetic Fingerprinting. J Clin Microbiol. **39:**154-161.

Lin, J. J., J. Kuo, and J. Ma. 1996. A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria. Nucleic Acids Res. **24:**3649-3650.

Lindsay, J. A., A. Ruzin, H. F. Ross, N. Kurepina, and R. P. Novick. 1998. The gene for toxic shock syndrome is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. Mol Microbiol. **29:**527-543.

Lindstedt, B. A., E. Heir, T. Vardund, and G. Kapperud. 2000. Fluorescent amplified-fragment length polymorphism genotyping of *Salmonella enterica* subsp. enterica serovars and comparison with pulsed- field gel electrophoresis typing. J Clin Microbiol. **38:**1623-1627.

Lior, H. 1989. Isolations of enteric pathogens from people in Canada. Safety Watch. **14:**3.

Lior, H., and R. Khakhria. 1985. Prevalence of *Salmonella* serotypes and phage types in Canada, pp. 332-333. *In* G. H. Snoeyenbos (ed.), Proceedings of the international symposium on *Salmonella*. American Association of Avian Pathologists, inc., New Orleans.

Lister, S. A. 1988. *Salmonella* enteritidis infection in broilers and broiler breeders. Vet Rec. **123:**350.

Loessner, M. J., R. B. Inman, P. Lauer, and R. Calendar. 2000. Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: implications for phage evolution. Mol Microbiol. **35:**324-340.

Loessner, M. J., G. Wendlinger, and S. Scherer. 1995. Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. Mol Microbiol. **16:**1231-1241.

Loutit, J. S. 1958. A transduction-like process within a single strain of *Pseudomonas aeruginosa*. Nature. **205:**754.

Low, J. C., M. Angus, G. Hopkins, D. Munro, and S. C. Rankin. 1997. Antimicrobial resistance of *Salmonella enterica* Typhimurium DT104 isolates and investigation of strains with transferable apramycin resistance. Epidemiol Infect. **118:**97-103.

Lown, J. W., D. Begleiter, D. Johnson, and A. R. Morgan. 1976. Studies related to antitumor antibiotics. Part V. Reaction of mitomycin C with DNA examined by ethidium fluorescence assay. Can J Biochem. **54:**110-119.

Ludwig, W., and K. H. Schleifer. 1994. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol Rev. **15:**155-173.

Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. van Alphen. 1975. Electrophoretic resolution of the "major outer menbrane protein" of *Escherichia coli* K-12 into four bands. FEBS Lett. **58:**254-258.

Lukashin, A., and M. Borodovsky. 1998. GeneMark.hmm: a new solution for gene finding. Nucleic Acids Res. **26:**1107-1115.

Lwoff, A. Lysogeny. 1953. Bacteriol Rev. **17:**269.

Madhubala, K. P., and B. K. Sharma. 1981. Change in the O serotype of *Pseudomonas aeruginosa* after lysogenisation with bacteriophages. Ind J Med Res. **73:**686-692.

Mahan, M. J., J. M. Slauch, and J. J. Mekalanos. 1993. Bacteriophage P22 transduction of integrated plasmids: single-step cloning of *Salmonella typhimurium* gene fusions. J Bacteriol. **175:**7086-7091.

Mahdi, A. A., G. J. Sharples, T. N. Mandal, and R. G. Lloyd. 1996. Holliday junction resolvases encoded by homologous rusA genes in *Escherichia coli* K-12 and phage 82. J Mol Biol. **257:**561-573.

Maher, K. O., J. G. Morris, E. Gotuzzo, C. Ferreccio, L. R. Ward, L. Benavente, R. E. Black, B. Rowe, and M. M. Levine. 1986. Molecular techniques in the study of *Salmonella typhi* in epidemiologic studies in endemic areas: comparison with Vi phage typing. Am J Trop Med Hyg. **35:**831-835.

Maniatis, T., Fritsch, EF and Sambrook, J. 1982. Molecular Cloning: A laboratory Manual. Cold Spring Harbor Laboratory.

Maniloff, J., and H. W. Ackermann. 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order Caudovirales. Arch Virol. **143:**2051-2063.

Mann, B. A., and J. M. Slauch. 1997. Transduction of low-copy number plasmids by bacteriophage P22. Genetics. **146:**447-456.

Margolin, P. 1987. Generalized transduction, pp. 1154-1168, *Escherichia coli* and *Salmonella typhimurium.* American Society for Microbiology, Washington, D.C.

Markel, D. E., and C. Eklund. 1974. Isolation, characterization and classification of three bacteriophage isolates for the genus *Levinea*. Int J Syst Bacteriol. **24:**230-234.

Maslow, J. N., M. E. Mulligan, and D. R. Arbeit. 1993a. Molecular epidemiology: Application of contemporary techniques to the typing of microorganisms. Clin Infect Dis. **17:**153-164.

Maslow, J. N., A. M. Slutsky, and R. D. Arbeit. 1993b. Application of pulsed-field gel electrophoresis to molecular epidemiology, pp. 563-572, American society for microbiology. ASM, Washington, D.C.

Matthews, D. H., J. Sabina, M. Zucker, and H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. J Mol Biol. **288:**911-940.

Matushek, M. G., M. J. Bonten, and M. K. Hayden. 1996. Rapid preparation of bacterial DNA for pulsed-field gel electrophoresis. J Clin Microbiol. **34:**2598-2600.

Mavris, M., P. A. Manning, and R. Morona. 1997. Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. Mol Microbiol. **26:**939-950.

McDaniel, T. K., K. G. Javis, M. S. Donnenberg, and J. B. Kaper. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc Natl Acad Sci. **92:**1664-1668.

Mecsas, J., and E. J. Strauss. 1996. Molecular mechanisms of bacterial virulence: type III secretion and pathogenicity islands. Emerg Infect Dis. **2:**271-285.

Messenger, S. L., I. J. Molineux, and J. J. Bull. 1999. Virulence evolution in a virus obeys a trade-off. Proc R Soc Lond B Biol Sci. **266:**397-404.

Miao, E. A., C. A. Scherer, R. M. Tsolis, R. A. Kingsley, L. G. Adams, A. J. Baumler, and S. I. Miller. 1999. *Salmonella typhimurium* leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. Mol Microbiol. **34:**850-864.

Michiels, T., P. Wattiau, R. Brasseur, J.-M. Ruysschaert, and G. R. Cornelis. 1990. Secretion of Yop proteins by *Yersinia*. Infect Immun. **58:**2840-2849.

Miller, H. 1987. Practical aspects of preparing phage and plasmid DNA: Growth, maintenance, and storage of bacteria and bacteriophage, pp. 145-170. *In* L. a. K. Begger, A. R. (ed.), Methods in enzymology: Guide to melecular cloning techniques, vol. 152. Harcourt Brace Jovanovich, Toronto.

Miller, J. H. 1992. A short course in bacterial genetics. Cold Spring Harbor Press, New York.

Miller, R. V. 1998. Bacterial gene swapping in nature. Scientific American. **278:**47-51.

Mills, D. M., V. Bajaj, and C. A. Lee. 1995. A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K- 12 chromosome. Mol Microbiol. **15:**749-759.

Mirold, S., W. Rabsch, M. Rohde, S. Stender, H. Tschape, H. Russmann, E. Igwe, and W. D. Hardt. 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. Proc Natl Acad Sci U S A. **96:**9863-9866.

Moazamie, N., H. W. Ackermann, and M. R. V. Murthy. 1979. Characterization of two *Salmonella newport* bacteriophages. Can J Microbiol. **25:**1063-1072.

Mokgatla, R. M., V. S. Brozel, and P. A. Gouws. 1998. Isolation of *Salmonella* resistant to hypochlorous acid from a poultry abattoir. Lett Appl Microbiol. **27:**1-4.

Morse, M. L. 1954. Transduction of certain loci in *Esherichia coli* K-12. Genetics. **39:**621-627.

Morse, M. L. 1959. Transduction by staphylococcal bacteriophage. Proc Natl Acad Sci. **45:**621.

Morse, M. L., E. M. Lederberg, and J. Lederberg. 1956. Transduction in *Escherichia coli* K-12. Genetics. **41:**142-156.

Mueller, U. G., S. E. Lipari, and M. G. Milgroom. 1996. Amplified fragment length polymorphism (AFLP) fingerprinting of symbiotic fungi cultured by the fungus-growing ant *Cyphomyrmex minutus*. Mol Ecol. **5:**119-122.

