Genome Sequence of Bacteriophage ϕAR29: a Basis for integrative Plasmid Vectors

Shawn Ginn Ming Seet

B. Sc., Hons. Biotech., MU.

This thesis is presented for the degree of

Doctor of Philosophy

at Murdoch University

March, 2005

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

_____

Shawn Ginn Ming Seet

ACKNOWLEDGEMENTS

I would firstly like to thank Prof. Keith Gregg for his never ending support throughout my study of philosophy. During these last four years, Keith has performed beyond his means to provide funds for the continuation of his students' research as well as consistently meeting datelines. His guidance has also allowed me to open up my mind and think outside the box. For all that his has done, I will always be indebted to him.

I would also like to say thanks to my partner, Vera Limadinata, for taking this long journey with me. Her presence has been a comforting light in times of darkness. Her patience, humour and love have, and always will, put a smile on my face.

My deepest thanks go to my family, especially to my mother, for being the pillar of strength and continuos love. I would also like to thank them for understanding my absence around house and showing me that family runs deeper than blood.

Being the last member of Rumen Biotech, I like to give a special big thankyou to Tobias Schoep for standing beside me as long as he can. I never forget your encouragements, brain storming sessions and much laughter. Finally, thanks to Scott McNeil, Tom La and Andrew Low for helping me trouble shoot some of my difficulties and providing many funny jokes.

Lastly, sincere thanks to all SABC people for making the night-life in the lab enjoyable.

Table of Content

Abstract ... 1
Commonly Used Abbreviations ... 2
Chapter 1: Introduction ... 3
1.0: Genetic Manipulation of Rumen Bacteria ... 3
1.1: Criteria for Genetic Manipulation ... 5
1.2: Difficulties in Genetic Manipulation of Rumen Bacteria ... 6
1.3: Transformation Systems ... 7
1.4: Bacteriophages and their Life Cycle ... 11
1.4.1: Lytic Cycle ... 11
1.4.2: Lysogenic Cycle ... 13
1.4.3: Mechanism of Phage λ and Lysogenic Cycle ... 14
1.4.3.1: The induction of phage λ lysogenic life cycle ... 16
1.4.3.1.1: Transcription of the Int gene ... 16
1.4.3.1.2: Maintenance of λ prophage ... 17
1.4.3.2: The role of protein Cro in the phage λ lytic cycle ... 19
1.4.3.2.1: Activation of the lytic cycle from the lysogenic state ... 19
1.4.3.2.2: Lytic cycle from infection ... 20
1.4.4: Phage Integration ... 21
1.4.4.1: attB site ... 23
1.4.5: Excision of Prophage DNA from the Bacterial Chromosome ... 26
1.5: Phage in the Rumen ... 28
1.6: Bacteriophage as Transformation Tools ... 29
1.6.1: Transformation Systems That are Based on Phage Integrative Recombination Process 30
1.6.1.1 Staphylococcus ... 31
1.6.1.2 Mycobacterium ... 32
1.6.1.3 Streptomyces ... 33
1.6.1.4 Pseudomonas ... 34
1.6.1.5 Streptococcus ... 35
1.6.1.6 Listeria ... 36
1.6.1.7 Enterococcus ... 36
1.6.1.8 Rhizobium ... 36
1.6.2 Current developments in integrative vectors ... 37
1.7: Ruminal *Bacteroides* ... 43
1.7.1: Characteristics of *Bacteroides* ... 43
1.7.2: *Bacteroides thetaiotaomicron* Strain AR29 and *Bacteroides uniformis* Strain AR20 ... 43
1.7.3: Difficulties in Transforming *B. thetaiotaomicron* AR29 and *B. uniformis* AR20 ... 44
1.8: Rumen Bacteriophage ϕAR29 and Development of a Transformation System for AR29 and AR20 ... 45
1.9: Summary ... 48
1.10: Aim of This Project ... 48
Chapter 2: Materials and Methods ... 50
2.0: General Chemicals ... 50
2.1: Specialised Chemicals, Enzymes and Laboratory Equipments ... 50
2.2: *In silico* Analysis ... 52
2.3: Bacterial Strains ... 53
2.4: Commonly Used Solutions ... 55
2.5: Bacterial Growth Media ... 56
2.6: General Methods ... 58
2.6.1: Agarose Gel Electrophoresis ... 58
2.6.2: Ethanol Precipitation ... 59
2.6.3: Phenol:Chloroform:Isoamyl Alcohol Extraction ... 59
2.6.4: Polymerase Chain Reaction (PCR) ... 59
2.6.5: Primer Design ... 60
2.6.7: Purification of DNA from Agarose Gels and Solution ... 62
2.6.8: Spectrophotometry for Nucleic Acid Quantification and Monitoring Bacterial Cell Growth ... 63
2.6.9: Nucleic Acid Isolation ... 64
2.6.9.1: Plasmid DNA mini-preparations ... 64
2.6.9.2: RNA isolation ... 65
2.6.9.3: Removal of DNA contaminants from RNA preparations ... 66
2.6.10: Restriction Enzyme reaction ... 67

2.6.11: Dephosphorylation of DNA 5' Terminus 67
2.6.12: Ligation of DNA 67
2.6.13: Preparation Electro-Competent Cells for Bacterial Transformation 67
2.6.13.1: Preparing of Escherichia coli Electro-Competent Cells 67
2.6.13.2: Preparing of Bacteroides Electro-Competent Cells 68
2.6.14: Bacterial Transformation Electroporation 69
2.6.14.1: E. coli Transformation by Electroporation 69
2.6.14.2: B. uniformis Strain* AR20 *and B. thetaiotaomicron Strain* AR29 *Transformation by Electroporation
2.6.15: Screening for Transformed E. coli and Bacteroides 69
2.6.16: Slide Preparation for *Bacteroides* Cells – The Indian Ink Test 70
2.6.17: DNA Sequencing 70
Chapter 3: Expression of Integrase and Excisionase Genes from Plasmid pBA 73
3.0: Introduction 73
3.0.1: Aim 74
3.1.0: Materials and Methods: 74
3.1.1 Bacterial strains and Plasmids 74
3.1.2: RNA isolation 74
3.1.3: Two-Step Reverse Transcription Polymerase Chain Reaction (RT-PCR) 75
3.1.4: Gene Sequence Analysis 76
3.2: Results: 79
3.2.1: Transformation of *Bacteroides thetaiotaomicron* strain AR29 with pBA. 79
3.2.2: Detection of Int and Xis RNA transcripts in *E. coli* strain SCS110. 81
3.2.3:Detection of Int and Xis RNA transcripts in *B. uniformis* strain AR20. 82
3.2.4: Detection of Int and Xis RNA transcripts in *B. thetaiotaomicron* strain AR29 83
***3.2.5:* Sequence analysis of proposed Int and Xis** 85
3.3: Discussion 86
3.3.1 Encapsulation of AR20 and AR29 86
3.3.2 *In silico* DNA sequence analysis of ϕAR29 Int and Xis with other known recombinase gene in the GenBank 87
3.3.3 Detecting the transcripts of Int and Xis gene in SCS110, AR20 and AR29 87
Chapter 4: Characterisation of the Bacteriophage ϕAR29 Genome 90
4.0: Introduction 90
4.0.1: Aim 91
4.2: Material and Methods 92
4.2.1: Bacterial strains 92
4.2.2: Phage Genomic DNA Preparation 92
4.2.3: Testing for the Presence of Cohesive Termini (cos site) 93
4.2.4: Sequencing of the ϕAR29 Genome 96
4.2.5: Predicting ORFs 96
4.2.6: *B. thetaiotaomicron* AR29 Genomic DNA Preparation and Sequencing 97
4.2.7: Predicting the Position of Regulatory Elements in ϕAR29 Genome 98
4.2.8: Predicting Translational Frameshifts in the ϕAR29 Genome 98
4.2.9: Identifying tRNA in ϕAR29 Genome 99
4.2.10: Cloning and Examination of Lytic Activity of ϕAR29 Amidase Gene 99
4.2.11: Western Blots 101
4.3: Results: 102
4.3.1: Nucleotide Sequence of the ϕAR29 Genome 102
4.3.2: Identification of ϕAR29 Open Reading Frames, Functional Prediction of Genome Organization 105
4.3.3: Identifiable ORFs 109
4.3.3.1: ORF5 (Terminase Large subunit) 109
4.3.3.2: ORF16 (DnaC protein) 111
4.3.3.3: ORF 25 (Transcription response regulator protein) 113
4.3.3.4: ORF 31 (cI protein) 114
4.3.3.5: ORF 35 (Integrase) 116
4.3.3.6: ORF40 (N-acetylmuramoyl-L-alanine amidase) 120
4.3.4: ORFs with significant matches to known genes and proteins 125
4.3.4.1: ORF4 125
4.3.4.2: ORF6 125
4.3.4.3: ORF17 125
4.3.4.4: ORF18 126

4.3.4.5: ORF22 ... 126
4.3.4.6: ORF 23 ... 127
4.3.4.7:ORF51 ... 127
4.3.5: Poorly Matched ORFs ... 127
4.3.6: tRNA Gene Prediction ... 128
4.3.7: Frameshift Analysis ... 129
4.3.8: Regulatory Elements in ϕAR29 Genome. ... 129
4.3.9: Comparison of ϕAR29 gene arrangement with other bacteriophages ... 132
4.3.10: Locating and Identifying the Integration Site ... 134
4.3.11: Identifying the Site and Orientation of Integration of ϕAR29 into *B. thetaiotaomicron* AR29 ... 139
4.3.12: Amidase Activity Studies ... 140
4.4: Discussion ... 142
4.4.1: DNA Sequence of ϕAR29 ... 142
4.4.2: Characterisation of the ϕAR29 Genome ... 143
4.4.3: Prediction of a Pseudo-tRNA (Leu) ... 146
4.4.4: Coding Frameshifts ... 150
4.4.5: Regulatory Elements ... 151
4.4.6: *Att* Sites ... 152
4.4.6.1: AttP ... 152
4.4.6.2 AttB site ... 153
4.4.6.3: Core attachment site ... 155
4.4.7: Determining the Function of ORF40 (amidase) ... 156
4.5: Conclusion ... 157
Chapter 5: Chromosomal integration of plasmid pBA in *B. thetaiotaomicron* AR29 ... 159
5.0: Introduction ... 159
5.0.1: Phage Integrases and Attachment (*att*) sites ... 160
5.0.1.1: Serine Integrases ... 160
5.0.1.2: Serine Integrase Attachment ... 164
5.0.1.3: Tyrosine Integrases ... 164
5.0.1.4: Tyrosine Integrase Att site ... 166
5.0.2: Plasmid pBA ... 166
5.0.3 Aim: ... 168
5.1: Materials and Methods ... 168
5.1.1: Bacterial strains and Plasmids, Sequence Analysis and RNA Isolation ... 168
5.1.2: Testing for the Expression of Prophage *Int* Genes from *B. thetaiotaomicron* AR29 ... 168
5.1.3: Testing for pBA Integration ... 169
5.1.3.1: Preparation of samples for PCR analysis ... 169
5.1.3.2: Tests for the presence of integrated plasmid pBA ... 169
5.1.3.3: Detection of free, non-integrated plasmid pBA in AR20 and AR29. ... 169
5.1.3.4: Tests for plasmid pBA integration adjacent to prophage in B. thetaiotaomicron AR29. ... 170
5.1.4: PCR Detection of Unrecombined *B. thetaiotaomicron* AR29 *attB* Site. ... 172
5.1.5: Approach to Curing Lysogenic Phage ϕAR29 from *B. thetaiotaomicron* AR29 ... 173
5.1.5.1: Monitoring the removal of ϕAR29 by determining the presence of intact attP, and integrated phageϕAR29 in AR29 with PCR ... 173
5.2: Results ... 173
5.2.1: Identification of *Int* and *att* site in pBA ... 173
5.2.2:Expression of *Int* from the ϕAR29 Prophage ... 173
5.2.3: Test for Integration of pBA into AR20 and AR29 Genomes ... 175
5.2.4: Characterisation of pBA Integration and Excision of Lysogenic Phage ϕAR29 ... 177
5.2.5: Identification of Potential *attB* Site in Other Bacterial Genomes ... 181
5.3: Discussion ... 184
5.3.1: Presence of *Int* in AR29 Untransformed with pBA ... 184
5.3.2: Absence of Co-Integrated pBA and ϕAR29 ... 185
5.3.3: Detection of intact *attB* and *attP* ... 186
5.3.4: Detection of integrated phage and integrated pBA ... 190
5.3.5: Mechanisms for Excision of ϕAR29, Integration of pBA, and Curing of AR29 ... 191
Model 1: ... 192
Model 2: ... 193
Model 3: ... 194
5.3.6: The Identification of Potential *attB* Sites in Other Organisms ... 195

5.4: Conclusions ... 196
Chapter 6: General Conclusion and Closing Remarks ... 197
6.0: Characterisation of the integration module of bacteriophage ϕAR29, cloned in pBA ... 198
6.1: Complexity of the Phage genomic sequence ... 199
6.1.1: Mosaic of the phage genome ... 199
6.2: Current approach of bacteriophage Integrative vectors in mammalian cells as tools for gene for gene therapy ... 201
6.2.1: Tyrosine integrase ... 201
6.2.2: Serine Integrase ... 202
6.3: Future Work ... 204
6.3.1: Genomic Analysis ... 205
6.3.2: Refinement of pBA ... 205
6.3.3: Application of pBA in Other Bacteria ... 208
6.4: Closing remarks ... 210
References ... 211
Appendix 1 ... 231
Figure 3.1A: ... 231
Figure 3.2A: ... 237
Appendix 2 ... 239
Table 4.1A: ... 239
ϕAR29 Genomic Sequence ... 242

ABSTRACT

The initial aim of this project was to characterise the integrative recombination mechanism of bacteriophage ϕAR29, to provide a better understanding for development of the shuttle plasmid pBA as a site-specific *Bacteroides* integration vector. RT-PCR showed that the previously identified ϕAR29 recombination genes, integrase (*Int*) and excisionase (*Xis*), were transcribed from pBA in *E. coli* SCS110, *B. thetaiotaomicron* AR29 and *B. uniformis* AR20. *In silico* derived amino acid sequences from both genes showed only very low levels of similarity to other known *Int* and *Xis* in GenBank. To improve understanding of the phage recombination system, the ϕAR29 genome was sequenced. This revealed a 35,558 bp double-stranded DNA genome with GC content of 39.11%. Bioinformatic analysis identified 53 open reading frames (>30 codons) and gene promoters and terminators that allowed the genome arrangement to be compared with other phages. Comparison of deduced gene products with proteins from other phages identified 6 reading frames, allowed tentative identification of 7 others, but left 40 ORFs unidentified. Those with strong homology to known genes were: large terminase subunit (44.66 kDa), *dnaC* (27.94 kDa), helix-turn-helix (*HTH*) transcription regulator (14.69 kDa), *cI* repressor (26.48 kDa), amidase (18.42kDa) and a novel integrase (54.22 kDa). The integrase gene is located 162 base-pairs downstream of the phage attachment (*attP*) core site, rather than the previously suggested location upstream of the integration site. The ϕAR29 *attP* was shown to include a 16-bp *att* core region, 117 bp upstream of the previously suggested location. Integration of ϕAR29 was found to occur at the 3'end of an arg-tRNA gene on the AR29 genome (*att*B). Imperfect direct repeats with a consensus sequence (ANGTTGTGCAA) were found surrounding the *att*P core. A review of pBA sequence showed that only the 5' end (435 bp) of the newly identified *Int* gene was cloned in pBA. Despite this, PCR analysis revealed integration of pBA into the AR29 genome. Serial subculturing of pBA-transformed AR29 was able to cure AR29 of the ϕAR29 prophage, providing an improved host for integrative plasmids, and for detailed studies of AR29 physiology and ϕAR29 life cycles.

COMMONLY USED ABBREVIATIONS

Abbreviation	Expansion
ABI	Applied Biosystems Inc.
AMV	Avian myeloblastosis virus
ATP	adenosine-5'-triphosphate
AR	analytical reagent
b	bases
bp	base pair
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic Acid
DNAse	deoxyribonuclease
dNTP	deoxynucleotides
ds	double stranded
EDTA	ethylenediaminetetra-acetic acid
IPTG	isopropyl-β-D- thiogalactopyranoside
MCS	multiple cloning site
OD	optical density
ORF	Open Reading Frame
PCR	polymerase chain reaction
pers. comm.	personal communication
RNAse	ribonuclease
SDS	sodium dodecyl sulphate
TAE	Tris-acetate EDTA buffer
TBE	Tris-borate EDTA buffer
X-Gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

Chapter 1: Introduction

Ruminants are animals that include cattle, sheep, goats, buffalos and reindeer. These animals possess a complex digestive system. The ruminant stomach system comprises four compartments, of which the rumen is the first and largest (Church, 1993). The rumen contains communities of symbiotic microorganisms that allow a continuous process of anaerobic fermentation of feed prior to further processing by the rest of the digestive organs (Hungate, 1966).

Some of the vital functions that rumen bacteria carry out include:

1) Providing nutrients to the ruminant, by fermentation of plant fibers and other polymeric plant materials, which cannot be degraded by the host, to volatile fatty acids, carbon dioxide, and methane.
2) Provision of proteins for the ruminant by synthesis of bacterial biomass.
3) Synthesis of vitamins as a supplement to the animals

(Hungate, 1966; Gregg, 1992; Nakamura *et al.*, 2003)

Rumen bacteria may also allow their host animals to consume plants that are poisonous to monogastric animals (Culvenor, 1987). An example of this is the ability to detoxify secondary plant metabolites, such as the compound 3-hydroxy,4-(1*H*)-pyridone (DHP: a metabolite of mimosine) that makes the legume *Leucaena* toxic to Australian livestock (Jones and Lowry, 1984; Hammond, 1995; Tan *et al.*, 1994).

1.0: Genetic Manipulation of Rumen Bacteria

The technology for enchancing the metabolic capabilities of rumen bacteria through the use of genetic engineering is still very much in its infancy, compared to

manipulation of well studied organisms such as *E.coli* (McSweeney *et al.,* 1999). Early strategies on the genetic manipulation of rumen bacteria focused on the enhancement of nutrient flow to the ruminant by improving digestion of plant tissues (Smith and Hespell, 1983; Gregg *et al.*, 1996; McSweeney *et al.*, 1999; Teather, 1985). A large proportion of the carbohydrates from plants is protected by lignin and ruminants in tropical regions are often fed lignified forages and crop residues that are less than 45% digestible (McSweeney *et al.,* 1999). Experiments have been carried out in breeding lignin-reduced plants and on the fermentation of plant fibre by microbial delignification prior to ingestion (McSweeney *et al.,* 1999). Although these methods of delignifying plant tissues have produced promising signs, improving the digestive capabilities of rumen bacteria could further enhance nutrient flow.

There have been a few genes cloned into ruminal bacteria since the potential of genetically modifying ruminal microorganisms was suggested in the mid 1980's. Table 1.1 provides a summary of recombinant ruminal and enteric microorganisms. Some of these cloned genes were able to increase the capacity of bacteria to carry out specific activities such as the break down of xlyan by xlyanase-transformed *B. fibrisolvens* (Xue *et al.*, 1995; Xue *et al.*, 1997) or broaden the capabilities of rumen bacteria, e.g. detoxification of fluoroacetate by *B. fibrisolvens* transformed with a dehalogenase gene (Gregg *et al.*, 1994a; Gregg *et al.*, 1998). The development of molecular genetic techniques for ruminal microorganisms has also been for applications that include:

1. Producing and storing proteins from newly inserted genes, which match the particular amino acid requirements of the ruminant under certain conditions (Teather, 1985).
2. Producing biological control agents to protect the host from parasites or to reduce populations of selected microorganisms within the rumen.

(Smith and Hespell, 1983; Teather, 1985; Gregg and Sharpe, 1991; Gregg *et al.*, 1996)

Table 1.1 : Rumen bacterial strains that have been genetically modified. Adapted from McSweeney *et al.*, 1999.

Host organism	Strains	Introduced genes	Reference
Butyrivibrio fibrisolvens	OB156	Xylanase	(Kobayashi *et al.,* 1995; Xue *et al.,* 1995; Xue *et al.,* 1997)
	H17c	"	
	LP1028	"	
	LP1309	"	
	AR9	"	
	AR11	"	
	210B	"	
	461A	"	
	VARIOUS	Cellulase	
	OB156	Cinnamoyl ester hydrolase	(Dalrymple *et al.*, 1996; Dalrymple and Swadling, 1997)
	OB156	Acetylxylan esterase	(Cybinski *et al.,* 1999; McSweeney *et al.,* 1999)
	OB156, AR14, OB291, OR85, S2/10, 149/33, 0/10, 10/1,	Fluoroacetate Dehalogenase	(Gregg *et al.*, 1994a; Gregg *et al.*, 1998) L. Kennedy, unpublished
Prevotella ruminocola	2202	Endoglucanase/xylanase	(Daniel *et al.,* 1995)
	B_14	CMCase	(Gardner *et al.,* 1996)
Streptococcus bovis	JB1	Endoglucanase	(Whitehead and Flint, 1995)
	WI-1	α-Amylase	(Brooker and McCarthy, 1997)
Bacteroides thetaiotaomicron	5402	Endoglucanase/xylanase	(Whitehead *et al.,* 1991)
	AR29	Fluoroacetate Dehalogenase	(Wong *et al.,* 2003)
	BTX	Endoglucanase/xylanase	(Whitehead *et al.,* 1991)
Bacteroides uniformis	AR20	Fluoroacetate Dehalogenase	(Wong *et al.,* 2003)

1.1: Criteria for Genetic Manipulation

Identifying an appropriate host as the recipient for new genetic traits requires certain characteristic. To be a suitable host, ruminal bacteria should be:

a) Organisms that are able to compete with indigenous microbes for growth substrates
b) Able to grow in the rumen at a rate faster than the dilution rate of the rumen contents
c) Able to withstand protozoal endocytosis, and

d) Resistant to inhibition by bacteriocins and bacteriophage infection

(Smith and Hespell, 1983; Teather, 1985)

However, to modify the genetics of rumen bacteria successfully, several additional criteria need to be fulfilled. These are:

1. Detailed biochemical analysis must be made of the process that genetic manipulation is intended to improve.
2. Understanding of the regulation and physiology of the processes that govern the biochemical pathway(s) should be obtained.
3. Development of genetic systems in the bacterium of choice and then the genetic components of the process must be analysed.

(Smith and Hespell, 1983)

1.2: Difficulties in Genetic Manipulation of Rumen Bacteria

Rumen bacterial molecular biology research has progressed relatively slowly since the realisation of the potential commercial value within the field. Genetic manipulation of rumen bacteria has proven difficult, due partly to the anaerobic requirements of microorganisms from the rumen microbial ecosystem (Gregg, 1992; Hamdorf, 1998). Furthermore, the limited knowledge of molecular biology procedures and transformation systems in this area has not helped fuel its progression. Unlike extensively studied species such as *E. coli*, groups have tended to work on a diverse array of rumen organisms. This has led to the development of transformation systems for one organism, which are not necessarily applicable to other or similar species. In addition, many procedures devised for manipulation of *E. coli* cannot be applied for use in rumen bacteria.

Biochemical barriers such as endonucleases have also limited gene manipulation studies on bacteria from the rumen. Highly active nuclease systems, like those found in

Bacteroides uniformis strain AR20 (previously designated *Bacteroides/Prevotella ruminicola*), have caused extremely low transformation efficiency (Hamdorf, 1998; Wong *et al.*, 2003). Similarly, endonucleases *Rf/*FI and *Rf/*FII, which are isoschizomers of *Sal*I and *Sca*I respectively (Morrison *et al.*, 1994; Morrison *et al.*, 1992), have prevented the introduction of non-indigenous plasmid DNA into *R.flavefaciens* strains by electroporation (Morrison *et al.*, 1994; Morrison *et al.*, 1992).

A much less well defined problem has also been noted. This involves a marked reduction in plasmid stability when DNA from different bacterial sources is combined in recombinant molecules. This phenomenon has created problems in cloning and manipulating genes from rumen bacteria, even in highly developed cloning systems such *E. coli* (Shoemaker *et al.*, 1991; Ware *et al.*, 1992; White, 1998; Gregg, 1992).

Despite these difficulties, research in this area is still currently in progress (Hermanova *et al.,* 2001; Mercer *et al.,* 2001; Nakamura *et al.,* 2001; Whitford *et al.,* 2001; Gobius *et al.,* 2002; Wong *et al.,* 2003; Klieve *et al.,* 2004).

1.3: Transformation Systems

The two major aims in developing transformation systems for ruminal bacteria are:

1. To provide tools for molecular genetic studies within a particular species, and
2. To be able to make stable genetic modifications to them, to impart useful new traits.

Numerous studies on cloning and DNA analysis have been reported for rumen bacterial genes (Berger *et al.*, 1989; Zhang and Flint, 1992; Cotta *et al.*, 1997; Flint *et al.*, 1993; Goodman and Woods, 1993; Paradis *et al.*, 1993; Zhu *et al.*, 1994; Forsberg *et al.*, 1997; Vercoe and White, 1997; Wyckoff and Whitehead, 1997; White, 1998).

Although these studies have helped in the understanding of gene structure and expression control signals in ruminal species, much of the application of genetic transfer systems to ruminal bacteria is still guided by information from other organisms with more completely understood genetic systems such as *E. coli*, *Bacillus* sp. and colonic *Bacteroides* (Gregg *et al.*, 1996; Smith and Hespell, 1983; White, 1998). However, much work has been carried out on the isolation and identification of genetic elements from rumen bacteria, such as plasmids; transposons and bacteriophages, which are essential for the development of genetic transfer systems (Gregg *et al.*, 1996; White, 1998; Wallace, 1994).

A number of cryptic plasmids have been isolated from various rumen bacteria (Table 1.2), some of which have been sequenced and characterized. These include

- *Butyrivibrio fibrisolvens* pOM1 (Mann *et al.,* 1986), pRJF2 (Kobayashi *et al.,* 1995), pRJF1 (Hefford *et al.,* 1993);
- *Bacteroides* pRRI2 (Thomson *et al.,* 1992), pRAM4 (Ogata *et al.,* 1996);
- *Ruminococcus* pAR67 (Ohara *et al.,* 1998);
- *Streptococcus* pSBO2 (Nakamura *et al.,* 2000), pSBO1 (Nakamura *et al.,* 2001); and
- *Selenomonas* pJDB23 (Attwood and Brooker, 1992), pJDB21 (Zhang and Brooker, 1993), pONE429 (Nakamura *et al.,* 1999), pONE430 (Nakamura *et al.,* 1999), pJJM1 (Fliegerova *et al.,* 2000), pJW1 (Fliegerova *et al.,* 2000).

Early rumen bacterial/*E. coli* shuttle vectors constructed from cryptic plasmids were poorly retained by the host cells or lacked efficient selectability (Shoemaker *et al.*, 1991; Ware *et al.*, 1992; White, 1998). Nevertheless, successful shuttle vectors have been constructed over the years. Most of these vectors were developed for *B. fibrisolvens* [pBH*erm* (Beard *et al.*, 1995), pYK4 (Kobayashi *et al.*, 1998)] or

Bacteroides [pVAL1 (Whitehead *et al.*, 1991), pRRI207 (Thomson *et al.*, 1992), pRH3 (Daniel *et al.*, 1995), pBA (Wong *et al.*, 2003)].

The efficiency of transformation systems for rumen bacteria was greatly improved by the introduction of electroporation. Electroporation techniques have been reported for the transformation of *Rumincoccus* (Cocconcelli *et al.,* 1992), *Selenomonas* (Lockington *et al.,* 1988), *Streptococcus* (Whitehead, 1992; Wyckoff and Whitehead, 1997), *Butyrivibrio* (Ware *et al.*, 1992; Whitehead, 1992; Beard *et al.*, 1995; Kobayashi *et al.*, 1995; Gregg *et al.*, 1998; Kobayashi *et al.*, 1998; Beard *et al.*, 2000) and *Bacteroides* (Thomson and Flint, 1989; Wong *et al.*, 2003).

Table 1.2: Some of the cryptic plasmids isolated from rumen bacteria and used for the development of shuttle vectors.

Organisms	Species/Strain	Cryptic plasmid	Sequenced (Acc. Num)	Other Vector	Derivative Plasmid
Ruminococcus	*R. flavefaciens*	pRF186(Asmundson and Kelly, 1987)	(N/A)		
	R. flavefaciens R13e2	pBAW301(May *et al.,* 1996)	(N/A)		
	R. albus F40	pRAB(Ohmiya *et al.,* 1989)	(N/A)		
		pRAC(Ohmiya *et al.,* 1989)	(N/A)		
	R. albus AR67	pAR67(Ohara *et al.,* 1998)	YES (D88665)		
Streptococcus	*S. bovis* JB1	pSBO1(Nakamura *et al.,* 2001)	YES (AB021464)		pSBE10(Nakamura *et al.,* 2001) pSBE11(Nakamura *et al.,* 2001)
		pSBO2(Nakamura *et al.,* 2000)	YES (AB021465)		pSBE2A(Nakamura *et al.,* 2000)
	S. anguis/E. coli		(N/A)	pVA838(Macrina *et al.,* 1982)	pTRW(Wyckoff and Whitehead, 1997)
Selenomonas	*S. ruminantium* HD4	pSR1 (Dean *et al.*, 1989; Martin and Dean, 1989)	(N/A)		
	S. ruminantium subspecies lactilytica	pJDB21(Zhang and Brooker, 1993)	YES (Z12102)		
		pJDB23(Attwood and Brooker, 1992)	YES (N/A)		
	S. ruminantium S20	pONE429(Nakamura *et al.,* 1999)	YES (NC_004986)		
		pONE430(Nakamura *et al.,* 1999)	YES (NC_004977)		
	S. ruminantium subspecies lactilytica JW13	pJJM1(Fliegerova *et al.,* 2000)	YES (Z49917)		
		pJW1(Fliegerova *et al.,* 2000)	YES (NC_004962)		
	S. ruminantium 18	pSRD181(Hermanova *et al.,* 2001)	Partially (N/A)		
		pSRD182(Hermanova *et al.,* 2001)	Partially (N/A)		
Butyrivibrio	*B. fibrisolvens*	pOM1(Mann *et al.,* 1986)	Yes (NC_002059)		
	B. fibrisolvens AR10	pBf1(Ware *et al.,* 1992)	(N/A)		pCW type(Ware *et al.,* 1992) pCK1(Ware *et al.,* 1992)
	B. fibrisolvens OB156	pRJF1(Hefford *et al.,* 1993)	Yes (M94552)		pBH*erm(Beard et al., 1995)*
	B. fibrisolvens OB157	pRJF2(Kobayashi *et al.,* 1995)	Yes (L31578)		pYK4(Kobayashi *et al.,* 1998)
Bacteroides/ Prevotella	*P.ruminicola* 223/M2/7	pRRI4(Flint *et al.,* 1988)	(N/A)		
		pRRI2(Thomson *et al.,* 1992)	Yes (AJ278872)		pRRI207(Thomson *et al.,* 1992) pRH3(Daniel *et al.,* 1995)
	P.ruminicola 223/M2/7 NCFV2202	pRRI7(Bechet *et al.,* 1993)	(N/A)		pKBR(Bechet *et al.,* 1993)
	Prevotella ruminicola	pRAM4(Ogata *et al.,* 1996)	Yes (NC_001760)		
	Bacteroides/E. coli		Yes (AF203972)	pBA(Wong *et al.,* 2003) pVAL1(van de Guchte *et al.,* 1994)	
Others					
	Staphylococcus			pUB110(Clark *et al.,* 1994) pUBLRS(Clark *et al.,* 1994)	
	E. coli/ Bacillus subtilis			pBS42(Whitehead, 1992)	
	Enterococcus facealis			pAMβ1(Hespell and Whitehead, 1991a; Hespell and Whitehead, 1991b)	

1.4: Bacteriophages and their Life Cycle

First described in 1915, bacteriophages are viruses that infect bacteria (Twort, 1915; Campbell, 2003). The ease with which they could be manipulated laid the foundations for studying mechanisms of inheritance and helped to establish the field of molecular biology during the period from 1940 to the 1970s (Luria *et al.*, 1943; Hershey, 1952; Benzer, 1957; Crick *et al.*, 1961; Jacob and Monod, 1961; Gellert, 1967; Ptashne, 1967; Okazaki and Okazaki, 1969; Okazaki, 1969; Crick, 1970; Hart and Ellison, 1970; Hedgpeth *et al.*, 1972; Mertz and Davis, 1972; Bigger *et al.*, 1973; Ellis *et al.*, 1973; Georgopoulos *et al.*, 1973). In addition to the growth of fundamental knowledge, the application of bacteriophage integrative mechanisms has aroused some interest as a tool for gene replacement technology and mutational studies.

In discussing how bacteriophage recombination mechanisms are utilised as molecular tools, it is necessary first to describe the bacteriophage life cycle, which can be directed into one of two phases: lytic cycle or lysogenic phase (Figure 1.1) (Prescott *et al.,* 1996).

1.4.1: Lytic Cycle

The lytic cycle is characteristic of a virulent phage and involves infection of the host, phage replication, and eventual release of progeny phage (Hendrix, 1983). There are several parasitic mechanisms used by different phages. In general, the process consists of 4 stages: Adsorption, Infection, Multiplication, and Release (Prescott et al., 1996). In brief:

1. Bacteriophage first adsorb to the host by attachment to receptor sites on the cell wall, such as specific lipopolysaccharides, proteins, teichoic acids, flagella or pili (Prescott *et al.,* 1996).

2. Infection occurs when the genetic material of the phage is inserted into the bacterium (Becker *et al.,* 1996).
3. After infection of the bacterial host, synthesis of bacterial products is halted and the host's cellular machinery is used by the phage to make essential viral constituents (Hendrix, 1983). Following the synthesis of viral genome and structural proteins, progeny phages are assembled and mature viruses accumulate within the host until released.
4. Release generally occurs through lysis of the cell, following the production of lysozyme by the phage (Friedman and Gottesman, 1983). Some phages are liberated by extrusion through the cell membrane without killing the host, for example filamentous fd (Hendrix, 1983). This mechanism involves the phage inserting its protein coat into the host membrane. As the phage DNA is secreted through the host plasma membrane, the coat then assembles around the viral DNA (Watson et al., 1987).

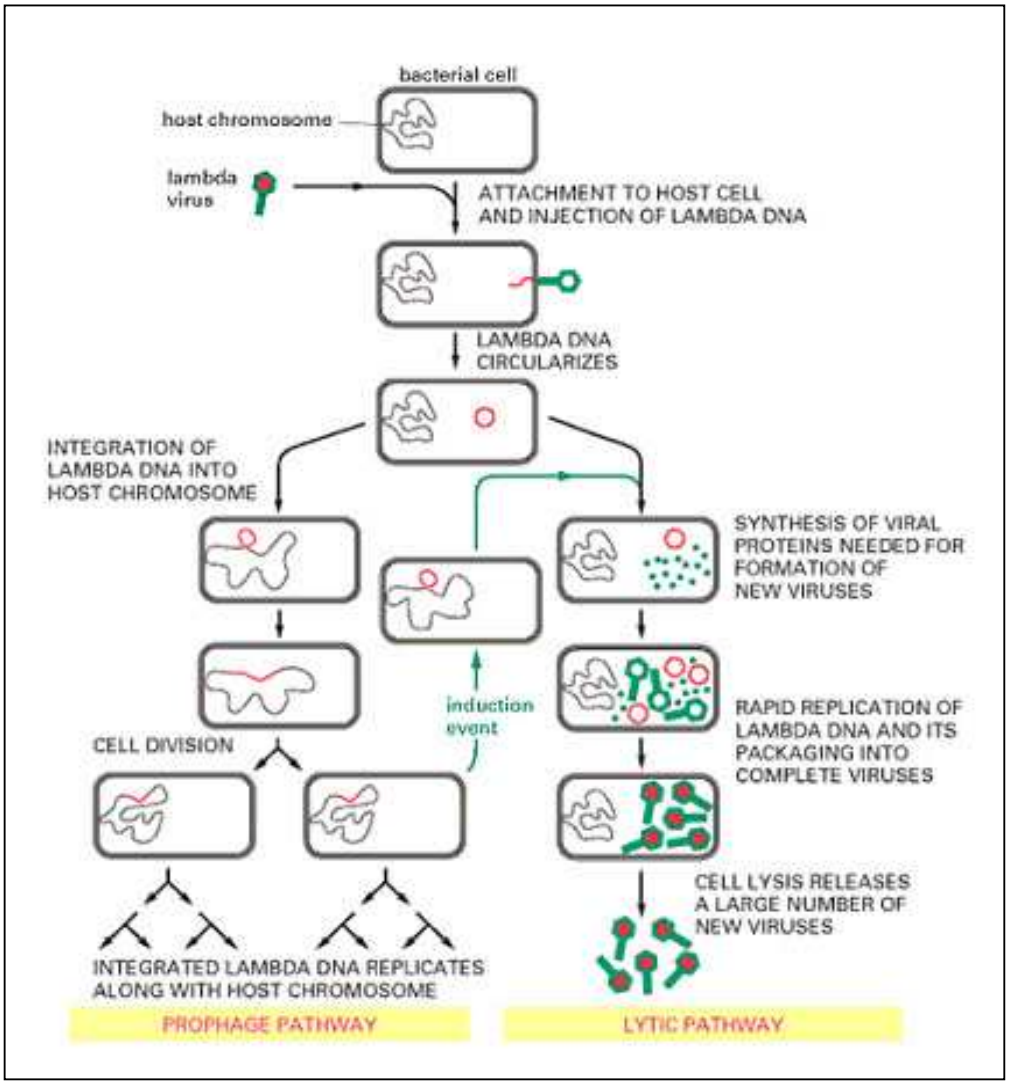


Figure 1.1: Life cycle of bacteriophage λ (Prescott *et al.*, 1996).

1.4.2: Lysogenic Cycle

In the lysogenic cycle, instead of reproducing large numbers of progeny and lysing the host cell, the bacteriophage integrates its genome into the host chromosome without causing apparent ill effect to the host (Voet & Voet 1990). In this form the integrated genetic material is referred to as a "prophage". The best known example of a bacterial virus that goes through this type of life cycle is the *E. coli* phage lambda (λ).

1.4.3: Mechanism of Phage λ and Lysogenic Cycle

The question of how bacteriophages "decide" to enter a lytic or lysogenic life cycle has been extensively studied in phage λ, which will be used as an example for the mechanisms of lysogeny.

Following infection, the λ genome is initially transcribed from bi-directional operons. Each operon is regulated by its respective promoter, designated P_L (promoter leftward) and P_R (promoter rightward). These promoters are responsible for the transcription of genes involved in the lysogenic and lytic cycles, respectively. Genes that are found in the leftward operon include:

- *cIII*, which facilitates the lysogenic pathway;
- *Xis* and *Int*, which code for excisionase and integrase proteins that are involved in site-specific recombination and excision reactions.

The rightward operon contains the "early" genes, including:

- the gene for the cro protein (required for the take-over of host cellular machinery);
- *cII*, O and P; and late gene Q (required for producing proteins needed in capsid construction and virus release) .

(Voet and Voet, 1990; Figure 1.2)

During early λ development, *N* and *cro*, are the first genes to be transcribed and translated from P_L and P_R, respectively. Protein *N* is involved in the antitermination of other early gene transcription, acting at three terminator sites located:

1. immediately after gene N (t_{L1}),
2. downstream of gene cro (t_{R1}), and
3. between genes P and Q (t_{R2}).

(Voet and Voet, 1990; Figure 1.2).

The decision of phage λ to enter one of two life cycles is governed by the presence of *cro* and *cII* protein (Watson *et al.*, 1987).

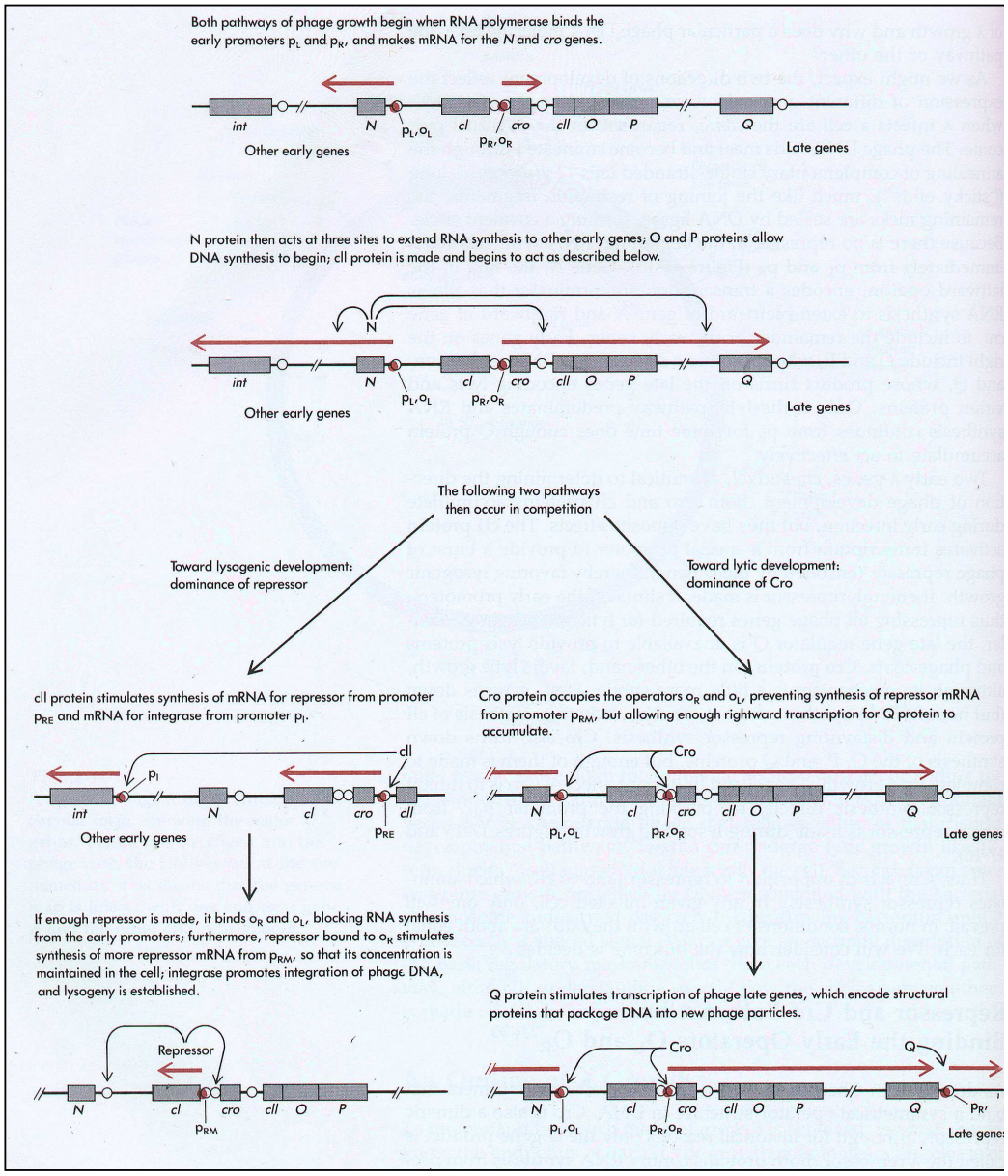


Figure 1.2: A flow diagram of regulation of genes involved in the lysogenic cycle of bacteriophage λ (from Watson et al.,1987).

1.4.3.1: The initiation of phage λ lysogenic life cycle

The initiation of the λ lysogenic cycle is triggered by a tetrameric activator protein, *cII* (Voet and Voet, 1990), which is expressed as a result of the *N* protein binding to the transcription terminator site, t_{R1}. The *cII* activator forms a complex with phage protein *cIII*, which helps to protect the activator from host protease cleavage (Abraham *et al.*, 1980; Echols and Guarneros, 1983).

The *cII* activator initiates phage lysogeny by its interaction at promoters P_I (I = integrase) and P_{RE} (RE = repressor establishment; (Echols and Guarneros, 1983), which initiates transcription. Transcription from these promoters produces the integrase protein (*Int*) and *cI* repressor, respectively (Abraham *et al.*, 1980; Echols and Guarneros, 1983).

1.4.3.1.1: Transcription of the Int gene

The enzyme Int catalyzes the integration of lambda DNA into the host chromosome by site specific recombination. Although the *Int* gene is located downstream of the excisionase gene (*Xis*: Figure 1.3), transcription from promoter P_I stimulates the production of Int protein but not the Xis *protein*, because P_I lies within the coding region of the *Xis* gene. As a consequence, elevated integrase concentrations allow the phage to become a lysogen (Watson *et al.,* 1987).