Murphy, F. A., C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers. 1995. Virus taxonomy. Sixth report of the international committe on taxonomy of viruses, Arch Virol, vol. Supp 10. Springer, Vienna.

Murphy, K. C., L. Casey, N. Yannoutsos, A. R. Poteete, and R. W. Hendrix. 1987a. Localization of a DNA-binding determinant in the bacteriophage P22 Erf protein. J Mol Biol. **194:**105-117.

Murphy, K. C., A. C. Fenton, and A. R. Poteete. 1987b. Sequence of the bacteriophage P22 anti-RecBCD (abc) genes and properties of P22 abc region deletion mutants. Virology. **160:**456-464.

Murray, C. J. 1987. *Salmonella* and *Escherichia coli* from veterinary and human sources in Australia during 1985 and 1986. Aust Vet J. **64:**256-257.

Muschel, L. H., and K. Schmoker. 1966. Activity of mitomycin C, other antibiotics, and serum against lysogenic bacteria. J Bacteriol. **92:**967-971.

Nagaraja, R., and R. A. Weisberg. 1990. Specificity determinants in the attachment sites of bacteriophages HK022 and lambda. J Bacteriol. **172:**6540-6550

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

Nam, K.-B., U. Blasi, M. T. Zagotta, and R. Young. 1990. Conservation of a dual-start motif in P22 lysis gene regulation. J. Bacteriol. **172:**204-211.

Neely, M. N., and D. I. Friedman. 1998. Functrional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role in toxin release. Mol Microbiol. **28:**1255-1267.

Newland, J. W., and R. J. Neill. 1988. DNA probes for Shiga-like toxins I and II and for toxin-converting bacteriophages. J Clin Microbiol. **26:**1292-1297.

Newland, J. W., N. A. Strockbine, S. F. Miller, A. D. O'Brien, and R. K. Holmes. 1985. Cloning of Shiga-like toxin structural genes from a toxin converting phage of *Escherichia coli*. Science. **230:**179-181.

Noble, R. T., and J. A. Fuhrman. 1998. Use of SYBR green I for rapid epifluorescense counts of marine viruses and bacteria. Aquat Microb Ecol. **14:**113-118.

Oboegbulem, S. I. a. O., I. 1990. *Salmonella* in the African great cane rat (*Thryonomys swinderianus*). J Wildl Dis. **26:**119-121.

Oberto, J., S. B. Sloan, and R. A. Weisberg. 1994. A segment of the phage HK022 chromosome is a mosaic of other lambdoid chromosomes. Nucleic Acids Res. **22:**354-356.

O'Brien, A. D., J. W. Newland, S. F. Miller, R. K. Holmes, H. W. Smith, and S. B. Formal. 1984. Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. Science. **226:**694-696.

Ochman, H., F. C. Soncini, F. Solomon, and E. A. Groisman. 1996. Identification of a pathogenicity island required for *Salmonella* survival in host cells. Proc Natl Acad Sci U S A. **93:**7800-7804.

Ogg, J. E., T. L. Timme, and M. M. Alemohammad. 1981. General transduction in *Vibrio cholerae*. Infect Immun. **31:**737-741.

Olsen, J. E., D. J. Brown, M. N. Skov, and J. P. Christensen. 1993. Bacterial typing methods suitable for epidemiological analysis: Applications in investigations of salmonellosis among livestock. Vet. Q. **15:**125-134.

Olsen, J. E., M. N. Skov, E. J. Threlfall, and D. J. Brown. 1994. Clonal lines of *Salmonella enterica* serotype Enteritidis documented by IS200-, ribo-, pulse-field electrophoresis and RFLP typing. J Med Microbiol. **40:**15-22.

Orbach, M. J., and E. N. Jackson. 1982. Transfer of chimeric plasmids among *Salmonella typhimurium* by P22 transduction. J Bacteriol. **149:**985-994.

Španová, A. 1992. Comparison of permuted region lengths in the genomes of related *Salmonella typhimurium* phages P22 and L. Folia Microbiol. **37:**188-192.

Pappenheimer, A. M. J., and J. R. Murphy. 1983. Studies on the molecular epidemiology of diphtheria. Lancet. **ii:**923-926.

Parker, J. M., D. Guo, and R. S. Hodges. 1986. New hydrophilicity scale derived from high-performance liquid chromotography peptide retension data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. **25:**5425-5432.

Parker, M. T. 1983. The significance of phage typing patterns in *Staphylococcus aureus*, pp. 33-62. *In* C. S. F. Easmon and C. Adlam (eds.), *Staphylococci* and Staphylococcal infections, vol. 1. Academic Press, London.

Paton, A. W., and J. C. Paton. 1996. *Enterobacter cloacae* producing a shiga-like toxin II-associated with a case of hemolytic-uremic syndrome. J Clin Microbiol. **34:**463-465.

Paton, A. W., J. C. Paton, M. W. Heuzenroeder, P. N. Goldwater, and P. A. Manning. 1992. Cloning and nucleotide sequence of a variant Shiga-like toxin II gene from *Escherichia coli* OX3:H21 isolated from a case of sudden infant death syndrome. Microb Pathog. **13:**225-236.

Perales, I., and A. Audicana. 1988. *Salmonella enteritidis* and eggs. Lancet. **2:**1133.

Person, W. R., T. Wood, Z. Zhang, and W. Miller. 1997. Comparison of DNA sequences with protein sequences. Genomics. **46:**24-36.

Pierson, L. S. d., and M. L. Kahn. 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. J Mol Biol. **196:**487-496.

Platt, D. J., D. J. Brown, D. C. Old, R. M. Barker, D. S. Munro, and J. Taylor. 1987. Old and new techniques together resolve a problem of infection by *Salmonella typhimurium*. Epidemiol Infect. **99:**137-142.

Popoff, M. W., and L. LeMinor. 1997. Antigenic formulas of the *Salmonella* serovars, 7th ed. World Health Organization Collaborating Centre for Reference and Research on *Salmonella*, Pasteur Institute, Paris, France.

Poppe, C., R. J. Irwin, S. Messier, G. G. Finley, and J. Oggel. 1991. The prevalence of *Salmonella enteritidis* and other *Salmonella sp.* among Canadian registered commercial chicken broiler flocks. J Clin Microbiol. **29:**2817-2823.

Poteete, A. R. 1988. Bacteriophage P22, pp. 647-682. *In* R. Calendar (ed.), The Bacteriophages, vol. 2. Plenum Press, New York.

Poteete, A. R., and A. C. Fenton. 1984. Lambda *red*-dependent growth and recombination of phage P22. Virology. **134:**161-167.

Poteete, A. R., R. T. Sauer, and R. W. Hendrix. 1983. Domain structure and quaternary organization of the bacteriophage P22 Erf protein. J Mol Biol. **171:**401-418.

Powell, N. G., E. J. Threlfall, H. Chart, S. L. Schofield, and B. Rowe. 1995. Correlation of change in phage type with pulse field and 16S rrn profile in *Salmonella enteritidis* phage types 4, 7 and 9a. Epidemiol Infect. **114:**303-411.

Prevelige, P. E., D. Thomas, and J. King. 1988. Scaffolding protein regulates the polymerization of P22 coat subunits into icosahedral shells in vitro. J Mol Biol. **202:**743-757.

Puohiniemi, R., T. Heiskanen, and A. Siitonen. 1997. Molecular epidemiology of two international sprout-borne *Salmonella* outbreaks. J Clin Microbiol. **35:**2487-2491.

Quiberoni, A., L. Rezaiki, M. El Karoui, I. Biswas, P. Tailliez, and A. Gruss. 2001. Distinctive features of homologous recombination in an 'old' microorganism, *Lactococcus lactis*. Res Microbiol. **152:**131-139.

Raettig, H. 1967. Bakteriophagie. 1957-1965 Part 1. Gustav-Fischer Verlang.

Ranade, K., and A. R. Poteete. 1993. Superinfection exclusion (*sieB*) genes of bacteriophages P22 and Lambda. J Bacteriol. **175:**4712-4718.

Rankin, S., and D. J. Platt. 1995. Phage conversion in *Salmonella enterica* serotype Enteritidis: implications for epidemiology. J Appl Bacteriol. **78:**402-408.

Reidl, J., and J. J. Mekalanos. 1995. Characterization of *Vibrio cholerae* bacteriophage K139 and use of a novel mini-transposon to identify a phage-encoded virulence factor. Mol Microbiol. **18:**685-701.

Relman, D. A. S., T. M. MacDermott, R. P. Falkow, S. 1992. Identification of the uncultured bacillus of Whipple's disease. N Engl J Med. **327:**293-301.