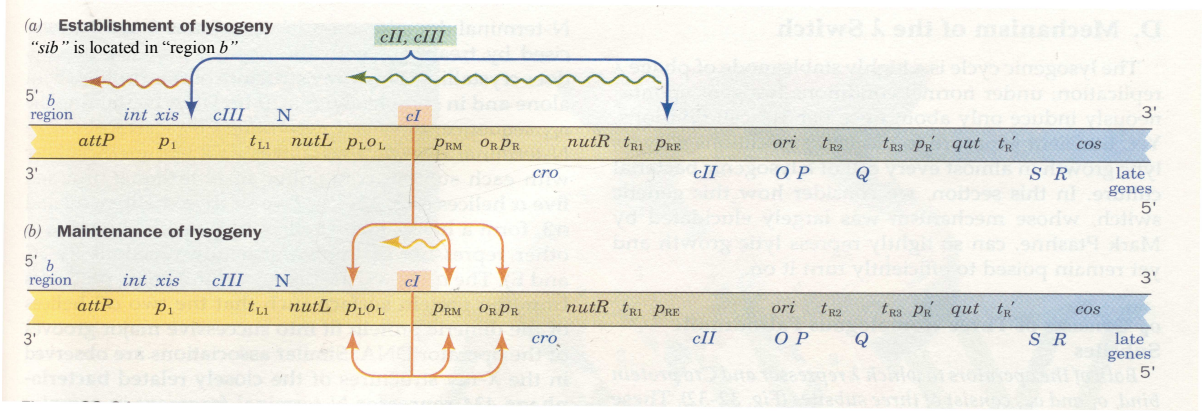


Figure 1.3: The establishment of λ lysogency by *cII* and *cI* (Voet and Voet, 1990).

The *Int* mRNA may also be transcribed from a distant upstream promoter P_L (Watson *et al.,* 1987). However, unlike promoter P_I, transcription from P_L results in translation of the *Xis* section of the mRNA but not of the *Int* sequences (Echols and Guarneros, 1983). The inhibition of *Int* translation from P_L is governed by a specific site, *sib*, in "region *b*" of the λ genome (*sib*: ***s***itio ***i***nhibidor en ***b*****;** Figure 1.3).

Sib is a cis-acting negative regulatory element that plays a role in the inhibition of *Int* expression from the P_L-initiated mRNA. Evidence suggests that regulation of *Int* expression by *sib* is post-transcriptional (Schindler and Echols, 1981; Guarneros *et al.*, 1982). Transcription of P_I mRNA stops at a terminator site 300 bases downstream of the *Int* gene, within the *sib* region (Watson *et al.,* 1987). As a result, a short stem loops structure is formed at the end of the P_I mRNA (Figure 1.4: (Echols and Guarneros, 1983). However, during the transcription of mRNA initiated from P_L promoter, the antiterminator protein *N* binds to the DNA and allows RNA polymerase to transcribe beyond the *sib* region and terminator site encountered in the P_I transcript (Echols and Guarneros, 1983). The presence of the complete *sib* region allows the formation of a longer stem structure at the 3' end of the P_L derived transcript, thus changing the stability of the mRNA (Figure 1:4: (Watson *et al.*, 1987). The extended form of the mRNA is then vulnerable to nuclease cleavage.

1.4.3.1.2: Mainteance of λ prophage

Once integrated, λ prophage is able to express the *Int* gene constitutively from P_I, at low levels (Shimada and Campbell, 1974b; Shimada and Campbell, 1974a; Echols, 1975). The continous expression of *Int* helps λ to maintain the lysogenic state. However, the maintenance of λ prophage is largely achieved by repressor *cI* (Echols and Guarneros, 1983; Wulff and Rosenberg, 1983). Repressor *cI* has a high affinity for both right and left operators (O_L and O_R) of promoters P_L and P_R, respectively.

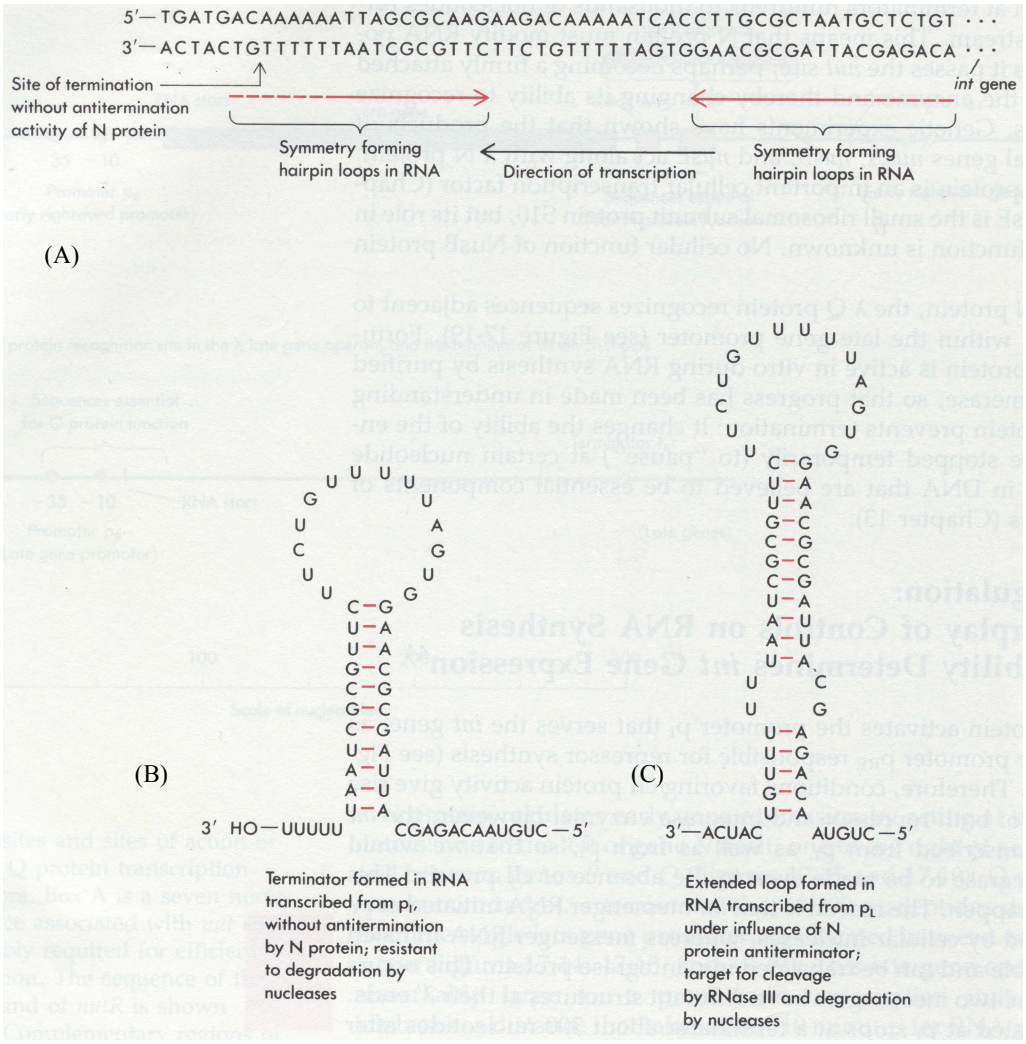


Figure 1.4: An illustration of *Int* mRNA structure resulting from transcription from P_I and P_L. (A) The 3′ end nucleotide sequence of the *Int* gene. (B) Transcription from P_I produces a shortened stem structure with terminator U residues at the 3′ end of the mRNA. (C) In contrast, transcription from P_L incorporates the poly U region into the stem structure, reducing their termination capability.(Watson *et al.*, 1987)

Binding of *cI* to these operators blocks transcription from the respective promoters, thus silencing most of the phage genes. The attachment of *cI* repressor to O_R also stimulates the production of *cI* from promoter P_{RM} (promoter for repressor maintenance). Therefore the maintenance of cellular *cI* protein level helps preserve the lysogenic state (Watson *et al.,* 1987).

1.4.3.2: The role of protein Cro in the phage λ lytic cycle

Cro protein is a dimeric repressor that initiates λ's irreversible lytic cycle (Voet and Voet, 1990; Little *et al.*, 1999), It also has a lower affinity for the same operator sites as *cI* protein (Gussin *et al.,* 1983). The protein *Cro* induces the transcription of λ genes involved in the lytic cycle, by competing against *cI* for these operator sites.

1.4.3.2.1: Activation of the lytic cycle from the lysogenic state

The induction of prophage lytic cycle requires the removal of repressor *cI*. This is achieved by *RecA* protease which is synthesised under conditions that induce the host SOS response mechanism. *RecA* abolishes *cI* affinity for the operators O_L and O_R by cleaving the *cI* monomers at the Ala 111 – Gly 112 bond (Voet and Voet, 1990). Subsequently, λ undergoes the three stages of the lytic cycle:

1. Early transcription
2. Delayed early transcription
3. Late transcription

During the early transcription stage, *N* and *cro* proteins are synthesised as a result of the inability of *cI* to bind at the operator sites (Figure 1.5: (Watson *et al.*, 1987).

The expression of antiterminator *N* allows the *Xis* protein to be expressed by allowing transcription from promoter P_L in the delayed early transcription stage. The presence of *Xis* promotes the excision of the viral genome from the bacterial chromosome. In addition, protein *N* also binds t_{R1} and $t_{R2,}$ thus allowing transcription from promoter P_R to progress beyond these terminator sites. As a result, proteins *O*, *P* and Q are synthesised (Figure 1.5). Proteins O and P are required for DNA replication. Protein Q is an antiterminator, which allows transcription of the structure/packaging and lytic genes for the final stages of the lytic cycle.

The transition of lysogenic to lytic cycle is an irreversible process (Voet and Voet, 1990; Little *et al.*, 1999). Following sufficient accumulation of *cro* factor during the early and delayed early transcription stages, *cro* binds to the O_L and O_R operators and to $P_{RM,}$ preventing any further opportunity for *cI* binding. The interaction of *cro* with both operators switches off the transcription of the *cI* repressor gene from promoter P_{RM} but also reduces the expression of genes from P_L *and* P_R (Watson *et al.,* 1987). However, the inhibition of gene expression by *cro* is not complete and allows sufficient residual transcription to occur for repression of the later genes required for lytic growth to be removed, allowing the lytic cycle to occur (Abraham *et al.,* 1980).

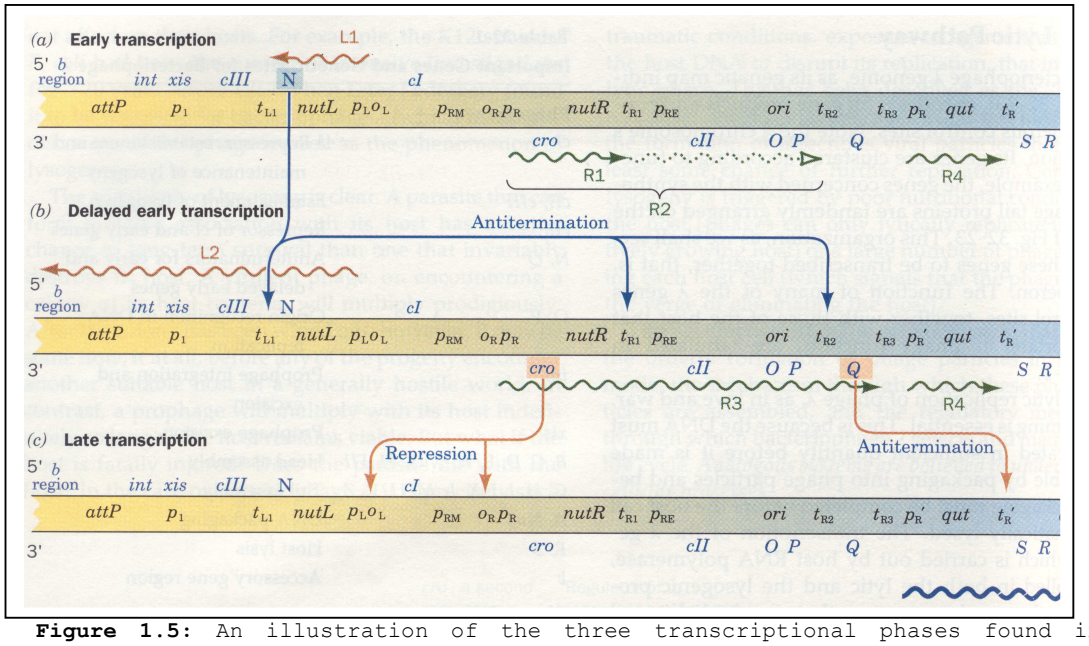


Figure 1.5: An illustration of the three transcriptional phases found in Lambda's lytic pathway: (a) early transcription, (b) delayed early transcription and (c) late transcription. Transcripts are represented by wiggly arrows pointing in the orientation of mRNA elongation. The site(s) in which regulatory proteins N, Cro and Q act on are denoted by arrows. (a) early transcription, (b) delayed early transcription and (c) late transcription. (Watson *et al.*, 1987)

1.4.3.2.2: Lytic cycle from infection

The wellbeing of the host bacterium is essential for λ to enter the lytic cycle, following infection, with the catabolite repression system in the host playing an important role in the lytic cycle (Voet and Voet, 1990). In conditions where sufficient glucose is available, the level of cyclic AMP is low. As a result, the host produces

elevated levels of proteases, which cause *cII* degradation (Watson *et al.,* 1987). The decrease in the level of *cII* leads to a decrease in cellular *cI* levels, allowing progression to the lytic cycle as described in section 1.4.3.2.1.

1.4.4: Phage Integration

Campbell (Campbell, 1962) was the first to propose a λ-based model that illustrates how a circular phage genome could be integrated into a bacterial chromosome. Bacteriophages integrate into their host's chromosome via an energy-independent, site-specific recombination process. Successful recombination of λ requires several factors, which include:

1. Integrase enzyme (*Int*)
2. An *E. coli* accessory protein, integration host factor (IHF),
3. Magnesium ions, and
4. A specific DNA region for recombination that is present as homologous sequences on both the host and phage DNA, called the attachment (*att*) site. For bacterium and phage the attachment sites are designated *attB/attB'* and *attP/attP'*, respectively.

(Crisona *et al.*, 1999; Weisberg and Landy, 1983)

During integrative recombination, the *Int* protein carries out site specific recombination by its ability to identify attachment sites, *attB* and *attP*. The length of the *attP* and *attB* sites are 240 bp and 20 bp respectively. Integrase is able to identify these *att* sites by the recognition of two distinct DNA sequences to which it binds: (1) core-type and (2) arm-type binding site (Figure 1.6).

The a*ttB* site consists only of a core-type binding site for the integrase. However, *attP* provides both core-type and arm-type binding regions. Furthermore, *attP* also

contain specific IHF binding sites (Figure 1.6: (Gottesman, 1981; Crisona *et al.*, 1999; Weisberg and Landy, 1983).

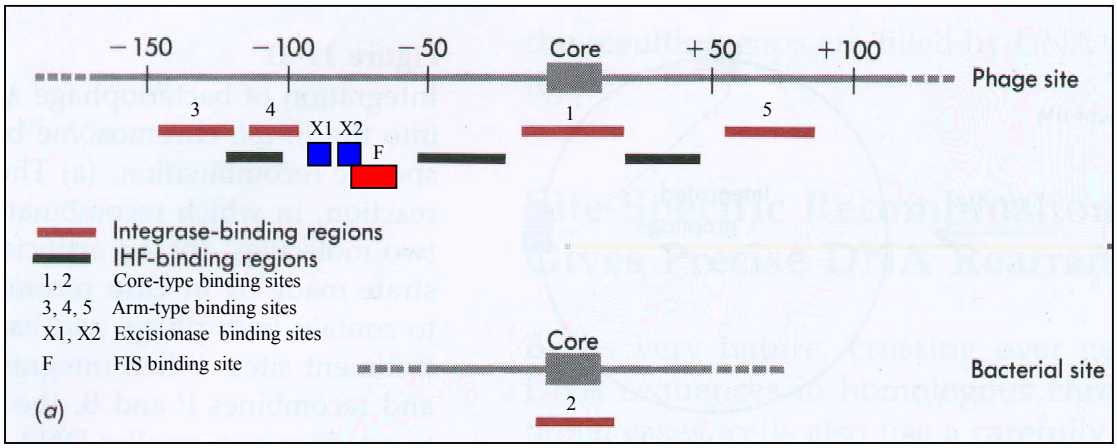


Figure 1.6: The binding sites of integrase, excisionase, IHF and FIS on the bacterial and phage *att* sites (Watson *et al.*, 1987; Swalla *et al.*, 2003).

The phage integration mechanism consists of 2 steps :

1. Synapsis, which is the juxtaposition of the two recombination sites on the recombinase, resulting in the formation of a synaptic complex.
2. Strand exchange, which involves the two cleaved recombination sites and ligation of opposite DNA strands to complete the recombination process.

The *Int* function is similar to that of a Type I topoisomerase. The interaction of IHF with integrase at the integration region forms a nucleoprotein complex, called an intasome. This complex allows integrase to relax supercoils at the *att* site and introduce a staggered cut with a seven-nucleotide overlap in both phage and bacterial *att* sites (Hsu and Landy, 1984). The breakage of DNA then induces a cross-over strand exchange and the ligation of the resulting strand exchange forms the "Holliday" structure (Figure 1.7: (Crisona *et al.*, 1999).

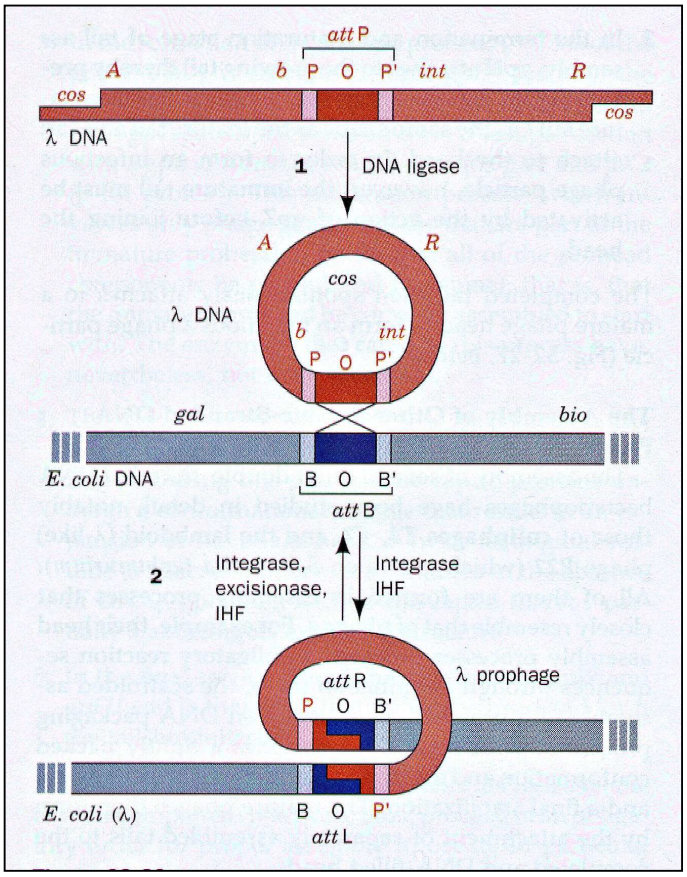


Figure 1.7: A schematic diagram showing integration of phage λ into the *E.coli* genome. (1) The circularisation of the linear λ genome. (2) Synapsis of phage λ genome and the *E. coli* genome, which is then followed by DNA cross-over strand-exchange.(Voet & Voet, 1990)

1.4.4.1: attB site

Most bacteriophages utilise a host tRNA gene as an *attB* site. Gabriel *et al.,* (1995) has established a set of general rules for the integration of phage genome into tRNA genes. These are:

1. The targeted tRNA site must be a functional gene
2. The 3' terminal end of the tRNA gene is part of the core sequence
3. Integration restores an intact tRNA gene
4. The anticodon is part of the core region.

(Gabriel *et al.*, 1995)

The insertion of the phage genome generally occurs in one of three sublocations on the tRNA: the anticodon-loop, T-loop, and at the asymmetric 3' end of the gene (Figure 1.8).

The position of *attB* can be classified into 4 classes based on the phage integration sublocation site. These classes are:

1. Classes IA and IB in which *attB* includes the anticodon loop and extends to the 3' end and 5' of the tRNA, respectively;
2. Class II in which the *attB* encompasses the T-loop without extending into the variable region; and
3. Class III in which *attB* resides in the further 3' end of the tRNA and does not encompass the T-loop.

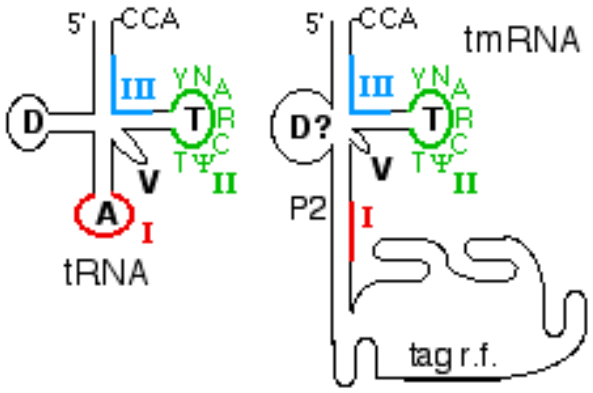


Figure 1.8: Secondary structures of a tRNA and tmRNA gene. Three presumed crossover sites used to classify *attB* are represented by Roman numerals. The 7-nucleotide anticodon and T loops of tRNAs are flanked by the symmetrical sequences that form 5-bp stems. The T-loop consensus sequence is shown in green. (adapted from (Williams, 2002)

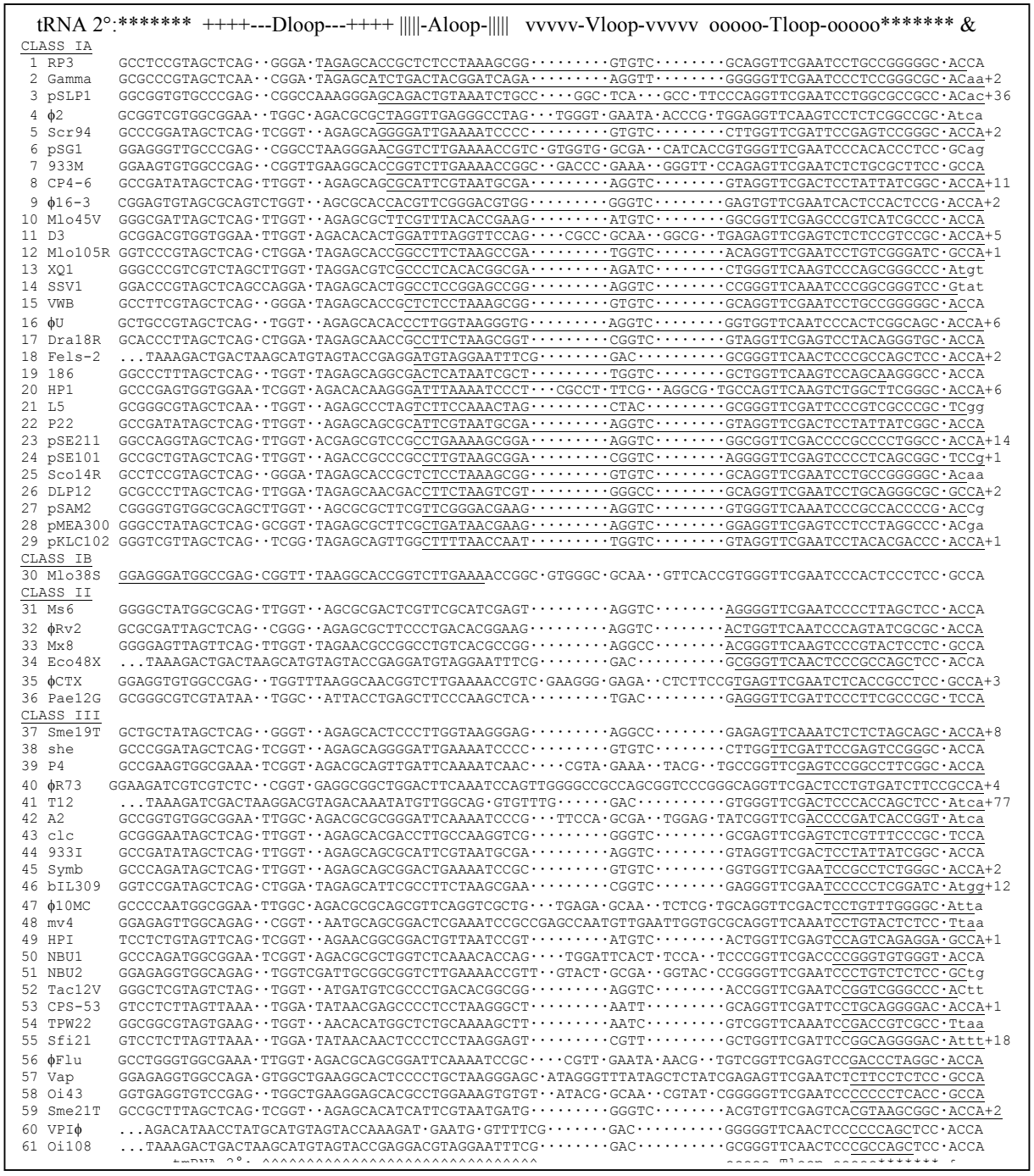


Figure 1.9: Examples of the different classes of tRNA and tmRNA gene *attB* sites used by integrative elements. Genes are divided into the 4 different classes (IA, IB, II, III) and aligned according to the secondary structure of the encoded RNA, indicated above for tRNA and below for tmRNA (discriminator position is marked by ampersand). The sharing of identical regions between *attB* and *attP* is indicated by the underlined sequences. The length of continued rightward extension of the identity block is given. Terminal positions where the gene does not encode the full CCA tail of the mature RNA is indicated by lower case. For references to these molecules see Appendix 6.0 (adapted from (Williams, 2002)

Some bacteriophages may use a tmRNA gene as their alternative *attB* site as frequently as tRNA. tmRNA is a bacterial RNA that has some structural similarity to tRNA but has a different physiological role (Figure 1.8). Although they are charged with alanine, tmRNA does not read any codon on the mRNA and the ribosome transfers that moiety to a nascent peptide (Karzai *et al.,* 2000). The main role of tmRNA is to resolve stalled ribosomes that may occur through the presence of rare codons, or at the end of mRNAs that have no stop codon. They achieve this by exchanging their alanyl moiety with the troubled mRNA which caused the stalling of ribosome. As a result, a peptide tag is added to the incomplete protein in the stalled ribosome. The ribosome is rescued by allowing the continuation of translation to the stop codon in the tmRNA and release of the ribosome for further rounds of translation. The tagged product is then subjected to proteolysis. An example of a phage using tmRNA as an *attB* site is Fels-2 isolated from *Salmonella typhimurium* (Williams, 2002). Fels-2 uses a 7-bp anticodon loop as its crossover site and has been shown to use a class IA *attB* site (Williams, 2002).

1.4.5: Excision of Prophage DNA from the Bacterial Chromosome

For prophage to re-enter a lytic life-cycle, its genome must be excised from the bacterial chromosome. Integrase enzymes cannot recognise the hybrid sequences at the *attB-attP'* and *attP-attB'* junctions, which are created by the recombination process. The excision of viral DNA from the bacterial chromosome is performed by excisionase, which is encoded by the *Xis* gene (Wulff and Rosenberg, 1983). The requirements for excisive recombination are the same as for the integration reaction, with an extra requirement for the *E. coli* accessory protein "factor for inversion stimulation" (FIS) (Crisona *et al.,* 1999).

Excisionase is able to bind to the *att*-site, break the B-P' and P-B' junction sequences, and re-ligate the P-P' and B-B' attachment sites. In the case of phage λ, *Xis* binds to the two direct, imperfect, 13-base repeat nucleotide sequences located in the arm-type region of *attP*, which are designated X1 and X2 (Figure 1.6) (Yin *et al.*, 1985; Cho *et al.*, 2002; Swalla *et al.*, 2003). Phage λ's *Xis* protein brings about excision by affecting the binding of other proteins to the *att* site. To achieve this, the enzyme participates in two different reaction.

First, *Xis* induces DNA curvature of 140° in the DNA, thereby altering the trajectory of *attR* within the intasome (Sam *et al.,* 2002; Swalla *et al.,* 2003). *AttR* is the junction of phage-to-bacterial DNA located on the right hand side of the integrated phage, as opposed to *attL*, which is the junction of bacterial-to-phage DNA at the left. The excisionase induces curvature by binding to the X1 site. The stability of this binding is maintained by a second *Xis* protein bound at the adjacent X2 site or, alternatively, by FIS attached to the F site that overlaps X2 (Sam *et al.,* 2002). The resulting interaction of *Xis*-*Xis* or *Xis*-FIS changes the spatial relationship between *Int* and *IHF* binding sites within the excisive intasome.

Secondly, the binding of *Xis* at the X1 site is cooperative with *Int* binding at the adjacent arm-type site (position 4, Figure 1.6). This protein-protein interaction encourages the formation of the correct *attR* intasome configuration, by stabilizing *Int* at the arm-type binding site adjacent to X1 (Swalla *et al.,* 2003). The interaction of these two reactions enhances the excision reaction, while simultaneously inhibiting integrative recombination by converting *attP* to a non-catalytic architecture (Lewis and Hatfull, 2001; Cho *et al.*, 2002; Sam *et al.*, 2002; Swalla *et al.*, 2003).

1.5: Phage in the Rumen

Bacteriophages are present in large quantities in the rumen, with a density of 1×10^7 to 1.6×10^{10} phage particles per millilitre of ruminal fluid (Klieve and Swain, 1993; Wells and Russell, 1996). Phages isolated from the rumen were found to belong to the viral families *Myoviridae, Siphoviridae and Podoviridae,* with 26 – 40 morphologically distinct types (Klieve *et al.*, 2004; Klieve and Bauchop, 1988). It has been suggested that the majority of phages in the rumen exist in a state of lysogeny or pseudolysogeny and that 25% of ruminal bacteria may be carriers of chromosomally stable, lysogenic prophages (Wells and Russell, 1996; Klieve *et al.*, 1989). Due to their lytic reaction, bacteriophage within the rumen are involved in the process known as protein recycling, which reduces feed utilization efficiency of the ruminant (Firkins *et al.*, 1992; Nolan and Stachiw, 1979).

Despite their large number and diversity, current work on phages from the rumen is limited. Early work in this field was focused in the isolation and morphological classification of bacteriophage. Most bacteriophage isolated during the 1960s to late '80s were from *Serratia* spp. (Adams *et al.*, 1966), *Streptococcus bovis* (Iverson and Millis, 1976a; Iverson and Millis, 1976b; Tarakanov, 1976), *Streptococcus durans* (Brailsford and Hartman, 1968), *Bifidobacterium ruminale* (Matteuzzi and Sozzi, 1971), *Magnoovum eadii* (Orpin and Munn, 1974), *Methanobrevibacter* sp. (Baresi and Bertani, 1984; Knox and Harris, 1986), *Fusobacterium* spp. (Tamada *et al.*, 1985)and *Selenomonas ruminantium* (Hazlewood *et al.*, 1983; Lockington *et al.*, 1988). However, the majority of isolates were from non-ruminal hosts. In 1989, Klieve *et al.* were the first to report the identification of phage from *Eubacterium, Bacteroides*, *Butyrivibrio* and *Ruminococus*. Since then, other research has included the isolation and characterisation of lytic phages from ruminal bacteria, Phages ϕBrb01, ϕBrb02, M1 and various phages isolated from *Prevotella*, have been described (Klieve *et al.*, 1991;

Cheong and Brooker, 1998; Jiang *et al.*, 1995; Ambrozic *et al.*, 2001). Recently, the isolation and characterisation of four phages (ϕRa01, ϕRa02, ϕRa03 and ϕRa04) from *R. albus* was the first reported isolation of lytic phage that infect cellulolytic bacteria of the rumen (Klieve *et al.,* 2004).

1.6: Bacteriophage as Transformation Tools

The integration mechanisms of lysogenic bacteriophages have been of some interest as tools for introducing novel genes into bacteria (Lee *et al.*, 1991a). Such a tool has several advantages over plasmid-based vectors. These include:

1. Integration of a single copy of DNA.

   Transformation studies with plasmids generally result in high plasmid copy-numbers. Although this may be useful for the identification of weak promoters and for studying gene structure-function, multicopy plasmids can result in high-copy-number artifacts, especially during physiological studies (Hoang *et al.,* 2000). In addition, gene expression studies on genes from high-copy-number plasmids may be difficult to control in situations where a single copy regulator gene is chromosomally encoded (Rossignol *et al.,* 2002). Unlike plasmids, integration vectors allow the introduction of a single copy of heterologous DNA into the bacterial genome. This allows expression studies to be conducted under conditions that are similar to those for single-copy chromosomal genes or operons (Rossignol *et al.,* 2002; Yang *et al.,* 2002).

2. Removing the need for selective mechanism for transformed organisms

   In plasmids, a selective marker is generally required for both the identication of transformed bacteria and for the maintenance of plasmid within the host. Such markers may have negative outcomes on the environment or host cells, i.e. the

possibility of spreading antibiotic resistance and the detrimental effects of maintaining several plasmids using multiple antibiotics (Rossignol *et al.,* 2002). The chromosomal insertion of a stable integrative vector removes the need for selective pressure to help maintain foreign DNA in the host cells (Shimizu-Kadota, 2001).

3. A greater size of insert can be used.

One of the major difficulties encountered through the use of plasmids is the cloning of large insert fragments (Lee *et al.*, 1991a; Shimizu-Kadota, 2001; Lauer *et al.*, 2002). Generally, larger inserts compromise the stability of plasmid in the host cells. In contrast, integrative vectors are able to insert fragments of DNA at least as large as a phage genome. For example, temperate bacteriophage SPβc2 is capable of stable integration, despite being approximately 134 kb in size (Lazarevic *et al.*, 1999; Yang *et al.*, 2002; Shimizu-Kadota, 2001). However, Yang *et al.,* (2002) proposed that in an integrative vector that does not require the packaging of DNA into the phage head, there is no upper size limit for the introduced DNA.

1.6.1: Transformation Systems That are Based on Phage Integrative Recombination Process

There have been a number of reports on the construction of vectors that utilise bacteriophage integration mechanisms in bacterial cells (Table 1.3). These include bacteria belonging to the genera: *Staphylococcus* (Lee *et al.*, 1991a), *Mycobacterium* (Lee *et al.*, 1991b; Freitas-Vieira *et al.*, 1998; Saviola and Bishai, 2004), *Streptomyces* (Kuhstoss *et al.,* 1991; Bierman *et al.,* 1992; Gabriel *et al.,* 1995; Van Mellaert *et al.,* 1998; Gregory *et al.,* 2003), *Pseudomonas* (Wang *et al.,* 1995; Hoang *et al.,* 2000),

Streptococcus (McShan *et al.,* 1998; Gindreau *et al.,* 2000), *Listeria* (Lauer *et al.,* 2002), *Enterococcus* (Yang *et al.,* 2002), *Rhizobium*(Hermesz *et al.,* 1992; Elo *et al.,* 1998; Semsey *et al.,* 1999; Semsey *et al.,* 2002; Ferenczi *et al.,* 2004) and *Lactobacillus* (Shimizu-Kadota, 2001).

Table 1.3: A list of integrative vectors developed for bacteria.

Bacteria	Specie/Stain	Phage	Plasmid
Staphylococcus	*S. aureus* RN4220	L54a	pCL55(Lee *et al.*, 1991a))
	S. aureus CYL316	L54a	pYL112Δ19/pCL83 or pCL84(Lee *et al.*, 1991a)1)*
Mycobacterium	*M. smegmetis*	L5	pMH94(Lee *et al.*, 1991b)1) pBluescriptint/pBS37 and pBS33(Saviola and Bishai, 2004)4)*
	M. smegmatis, M. vaccae, M. bovis BCG, *M. tuberculosis* H37Ra	Ms6	pAV1(Freitas-Vieira *et al.,* 1998)
Streptomyces	Various *Streptomyces* speices	ϕC31	pKC796(Kuhstoss *et al.,* 1991) pSET152(Bierman *et al.,* 1992)
	S. rimosus R6-554	RP3	pKG2(Gabriel *et al.,* 1995)
	S.venezuelae ETH14603	VWB	pKT02(Van Mellaert *et al.,* 1998)
	Various *Streptomyces* speices	ϕBT1	pRT801(Gregory *et al.,* 2003)
Pseudomonas	*P. aeruginosa*	ϕCTX	pIBH/pTABF, p1000, p400 (Wang *et al.,* 1995)* mini-CTX1, mini-CTX2 with pFLP2(Hoang *et al.,* 2000)
Streptococcus	*S. pyogenes*	T12	pWM139, pWM245, p7INT(McShan *et al.,* 1998)
	S. pneumoniae	MM1	pIAPU1(Gindreau *et al.,* 2000)
Listeria	*L. monoctogenes*	U153	pPL1(Lauer *et al.,* 2002)
		PSA	pPL2(Lauer *et al.,* 2002)
Enterococcus	*E. faecalis KBL707*	ϕFC1	pEMJ1-1(Yang *et al.,* 2002)
Rhizobium	*R. meliloti 41*	*16-3*	patt164, patt202(Hermesz *et al.,* 1992) pSEM102/pEP226 pEP227, pEP228,pEP181, pEP184(Elo *et al.,* 1998)* pSEM167(Semsey *et al.,* 1999; Semsey *et al.,* 2002) pGSB1(Ferenczi *et al.,* 2004)
Lactobacillus	*Lactobacillus casei* YIT9029	ϕFSW	pMSK761(Shimizu-Kadota, 2001)1)
Escherichia	*E. coli*	HK022	pHK-Int/pHK11(Rossignol *et al.,* 2002)
Bacteroides	*B. thetaiotaomicron* AR29	ϕAR29	pBA(Wong *et al.,* 2003)

*note *: require two plasmids to insert plasmid, 1) plasmid carrying Int and 2) plasmid carrying attP site. (Cloned Integrase plasmid/plasmid with attP)*

1.6.1.1 Staphylococcus

One of the early developments of integrative vectors was pCL55, which was designed for delivery of exogenous gene into *S. aureus* RN4220 (Lee *et al.*, 1991a). This phage L54a-based plasmid is a shuttle vector that carries an *ori* and *Amp*^{r} from pBR322 for replication and selection in *E. coli*, and *Clin*^{r} for the selection for transformed *S. aureus*. Although the plasmid was successfully integrated into the

genome of strain RN4220 the transformation efficiency was low, with a range of 0 to 29 transformants per μg of DNA. This vector system was improved by cloning the *attP* and *Int* gene in two separate plasmids, pCL84 and pYL112Δ19, respectively. The resulting transformation efficiency ranged from 15 to 223 transformants per μg of DNA.

1.6.1.2 Mycobacterium

The integrative mechanism of mycobacteriophage L5 has been used for the construction of vector pMH94 (Lee *et al.*, 1991b). The stability of integration was demonstrated when pMH94-transformed *Mycobacterium smegmatis* were able to grow on antibiotic plates following 30 generations of growth in non-selective medium.

Plasmid pAV1 is also an integrative vector, constructed for use in mycobacterial cells by ligating plasmid pCR3 with the *attP-Int* region from phage Ms6 (Freitas-Vieira *et al.,* 1998). The vector was able to replicate extrachromosomally in *E. coli* but not in mycobacterial cells due to the absence of a mycobacterial origin of replication. Following transformation into *M. smegmatis*, *M. vaccae*, *M. bovis* BCG and *M. tuberculosis* H37Ra, pAV1 showed stable integration (Freitas-Vieira *et al.,* 1998).

Recently, a method for integrating multiple plasmids into the chromosome of *Mycobacterium* was reported by Saviola and Bishai (2004). This was achieved by cloning the *attB* site into the integrative plasmid pBS20, which carries kanamycin resistance, to form pBS29. Following the integration of pBS29, the provision of *attB* in the vector allowed the insertion of pBS11 (a hygromycin resistance integrative plasmid) into the cloned *attB* (Saviola and Bishai, 2004). Although integration of both plasmids occurred, they were found to be unstable in *M. smegmatis* with 28% and 8% of the bacteria becoming kanamycin and hygromycin sensitive, respectively, over 20 generations. An improvement to <1% reversion to antibiotic sensitivity for both vectors was made possible when the cloned *Int* was removed from pBS29 and pBS11 to form

pBS37 and pBS33, respectively (Saviola and Bishai, 2004). The source of the *Int* that allowed the insertion of pBS37 and pBS33 was pBluescript*Int*, which was introduced after transformation with pBS37 and pBS33 (Saviola and Bishai, 2004).

1.6.1.3 Streptomyces

The well characterised *Streptomyces* phage ϕC31 has commonly been used in the construction of phage-based integration vectors. One of the early developments of ϕC31-based vectors was pKC796 (Kuhstoss *et al.,* 1991). Plasmid pKC796 was designed to demonstrate the efficiency of transformation of phage-based integrative vectors in comparison to integrative plasmid (pKC824), which was derived from a naturally occurring integrative plasmid, pSAM2 (Kuhstoss *et al.,* 1991). Kuhstoss *et al.* (1991) showed that pKC796 produced a transformation efficiency of 3.8×10^6 transformants per μg of DNA compared to 6.4×10^3 for pKC824. Another ϕC31 based integrative vector, pSET152 (Bierman *et al.,* 1992), was later used in a study that demonstrated that the ϕC31 integration mechanism is capable of integration in various *Streptomyces* species via pseudo-*attB* sites (Combes *et al.,* 2002).

Three other phages that have been used in the construction of integrative plasmids were actinophage RP3, bacteriophage VWB and ϕBT1. Actinophage RP3 integrates specifically into the 3' terminal end of the *S. rimosus* R6-554 chromosomal arginine-tRNA (AGG) gene. The phage RP3 integrative module was used in the construction of plasmid pKG2, which is based on the *E. coli* vector pIC20R (Marsh *et al.,* 1984), with the thiostrepton resistance gene from plasmid pIJ702 (Katz *et al.,* 1983) as a selection marker. The transformation efficiency of pKG2 into *S. rimosus* was low (10-100 transformants per μg of DNA). Nevertheless, pKG2 showed stable integration into the chromosome of *S. rimosus*, which retained thiostrepton resistance after six sporulation cycles in the absence of antibiotic.

Isolated from *S. venezuelae* ETH14603, phage VWB was also found to integrate into an argnine-tRNA (CCU) gene (Van Mellaert *et al.,* 1998). The VWB *Int* and *attP* locus were cloned into the same plasmids as the RP3 integration module, producing plasmid pKT02. Although, the plasmid was able to produce stable integrants, the transformation efficiency of pKT02 was lower than ϕC31-based vectors (10 – 50 transformed *S. venezuelae* ETH14630 per μg vector and 300 – 500 transformed *S. lividan* TK24 per μg vector).

The development of phage ϕBT1-based integration vector, pRT801, was achieved by constructing integration vectors similar in design to pSET152, except that ϕBT1 *att-Int* integration module was used, rather the recombination region of ϕC31 (Gregory *et al.,* 2003). Plasmid pRT801 has a broad host-range, with successful transformants generated by conjugation into *S. avermitilis*, *S. cinnamonensis*, *S. fradiae*, *S. lincolnensis*, *S. nogolater*, *S. roseosporus* and *S. venezuelae* (Gregory *et al.,* 2003).

1.6.1.4 Pseudomonas

Bacteriophage ϕCTX was used in the initial attempts at construction of phage-based vectors for efficient and stable integration of exogenous sequences into the *Pseudomonas aeruginosa* chromosome (Wang *et al.,* 1995). Despite the successful integration of *attP*- containing plasmids (pTABF, p1000 and p400) into the host genome, the phage-based system was not user-friendly as it required the co-transformation of an integrase-expressing plasmid, pIBH. In addition, the *attP*-containing plasmid lacked efficient multiple cloning sites (MCS) and did not provide mechanisms for the removal of unwanted plasmid-backbone sequences (Hoang *et al.,* 2000).

In 2000, Hoang *et al*. reported the development of two integration plasmids, mini-CTX1 and mini-CTX2, which were improvements on the previous phage ϕCTX-

based systems. Both plasmids contain a tetracycline resistance gene, *oriT* for conjugation-mediated plasmid transfer, an origin of replication from plasmid pMB1, ϕCTX integrase, a versatile MCS which was flanked by T4 transcription termination elements, and the ϕCTX *attP* site (Hoang *et al.,* 2000). These plasmids also carry yeast Flp recombinase target sites (FRT), which flank the MCS and T4 transcription termination elements. The presence of Flp recombinase from helper plasmid pFLP2, allows the excision of unwanted plasmid-backbone sequences following the integration of mini-CTX-based vectors. The difference between the two mini-CTX vectors is that the transcription of the *Int* gene in mini-CTX2 is driven by a strong *trc* promoter which is regulated by the *Lac* repressor ($lacI^q$).