Rhoades, M., L. A. Machattie, and C. A. J. Thomas. 1968. The P22 bacteriophage DNA molecule. I. The mature form. J Mol Biol. **37:**21-40.

Ridley, A. M., E. J. Threlfall, and B. Rowe. 1998. Genotypic characterization of *Salmonella enteritidis* phage types by plasmid analysis, ribotyping, and pulsed-field gel electrophoresis. J Clin Microbiol. **36:**2314-2321.

Riley, M., and A. Anilionis. 1980. Conservation and variation of nucleotide sequences within related bacterial genomes: enterobacteria. J Bacteriol. **143:**366-376.

Rodrigue, D. C., D. N. Cameron, N. D. Puhr, F. W. Brenner, M. E. St Louis, I. K. Wachsmuth, and R. V. Tauxe. 1992. Comparison of plasmid profiles, phage types, and antimicrobial resistance patterns of *Salmonella enteritidis* isolates in the United States. J Clin Microbiol. **30:**854-857.

Rountree, P. M. 1949. The phenomenon of lysogenicity in staphylococci. J Gen Microbiol. **3:**153.

Rowe, B. 1984-1987. *Salmonella* surveillance, Annual reports from centres participating in the WHO programme. WHO collaborating centre for phage typing and resistance Enterobacteriaceae, London.

Rubin, R. H., and L. Weinstein. 1977. Salmonellosis: Microbiologic, pathologic and clinical features. Stratton Intercontinental Medical Book Corporation, New York.

Rundell, K., and C. W. Shuster. 1975. Membrane-associated nucleotide sugar reactions: influence of mutations affecting lipopolysaccharide on the first enzyme of O-antigen synthesis. J Bacteriol. **123:**928-936.

Russel, M. 1991. Filamentous phage assembly. Mol Microbiol. **5:**1607-1613.

Sallen, B., A. Rajoharison, S. Desvarenne, F. Quinn, and C. Mabilat. 1996. Comparative analysis of 16S and 23S rRNA sequences of *Listeria* species. Int J Syst Bacteriol. **46:**669-674.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2 ed. Cold Spring Laboratory Press, New York.

Sander, M., and H. Schmieger. 2001. Method for host-independent detection of generalized transducing bacteriophages in natural habitats. Appl Environ Microbiol. **67:**1490-1493.

Sanderson, K. E., A. Hessel, and K. E. Rudd. 1995. Genetic map of *Salmonella typhimurium*, edition VIII. Microbiol Rev. **59:**241-303.

Sanderson, K. E., and J. R. Roth. 1988. Linkage map of *Salmonella typhimurium*, edition VII. Microbiol Rev. **52:**485-532.

Sanger, F., A. R. Coulson, G. H. Hong, D. F. Hill, and G. B. Petersen. 1982. Nucleotide sequence of bacteriophage λ DNA. J Mol Biol. **162:**729-773.

Sauer, R. T., W. Krovatin, J. DeAnda, P. Youderian, and M. Susskind. 1983. Primary structure of the *immI* region of bacteriophage P22. J Mol Biol. **168:**699-713.

Saunders, N. A., T. G. Harrison, A. Haththoutuwa, N. Kachwalla, and A. G. Taylor. 1990. A method for typing strains of *Legionella pneumophila* serogroup 1 by analysis of restriction fragment legnth polymorphisms. J Med Microbiol. **31:**45-55.

Savelkoul, P. H., H. J. Aarts, J. de Haas, L. Dijkshoorn, B. Duim, M. Otsen, J. L. Rademaker, L. Schouls, and J. A. Lenstra. 1999. Amplified-fragment length polymorphism analysis: the state of an art. J Clin Microbiol. **37:**3083-3091.

Sawadogo, M., and M. W. Van Dyke. 1991. A rapid method for the purification of deprotected oligodeoxynucleotides. Nucleic Acids Res. **19:**674.

Schanda-Mulfinger, U. E., and H. Schmieger. 1980. Growth of *Salmonella* bacteriophage P22 in *Escherichia coli* dna(Ts) mutants. J Bacteriol. **143:**1042-1045.

Schicklmaier, P., E. Moser, T. Wieland, W. Rabsch, and H. Schmieger. 1998. A comparative study on the frequency of prophages among natural isolates of *Salmonella* and *Escherichia coli* with emphasis on generalized transducers. Antonie Van Leeuwenhoek. **73:**49-54.

Schicklmaier, P., and H. Schmieger. 1995. Frequency of generalized transducing phages in natural isolates of the *Salmonella typhimurium* complex. Appl Environ Microbiol. **61:**1637-1640.

Schicklmaier, P., and H. Schmieger. 1997. Sequence comparison of the genes for immunity, DNA replication, and cell lysis of the P22-related *Salmonella* phages ES18 and L. Gene. **195:**93-100.

Schicklmaier, P., T. Wieland, and H. Schmieger. 1999. Molecular characterization and module composition of P22-related *Salmonella* phage genomes. J Biotechnol. **73:**185-194.

Schmidt, C., and H. Schmieger. 1984. Selective transduction of recombinant plasmids with cloned pac sites by *Salmonella* phage P22. Mol Gen Genet. **196:**123-128.

Schmidt, H., M. Bielaszewska, and H. Karch. 1999. Transduction of enteric *Escherichia coli* isolates with a derivative of Shiga toxin 2-encoding bacteriophage phi3538 isolated from *Escherichia coli* O157:H7. Appl Environ Microbiol. **65:**3855-3861.

Schmieger, H. 1970. The molecular structure of the transducing particles of *Salmonella* phage P22. II. Density gradient analysis of DNA. Mol Gen Genet. **109:**323-337.

Schmieger, H. 1972. Phage P22-mutants with increased or decreased transduction abilities. Mol Gen Genet. **119:**75-88.

Schmieger, H. 1982. Packaging signals for phage P22 on the chromosome of *Salmonella typhimurium*. Mol Gen Genet. **187:**516-518.

Schmieger, H. 1999. Molecular survey of the *Salmonella* phage typing system of Anderson. Proc Natl Acad Sci U S A. **96:**2192-2197.

Schmieger, H., and H. Backhaus. 1976. Altered cotransduction frequencies exhibited by HT-mutants of *Salmonella*-phage P22. Mol Gen Genet. **143:**307-309.

Schmieger, H., and P. Schicklmaier. 1999. Transduction of multiple drug resistance of *Salmonella enterica* serovar Typhimurium DT104. FEMS Microbiol Lett. **170:**313-317.

Schulte, T. 1994. The 1990 *Salmonella* epidemic in the commercial city of Lubeck: An epidemiologic study as a contribution to determining the aetiology of the *Salmonella* outbreak. Gesundheitswesen. **56:**606-610.

Seginkova, Z., V. Krcmery, and H. Knothe. 1986. Ceftazidime resistance in *Pseudomonas aeruginosa*: transduction by a wild-type phage. J Infect Dis. **154:**1049-1050.

Semerjian, A. V., D. C. Malloy, and A. R. Poteete. 1989. Genetic structure of the bacteriophage P22 pL operon. J Mol Biol. **207:**1-13.

Seyedirashti, S., C. Wood, and J. M. Akagi. 1991. Induction and partial purification of bacteriophage from *Desulfovibrio vulgaris* (*Hildenborough*) and *Desulfovibrio desulfuricans* ATCC 13541. J Gen Microbiol. **137:**1545-1549.

Shanado, Y., J.-I. Kato, and H. Ikeda. 1997. Fis is required for illegitimate recombination during formation of bio transducing phage. J Bacteriol. **179:**4239-4245.

Sharp, R., I. S. Jansons, E. Gertman, and A. M. Kropinski. 1996. Genetic and sequence analysis of the *cos* region of the temperate *Pseudomonas aeruginosa* bacteriophage, D3. Gene. **177:**47-53.

Shea, J. E., M. Hensel, C. Gleeson, and D. W. Holden. 1996. Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. Proc Natl Acad Sci U S A. **93:**2593-2597.

Silliker, J. H. 1980. Status of *Salmonella*: ten years later. J Food Prot. **43:**307-313.

Simango, C., and C. Mbewe. 2000. *Salmonella enteritidis* diarrhoea in Harare, Zimbabwe. Trop Med Int Health. **5:**503-506.

Simon, M. N., R. W. Davis, and N. Davidson. 1971. Heteroduplexes of DNA molecules of lambdoid phages: Physical mapping of their base sequence by electron microscopy, pp. 313-328, The Bacteriophage Lambda. Cold Spring Harbor Laboratory, New York.

Smith, G. R. 1983. General recombination, pp. 175-210, Lambda II. Cold Spring Harbor, New York.

Smith, H. O., and M. Levine. 1967. A phage P22 gene controlling integration of prophage. Virology. **31:**207-216.

Smith, H. W., and Z. Parsell. 1974. The effect of virulence of converting the O antigen of *Salmonella cholerae-suis* from 627 to 617 by phage. J Gen Microbiol. **81:**217-224.

Smith, M. C., R. N. Burns, S. E. Wilson, and M. A. Gregory. 1999. The complete genome sequence of the *Streptomyces* temperate phage straight phi-C31: evolutionary relationships to other viruses. Nucleic Acids Res. **27:**2145-2155.

Smith-Mungo, L., I. T. Chan, and A. Landy. 1994. Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. J Biol Chem. **269:**20798-20805.

Sonnhammer, E. L. L., G. von Heijne, and A. Krogh. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences, pp. 175-182, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology. AAAI Press, Menlo Park, California.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol. **98:**503-517.

Staedler, P., and J. W. Nesbit. 1990. Salmonellosis in an adult dairy cow. J S Afr Vet Assoc. **61:**65-67.

Stahl, F. W., C. E. Shurvinton, L. S. Thomason, S. Hill, and M. M. Stahl. 1995. On the clustal exchanges of the RecBCD pathway operating on phage lambda. Genetics. **139:**1107-1121.

Stanley, J., M. Goldsworthy, and E. J. Threlfall. 1992. Molecular phylogenetic typing of pandemic isolates of *Salmonella enteritidis*. FEMS Microbiol Lett. **69:**153-160.

Steinberg, B. M., and M. Gough. 1976. Bacteriophage P22 lysogenises efficiently at high multiplicities of infection because *Salmonella typhimurium* DNA synthetic capacity is limited. Nature. **263:**54-56.

Strobel, E., W. Behnisch, and H. Schmieger. 1984. In vitro packaging of mature phage DNA by *Salmonella* phage P22. Virology. **133:**158-165.

Stubbs, A. D., F. W. Hickman-Brenner, D. N. Cameron, and J. J. Farmer. 1994. Differentiation of *Salmonella enteritidis* phage type 8 strains: evaluation of three additional phage typing systems, plasmid profiles, antibiotic susceptibility patterns and biotyping. J Clin Microbiol. **32:**199-201.

Sun, J., M. Inouye, and S. Inouye. 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage phi R73) integrated into the selenocystyl tRNA gene of *Escherichia coli*. J Bacteriol. **173:**4171-4181.

Susskind, M. M. 1980. A new gene of P22 which regulates synthesis of antirepressor. J Mol Biol. **138:**685-713.

Susskind, M. M., and D. Botstein. 1975. Mechanisms of action of *Salmonella* phage P22 antirepressor. J Mol Biol. **98:**413-424.

Susskind, M. M., and D. Botstein. 1978a. Molecular genetics of bacteriophage P22. Microbiol Rev. **42:**385-413.

Susskind, M. M., and D. Botstein. 1978b. Repression and immunity in *Salmonella* phages P22 and L: phage L lacks a functional secondary immunity system. Virology. **89:**618-622.

Susskind, M. M., and D. Botstein. 1980. Superinfection exclusion by lambda prophage in lysogens of *Salmonella typhimurium*. Virology. **100:**212.

Susskind, M. M., A. Wright, and D. Botstein. 1971. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. II. Genetic evidence fro two exclusion systems. Virology. **45:**638-652.

Susskind, M. M., A. Wright, and D. Botstein. 1974. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium*. Virology. **62:**367-384.

Susskind, M. M., and P. Youderian. 1982. Transcription *in vitro* of the bacteriophage P22 antirepressor gene. J Mol Biol. **154:**427-447.

Suzuki, Y., M. Ishihara, M. Matsumoto, S. Arakawa, M. Saito, N. Ishikawa, and T. Yokochi. 1995. Molecular epidemiology of *Salmonella enteritidis*. An outbreak and sporadic cases studied by means of pulsed-field gel electrophoresis. J Infect. **31:**211-217.

Swofford, D. 1993. PAUP - Phylogenetic Analysis Using Parsimony, 3.1 ed. Natural History Survey, Champaign, Illinois.

Szalewska-Palasz, A., C. Weigel, C. Speck, S. Srutkowska, G. Konopa, R. Lurz, K. Marszalek, K. Taylor, W. Messer, and G. Wergzyn. 1998. Interaction of the *Escherichia coli* DnaA protein with bacteriophage lambda DNA. Mol Gen Genet. **259:**679-688.

Takahashi, I. 1961. Genetic transduction in *Bacillus subtilis*. Biochem Biophys Res Commun. **5:**615, 621.

Takahashi, I. 1963. Transducing phages for *Bacillus subtilis*. J Gen Microbiol. **31:**614, 621.

Takeda, Y., K. Okamoto, and T. Miwatani. 1979. Mitomycin C stimulates production of a toxin in *Shigella* species that causes morphological changes in chinese hamster ovary cells. Infect Immun. **23:**178-180.

Tamada, Y., Y. Nakaoka, K. Nishimori, A. Doi, T. Kumaki, N. Uemura, K. Tanaka, S.-I. Makino, T. Sameshima, M. Akibo, M. Nakazawa, and I. Uchida. 2001. Molecular typing and epidemiological study of *Salmonella enterica* serotype Typhimurium isolates from cattle by fluorescent amplified fragment length polymorphism fingerprinting and pulsed-field gel electrophoresis. J Clin Microbiol. **39:**1057-1066.

Taylor, K., and G. Wegrzyn. 1995. Replication of coliphage lambda DNA. FEMS Microbiol Lett. **17:**109-119.

Tenover, C. F., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan. 1995. Interpreting chromosomal DNA restriction patterns produced by pulse field gel electrophoresis: Criteria for bacterial strain typing. J Clin Microbiol. **33:**2233-2239.

Tétart, F., C. Desplats, M. Kutateladze, C. Monod, H. W. Ackermann, and H. M. Krisch. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J Bacteriol. **183:**358-366.

Thong, K., Y. Cheong, S. Puthucheary, C. Koh, and T. Pang. 1994. Epidemiologic analysis of sporadic *Salmonella typhi* isolates and those from outbreaks by pulsed-field gel electrophoresis. J Clin Microbiol. **32:**1135-1141.

Thong, K. L., S. Nair, R. Chaudhry, P. Seth, A. Kapil, D. Kumar, H. Kapoor, S. Puthucheary, and T. Pang. 1998. Molecular analysis of *Salmonella* paratyphi A from an outbreak in New Delhi, India. Emerg Infect Dis. **4:**507-508.

Thong, K. L., Y. F. Ngeow, M. Altwegg, P. Navaratnam, and T. Pang. 1995a. Molecular analysis of *Salmonella enteritidis* by pulsed-field gel electrophoresis and ribotyping. J Clin Microbiol. **33:**1070-1074.

Thong, K. L., S. Puthucheary, R. M. Yassin, P. Sudarmono, M. Padmidewi, E. Soewandojo, I. Handojo, S. Sarasombath, and T. Pang. 1995b. Analysis of *Salmonella typhi* isolates from Southeast Asia by pulsed- field gel electrophoresis. J Clin Microbiol. **33:**1938-1941.

Threlfall, E. J., and J. A. Frost. 1990. The identification, typing and fingerprinting of *Salmonella*: laboratory aspects and epidemiological applications. J Appl Bacteriol. **68:**5-16.

Threlfall, E. J., L. R. Ward, and B. Rowe. 1978a. Epidermic spread of a chloramphenicol-resistant strain of *Salmonella typhimurium* phage type 204 in bovine animals in Britain. Vet Rec. **103:**438-440.

Threlfall, E. J., L. R. Ward, and B. Rowe. 1978b. Spread of multiresistant strains of *Salmonella typhimurium* phage types 204 and 193 in Britain. Br Med J. **2:**997.

Todd, E. 1990. Foodborne illness. The lancet. **336:**788-790.

Tompkins, L. S., N. Troup, A. Labaigne-Roussel, and M. L. Cohen. 1986. Cloned, random chromosomal sequences as probes to identify *Salmonella* species. J Infect Dis. **154:**156-162.

Tribe, I. G., H. Tsimogiannis, P. Mmolawa, and D. Davos. 2001. An outbreak of *Salmonella* Typhimurium PT 29 linked to a restaurant in South Australia. Communicable Disease Intelligence. **25.**

Trun, N. J., and T. J. Silhavy. 1987. Characterization and in vivo cloning of *prlC*, a suppressor of signal sequence mutations in *Escherichia coli* K12. Genetics. **116:**513-521.