1.6.1.5 Streptococcus

Temperate phage T12 has been used in the development of a shuttle vector for *Streptococcal* species. Isolated from *S. pyrogenes*, phage T12 is capable of genomic integration into a serine-tRNA (CGA) gene (McShan *et al.,* 1997). The integrative region of phage T12 DNA was cloned into a *Streptococcal* suicide vector, p7ERM-1B, resulting in the stable integrative plasmid, pWM139. The newly constructed vector was capable of replication in *E. coli* but not *S. pyogenes* and its erythromycin resistance marker was used for selecting transformants. However, it was hypothesised that the *lacZ* promoter located upstream of the *Int* gene induced an over-expression of phage integrase in *E. coli*, which is toxic to the bacterium (McShan *et al.,* 1998). To overcome this problem the pWM139 derivative, pWM245, was constructed (McShan *et al.,* 1998). Due to the relatively few unique cloning sites and the absence of blue/white screening of clones, *lacZ* was cloned into pWM245 to form the plasmid p7INT (McShan *et al.,* 1998).

Recently, the site-specific integration system of temperate phage MM1, has been structurally analysed. The *Int-attP* cassette of MM1 was ligated into plasmid pUCE191, resulting in the lincomycin resistance vector, pIAPU1 (Gindreau *et al.,* 2000).

1.6.1.6 Listeria

Listeriophage U153 has been used for the development of integration vector pPL1, which formed single copy integrants in various *L. monocytogenes* strains (Lauer *et al.*, 2002) at a frequency of $\approx 10^4$ per donor cell, in the *comK* gene of the *L. monocytogenes* chromosome. Bacteriophage PSA integrase and *attP* cassette was used in the construction of integrative plasmid, pPL2, which is similar to pPL1 (Lauer *et al.,* 2002). Plasmid pPL2 was shown to integrate into a *L. monocytogenes* arginine-tRNA (UCU) gene with the same frequency as pPL1(Lauer *et al.,* 2002).

1.6.1.7 Enterococcus

The *Int-attP* region of temperate bacteriophage ϕFC1 was ligated into a 3.7-kbp *Hind*III fragment of a pUC19-derived plasmid, pESH2.8 (Yang *et al.,* 2002). The resulting plasmid pEMJ1-1 was able to insert into the chromosome of *E. faecalis* KBL707, following electroporation, with a transformation efficiency of 6×10^3 colonies per μg of DNA (Yang *et al.,* 2002).

1.6.1.8 Rhizobium

The integrative mechanism of temperate phage *16-3* of *Rhizobium meliloti 41* has been characterised extensively (Dorgai *et al.,* 1993; Papp *et al.,* 1993; Semsey *et al.,* 1999; Semsey *et al.,* 2002). Hermesz *et al.* (1992) constructed two non-replicating integrative plasmids by cloning the phage *16-3 attP* site into plasmids pLAFR1 and pSUP202, resulting in vectors patt164 and patt202, respectively (Hermesz *et al.,* 1992).

Both patt164 and patt202 formed stable integrants in the presences of helper phage, which provided integrase enzyme (Hermesz *et al.,* 1992).

Elo *et al.* (1998) reported the construction of a set of integrative promoter-cloning vectors for translational (pEP226, pEP227, pEP228) and transcriptional (pEP181 and pEP184) fusion studies in *R. meliloti*. The set of vectors was based on a broad host range, low copy number plasmid pRK290(*IncP1*) carrying the phage *16-3 attP* site and *E. coli lacZ* gene. The integration of these vectors was achieved by the provision of *Int* protein in *trans* from helper plasmid pSEM102, which carries the phage *16-3 Int* gene (Elo *et al.,* 1998).

The development of a more efficient vector that removed the need for helper phage, or plasmid, and which provides the *Int* protein in *trans*, was first achieved by Semsey *et al.* (1999). The *attP-Int* cassette of phage *16-3* was cloned into an *E. coli/ R. meliloti* expression vector, pSEM91. The resulting vector pSEM167 was successfully integrated into a number of bacterial genera that include *Rhizobium*, *Bradyhizobium*, *Azorhizobium* and *Agrobacterium* (Semsey *et al.,* 1999; Semsey *et al.,* 2002).

Recently, a new integrative plasmid vector, pGSB1, has been constructed carrying both *attP* and an *Int* gene (Ferenczi *et al.*, 2004). Plasmid pGSB1, when introduced into *R. meliloti* and *Azorhizobium tumefaciens*, was able to form single copy stable integrants.

1.6.2 Current developments in integrative vectors

The cloning of *attP*, *Int* genes and the removal of origins of replication, has allowed phage-based vectors to insert as a single copy of exogenous DNA into the targeted host bacterium. However, the majority of these vectors retain their plasmid-backbone sequences. The current focus on integrative vectors is on the development of a single copy, stable transformation system, with the ability to remove unwanted

plasmid backbone sequences (Hoang *et al.*, 2000; Shimizu-Kadota, 2001; Rossignol *et al.*, 2002).

Recently, three methods of obtaining such vectors were developed for *P. aeruginosa* (Hoang *et al.,* 2000), *Lactobacillus casei* (Shimizu-Kadota, 2001) and *E. coli* (Rossignol *et al.,* 2002).

As described above, the *P. aeruginosa* integrative vectors mini-CTX1 and mini-CTX2, carry yeast *FRT* sites and were effective in removal of the unwanted plasmid sequences that include the antibiotic resistance gene and the *Int* gene (Figure 1.10). The helper plasmid, pFLP2, which provided the yeast FLP recombinase, catalyses the *in vivo* excisive reaction of the undesirable sequences. Plasmid pFLP2 also carries a negative selective marker, the sucrase (*sacB*) gene (Hoang *et al.,* 1998). *P. aeruginosa* strains containing the *sacB* gene in single- or multiple copy were highly sensitive to 5% sucrose in rich medium (Schweizer, 1992). After integration of mini-CTX-based vector, helper plasmid pFLP2 was removed by growing the transformants on sucrose based medium (Hoang *et al.,* 1998; Hoang *et al.,* 2000).

The mechanisms used with plasmid pMSK761, to achieve chromosomal integration and removal of unwanted vector sequences, involved the use of the *Lactobacillus casei* phage ϕFSW site-specific integration module and homologous recombination (Figure1.11). Plasmid pMSK761 has a cloned *ori* that is designed to allow replication in *E. coli* but not in *L. casei* (Shimizu-Kadota, 2001). Following the insertion of pMSK761 into the *L. casei* genome via phage ϕFSW *Int* gene and *attP* site, endogenous homologous recombination was used to maintain the stability of the integration by eliminating sequences on the vector that were not required in the final integrant (Figure 1.11). This was achieved by cloning two copies of the gene that was to be integrated.

The third integration system, used for *E. coli* transformation, requires two components:

1. a helper plasmid, pHK-Int, to provide the *Int* gene, which has a thermosensitive replicon that is lost after prolonged exposure to 42°C. Expression of the *Int* gene is controlled by the presence of the λ P_R promoter and temperature sensitive *cI857* repressor. At low temperature (≈32°C), cI^{857} repressor binds to the operator of P_R promoter, thus inhibiting expression of *Int*. However, at 37°C, cI^{857} denatures, thus allowing basal level of *Int* gene expression.
2. A non-replicating integrative plasmid (pHK11), which delivers the genetic material into the genome of its host. The plasmid also contains an *attP* that allows integration, and antibiotic resistance gene for hygromycin or apramycin.

(Rossignol *et al.,* 2002)

The system was used to generate genetically engineered bacteria by first transforming *E. coli* with helper plasmid, pHK-Int (Rossignol *et al.,* 2002) (Figure 1.12). Transformants were then allowed to grow at 37°C to accumulate high pHK-Int copy numbers and provide a basal level of *Int* expression, prior to transformation of competent cells with pHK-11. After the introduction of plasmid pHK11 into pHK-Int transformed cells, they were allowed to grow at 30°C to ensure the suppression of the *Int* gene (Rossignol *et al.,* 2002). The basal level of *Int* that was produced prior to suppression allowed pHK11 to be integrated into the *E. coli* chromosome. The unwanted helper plasmid was then removed by incubating transformants at 42°C overnight (Rossignol *et al.,* 2002).

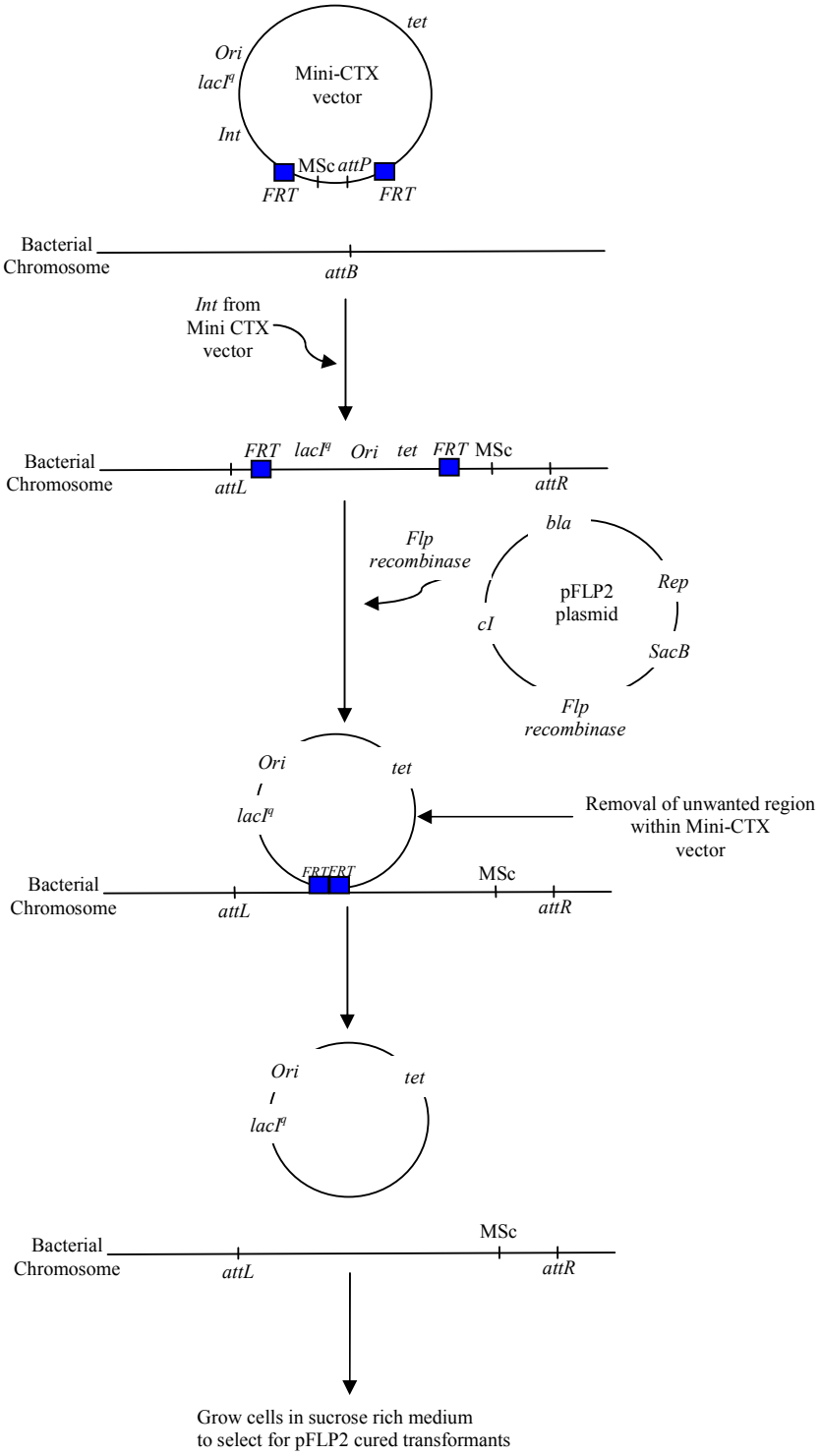


Figure 1.10: A flow diagram illustrating the integrative procedure that allowed mini-CTX vector to achieve stable, single copy integration into *P. aeruginosa*. (Hoang *et al.*, 2000; Hoang *et al.*, 1998)

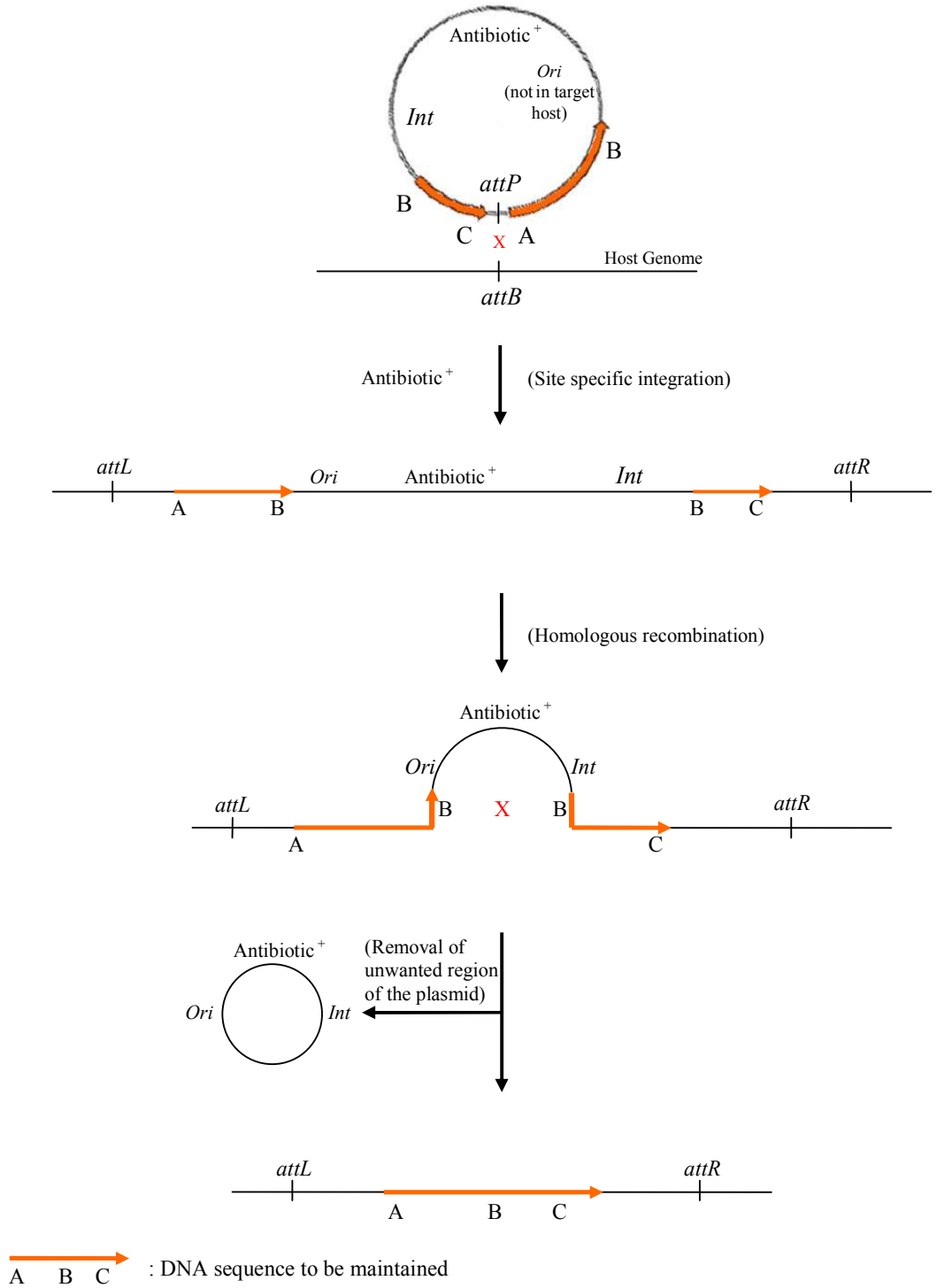


Figure 1.11: A flow diagram depicting the procedure that Shimizu-kadota (2001) used to achieve stable, single copy integration of plasmid pMSK61. Following integration, the introduced DNA sequence, ABC, forms a complete gene by endogenous homologous recombination with two copies of sequence B. The resulting process allows the removal of unwanted regions (*ori*, antibiotic gene and *Int*).

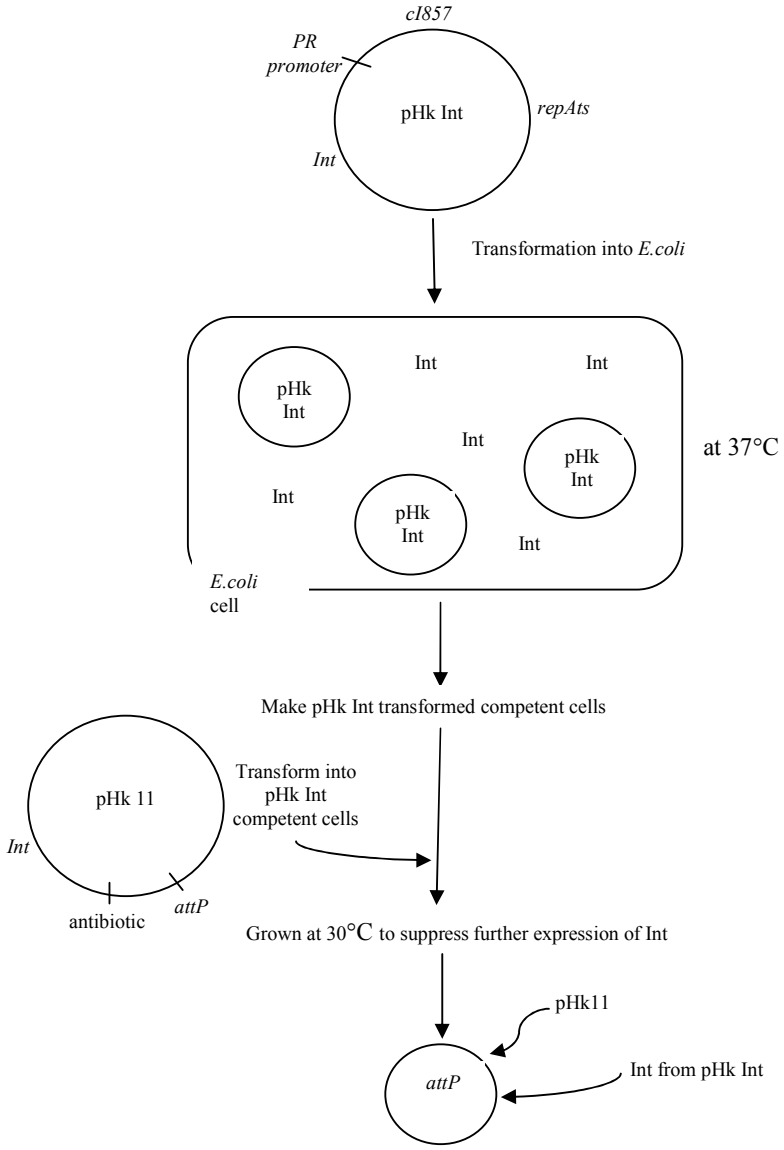


Figure 1.12: A flow diagram depicting the use of phage HK022-based integrative vector by Rossignol *et al.* (2001).

Despite the progress in developing integrative systems, there has been a lack of work for this application in rumen bacteria. Although rumen phages have been phenotypically characterized (Klieve *et al.*, 1991; Klieve and Bauchop, 1988), attempts at isolating recombinase genes or enzymes from these phage have been reported only on phages ϕAR29 (Gregg *et al.*, 1994b) and M1 (Cheong and Brooker, 1998). The report found on the construction of a phage-based transformation system was on plasmid

vector pBA, which was derived from phage ϕAR29 (Wong *et al*., 2003) and is discussed below.

1.7: Ruminal *Bacteroides*

1.7.1: Characteristics of *Bacteroides*

Bacteroides are obligately anaerobic, Gram-negative, nonsporulating, pleomorphic rod-shaped bacteria (Avgustin *et al.,* 1994). A widely adapted species, *B. ruminicola* is one of the most numerous species of bacteria that inhabits the rumen and the hind gut of non-ruminants, including pigs and humans (Comstock and Coyne, 2003; Xu *et al.*, 2003). Studies conducted by VanGylswyk (1990) concluded that *B. ruminicola* accounted for as much as 60% of the total bacteria isolated from rumen contents of silage-fed cows. *Bacteroides* has been shown to be an extremely diverse group at the genotypic level (Comstock and Coyne, 2003). Due to this, many organisms that were formerly classified as *Bacteroides* have been reclassified under the genus *Prevotella* (Shah and Collins, 1990). Two example of ruminal *Bacteroides* that were isolated from the sheep rumen were used in this project.

1.7.2: *Bacteroides thetaiotaomicron* Strain AR29 and *Bacteroides uniformis* Strain AR20

Bacteroides thetaiotaomicron AR29 and *Bacteroides uniformis* AR20 were first described by Hudman and Gregg (1989) in a study on genetic diversity among strains of ruminal bacteria. Although originally classified as *Bacteroides ruminicola* sub-sp. *brevis* (Hudman and Gregg, 1989) among a bacterial group that was subsequently reclassified as *Prevotella* (Shah and Collins, 1990; Shah and Gharbia, 1993), 16S rRNA gene sequence analysis showed that they were more closely related to *Bacteroides* than to *Prevotella*. Consensus tree construction indicated that AR29 (16S sequence accession

number AF139525) and AR20 (AF139524) are ruminal strains of the common colonic species *Bacteroides thetaiotaomicron* and *B. uniformis*, respectively (C. Wong, unpublished).