Tsaneva, I. R., and S. C. West. 1994. Targeted versus non-targeted DNA helicase activity of the RuvA and RuvB proteins of *Escherichia coli*. J Biol Chem. **269:**26552-26558.

Tschape, H., R. Prager, W. Streckel, A. Fruth, E. Tietze, and G. Bohme. 1995. Verotoxinogenic *Citrobacter freundii* associated with severe gastroenteritis and cases of haemolytic uraemic syndrome in a nursery school: green butter as the infection source. Epidemiol Infect. **114:**441-450.

Tuma, R., P. E. Prevelige, Jr., and G. J. Thomas, Jr. 1998. Mechanism of capsid maturation in a double-stranded DNA virus. Proc Natl Acad Sci U S A. **95:**9885-9890.

van Belkum, A. 1997. Amplification-based DNA fingerprinting: from artifactual to definitive typing and in between. J Clin Microbiol. **35:**3008-3009.

van Belkum, A., M. Struelens, A. de Visser, H. Verbrugh, and M. Tibayrenc. 2001. Role of genomic typing in taxonomy, evolutionary genetics and microbial epidemiology. Clin Microbiol Rev. **14:**547-560.

van der Wilk, F., A. M. Dullemans, M. Verbeek, and J. F. van den Heuvel. 1999. Isolation and characterization of *AP*SE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrthosiphon pisum*. Virology. **262:**104-113.

Vander Byl, C., and A. M. Kropinski. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J Bacteriol. **182:** 6472-6481.

Vaneechoutte, M. 1996. Dna fingerprinting techniques for microorganisms. A proposal for classification and nomenclature. Mol Biotechnol. **6:**115-142.

Verma, N. K., D. J. Verma, P. Huan, and A. A. Lindberg. 1999. Cloning and sequencing of the glucosyl transferase-encoding gene from converting bacteriophage X (SfX) of *Shigella flexneri*. Gene. **129:**99-101.

Vogel, W., and H. Schmieger. 1986. Selection of bacterial *pac* sites recognized by *Salmonella* phage P22. Mol Gen Genet. **205:**563-567.

von Heijne, G. 1981. On the hydrophobic nature of signal sequences. Eur J Biochem. **116:**419-422.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman and M. Kuiper 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:**4407-4414.

Waldor, M. K. 1998. Bacteriophage biology and bacterial virulence. Trends Microbiol. **6:**295-296.

Walker, G. C. 1987. The SOS response of *Escherichia coli*, pp. 1346-1357, *Escherichia coli* and *Salmonella typhimurium*. American society for microbiology, Washington, D.C.

Ward, L. R., J. D. de Sa, and B. Rowe. 1987. A phage-typing scheme for *Salmonella enteritidis*. Epidemiol Infect. **99:**291-294.

Watanabe, T., Y. Ogata, R. K. Chan, and D. Botstein. 1972. Specialized transduction of tetracycline resistance by phage P22 in *Salmonella typhimurium*. I. Transduction of R factor by phage P22. Virology. **50:**874-882.

Watanabe, T., and M. Watanabe. 1959. Transduction of streptomycin resistance in *Salmonella typhimurium*. J Gen Microbiol. **21:**16-29.

Watarai, M., T. Tobe, M. Yoshikawa, and C. Sasakawa. 1995. Contact of *Shigella* with host cells triggers release of Ipa invasins and is an essential function of invasiveness. Embo J. **14:**2461-2470.

Watson, J. D. and F. H. C. Chase. 1953. The structure of DNA. Cold Spring Harbor Symp. Quant. Biol. 18. 123.

Wegener, H. C., D. L. Baggesen, and K. Gaarslev. 1994. *Salmonella typhimurium* phage types from human salmonellosis in Denmark 1988-1993. APMIS. **102:**521-525

Weide-Botjes, M., B. Kobe, and S. Schwarz. 1998. Inter- and intra-phage type differentiation of *Salmonella enterica* subsp. enterica serovar enteritidis isolates using molecular typing methods. Zentralbl Bakteriol. **288:**181-193.

Weintraub, A., B. Johnson, B. A. Stocker, and A. A. Lindberg. 1992. Structural and immunochemical studies of the lipopolysaccharides of *Salmonella* strains with both antigen O4 and antigen O9. J Bacteriol. **174:**1916-1922.

Williams Smith, H., and M. A. Linggood. 1971. The transmissible nature of enterotoxin production in a human enteropathogenic strain of *Escherichia coli*. J Med Microbiol. **4:**301-305.

Winston, F., and D. Botstein. 1981. Control of lysogenization by P22. II. Mutations (*clyA*) in the *c1* gene that cause increased lysogenization. J Mol biol. **152:**233.

Wray, C., I. M. McLaren, and Y. E. Jones. 1998. The epidemiology of *Salmonella typhimurium* in cattle: plasmid profile analysis of definitive phage type (DT) 204c. J Med Microbiol. **47:**483-487.

Yee, A. J., S. De Grandis, and C. L. Gyles. 1993. Mitomycin-induced synthesis of a Shiga-like toxin from enteropathogenic *Escherichia coli* H.I.8. Infect Immun. **61:**4510-4513.

Youderian, P., S. J. Chadwick, and M. M. Susskind. 1982. Autogenous regulation by the bacteriophage P22 arc gene product. J Mol Biol. **154:**449-464.

Youderian, P., and M. M. Susskind. 1980. Bacteriophage P22 proteins specified by the region between genes *9* and *erf*. Virology. **107:**270-282.

Young, B. G., Y. Fukazawa, and P. Hartman. 1964. A P22 bacteriophage mutant defective in antigen conversion. Virology. **23:**279-283.

Young, R. 1992. Bacteriophage lysis: mechanism and regulation. Microbiol Rev. **56:**430-481.

Young, R., and U. Blasi. 1995. Holins: form and function in bacteriophage lysis. FEMS Microbiol Rev. **17:**191-205.

Zabriskie, J. B. 1964. The role of temperate bacteriophage in the reproduction of erythrogenic toxin ragment A streptococci. J Exp Med. **119:**761-780.

Zhang, N., and R. Young. 1999. Complementation and characterization of the nested *Rz* and *Rz1* reading frames in the genome of bacteriophage lambda. Mol Gen Genet. **262:**659-667.

Zinder, N. D. 1957. Lysogenic conversion in *Salmonella typhimurium*. Science. **126:**1237.

Zinder, N. D., and J. Lederberg. 1952. Genetic exhange in *Salmonella*. J Bact. **64:**679-699.

Zissler, J., E. R. Singer, and F. Shaefer. 1971. The role of recombination in the growth of bacteriophage lambda. I. The gamma gene, p 469, The bacteriophage lambda. Cold Spring Harbor Laboratory, New York.

Appendices

APPENDIX I

Oligonucleotide primers used for sequencing ST64T genome

Primer	DNA sequence	Application
TFLR	5' –GGC GTA GGT GCG TGG CGT CTG G –3'	Sequencing of bacteriophage ST64T (normal)
TFKR	5' –CGC CGA CAT CGT GGC CCG CAC C –3'	"
TFJR	5'–CAG TTA AAC GCG GAG AAT TGC CGG –3'	"
TFHR	5' –GCA ATG GTT CAG GCT CAG GGC –3'	"
TFGR	5' –TAC CGG TGA TGT CAG CGC AGC –3'	"
TFGFR	5' – CGT ACT TCT TCA ACC CAC AGG – 3'	"
TRGCR	5' –CGG CCT TAA CGG TAT CTT CGC –3'	"
TFGAR	5' –GAT GGC TGA GTG GTA TCA GGA CGG –3'	"
TRGA	5' –GCT GAT CGC AAT CAG GCA ATG GG –3'	"
TRGB	5' –GGT CGT TCA GCT TGC AGA TGG –3'	"
TRGC	5' –GCA GGT TGA CTA CTC AAG CCC –3'	"
TRGD	5' –TCA GCC ATC GCC TCG CGT CAC G –3'	"
TRGE	5' –GCA CTA TGC GAG CTA ATC AGC –3'	"
TRGF	5' –GAA TCT GTA CCG GTC AGT CGG –3'	"
TRH	5' –AGT CTA CCA ACT ACG ATG CCG –3'	"
TRH2	5' –GTA TCC GAT TAC AGG CGT AGG C –3'	"
TRJ	5' –CCG CTA ATT ATC AAC GCA GCC GG –3'	"
TRK	5' –ATG GGT TAT CAA GCC TTG GAC GG –3'	"
TRL	5' –CCA AGC GGA GGT AAA GAT GGC –3'	"
TRM	5' –CTC TGG CGC GTA CCT ACT CAC C –3'	"
TRN	5' –CCA TTT GTC CCT GAC ATC GCG –3'	"
TRO	5' –ACC AAT CAA TAC CCT CGG TGC G –3'	"
TRP	5' –ATG CTG TGA TGT AGT CAT CTC C –3'	"
TRQ	5' –TTG AAT TTC GAA TCA CCG GCC –3'	"
TFAS	5' – CGG AAA GGT CTG GAG TGT AGC –3'	"
TFAT	5' –GCG CAT CTT CCT CAC TAC AGG –3'	"
TFAU	5' –CAT GAA GCT CTG TGG CCT GCC –3'	"
TFAW	5' –GCG ATG TCG TAG AAT TTC CCG –3'	"
TFAX	5' –CTT AAC ATC CTC ACC CTC TAC C –3'	"
TFAY	5' –CAT CGT CGT ATG TAG CAA CG –3'	"
TFAZ	5' –GCA GGC GTT ATA GCC TTG CGC –3'	"
TFAZA	5' –GCG TAT CCG GTT GCG TTC TGG –3'	"
TFZB	5' –GTA TTT GCG CAG GTC ATA CAG G –3'	"
TFZC	5' –CCG ACA GAA TGC TTT CCA CGC –3'	"
TFZD	5' –AGG TTG ACC TTC CGC GCC TCC –3'	"
TFZE	5' –CCG CAT AAT CAC TAT GAC GCA CC –3'	"
TFZF	5' –CCT CCG ATA CGG AAT TGC CGG –3'	"
TFZG	5' –GGC AAG CTA CCA TGC GAC CGC –3'	"
TFZH	5' –CAT TGC GCC GGA GCG ACC ACC –3'	"
TFZI	5' –GGC ACT CCT GGT TGA TTC AGG –3'	"
TFZJ	5' –TTG CTC ACA TAG CAG ACT CG –3'	"

TFZK	5'–CTC TCA CGC ATT GAG GTA ATC C –3'	Sequencing of bacteriophage ST64T (normal)
TFZL	5' –CCG CGC TGC TTC TGT ACG GCG –3'	"
TFZL2	5' –TTT CTG CGC CGC TCA TGA CGG –3'	"
TFZM	5' –TAG CCA GGC GAC TAA CCC TGA –3'	"
TFZN	5' –ATG ACA ACA CCG ATA ATG CGG –3'	"
TFZO	5' –GGC AAT TTC CGT TTA TGG GG –3	"
TFZO2	5' –GCG GCC ATT AAG CCA ATG ACC –3'	"
TFZP	5' –TAT CCC AGG CCG TTC TGG CCG –3'	"
TARUR2	5' –GCA TGA TGC TTG CGG TTC TGG –3'	"
TARRR	5' –ACA GCT AAT TAC CGG ACG GCC –3'	"
TARRR2	5' – CGC GCC GCC ACA TAG CAT CGA GG – 3'	"
TARPR	5' –CCC GTC AGG CGA ACA ACG AGG –3'	"
TARPR2	5' –CTT GCC TAC GGC ATG ATC ACC –3'	"
TAROR	5' –CCC AGC ACA CTG ATG GAG AGG –3'	"
TARPR3	5' –GCA CGG TGT ATT GCA TTC AGG C –3'	"
TARPR4	5' –TAA GCC TTC TTG GTT ACG CC –3'	"
TARPR5	5' –AGG ATG AGT TCG ACA AAG GC –3'	"
TAROR3	5' –CGT ACC AGG TAT TGA TGG CGG –3'	"
TARJR	5' –GTT AAT GGC AGA CCC TAA CGG –3'	"
TARJR2	5'–GGC ACA TTA CCG GAT GGC TGG –3'	"
TARJR3	5' –CTG TCA GGC GAT TAA AGG CGG –3'	"
TARJR4	5' –CTA AAG AGC AGC ATG ACT GGC –3'	"
TARHR2	5' –TCG AGT CCA CGC TGT CCC GCC –3'	"
TARHR3	5' –TGA TAA GCG TTG AGG GTG GGC –3'	"
TARER	5' –AAG CAT GAT CCG TCA CTG CCC –3'	"
TARER3	5' –GGA AGA GTG GGA TCA GGC TTG C–3'	"
TARER4	5' –AAC GCA GAA TTC GAC GGC CC –3'	"
TARER5	5' –GCA ACC GAT TCT GGG AGG CCC –5'	"
TARER6	5' –TGG ACG CGA TTC TTG ATA ACC –5'	"
T1R	5' –GGG ATG TCT GAC GTT GTT ACC –3'	"
T2R2	5' –CCA ACG ATG ATG TGC TCG ATG C –3'	"
TFM	5' – GGT ATG TAT CGT ACA GAC ATG CG –3'	Sequencing of bacteriophage ST64T (complementary)
TFL	5' – GGC GAG CAT ATC CCA ATT GTT –3'	"
TFK	5' – GGT AGC GAC GCT CGG CGT TGA – 3'	"
TFJ	5' – CGC CTG AAG AGC AGC AAT GGT –3'	"
TFI	5' – GGC GAT CAT GCA GCG GCA TCT – 3'	"
TFH	5' – GCG CCT GTT CCC GGA GAA GTC –3'	"
TFGF	5' – CGT CGC GGT AAA CAT GGG AGA G –3'	"
TFGE	5' – CCT GAC TAA GCG CGA TAT CTT C –3'	"
TRGD	5' – CGG CCT TAA CGG TAT CTT CGC –3'	"
TRGC	5' – CGG CCT TAA CGG TAT CTT CGC –3'	"
TRGB	5' –GCG TAA GCT TGG TGT AGC ATC A – 3'	"
TRGA	5' – CGG GTT ATT TAC GCT CAT TCC C –3'	"
TFF	5' – CGC GCA GCC ATC GTT AAT GGT – 3'	"
TFE	5' – GCT GAC CGC CTG TTC CTG TCT – 3'	"

TFD	5' – GCC TGT CGT GGT TCA GGC TAC – 3'	Sequencing of bacteriophage ST64T (complementary)
TFC	5' – GCC AAT GCT CTA ATT GCT GAG C –3'	"
TFB	5' – GCC CTA GCC CGA TGA TGT AGA CG –3'	"
TFA	5' – GCA GCA GGC CAG GCA TCG GCA GG –3'	"
TRGAR	5' – CGA AGG TAT TGA CGA TAA CGG C – 3'	"
TRGCR2	5' – GGA CCG GCG CAG CCA CAG CAA – 3'	"
TRLR7	5' – GCC ATC GCC TCG CGT CAC GAC –3'	"
TRLR6	5' – GGG AAC CGT CTT GAG CAA GTT – 3'	"
TRLR5	5' – GGA CGC GCT CAG ATG CGT AAC – 3'	"
TRHR2	5' – GGC TAT GAC CTG ATT AGC ACT GG – 3'	"
TRLR3	5' – GCT CAG CCG CTA ATT ATC AAC –3'	"
TRHR	5' – CTC ACG TCC AGA TTG CTC AGC C –3'	"
TRKR	5' – GGC GGA CGA ACC AGC TAT GGA – 3'	"
TRMR2	5' – GGC GCG TAC CTA CTC ACC CAT – 3'	"
TRMR	5' – GCG CAG CGA TTA AAC TGC GGA – 3'	"
TRNR2	5' – GGT AAC ACA CCT GAC GCT GTA – 3'	"
TRNR	5' – CCC CTA CAA CAC CAA TAT CG – 3'	"
TROR2	5' – GCG CAA AGA ATA GCG AGT AG – 3'	"
TROR	5' – GCG TAA GTA TTC AGC GCA GCC – 3'	"
TRRR3	5' – CCA TTG CCT GAC CAACCC AAC –3'	"
TRRR2	5' – CCG CAC GGC TCA TTA ACC GAA – 3'	"
TRRR	5' – GCG AAG CGT GGT GGT GGT GGC – 3'	"
TFATR2	5' – CGA GCA TGT TCA GTC AGG TGA –3'	"
TFATR	5' – CGC GTT CAG CAG CCC GGA GGA – 3'	"
TFAWR2	5' – CCC GCG CTC ACC TGG ACA GTA – 3'	"
TFAWR	5' – CCA CGG CCC GAT TGA CCA CCA – 3'	"
TFAXR2	5' – CCC TCT ACC TGT AGT AGT CG –3'	"
TFAXR	5' – GGG AGG ATC TGG TAG CTG CAT –3'	"
TFAYR	5' – GGG AGG ATC TGG TAG CTG CAT –3'	"
TFZAR2	5' – CGG TTG CGT TCT GGT GCA GGC – 3'	"
TFZAR	5' – GTA TTT GCG CAG GTC ATA CAG G – 3'	"
TFZBR	5' – CCG ACA GAA TGC TTT CCA CGC – 3'	"
TFZCR	5' – GGT TGA CCT TCC GCG CCT CCA –3'	"
TFZDR	5' – CCG CGT CAC GTT TCT CAA CAG – 3'	"
TFZER	5' – GGG TTT GAA GCG GTG TCG AAT – 3'	"
TFZFR	5' – GCC TGA TGA GAT AAG CGC C – 3'	"
TFZHR2	5' – CCT GAA CGC ATC TTG TCT GCG C – 3'	"
TFZHR	5' – CGC CAC CGT TGT TGG GAA TAT –3'	"
TFZIR	5' – CGA TAC CCA TGA CGG CAC CGC – 3'	"
TFZJR	5' – CCG GGC GTT CAC CTT TAC CAA – 3'	"
TFZKR	5' – CCT GAG GAT GCT CTT CGA ACT – 3'	"
TARY	5' – GCC TTC AGG GCG ACG CTT TAA – 3'	"
TARW	5' – CGG TTC GAT AGC CGG ATC AAC – 3'	"
TARV	5' – CGA ATC AGG GCG ACG CCA TCC – 3'	"
TARU	5' – CCA GCT TCA AAC AGC CTG GTC – 3'	"
TART	5' – CCG ACA TTG GTG AGA TGT TGG – 3'	"