Both bacteria were involved in the Fluoroacetate Dehalogenase Project. Initiated by Applied Biotechnology Pty. Ltd. and largely funded by Meat and Livestock Australia and its precursor organisations, in which the primary aim was to express a fluoroacetate dehalogenase enzyme in rumen bacteria, for the detoxification of the natural plant toxin fluoroacetate in livestock. This problem was estimated to cost cattle producers approximately AU$140,000 per property each year (Gregg and Sharpe, 1991). Interestingly, although strains of *B. fibrisolvens* modified with the same gene were able to protect the host ruminant against the toxin (Gregg *et al*., 1998), the transgenic *Bacteroides* strain could not detoxify fluoroacetate from the surrounding medium. This was attributed to the intracellular retention of the detoxifying enzyme and the extracellular location of the toxin (Wong *et al*., 2003). Since their isolation, molecular work on AR29 and AR20 has included the development of transformation vectors and the introduction of the heterologous dehalogenase gene, which demonstrated the ability of the bacteria to express genes from other species.

1.7.3: Difficulties in Transforming *B. thetaiotaomicron* AR29 and *B. uniformis* AR20

Progress on genetic modification of both *Bacteroides* strains has been slow, owing to the poorly understood physiology and enzymology of the bacteria. The largest hurdle for attempts to transform both strains has been their formidable restriction barrier which was most extensively studied in AR20 (Hamdorf, 1998). Investigation on DNase activities in *Bacteroides* showed the restriction barrier to be responsible for the reduction of transformation efficiency by up to 10^5-fold (Hamdorf, 1998). Among the

aspects of genetic transformation that have been investigated, the lysogenic phage from *B. thetaiotaomicron* AR29 (ϕAR29) was tested as a component of genetic transfer mechanisms in AR29 and AR20 (Klieve et al., 1991; Gregg et al., 1994; Wong et al., 2003).

1.8: Rumen Bacteriophage ϕAR29 and Development of a Transformation System for AR29 and AR20

Bacteriophage ϕAR29 was isolated by Klieve *et al.,* (1989) from *Bacteroides* strain AR29. Morphologically, ϕAR29 consists of a head of 60 nm diameter, and a tail length and width of 120 nm × 12 nm respectively. Electron-microscopy studies did not observe the presence of a base plate or collar (Klieve *et al.,* 1989). Despite extensive host sensitivity tests conducted on ϕAR29, it did not cause lytic infection of any other ruminal bacteria (Klieve *et al.,* 1989). In strain AR29, ϕAR29 has been shown to be a highly stable lysogen. The use of Mitomycin C, acridine orange, and UV illumination treatments was investigated as a means to induce excision of the prophage, with the ultimate goal of curing the phage from its host. All attempts failed to produce unlysogenised forms of AR29 (Klieve *et al.,* 1989).

Due to the stability of ϕAR29 lysogeny, it was proposed as a source of material for the construction of an integrative shuttle vector (Figure 1.13: (Klieve *et al.*, 1989; Gregg *et al.*, 1994b). Kennedy (Hons thesis, 1989) was able to locate the DNA fragment that contains the phage's integration site, by restriction digestion and Southern blotting of free phage DNA and AR29 genomic DNA containing the prophage. The integration fragment was cloned in pTZ19U as a *Hin*dIII/*Eco*RV fragment and was sequenced (Gregg *et al*., 1994). The sequence contained two open reading frames (ORFs) of approximately 0.8 kb and 0.6 kb, which encoded *Int*- and *Xis*-like proteins, respectively. This DNA fragment was included in plasmid pBA, which was constructed as a shuttle

vector for transformation of *E. coli* and of *Bacteroides* strains AR20 and AR29. The construction of pBA is illustrated in figure 1.14. The integration module from bacteriophage ϕAR29 was inserted into plasmid pTZ19U to form pIF, which was then ligated to *Bacteroides* plasmid pBI191 (Wong *et al.,* 2003). The final plasmid, pBA, was successfully inserted into *E. coli, B. uniformis* AR20, and *B. thetaiotaomicron* AR29 (Wong *et al.,* 2003). In addition, other vectors have been derived from pBA, including promoter rescue plasmid, pPPR, the smaller cloning vector pBAT, and dehalogenase gene expressing vector pBAC (Wong *et al.*, 2003). However, integration of pBA into the chromosome of AR29 or AR20 was not observed.

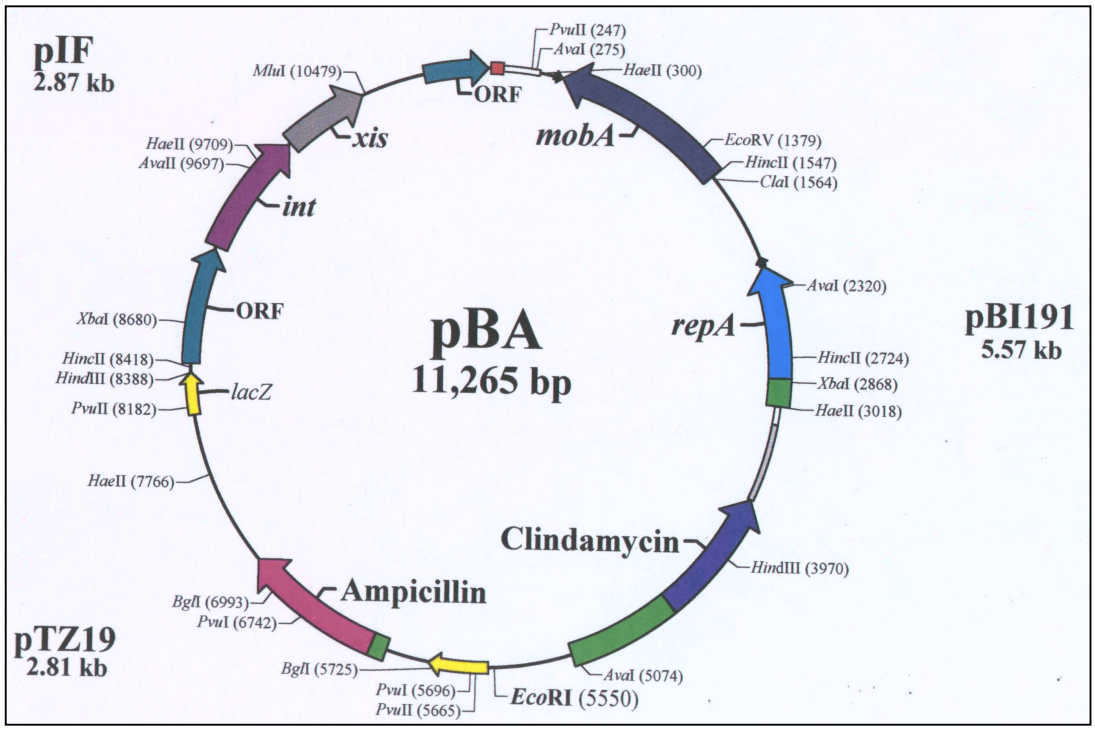


Figure 1.13: Diagram of pBA shuttle vector (Wong *et al.*, 2003).

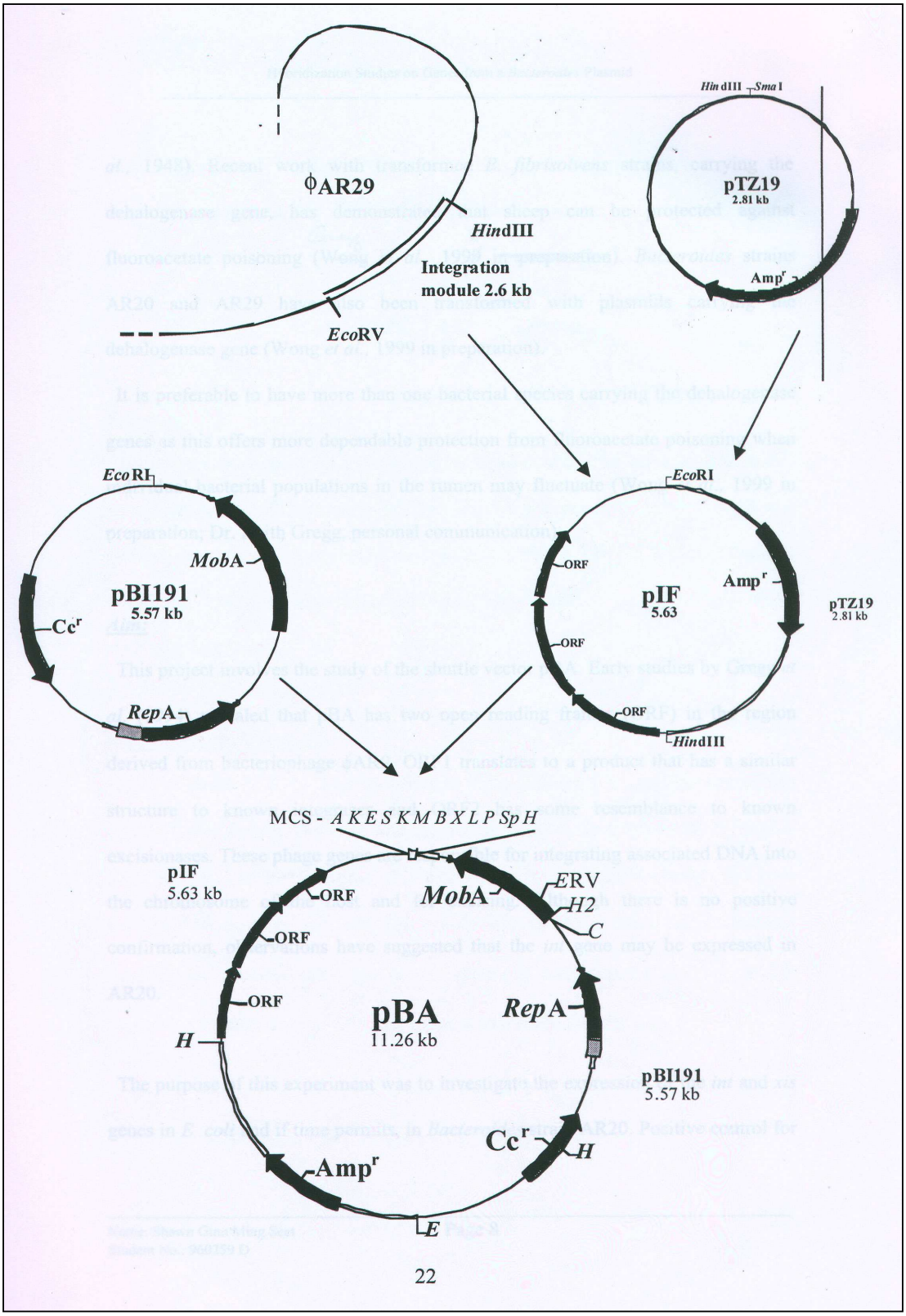


Figure 1.14: A schematic diagram showing the construction of pBA. (Wong *et al.*, 2003)

1.9: Summary

In addition to the knowledge of how intrinsic and extrinsic factors affect gene expression, the use of lambda phage site-specific integration mechanisms as a model have helped understand how bacteriophage integration genes, *Int* and *Xis*, function and are expressed (Wulff & Rosenberg, 1983; Echols & Guarneros, 1983). Stable genetic modification of rumen bacteria remains difficult, despite the advances in molecular tools and techniques. The success in development of shuttle plasmids using integration genes from bacteriophages in non-ruminal bacteria (described above) suggests that similar developments for ruminal species could help to advance this research field. Construction of such a plasmid has been a focus in developing gene transfer mechanisms for *B. thetaiotaomicron* AR29 and *B. uniformis* AR20 (Klieve *et al.*, 1989; Gregg *et al.*, 1994b; Wong *et al.*, 2003).

1.10: Aim of This Project

The main aim of this study was to establish the factors required for a plasmid such as pBA to function as an integrative transformation system for AR20 and AR29. The approach was firstly to determine which integration-related genes encoded within pBA are expressed in *E. coli*, AR29 and AR20 and, if necessary, to modify the plasmid to ensure expression of the integrase gene. Other objectives included:

- determining the precise location of *attP* and *attB* site for ϕAR29 site-specific recombination, using data from previous studies and DNA sequencing to identify attachment sites..
- The sequencing and mapping of the ϕAR29 genome to obtain a more complete understanding of the biology of ϕAR29 and to investigate further the mechanisms of phage integration, such as promoters and regulatory proteins, and

- Finally, to examine the capability of pBA for host integration by the mechanism of site-specific integration

Chapter 2: Materials and Methods

2.0: General Chemicals

Chemicals used in this work were analytical reagent grade and, unless specified, were used without further purification. All percentage values are expressed as weight/volume for dissolved solids, or vol/vol for mixed liquids.

Table 2.0: General Chemicals used throughout the project.

Chemicals	% (w/v) or (v/v)	Supplier
Chloroform	99	AJAX CHEMICALS
Ethanol	95	AJAX CHEMICALS
Isoamyl alcohol	95	AJAX CHEMICALS
Phenol	95	AJAX CHEMICALS

2.1: Specialised Chemicals, Enzymes and Laboratory Equipments

Table 2.1: Chemicals/reagents/kits/laboratory equipment used in this study.and their suppliers

Compounds	Supplier
100 bp DNA Marker	Promega
Agarose	Promega
Agar	BBL
Ampicillin	Sigma/Roche
BRESA-CLEAN™ DNA Purification Kit	Bresatec (GeneWorks)
Calcium Chloride	Merck
Cellobiose	ICN Biomedicals Inc
Clindamycin	Sigma/Roche
Cysteine.HCl	Sigma/Roche
D-Glucose anhydrous	Asia Pacific Specialty Chemical Limited
dATP, dGTP, dCTP, dTTP	Promega
Diethyl Pyrocarbonate	Sigma/Roche
DIG DNA Labelling and Detection Kit	Boehringer Mannheim
DIG Easy Hyb Granules	Boehringer Mannheim
DIG Labelled DNA Molecular Weight Marker II	Boehringer Mannheim
DIG Wash and Block buffer Set	Boehringer Mannheim
Dithiothreitol (DTT)	Promega
Ethidium Bromide	ICN Biomedical Inc.
Haemin	Sigma/Roche
Isopropyl-β-D-thiogalactopyranoside	Sigma/Roche
λ DNA digested with *Hind*III	Biotech
Maltose	Sigma/Roche
Mitomycin C	Sigma/Roche
NBT/BCIP Stock Solution	Boehringer Mannheim
UtraClean™ PCR Clean-up™ Kit	Mo Bio Laboraties, Inc
K_2HPO_4	Ajax Laboratory Chemical
KH_2PO_4	Merck
PCR reaction Buffers and Enzymes	Sigma/Roche
*Hind*III restriction endonuclease	Promega
Sodium Carbonate	Sigma/Roche
Sodium Chloride	Merck
Trishydroxymethylaminomethane (Tris) Base	Gibco BRL
Tryptone peptone	DIFCO
Wizard® *Plus* SV Minipreps DNA Purification	Promega

Systems	
Yeast Extract	BBL
Enzymes	
AMV Reverse Transcriptase (AMV RT)	Bresatec (GeneWorks) and Promega
BamHI restriction endonuclease	Promega
Cal	Promega
Calf Intestinal Alkaline Phosphatase	Promega
Calf Thymus DNA Standard	Amersham Biosciences
DNA Ligase	Promega
DNA polymerase (Klenow)	Fisher Biotec
EcoRI restriction endonuclease	Promega
EcoRV restriction endonuclease	Promega
KnpI restriction endonuclease	Promega
Lysozyme	Boehringer Mannheim and Sigma
Proteinase K	Merck
RNase ONE™	Promega
RNasin® Ribonuclease Inhibitor	Promega
RQ1 Rnase-Free DNase	Promega
Taq DNA Polymerase	Gibco BRL
Laboratory Equipment	
Camera	
SC35	Olympus
Centrifuge	
Capsule HF-120	Tomy
Eppendorf 5417C	Eppendorf
Sigma 1-15	Sigma
Avanti J-25I	Beckman
SpeedVac Concentrator	Savant
Electroporation	
Capacitance Extender Plus	Bio-Rad
Gene Pulser II	Bio-Rad
Pulse Controller Plus	Bio-Rad
Gel imaging	
Gel Doc 1000	Bio-Rad
Heat Block	
Single Dry Block Heater DBH10D	Ratek Instruments
Dry Block Heater DBH20D	Ratek Instruments
Microscope	
Olympus BH2 microscope	Olympus
PCR	
Master Gradient	Eppendorf
GeneAmp PCR System 2400	PerkinElmer
Scanner	
Astra 610S	UMAX
Spectrophotometer	
Lambda 25 UV/Vis	PerkinElmer
UV-1201 UV-VIS	Shimadzu Scientific
DyNA Quant 200	Hoefer
Ultraviolet Chamber	
GS Gene Linker™ UV Chamber	Bio-Rad

2.2: *In silico* Analysis

Table 2.2: Description of the programs and databases used in this study.

Program/Database	Note	Reference	Website
Bioedit5.0.9	DNA and protein analysis program used for analysis of alignment of protein and DNA sequences	(Hall, 1999)	http://www.mbio.ncsu.edu/BioEdit/bioedit.htmL
Bionagivator GCG Terminator	Web-based terminator prediction program	(Brendel and Trifonov, 1984)	http://www.angis.org.au/htmL/index.htmL
European Molecular Biology Laboratory - European Bioinformatics Institut (EMBL-EBI)	A web-based database search of protein homology using IntePro PROSITE PFAM	(Mulder *et al.*, 2005; Hulo *et al.*, 2004; Bateman *et al.*, 2004)	http://www.ebi.ac.uk/
FGENESV	A trained Pattern/Markov chain-based viral gene prediction	(Xu *et al.*, 2003b)	http://www.softberry.com/berry.phtmL?topic=virus&group=programs&subgroup=gfindv
FSFinder	Predicts (-/+1) programmed frameshift	(Moon *et al.*)	http://wilab.inha.ac.kr/FSFinder/
GeneTool Lite Version 1.0	An analytical DNA and primer design software by Biotools Incorporated.	(Wishart *et al.*, 2000)	http://www.biotools.com/downloads/productinfo.html
GeneMark Heuristic model	Open reading frame (ORF) prediction	(Besemer and Borodovsky, 1999)	http://opal.biology.gatech.edu/GeneMark/heuristic_hmm2.cgi
Mega3	Molecular Evolutionary Genetics Analysis	(Kumar *et al.*, 2004)	http://www.megasoftware.net/mega3/index.htmL
Motif Scan	Search for all known motifs that occur in a sequence	(Falquet *et al.*, 2002)	http://myhits.isb-sib.ch/cgi-bin/motif_scan
National Center for Biotechology Information (NCBI)	A web-based program applied in BLAST search for protein and nucleotides sequences for homology.	(Altschul *et al.*, 1990; Gish and States, 1993; Madden *et al.*, 1996; Altschul *et al.*, 1997; Zhang and Madden, 1997; Zhang *et al.*, 2000)	http://www.ncbi.nlm.nih.gov/
Neural Network Promoter Prediction version 2.2	Web-based promoter prediction program	(Reese, 2001)	http://www.fruitfly.org/seq_tools/promoter.htmL
Oilgonucleotide Properties Calculator	A web-based primer analytical program.		www.basic.nwu.edu/biotools/oligocalc.htmL
PHIRE version 1.0	Predicts regulatory elements (promoter/terminators) in phage genome	(Lavigne *et al.*, 2004)	http://www.agr.kuleuven.ac.be/logt/PHIRE.htm
Programmed Frameshift Finder	Predicts (-1) programmed frameshift	(Xu *et al.*, 2004)	http://chainmail.bio.pitt.edu/~junxu/webshift.html

Rnadraw version 1.1	Aprogram used to construct diagrams with predicted stem loops in RNA	(Matzura and Wennborg, 1996)	http://www.rnadraw.com/
tRNAscan-SE Search Server	Seach for tRNA sequence in genome	(Lowe and Eddy, 1997)	http://www.genetics.wustl.edu/eddy/tRNAscan-SE/
Vector NTI Advance™	DNA/protein sequence analysis program used in the construction of genetic/plasmid maps, identifying open reading frames and translation of nucleotides sequences to amino acids sequences. Also used in sequence editing, alignment tools and the construction of primers	(Lu and Moriyama, 2004)	http://www.invitrogen.com/content.cfm?pageid=10373

2.3: Bacterial Strains

Bacterial strains: *E. coli* SCS110, PMC112, *B. uniformis* AR20 and *B. thetaiotaomicron* AR29 were used in this study. Their characteristics are described in Table 2.3, whilst the plasmids used in this study are described in Table 2.4.