TARR	5' – GGT TCA CGA TGC AGG ACT GGA – 3'	Sequencing of bacteriophage ST64T (complementary)
TARQ	5' – CCC GTC AGG CGA ACA ACG AGG A – 3'	"
TARP	5' – CCA AGT GAC TCA CGA GAT ACA GG – 3'	"
TARO	5' – CCT TGA AGA GTC AGG AGC ACG – 3'	"
TARN	5' – GGC TTC AGC AGT ATG GCT GTC – 3'	"
TARM	5' – GCT TCA TGA TGT CTC CCG TCA –3'	"
TARL	5' – GGT GTA TCA GAT GTG ATC GGT C – 3'	"
TARK	5' – GCA TGT GCT GGA AGT TCA CGT G – 3'	"
TARJ	5' – CGC GTG TCA CGA TGA AAT CGA – 3'	"
TARI	5' – GCC AGG CGA TTG TCG TGC CAT – 3'	"
TARH	5' – GCG CAT GAG CAG CGT AAT AGC – 3'	"
TARG	5' – CCC GCT GGC AGG GAT TGC AGG – 3'	"
TARF	5' – GCG AGA AGA ACC TGA ACG GAG – 3'	"
TARE	5' – GGA TGA ATT GGC GCG AGC GGA – 3'	"
TARD	5' – CCG GAG CCA ACA CAA TGG CAG – 3'	"
TARC	5' – CCC AAA CTT CAC GCC ATC ATG GCC – 3'	"
TARB	5' – GCC TAT TGC CAA GAT GAG GGC – 3'	"
TAR	5' – CGC ATC CTG TGC GGG CGT ATC – 3'	"
TFO	5' – GGA AGA AGT CTC CGT TCT TCC C – 3'	"
TFN	5' – GGC CTC ATG ACG GTC ACC AAC – 3'	"

APPENDIX II

Oligonucleotide primers used for sequencing ST64B genome

Primer	DNA sequence	Application
F3H	5'– GTT CTA TGC TGG GCC TTG ACC –3'	Sequencing of bacteriophage ST64B (normal)
F3I	5'– GGT GCG GAA GTA TAC TCC GGC –3'	"
B2	5'– GAC GCG AAG TGA TTG AGA TGC –3'	"
F2C	5'– CGC TAC ATC CTT GAA GA TGCC –3'	"
F2D	5'– TGA CCA CAT TGA GTC CTA CGG –3'	"
F2E	5'– TGG TGG TCA CCT GTC TGT GCC –3'	"
F2F	5'– TTG CAC ATG GTG CAG AAC ACC –3'	"
F2F2	5'– TAA ATC TGT AAG TGA CTC CGG –3'	"
F2I	5'– GCG TAA GGC GTT GCA GGT GGG C –3'	"
F2J	5'– TAA GGC GCA TCG TAC CGG CGC G –3'	"
F6A	5'– TGC ACA TGC GTG ATG CGC GGC –3'	"
F6B	5'– ACC GTG TGG GTG GCG CAG CGC –3'	"
F6C	5'– GCG ACA TCG AGC ACC ACA CCG –3'	"
F4RA	5'– GGT GTC ATC GGT TGA TTA TGC CCG –3'	"
F4RB	5'– GGC ATC GGT GAA CTC GAT GGC –3'	"
F4RC	5'– TGT GGA AAG CGG CGT GCT GCG –3'	"
F4RD	5'– ACA GCA ACT GTC GGT AAC CGG –3'	"
F4RE	5'– CAT ACT TCT GGC GAA TAA ACC C –3'	"
B5A	5'– CGG CGC GGC GCT GCT GGC –3'	"
B5	5'– ATG ACG CTG ACA ATT GGC GTG G –3'	"
F5C	5'– TTG GTG CCA TTG AAA CCG CAG G –3'	"
F5D	5'– AGC CTG GCG CTC AGG TGA CAG GC –3'	"
F5E	5'– GCG GAC GGA TCT CAA TCG CG –3'	"
F5RA	5'– ACG GCT GGT TCT CGG CAG GCC G –3'	"
F5RB	5'– CGC CGA AGT GAC GTA CAG CCA GG –3'	"
F5RC	5'– CGG AAT AGT GGT AAA TGG CGG CGG –3'	"
F5RD	5'– CCG GGC GGA GCG GGA TGA CAG G –3'	"
F5RE	5'– GCC GGG TAC AGG CTG CTG CCG –3'	"
F5RF	5'– GGC GTT ACG CCT TGG CAC GCC –3'	"
F5RG	5'– GCA CCG TTG GTG TGA TGG TGG C –3'	"
F5RH	5'– GCC TGA TTC CTG TGC ACC GG –3'	"
F5RI	5'– AGC TTT GTC CGT CAC ATA CGG –3'	"
F5RLJ	5'– CCT GAA CAG GCG GCA ATC AGC G –3'	"
F5RLK	5'– TAA TGG TGC CGT ACG CGC AGG C –3'	"
F5RLL	5'– GAC TAC CTG GAA CAA TGC AGG –3'	"
F5RLM	5'– TGC CCT GAT GAT TGC CGT GGC G –3'	"
F5RLN	5'– TTT GCC GAA ACA TTG CTC GCG –3'	"
F5RLO	5'– CCC AAG CGT TTA TCA TGG GAC G –3'	"
F5RLP	5'– GGC AGA GTA ATA GCT GGA GTT GC –3'	"
F5RLP2	5'– GGT TGT GGC ATC CGG TGA ACC –3'	"
F5RLS	5'– GGA AAG TGC AGT CAA GAG GAC G –3'	"
F5RLT	5'– GAT CAG CAT AAT CGG CAC TGG C –3'	"
F5RLU	5'– CCA TCG TGC GGA CGT CAT GAT CG –3'	"

F5RLW	5'– GCT TTC GCA TTC TTC GTC GGC –3'	Sequencing of bacteriophage ST64B (normal)
F5RLX	5'– GCA CCA TGA AGC AGG GCA GCG CG –3'	"
F5RLY	5'– CAA ATA GTC CAT TGC GGC ACC G –3'	"
F5RLZ	5'– ATT CGA CTG GCA GAC ATG CGC –3'	"
F5RZA	5'– CCG GTG CGC GCG TCG AAT ACC –3'	"
F5RZB	5'– GCG GAT CAG TTG ATC GGT ACG –3'	"
F32	5'– GGC GTT CAG TTG CCA GCA TGG –3'	"
F33	5'– TTG CTG ATA AGC AGG TCG CGG –3'	"
F34	5'– GCG AGC TGT CAT ATG GTC AGC –3'	"
F35	5'– ACT ACC TTG CCT TTA TTG CCG –3'	"
F51	5'– CTG TTC GCT GCT CTA TGA GTG C –3'	"
F52	5'– TGA GTC CAT TAC CGG CGC TTG C –3'	"
F53	5'– TCC ATC GGT GCT GTG TTC TTC G –3'	"
F3LSR1	5'– AGA GCT TGC TGA GCG CCT GGC –3'	"
F3LSR2	5'– GTG CAG CTC CGG GCG CCG TGG CG –3'	"
F3LSR4	5'– GAC GCT GTG GCG CTT GCG AGC C –3'	"
F3LSR	5'– GGG TAG TGA AGG AGA TTC ACC –3'	"
F3NR	5'– GCA GGC CTG GTT GGC ATA TCC –3'	"
FS3	5'– TTA TGG GCT GTT CCT GCT GCA CC –3'	"
F3LMR	5'– TCG GAA CCA ACG ACC CGC AGG –3'	"
F3LLR1	5'– CAC CGG CAT GAC CCA CAT CAT GG –3'	"
F3LLR2	5'– TAA GTC ATG GCG CGG CGA GCG –3'	"
F3LJ	5'– TCG CGG TGG TGC TGA ACA AGG –3'	"
F3LJR2	5'– CTG CTG AAG CTG CTC GAT ACC G –3'	"
RF13B	5'– CTA TCG TGC TGG TGG TCT GGC –3'	"
RF13C	5'– GGG TTG ATC CGG AAT CGC CGG –3'	"
RF13D	5'– AAT GAC ACC ACG TCA ACG CCG –3'	"
RF13E	5'– TTA ACC CTG GCA TGA CGC CCG –3'	"
RF13F	5'– AAG CCG CCC TTG AGG AAC GGC –3'	"
3LR	5'– GGT GAA TAT CTG TCC GAT TCG –3'	"
B3	5'– TCA GCA GGC GGC TGG TAT TAG C –3'	"
F3	5'– CGG TGC TGG TAT GAA TCT GTT –3'	"
F3D	5'– AGG ATG ACC CTT GAG GAT GCC –3'	"
F3E	5'– GCT TCA CGT TGC TCG TGA GCG –3'	"
F3F	5'– CAA CTG GTG TGC CGA AGC GGG –3'	"
F3G	5'– TCA AAT CCC TGT GAC CTG ACG –3'	"
		Sequencing of bacteriophage ST64B (complementary)
B3R	5'– GGC CTG CCA GCG TCA CAT TGA –3'	"
F2RF	5'– GGC GTT CGA TCT GGC TCG TAA –3'	"
F2RE	5'– CGC TGT CTG ACC ACA TTG AGT –3'	"
F2RD	5'– GGG AGC TGG TGG TCA CCT GTC –3'	"
F2RC	5'– GGA CAG CCA GTT CCT GGA AAC –3'	"
F2RB	5'– GGC TCC GTC TTT GGT GTA AAG –3'	"
F2RA	5'– GCT GAG GAT GTG GGC TCT GCA –3'	"
B2R	5'– GCC GGA TAT CGC CGC CGA TGC GA –3'	"
F6RC	5'– ACC AGC GCA TTG TTA TCC GGC –3'	"
F6RB	5'– CGC GGC GAC TGG TCA TGA AAC G –3'	"
F6R	5'– CCA GCC GCG TAT GTT GTG CC –3'	"
B4	5'– CCC GCG TTC AGG CTC CCG TGA CC –3'	"
B4F	5'– CGA CGA CTG AAC AGC AGA CG –3'	"

F4D	5'– GCA ACT GTC GGT AAC CGG AGG –3'	Sequencing of bacteriophage ST64B (complementary)
F4E	5'– CCC GGT CGG AGA CGG AGG AGG –3'	"
F4F	5'– GGC GGG CAG GCA GGC ATT GCC –3'	"
F4G	5'– CAG GCC GCC GTC ATG AAT GCG –3'	"
F5RE	5'– GCG TGA ACG CGG GTA TAA CGG –3'	"
F5R	5'– CGC AGA CAG GCG GGA TCG TCT –3'	"
B5R	5'– GGT ATC ACG CTT GAT CCC GGT –3'	"
RF15E	5'– GCA GCC ACG GCA CTG GTA CTT G –3'	"
RF15D	5'– CGC GCG CAG TCA TTA CCG CGA –3'	"
RF15C	5'– GCC GAT CGA ATC CAC TGG CGA –3'	"
RF15B	5'– GGC GTG GCG GCA CGT ATT GAT G –3'	"
F15E	5'– GGA TGG TAT TGC CGG AAC GCC G –3'	"
F5RGL	5'– GCA CCG TTG GTG TGA TGG TGG C–3'	"
5RL2	5'– CCG TTA TGA ACA GCT GTA TGG C –3'	"
F5LIR2	5'– CGA ATG GTG ATC CCG CCA CGG GC –3'	"
F5LIR	5'– CGC AGG CTA CAT CAA CTG CGG –3'	"
F5LLR2	5'– GGT TCA CCT GTT GGA ATA ACC GCC –3'	"
F5LLR	5'– CCG GGC TGC ACA TGT CAG TAG –3'	"
F5LMR	5'– CGG GAG TCA GTG GTA TAC CGT –3'	"
F5LNR	5'– CGG GAG TCA GTG GTA TAC CGT –3'	"
F5LOR	5'– GCC ACC TCC AGG TTG CAG GTG A –3'	"
5PR2	5'– CGT GCC ATA GCT ATC TGA AAC G –3'	"
F5LOR2	5'– GCT GTC GAT ACA CGG GCT GTA C –3'	"
F3ZD	5'– CGT GCG GAC GTC ATG ATC GTT CC –3'	"
F3ZC	5'– GCC TTA ATC GTC GTC CAT CTG A –3'	"
F3ZB	5'– GCC TCT TCC GGC ACT ACC GGC GC –3'	"
F3ZA	5'– CCT CAG ACT CAG CCA GTC CGT G –3'	"
F3LZ	5'– GGT GAG ACG CTG CTT ACC GTC –3'	"
F3LY	5'– CCG ATC CGT CCA GGG TGT ACA –3'	"
F3LX	5'– GCG TCG TGC ATC AGG GCT TCA AA –3'	"
F3LW	5'– GGC GAT CGC TTC TTC CTG CGC TTC –3'	"
F3LV	5'– GGT GGT AGT CAT GGC GCC GCG –3'	"
F3LU	5'– GGC ACT ACG GTC ATG CGG TCA TC –3'	"
F3LST	5'– GCG AAT CAT CCG GTC ATT CGT A –3'	"
F3LP	5'– CCG CCC TGA ATG AGT CCA TTA CC –3'	"
F3LQ	5'– CGA TTG GGT GGG CGC TGG TA –3'	"
F3LQ2	5'– GGA GTT GTA TCA GGC GAT CGC –3'	"
F3LO	5'– CCA TCG GTG CTG AGC GAT GCA G –3'	"
F3LN	5'– GCC AGC GAT CAG GGT GAG TGC –3'	"
F3LM	5'– CGT GAA ACA GGC CAC CAC GTT –3'	"
F3KL	5'– GGT GTT TGC GGA AGA GCA GAA –3'	"
F3IL	5'– GGC CGC TTG AGG TTG GTC TGG T –3'	"
F3GL	5'– CGC CGC GAC GTT CAT GGG CGC –3'	"
F3HL	5'– GGA TCA GCC CGT CTT TCC ATC –3'	"
F3F	5'– CCG GCT GGC GCG CTG GAT GGT T –3'	"
F3E	5'– GCG GCG CGG TCC CGG CGG TAT C –3'	"
F3D	5'– CCC TCG GTG CCG GCC ACC AGC –3'	"
F3C	5'– GCC GCG CCG CGC AAG AGC TGG –3'	"
F3B	5'– CCG TGT GCG ATG GTG TGT GAA –3'	"
F3A	5'– CGC TTC CGA GAA ATC GTC AAC –3'	"

B3	5'– TCA GCA GGC GGC TGG TAT TAG C –3'	Sequencing of bacteriophage ST64B (complementary)
F31	5'– CGG TGC TGG TAT GAA TCT GTT –3'	"
F3C	5'– AGG ATG ACC CTT GAG GAT GCC –3'	"
F3E	5'– GCT TCA CGT TGC TCG TGA GCG –3'	"
F3F	5'– CAA CTG GTG TGC CGA AGC GGG –3'	"
F3G	5'– TCA AAT CCC TGT GAC CTG ACG –3'	"