Table 2.3: The characteristics of the bacterial cells used in this study

Species	Genotype and/or Phenotype	Reference
Escherichia coli Strain PMC112	*sup*E, *hsd*, Δ5, *thi* Δ(*lac-pro*A,B), F'[*tra*D36, *pro*A,B^{+}, *lac*I^{q}, *lac*Z, ΔM15],*mcr*A^{-}B^{-}	(Gibson, 1984)4); Peter McCallum Cancer Institute)
Escherichia coli Strain SCS110	*rpsL*, (Strr), *thr*, *leu*, *endA*, *thi-1*, *lacY*, *galK*, *galT*, *ara*, *tonA*, *tsx*, *dam*, *dcm*, *supE44* Δ(*lac-proAB*) [F´ *traD36 proABlacI*^{q}*ZΔM15*]	Stratagene(Cat# 200275)
Bacteroides uniformis strain AR20 (Wild type strain)	Gram-negative, coccoid to oval rod (shape depends on the stages of growth), exists singly, pairs or conjugate clusters of network chains, capable of encapsulation, digests arabinose, cellobiose, glucose, xylose, xylan, fucose, galactose, lactose, maltose, melibiose and able to hydrolyse para-nitrophenol (p-np) from the glycoside, p-np-β-D-glucopyranoside and p-n-β-D-xylanopyroside	(Hudman and Gregg, 1989; Wong *et al.*, Unpublished)d)
Bacteroides thetaiotaomicron strain AR29 (Wild type strain)	Gram-negative, coccoid to oval rod (shape depends on the stages of growth), exists singly, pairs or conjugate clusters of network chains, capable of encapsulation, digests arabinose, glucose, fucose, galactose, lactose, maltose, mannose, melibiose	(Hudman and Gregg, 1989; Wong *et al.*, Unpublished)d)

Table 2.4: The characteristics of the plasmids used in this study

Plasmid	Description	Reference
pUK21	similar to pUC18, containing novel MCS, Km^{r} and the *lac*Z under control of a weak promoter.	(Vieira and Messing, 1991)
pBA	containing ϕAR29 integration module, repA, mobA^{r}, Clin^{r}, Amp^{r}	(Wong *et al.*, 2003)

2.4: Commonly Used Solutions

All chemicals and reagents were dissolved in MilliQ reverse osmosis purified water

Antibiotics	The concentration of antibiotic stock solutions was 10 mg/mL for both ampicillin and clindamycin, dissolved in 70% v/v ethanol and in water respectively. Working concentrations were 10 μg/mL for both antibiotics. Stock of Clindamycin was stored at 4°C and ampicillin at –20°C.
Cryoprotected bacterial cultures	0.1 mL of resazurin (0.1%) 43 mL sterile water 50 mL of glycerol After autoclaving, 2 mL of cysteine.HCl solution (2.5%) and 5 mL sodium carbonate solution (8%) were added and the final solution was mixed with bacterial culture at a ratio of 1 mL per 1.5 mL culture. For ruminal bacteria, cryoprotected stocks were made by combining equal volumes of 40% glycerol in RF medium with a dense bacterial suspension (final concentration of glycerol was 20%)
0.5 M EDTA (pH 8.0)	0.5 M Ethylenediaminetetra-acetic acid (diodium salt) pH adjusted to 8.0 with 10 M NaOH
Gel loading buffer	1 mM EDTA, pH 8.0 0.25% bromophenol blue 0.25% xylene cyanol 50% glycerol
1M Glucose	18 g of Glucose dissolve in 90 mL of water Final volume to 100ml, filter sterilized (0.22 μm filter)
GTE buffer (Glucose/Tris/EDTA)	50 mM glucose 25 mM Tris.HCl, pH 8.0 10 mM EDTA
100 mM IPTG	0.24 g of IPTG 10 mL water Filter sterilized and store at –20 °C (0.22 μm filter)
3 M Potassium Acetate	60 mL of 5 M potassium acetate was combined with 11.5 mL of glacial acetic acid, the pH was adjusted to 4.4 and 28.5 mL of water was added.
Proteinase K (10 mg/mL)	Crystalline Proteinase K dissolved in sterile water
0.89% Saline	0.89% w/v NaCl
SDS (10%)	10 g of SDS was dissolved to a final volume of 100 mL

	in water and heated to 68°C until dissolved
3 M Sodium Acetate	246 g sodium acetate was dissolved in 500 mL of water and the pH was adjusted to 5.2 with acetic acid. Water was added to make the final volume 1 litre
Sodium hydroxide/SDS solution	0.2 M NaOH
	1.00 % SDS in water
20 × SSC	3 M NaCl
	0.3 M trisodium citrate
	adjusted pH to 7 with HCl
20 × TBE electrophoresis buffer	2.6 mM EDTA (disodium salt)
	44 mM boric acid
(Tris borate EDTA)	134 mM Tris Base
	pH 8.8
20 × TAE electrophoresis buffer	40 mM Tris Base
	20 mM sodium acetate
(Tris acetate EDTA)	1 mM EDTA (disodium salt)
	pH 8.2
TE buffer	10 mM Tris.HCl, pH 8.0
	1 mM EDTA

2.5: Bacterial Growth Media

All anaerobic media and media for *E.coli* were autoclaved at 121 °C for 30 minutes.

Luria-Bertani medium (LB broth)	1.00 % bacteriological peptone
	1.00 % tryptone
	0.50 % sodium chloride
Salt solution A	0.30 % potassium di-hydrogen phosphate
	0.60 % sodium chloride
	0.30 % ammonium sulfate
	0.03 % calcium chloride
	0.03 % magnesium sulfate
Salt solution B	0.3 % di-potassium-hydrogenorthophosphate
Rumen-Fluid (RF) medium (100 mL)	16.50 mL of salt solution A
	16.50 mL of salt solution B
	33.00 mL of clarified rumen fluid
	(clarification of rumen fluid was achieved

	by centrifugation at 25931 × g for 10 min)
	0.10 g peptone
	0.10 g yeast extract
	0.50 g $NaHCO_3$
	0.20 g cellobiose
	0.20 g glucose
	0.10 mL resazurin (0.1%)
	20.00 mg cysteine.HCl
	34.00 mL water
	Mixture was boiled for 20-30 min under a stream of CO_2/H_2 (96/4), chilled on ice, solid cysteine.HCl was then added and the bottles were sealed anaerobically
Mineral solution	6.00 g KH_2PO_4
	12.00 g NaCl
	6.00 g $(NH_4)_2SO_4$
	1.60 g $CaCl_2.2\ H_2O$
	2.50 g $MgSO_4.7\ H_2O$
	1000 mL water
Volatile Fatty Acid Mixutre	4.25 mL acetic acid
	1.50 mL propionic acid
	1.00 mL butyric acid
	0.25 mL n-valeric acid
	0.25 mL iso-butyric acid
	0.25 mL D-L-2-methyl butyric acid
	0.25 mL iso-valeric acid
Defined Rumen Bacterial (DF) Medium (100mL)	3.80 mL mineral solution
	0.03 g K_2HPO_4
	0.20 g tryptone peptone
	0.05 g yeast extract
	0.31 mL volatile fatty acid mixture
	1.00 mg haemin
	0.05 g glucose
	0.05 g cellobiose
	0.08 mL Glycerol (≈ 0.1 g)

	0.40 g Na_2CO_3 0.10 mL resazurin (0.1 %) 0.25 g Cysteine.HCl 95.70 mL water added to make up 100 mL of solution and adjusted to pH 6.7 – 6.8. Mixture was boiled for 20 – 30 min under a stream of CO_2/H_2 (96/4), chilled on ice, solid cysteine.HCl was then added and the bottles were sealed anaerobically
Agar Culture Plates *E.coli*	1.8% agar was dissolved in LB broth by autoclaving. When necessary, the appropriate antibiotic was added after cooling the medium below 50°C and mixed gently by inversion. The medium was poured into polystyrene petri dishes.
Bacteroides	1.5% agar was dissolved in RF or in DF medium by autoclaving. Solidified medium was remelted in boiling water, cooled to 55°C and appropriate antibiotics were added and mixed by inversion. The medium was poured into plastic petri-dishes and left to stand for at least one day in the anaerobic chamber before use.

2.6: General Methods

2.6.1: Agarose Gel Electrophoresis

Nucleic acids were analysed using agarose gel electrophoresis as described by Sambrook *et al*., (1989). All electrophoresis gels contained TAE buffer and the percentage of agarose used in the gels ranged from 1.0 – 1.5%. Electrophoresis was carried out using either a Bio-Rad Mini Sub™ or a Bio-Rad Wide Mini™ Sub and the gels electrophoresed at 5 Volts/cm and 10 Volts/cm respectively, until the bromophenol blue dye had migrated 2/3 the length of the gel. The sizes of nucleic acid fragments in each band were estimated by comparison with DNA markers of precisely known fragment size (λ DNA digested with *Hin*dIII and 100 bp ladder). DNA was detected by staining the gel in a 0.5 μg/mL solution of ethidium bromide. Visualisation and

recording of gel results was achieved using Biorad's GEL Doc 1000 system, with UV illumination (peak wavelength 302nm) and Molecular Analyst Software, Version 1.4 (Biorad).

2.6.2: Ethanol Precipitation

Nucleic acid samples were precipitated by adding 2.5 volumes of ice-cold ethanol and allowing them to precipitate for 1 hour at -20°C. Subsquently, the samples were centrifuged at 21,000 × g for 5 minutes, the pellets rinsed with 70% ethanol and then vacuum dried. Samples were redissolved in water or TE buffer. Re-precipitation of DNA was by adding 0.1 volume of 3 M sodium acetate (pH 5.2), followed by 2.5 volumes of cold ethanol. Samples were chilled at –20°C for at least 30 minutes, and centrifuged at 21,000 × g for 10 – 15 minutes. The supernatant was discarded, the pellet washed in ice cold 70% ethanol, vacuum dried, and the DNA dissolved in sterile water or TE buffer.

2.6.3: Phenol:Chloroform:Isoamyl Alcohol Extraction

Phenol extractions were performed by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) to samples and vortexing for 15 seconds. Samples were then centrifuged at 21,000 × g for 5 minutes and the aqueous phase was retained. An equal volume of chloroform:isoamyl alcohol (24:1) was then added, mixed thoroughly and centrifuged at 21,000 × g for 5 minutes. The top, aqueous phase was retained.

2.6.4: Polymerase Chain Reaction (PCR)

PCR was performed as described by the supplier of *Taq* polymerase (Gibco BRL). The composition of the reaction mixture and PCR conditions are shown in

Tables 2.5 and 2.6, respectively.

Table 2.5: The components and their final concentration in a PCR mixture.

Stock solution of reaction component	Final concentration of component
10 × PCR buffer minus Mg	1 ×
10 mM dNTP	0.2 mM
50 mM $MgCl_2$	2.5 mM
Primer mix	0.5 μM
Taq DNA Polymerase	2.5 units

Table 2.6: Thermal cycle conditions used for PCR

Cycle	1 cycle	25 cycles			1 cycle	
Temperature (°C)	94	94	55-60	72	72	14
Duration	5.00 min	30 sec	30 sec	30 sec	7 min	∝

The annealing temperature and elongation time of the PCR varied according to the primers and length of the product.

2.6.5: Primer Design

Primers were designed using the program Gene Tool Lite version 1.0., Vector NTI Advance™ and Oligonucleotide Properties Calculator were then applied to cross-check primer sequences for secondary structures and self priming. The characteristics that were taken into account when designing primers are shown in Table 2.7.

Table 2.7: Factors taken into account for the designing of primers.

Length	18-30 bases
% G+C	45-55
Tm (°C)	60-65

The annealing temperatures (Tm) of the primers were calculated using the formula:

$$Tm = 22 + 1.43\ (n+n^{G+C})$$

(Wu *et al.* 1991)

Where n is the total number of bases in the oligonucleotide and n^{G+C} is the sum of the number of G and C residues within the oligonucleotide.

2.6.6: Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

The reverse transcription reaction of RT-PCR used Avian Myeloblastosis Virus (AMV) reverse transcriptase. In short, the RT reaction was performed by heating the RNA sample and primer to remove any secondary structure. Deoxynucleotides and reaction buffer were added, together with water to make up the required reaction volume. Finally, RNasin and reverse transcriptase were added, mixed gently, and incubated at 42°C for 40 minutes. Thereafter, a standard PCR was performed using products of the RT reaction as template. The final concentration of reagents in the RT reaction is summarised in Table 2.8.

Table 2.8: Final concentration of each component in RT reaction.

Compounds	Final concentration
Tris.HCl (pH 8.3)	25 mM
KCl	50 mM
DTT	2.0 mM
$MgCl_2$	5.0 mM
dGTP, dTTP, dCTP, dATP	1.0 mM each
RNasin™	1 U/μL
Reverse Transcriptase	0.2 U/μL

2.6.7: Purification of DNA from Agarose Gels and Solution

Two methods were used in the purification of DNA from Gels, as follows:

1. BRESA-CLEAN™

DNA purification from agarose gels was performed using the BRESA-CLEAN™ Purification Kit. In brief, DNA bands were excised from the gel and three volumes of BRESA-SALT™ (chaotropic salt solution, NaI) was added. For TBE gels, 0.5 volume of TBE-MELT™ was added (not necessary for TAE gels), followed by 4.5 volumes of BRESA-SALT™). The mixture was heated at 55°C until the gel was completely dissolved.

An aliquot of BRESA-BIND™ (silica matrix) was added to the tube, which was agitated gently for at least five minutes and centrifuged at 21,000 × g for 5 seconds. Following the removal of the supernatant, the pellet was washed with BRESA-WASH™ (a solution containing Tris.HCl buffer, EDTA, NaCl, and 50% ethanol). The DNA was eluted from the matrix by the addition of water or TE buffer and heating at 55°C for at least five minutes. Finally, two additional centrifugation steps (21,000 × g for 5 minutes) were carried out to ensure most of the BRESA-BIND™ silica residue was removed from the purified DNA.

2. Freeze Squeeze and UltraClean™ PCR Clean-UP™ Kit

DNA fragments separated on 1% agarose were excised and frozen at 20°C for 30 minutes. The frozen gel was then quickly squeezed between layers of parafilm and the liquid was collected.

The DNA dissolved in the liquid was purified using an UltraClean™ PCR Clean-UP™ Kit. In brief, 5 volumes of SpinBind was added, mixed and transferred to a spin filter unit and centrifuged for 30 seconds at 21,000 × g. 300 μL of SpinClean buffer

was added and centrifuged for 3 minutes at 21,000 × g. DNA was recovered by adding 30 – 50 μL of water to the filter and centrifuging for 1 minute at 21,000 × g.

DNA was purified from solution using UltraClean™ PCR Clean-UP™ Kit and BRESA-CLEAN™ as described above. For this use of BRESA-CLEAN™, TBE-MELT™ was not required.

2.6.8: Spectrophotometry for Nucleic Acid Quantification and Monitoring Bacterial Cell Growth

DNA and RNA concentration was measured using a Perkin Elmer Lambda 25 UV/VIS spectrophotometer. Absorbance readings were substituted into the equation:

$$[\text{DNA or RNA in } \mu g/mL] = OD_{260} \times \ 40(\text{RNA}) \text{ or } 50(\text{DNA})$$

Protein impurities in DNA/RNA preparations were determined by calculating the ratio of absorbance at 260/280 nm. A nucleic acid preparation was assumed sufficiently clean of impurities when the ratio was between 1.9 and 2.

Bacterial growth was monitored by measuring the turbidity of the culture with a Shimadzu Scientific spectrophotometer UV-1201 UV-VIS at a wavelength of 600nm. To eliminate the possibility that production of extracellular polysaccharides might interfere with the spectrophotometer readings, samples were centrifuged at 14,000 × g for 1 minute and the cells were resuspended in 1 ml saline. Alternatively, cultures were diluted 1:10 with saline solution.

2.6.9: Nucleic Acid Isolation

2.6.9.1: Plasmid DNA mini-preparations

The plasmids used in this study were purified from bacteria using one of two methods. A 5 mL overnight bacterial culture was harvested by transferring the culture to 1.5 mL microcentrifuge tubes and centrifuging at 21,000 × g for 1 minute. The supernatant was discarded and pellets were pooled within a single tube. Brief descriptions for each method follow:

1) Alkaline Lysis

The alkaline lysis method (Birnboim and Doly, 1979) is described in *Promega's Protocols and Application Guide* (Doyle, 1996). The harvested cells were resuspended in GTE (10 μL). Subsequently, 200 μL lysis solution (1% SDS/0.2 M NaOH) was added, mixed by inversion, and proteins from the lysed cells were precipitated with the SDS through the addition of (150 μL) potassium acetate (3 M). Precipitates were then removed by centrifugation for 5 minutes at 21,000 × g and the supernatant was collected and transferred to a fresh microcentrifuge tube. RNA was digested with RNase A (100 μg/μL) and the sample was extracted with phenol:chloroform:isoamyl alcohol (section 2.6.3). DNA was recovered by ethanol precipitation (section 2.6.2).

2) Wizard™ Plus Minipreps DNA Purification System

Harvested cells were resuspended in 250 μL Cell Resuspension Solution (50 mM Tris.HCl pH 7.5, 10 mM EDTA, 100 μg/mL RNase) and lysed with 250 μL volumes of Lysis solution. Alkaline protease solution (10 μL) was added and the sample was allowed to incubate at room temperature for 5 minutes. Neutralization Solution (350 μL of 4.09 M guanidine hydrochloride, 0.759 M potassium acetate,

2.21 M glacial acetic acid) was added, mixed by inversion 4 times and then centrifuged for 5 minutes at 21,000 × g at room temperature. The supernatant was collected and transferred to a minicolumn filter. The column containing the sample was then centrifuged and the filter was washed with "Column Wash Solution" (60 mM potassium acetate, 10 mM Tris.HCl pH 7.5, 60% ethanol). The remaining solution was removed by centrifugation at 21,000 × g for 2 minutes. Sterile water was added to the minicolumn and centrifuged at 21,000 × g for 1 minute to collect the purified plasmid. Isolated plasmids were stored at –20°C

2.6.9.2: RNA isolation

A hot acid/Phenol RNA extraction protocol (Beard *et al.,* 1995), was used to isolate RNA from *E. coli* and *Bacteroides*. Cells were prepared by growing a 5-mL starter culture containing the appropriate antibiotic. The culture was allowed to grow overnight in a shaker at 37°C (39°C for *Bacteroides*). The freshly grown culture was used to inoculate 5 or 40 mL of fresh broth, and the new culture was grown to mid-log phase with shaking at 37°C (39°C for *Bacteroies*). Cells were then chilled in an ice slurry and harvested by transferring them to a prechilled Oakridge centrifuge tube and centrifuged in a Beckman's Avanti™ J-25I centrifuge (rotor JA-25.50) at 10,000 × g for 3 minutes at 4°C. The supernatant was discarded.

Harvested cells were resuspended in 200 µL prechilled RNase-free water, followed by the addition of 400 µL of ice-chilled 50 mM Sodium Acetate (pH 4.7). Resuspended cells were then transferred to a 1.5 mL microcentrifuge tube. 400 µL of phenol was added and mixed by inversion 5 times. The samples were incubated at 55°C for 5 minutes and the aqueous/phenol phases were separated by centrifugation at 21,000 × g. The aqueous (top) phase was transferred to 1.5 mL microcentrifuge tubes and the bottom phase discarded. 400 µL of phenol was added again, incubated at 55 °C for 5

minutes and centrifuged at 21,000 × g. After the collection of the top phase, 400 μL of chloroform was added to remove the phenol and the water/chloroform phases were separated by centrifugation at 21,000 × g for 5 minutes. The aqueous phase was transferred to a fresh 1.5 mL centrifuge tube and RNA was precipitated by adding 2 volumes of ethanol. RNA samples were stored at -80°C.

2.6.9.3: Removal of DNA contaminants from RNA preparations

DNA was removed from RNA samples prior to Reverse Transcription (RT) reactions by lithium chloride precipitation, followed by DNase I digestion. The procedure involved adding an equal volume of LiCl solution (8 M) to RNA samples dissolved in RNase-free water. The solutions were mixed thoroughly and immediately chilled at 4°C or frozen in liquid nitrogen overnight. The RNA/LiCl mixture was thawed and centrifuged for 15 minutes at 21,000 × g. The supernatant was discarded and the pellet was redissolved in RNase-free water (20 μL). To an aliquot (10 μL) of the redissolved RNA preparation, DNase I reaction buffer was added (2 μL of 10 × concentrate: 400 mM Tris.HCl, pH 8.0; 100 mM $MgSO_4$; 10 mM $CaCl_2$) together with 7.5 U of RNase-free DNase I, and the reaction was incubated at 37°C for 30 to 45 minutes. The RNA sample was extracted twice with an equal volume of phenol and once with an equal volume of chloroform. Two volumes of ethanol were added to the aqueous phase, which was stored at -80°C as an ethanol precipitate.

When RNA samples were to be used, a 100 μL, aliquot of the ethanol-suspended RNA preparation was centrifuged at 21,000 × g for 15 minutes at 4°C, washed with 70% ethanol and vacuum dried. The pellet was redissolved in RNase-free water (15 μL) and RNase inhibitor (RNasin™, 1 unit) was added to prevent RNA degradation.

2.6.10: Restriction Enzyme reaction

Restriction endonuclease reactions were performed according the instructions recommended by the manufacturers.

2.6.11: Dephosphorylation of DNA 5' Terminus

Dephosphorylation with calf intestinal alkaline phosphatase (CIAP) was performed as described in Promega Protocols and Applications Guide (Doyle, 1996). Briefly, the digested DNA was incubated with CIAP (0.01 U/pmole of ends) for 30 minutes at 37°C. For DNA with 5' recessed termini, the process was repeated once under identical conditions. The reaction was terminated by addition of 2 μL of 0.25 M EDTA, followed by a 65°C incubation for 30 minutes. The only modification to the Promega method was that purification of DNA was carried out as described in section 2.6.7.

2.6.12: Ligation of DNA

Ligation of DNA was performed using T4 DNA ligase. Vector and insert DNA (1:3 molar ratio) were added to the appropriate volume of ligation buffer, ligase and water. Reactions for ligation of overhanging termini and blunt-ended molecules were incubated for 3 hours at room temperature and 14°C overnight, respectively.

2.6.13: Preparation Electro-Competent Cells for Bacterial Transformation

2.6.13.1: Preparing of Escherichia coli Electro-Competent Cells

Electro-competent *E. coli* cells were prepared by growing cells in 500 mL of SOB medium, with 0.2% glucose, at 37°C with shaking. When the optical density of the cell culture reached between 0.3 – 0.8 (A_{600nm}), cells were harvested by centrifugation at 3,800 × g for 10 minutes at 4°C in a Beckman Avanti ™ J-25I centrifuge. The cell pellet was resuspended in an equal volume (to that of the original culture) of ice cold sterile

water and was re-centrifuged. Half the volume of ice cold sterile water was applied as a second wash and the cells were re-centrifuged at the same speed and duration. The pelleted cells were resuspended in 20 mL of sterile ice-chilled 10% glycerol and were re-centrifuged at 5,927 × g for 10 minutes at 4°C. Following the removal of the supernatant, the pellet was re-suspended in 1.5 mL of 10% glycerol. Aliquots of 50 µL were transferred to microcentrifuge tubes, and snap frozen with liquid nitrogen. The electro-competent cells were stored at -80°C.

2.6.13.2: Preparing of Bacteroides Electro-Competent Cells

Bacteroides competent cells were prepared freshly prior to transformation. Cells were inoculated into 5 mL of rumen fluid broth and grown overnight at 39°C. The overnight culture was used to inoculate a 40 mL flask of rumen fluid medium and cells were allowed to grow for no more than 18 hours. Optical density was determined by eye because the presence of extracellular polysaccharide in the medium prevented accurate spectrophotometer readings.

The *Bacteroides* culture was chilled in an ice slurry, transferred to a 40 mL Sorvall centrifuge tube, and centrifuged for 10 minutes in a Sorvall RC-5 Superspeed Centrifuge (HB-6 rotor) at 6,000 rpm at 4°C. The supernatant was removed and the cells resuspended in ice cold 10% glycerol The washed culture was re-centrifuged and the pellet was resuspended in 200-300 µL of ice cold 10% glycerol. Aliquots of 40 µL were transferred to 1.5 mL microcentrifuge tubes and kept on ice until they were used for transformation.

2.6.14: Bacterial Transformation Electroporation

2.6.14.1: E. coli Transformation by Electroporation

A mixture of 1 – 2 μL of DNA and 50 μL of competent cells was transferred to an ice-chilled 2-mm pathlength electroporation cuvette. Cells were pulsed with 2.5 kV from a 25-μF capacitor with a by-pass resistance of 200 Ω. Room temperature LB broth (1 mL) was quickly added, mixed, and the cells were transferred into a sterile tube with a screw cap. The electroporated cells were allowed to recover for 1 hour at 37°C, after which 100 μL of electroporated cultures was spread evenly on antibiotic LB plates and incubated at 37°C.

2.6.14.2: B. uniformis Strain AR20 *and B. thetaiotaomicron Strain* AR29 *Transformation by Electroporation*

An aliquot of 1 – 2 μL of plasmid was added to a 40 – 45 μL aliquot of competent cells and transferred to an ice-chilled 1-mm pathlength cuvette. Cells were pulsed with 1.5 kV from a 25-μF capacitor with a by-pass resistance of 200 Ω. Electroporated cells were resuspended in 1 mL of rumen fluid medium chilled for 10 minutes in an ice slurry and quickly transferred to a sterile Hungate tube. Cells were allowed to recover for 2 – 3 hours at 39°C.

After the recovery period, 1.5 mL of molten rumen fluid agar (45°C) was added to the electroporated cells, mixed and poured as top agar onto rumen fluid plates containing 10 μg/mL clindamycin. Plated cells were incubated at 39°C for 2 - 3 days or until colonies were formed.

2.6.15: Screening for Transformed *E. coli* and *Bacteroides*

E. coli and *Bacteroides* transformed with pBA were screened by PCR with the appropriate primers to confirm the presence of pBA. PCR products were analyzed by

gel electrophoresis. Transformed *E. coli* were screened by blue/white selection, by growth on selective LB agar plates containing 0.1mM IPTG and X-Gal (40 μg/mL)

2.6.16: Slide Preparation for *Bacteroides* Cells – The Indian Ink Test

Slides for microscopic examination of *Bacteroides* cells were prepared by centrifuging 200 μL of *Bacteroides* culture for 3 minutes at 10,000 × g. After removing the supernatant, cells were resuspended with 3 μL of Indian ink and transferred to a microscope slide. A cover slip was placed on top of the sample and pressed down firmly using a dry tissue. Nail polish was used to seal the cover slips to the slide. Cells were observed through an Olympus BH2 phase contrast microscope under 1000× magnification. Photographs were taken with an Olympus SC35 camera using Kodak Ektachrome 160T tungsten film.

2.6.17: DNA Sequencing

DNA sequencing was carried out using Big Dye version 3.1 dye terminators from Applied Biosystems with the protocol as described by (Sanger and Coulson, 1975) .

Composition of the sequencing reactions is summarized in table 2.9. The amount of template used depended on the type and size of the DNA (Table 2.10).

Table 2.9: Composition of Big Dye version 3.1 10 μL sequencing reaction.

Composition	Volume (μL) For sequencing plasmid and PCR template
Primer (3.2 pmole/μL)	1
Big Dye terminator version 3.1 mix	4
DNA (see Table 2.10 for amount of DNA)	X
Water	Volume to total of 10 μL

Table 2.10: The amount of template required per 10-µL sequencing reaction

Template	DNA required for a 10 µL reaction (ng)
Plasmid 3-10 kb	150-300
Plasmid 10-20 kb	400
Cosmids 30-45 kb	500-1000
Bacterial genomic DNA	2000-3000
M13 single stranded DNA	25-50
PCR product 100-200 bp	1-3
PCR product 200-250 bp	3-10
PCR product 500-1000 bp	5-20
PCR product 1000-2000 bp	10-40
PCR product >2000 bp	20-50

Thermal cycling conditions for Big Dye terminators version 3.1 are shown in Table 2.11.

Table 2.11: Thermal cycle characteristics used during PCR

Cycle	1 cycle	25 cycles			1 cycle
Temperature (°C)	96	96	55-60	60	14
Duration (min or sec)	2.00 mins	10 sec	20 sec	4 mins	until removal

After thermal cycling, the sequencing reaction was ethanol precipitated by adding 1 µL of 3M sodium acetate pH 5.2 and 1 µL of 125 mM EDTA. Subsequently 25 µL of ethanol was added and mixed thoroughly prior to chilling on ice for 20 mins. The chilled reaction was then centrifuged for 30 mins at 21,000 × g. Following this, the supernatant was discarded, the pellet rinsed with 125 µL of 80% ethanol, and

recentrifuged at 21,000 × g for 5 mins. Finally the supernatant was removed and the pellet dried before analysis using an ABI 373 or 377 Sequencer, within Murdoch University's State Agricultural Biotechnology Centre (SABC) sequencing facility.

Chapter 3: Expression of Integrase and Excisionase Genes from Plasmid pBA

3.0: Introduction

For a bacteriophage to enter a lysogenic cycle, a specific protein called integrase is required. In the presence of integrase and host-cell encoded integration factors, the phage is able to insert its DNA into the bacterial chromosome in a process that is precise and site specific (Groth and Calos, 2004).

The site specific integration mechanisms of lysogenic bacteriophages have been useful research tools for introducing novel genes into bacteria (Rossignol *et al.*, 2002; Shimizu-Kadota, 2001; Lauer *et al.*, 2002; Groth and Calos, 2004; Groth *et al.*, 2000). Such a tool allows the permanent transformation of bacteria, removing the need for the selective mechanisms that are usually required to maintain plasmids in bacteria (Rossignol *et al.*, 2002; Shimizu-Kadota, 2001)

Construction of an integration vector generally involves the cloning of the attachment site of a phage (*attP*) and the integrase gene (*Int*) into a vector suited for transformation of the intended host bacterium. In general, the process is designed for a host organism that is free of the bacteriophage. Attempts have been made at constructing an integrative vector for ruminal *Bacteroides* spp. (*B. uniformis* AR20 and *B. thetaiotaomicron* AR29), i.e. plasmid pBA. The isolation of the bacteriophage ϕAR29 "integration module" and construction of the vector pBA has been described in Chapter 1 and will not be repeated here. The main reason for constructing pBA was to create an integrative vector that would overcome some of the stability problems associated with genetically modifying ruminal bacteria with plasmids (Wong *et al*., 2003).

Although pBA was successfully introduced into the *Bacteroides* species, earlier experiments did not detect the integration of plasmid into the bacterial genome (Wong *et al.,* 2003). The cloned phage fragment in pBA carries two open reading frames which

were proposed to be recombinase genes, *Int* and *Xis*, although alignment of the conceptual translation of these genes to other known *Int* and *Xis* protein sequences did not show strong homology (Gregg et *al.,* 1994).The absence of evidence for plasmid integration suggested that the recombination genes may not be expressed in the two *Bacteroides* species.

3.0.1: Aim

In this project, Reverse Transcription Polymerase Chain Reaction (RT-PCR) was used to determine whether the *Int* and *Xis* genes in plasmid pBA are transcribed in *E. coli* SCS110, *B. uniformis* AR20 and *B. thetaiotaomicron* AR29. In addition, to take advantage of the far greater databank resources now available, the proposed *Int* and *Xis* genes in pBA were compared with other recombination related genes, by sequence alignment of nucleotide and deduced protein sequences.

3.1.0: Materials and Methods:

3.1.1 Bacterial strains and Plasmids

The phenotypes of *E. coli* strain SCS110, *B. uniformis* strain AR20, and *B. thetaiotaomicron* strain AR29 are summarized in Chapter 2.3. All bacteria were transformed with shuttle vector pBA by electroporation. Ampicillin and clindamycin were used to select and maintain *E. coli* and *Bacteroides* transformants respectively.

3.1.2: RNA isolation

RNA was isolated according to the protocol described in section 2.6.9.2 and 2..6.9.3. Extraction of RNA was carried out during the middle of the logarithmic phase of bacterial growth, determined using a growth curve that was constructed by measuring culture turbidity with a spectrophotometer at a wavelength of 600 nm. This involved:

- Growing a starter culture overnight and using 10 µl of the overnight culture to inoculate 5 ml of fresh medium.
- Spectrophotometer readings were carried out on 200 µl aliquots of bacterial culture, which were collected every 30 minutes for the first 2 hours and then every 1 hour until the stationary phase was reached.

The times selected for RNA extraction from SCS110, AR20 and AR29, were 4, 6 and 7 hours from inoculation, respectively.

3.1.3: Two-Step Reverse Transcription Polymerase Chain Reaction (RT-PCR)

The RT-PCR protocol described in Chapter 2, section 2.6.6 was applied to detect mRNA transcript derived from the two recombinase genes. However, the protocol proposed by Koo and Jaykus (2000), which uses "mismatched" primers, was later used in this study. The procedure for the latter protocol is as follows:

- A 100-µl aliquot of ethanol-suspended total RNA preparation was centrifuged at 20,000 × g for 15 minutes at 4°C, washed with 70% ethanol and vacuum dried.
- The pellet was redissolved in RNase-free water (15 µl) and RNase inhibitor (RNasin™, 1 unit) was added to prevent RNA degradation.
- The reaction master mix consisted of 1 µl of 20 µM reverse primers that had been heated at 70 °C for 5 minutes and chilled on ice, 7 µl of 4 mM dNTPs mix, 6 µl reaction buffer and 20 Units of RNasin™.
- RNA sample and master mix were both prewarmed to 42 °C before they were mixed together, and 5 units of AMV reverse transcriptase was added.
- The RT reaction was incubated for 1 hour at 42 °C.
- An aliquot of the reverse transcriptase reaction (1µl) was used as template for a 20 µl PCR which contained 2 µl of 10 × buffer, 2.5 mM $MgCl_2$, 10 µM dNTP

mix, 0.2 μM of each forward and reverse primer and 1 unit of Taq DNA polymerase.

- The PCR mixture was heated at 95 °C for 5 minutes to inactive Reverse Transcriptase, followed by 35 cycles of amplification, with a final extension at 72 °C for 7 minutes.

Annealing temperatures for PCR differed according to the requirements of the different primer sets used. The annealing temperature for each set of primers was determined by amplifying from a plasmid pBA template, with the "mismatched" primers, using a thermal gradient of annealing temperatures. Following this, products of the amplification were extracted and reamplified with the same "mismatched" primers. Products of the PCR were then compared to those from the initial gradient amplification to determine which annealing temperature prevents amplification from plasmid DNA, which might be present as a contaminant in the RT-PCR mixture.

The primers used for the detection of *Int* and *Xis* transcription by RT-PCR are listed in Table 3.1. The positive control targets for RT-PCR studies were the clindamycin gene and RepA gene (*Bacteroides*); and the ampicillin gene (*E.* coli).

3.1.4: Gene Sequence Analysis

The nucleotide sequences of the proposed *Int* and *Xis* genes within pBA were retrieved from GenBank, (accession number S75733) and were used for BLAST analysis of sequences within GenBank database. The notionally translated protein sequence of ϕAR29 *Int* and *Xis* were also aligned against other known phage *Int* and *Xis* proteins, Table 3.2 and Table 3.3.

Table 3.1: Primers used for RT-PCR studies. The red bases indicate mismatched regions.

Primers name	Sequence 5'→ 3'	Length (b)	Tm (° C)	Product (b)
SRep for	TTAGCCGACACTGAAACTGGAGAAG	25	72.3	294
SRep rev	GGCTGGTCCTCGGGCTATGAT TCG	21		
SClin for	TGGGGCAGGCAAGGGGTT	18	69	571
SClin rev	TTCCGAAATTGACCTGACCTGACTT TCT	25		
SAmp for	TTGGGTGCACGAGTGGGTTACA	22	69	635
SAmp rev	GGCCCCAGTGCTGCAATGATA GTC	21		
SInt2 for	GGGACGAGATGGAGGGAAAGTGATA	25	69	375
SInt2 rev	CCGTTCAGCCTGCGACCATCT CAT	21		
SXis for	CTGCGGAAGCCATTGATACTATAAA	25	62	182
SXis rev	ATCGCCCATCGCATTTTGAA CCC	20		

Table 3.2: Recombinase genes from Phage included in protein alignment.

Recombinase	Size	Organism	Accession#	Ref
PL-2	383		BAB83530	N/A
SF6	385	*Bacillus subtilis*	P37317	(Clark *et al.*, 1991)
21	356	*Escherichia coli*	AAC48884	(Wang *et al.*, 1997)
186	336	*Escherichia coli*	P06723	(Kaneko *et al.*, 1998)
434	356	*Escherichia coli*	P27078	(Baker *et al.*, 1991)
933W	444	*Escherichia coli*	NP_049461	(Plunkett *et al.*, 1999)
DLP12	387	*Escherichia coli*	P24218	(Lindsey *et al.*, 1989)
e14	375	*Escherichia coli*	P75969	(Blattner *et al.*, 1997)
EH297	428	*Escherichia coli*	CAD24081	N/A
HK022	357	*Escherichia coli*	P16407	(Yagil *et al.*, 1989)
HK97	356	*Escherichia coli*	NP_037720	(Juhala *et al.*, 2000)}
Lambda INT	356	*Escherichia coli*	P03700	(Davies, 1980)
P2	337	*Escherichia coli*	P36932	(Yu *et al.*, 1989)
P22	387	*Escherichia coli*	P04890	(Pedulla *et al.*, 2003)
P27	437	*Escherichia coli*	CAC83519	(Recktenwald and Schmidt, 2002)
Phi-80	402	*Escherichia coli*	P06155	(Leong *et al.*, 1986)
P4	439	*Escherichia coli*	P08320	(Haldimann and Wanner, 2001)
VT2-Sa	444	*Escherichia coli*	NP_050500	(Miyamoto *et al.*, 1999)
HP1 INT	337	*Haemophilus influenzae*	P21442	(Goodman and Scocca, 1989)
HP2	337	*Haemophilus influenzae*	NP_536807	(Williams *et al.*, 2002)
A2	385	*Lactobacillus casei*	NP_680502	(Proux *et al.*, 2002)
Phig1e	391	*Lactobacillus*	NP_695147	(Kodaira *et al.*, 1997)
bIL285	374	*Lactococcus lactis subsp. lactis*	NP_076573	(Chopin *et al.*, 2001)
bIL286	359	*Lactococcus lactis*	NP_076635	(Chopin *et al.*, 2001)

bIL309]	377	*Lactococcus lactis*	NP_076696	(Chopin *et al.,* 2001)
bIL311	399	*Lactococcus lactis*	NP_076780	(Chopin *et al.,* 2001)
bIL312	382	*Lactococcus lactis*	NP_076801	(Chopin *et al.,* 2001)
phi LC3	374	*Lactococcus lactis subsp. cremoris*	AAA32254	(Lillehaug and Birkeland, 1993)
TP901-1	485	*Lactococcus lactis*	AAK38018	(Brondsted and Hammer, 1999)
Tuc2009	374	*Lactococcus lactis*	NP_108680	(van de Guchte *et al.,* 1994)
phage 2389	384	*Listeria monocytogenes*	NP_511007	(Zimmer *et al.,* 2003)
D3	369	*Pseudomonas aeruginosa*	NP_061531.	(Kropinski, 2000)
phi CTX	389	*Pseudomonas aeruginosa*	NP_490644	(Nakayama *et al.,* 1999)
PSP3	349	*Pseudomonas putida*	NP_958084	N/A
pMLP1	392	*Micromonospora carbonacea*	AAO46045	(Alexander *et al.,* 2003)
mycobacteriophage D29	333	*Mycobacteria*	Q38361	(Ribeiro *et al.,* 1997)
Mx8	533	*Myxococcus xanthus*	NP_203426. 15320656	(Tojo *et al.,* 1996)
16-3	371	*Rhizobium*	CAB54831	(Semsey *et al.,* 1999)
SfX	386	*Shigella flexneri*	AAD10295	N/A
V	387	*Shigella flexneri*	NP_599058	(Allison *et al.,* 2002)
L54a	354	*Staphylococcal aureus*	P20709	(Ye and Lee, 1989)
Phi 11	348	*Staphylococcus aureus*	NP_803254	(Iandolo *et al.,* 2002)
phi 42	345	*Staphylococcus aureus*	AAA91615	(Carroll *et al.,* 1995)
EJ-1	380	*Streptococcus pneumoniae*	CAE82083	(Romero *et al.*, 2004b)
MM1	375	*Streptococcus pneumoniae*	NP_150133	(Obregon *et al.,* 2003)
T12	362	*Streptococcus pyogenes*	AAC48867	(McShan *et al.,* 1997)
TPW22	355	*Streptococcus thermophilus*	AAF12706	(Petersen *et al.,* 1999)
phi-BT1	594	*Streptomyces lividans*	CAD80152	(Gregory *et al.,* 2003)
phi-C31	605	*Streptomyces*	NP_047974	(Smith *et al.,* 1999)
K139	345	*Vibrio cholerae O139*	NP_536628	(Kapfhammer *et al.,* 2002)

Table 3.3: Excisionase sequences included in protein alignment.

Execisionase	Size	Organism	Accession#	Ref
21	78	*Escherichia coli*	AAA32335	(Baker *et al.,* 1991)
186	87	*Escherichia coli*	NP_052280	N/A
434	72	*Escherichia coli*	P11683	(Yagil *et al.,* 1989)
933W	99	*Escherichia coli*	NP_049462	(Plunkett *et al.,* 1999)
EH297	84	*Escherichia coli*	CAD24080	N/A
HK97	72	*Escherichia coli*	NP_037721	(Juhala *et al.,* 2000)
Hk022	72	*Escherichia coli*	NP_037687	(Juhala *et al.,* 2000)
Lambda *Int*	72	*Escherichia coli*	NP_040610	N/A
P27	78	*Escherichia coli*	NP_543054	(Recktenwald and Schmidt, 2002)
Phi-80	65	*Escherichia coli*	RSBPX8	(Leong *et al.,* 1986)

VT2-Sa	99	*Escherichia coli*	NP_050501	(Miyamoto *et al.,* 1999)
A2	76	*Lactobacillus casei*	NP_680503	(Alvarez *et al.,* 1998)
bIL312	86	*Lactococcus lactis*	NP_266665	(Bolotin *et al.,* 2001)
TP901-1	64	*Lactococcus lactis*	NP_112670	(Brondsted *et al.,* 2001)
pMLP1	141	*Micromonospora carbonacea*	AAO46044	(Alexander *et al.,* 2003)
P22	116	*Salmonella typhimurium*	NP_059585	(Pedulla *et al.,* 2003)
ST64T	116	*Salmonella typhimurium*	NP_720279	(Moitoso de Vargas and Landy, 1991)
SfX	115	*Shigella flexneri*	AAD10294	N/A
V	147	*Shigella flexneri*	NP_599059	(Allison *et al.,* 2002)
X	115	*Shigella flexneri*	CAC43409	(Dobrindt *et al.,* 2002)
L54a	354	*Staphylococcus aureus*	XSBPL5	(Ye and Lee, 1989)
Phi 11	66	*Staphylococcus aureus*	AAA32197	(Ye *et al.,* 1990)
315.3	103	*Streptococcus pyogenes MGAS315*	NP_795493	(Beres *et al.,* 2002)
T12	72	*Streptococcus pyogenes*	AAC48866	(McShan *et al.,* 1997)
EJ-1	103	*Streptococcus pneumoniae*	NP_945255	(Romero *et al.*, 2004b)
16-3	140	*Rhizobium*	CAB54832	(Semsey *et al.,* 1999)

3.2: Results:

3.2.1: Transformation of *Bacteroides thetaiotaomicron* strain AR29 with pBA.

Transformation of *B. thetaiotaomicron* strain AR29 with plasmid pBA proved technically difficult. Due to high endonuclease activity in AR29, pBA extracted from AR20 was used to improve the transformation efficiency. However, despite numerous attempts, transformation of AR29 was not achieved.

Microscopic examination of cells suspended in an India Ink solution revealed a clear halo around the majority of cells. It was concluded that AR29 cells were encapsulated. Phase contrast microscopy on AR29 cells showed encapsulated cells (Figure 3.2). It was concluded that the greatly thickened cell wall could lower the transformation efficiency of AR29. Similar microscopy study on AR20 showed that current cultures of AR20 were more heavily encapsulated than AR29 (Figure 3.3).

To allow transformation of AR20 and AR29, experiments were performed to remove or reduce the capsule surrounding both species. Attempts to select uncapsulated

bacterial colonies from plates were unsuccessful, because no uncapsulated colonies were found. It was thought that the diversity of nutrients present in rumen fluid medium might induce the production of capsule. An alternative rumen bacterial medium was obtained from DSMZ (Medium 330). Each of the principal carbon sources present in Medium 330 (Glucose, Maltose and Cellobiose) was examined individually for effects on the growth and physiology of AR29 and AR20. The tests showed that maltose appeared to induce thicker encapsulation of cultures and, as it was not essential for bacterial growth, it was removed from the medium. In contrast, the removal of glucose or cellobiose inhibited AR29 growth. Other changes made were

- the reduction of glycerol by 4-fold and
- the substitution of trypticase peptone with tryptone peptone

. The new medium, Defined Rumen Bacterial (DF) Medium, was then used for growing AR20 and AR29 cultures until uncapsulated populations were obtained.

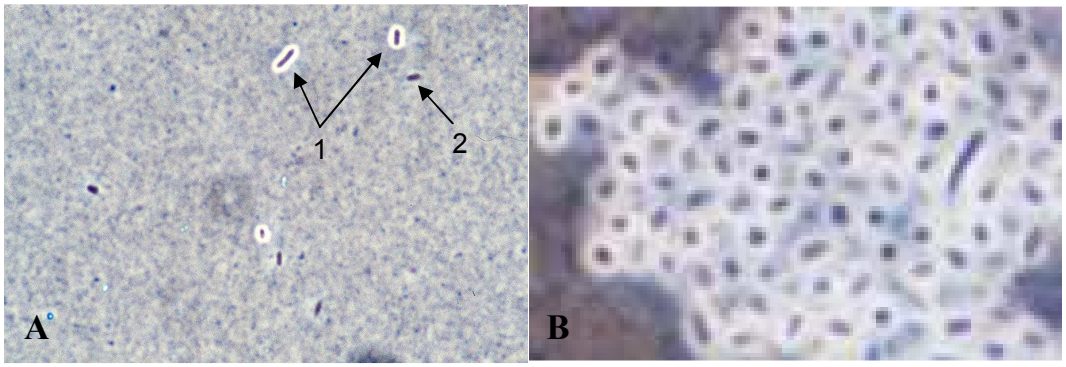


Figure 3.2: Phase contrast microscopy of *B. thetaiotaomicron* strain AR29 suspended in an India Ink solution (A) with capsule (1) and unencapsulated (2); and *B.uniformis* AR20 showing mass encapsulation (B) at 1000 X magnification.

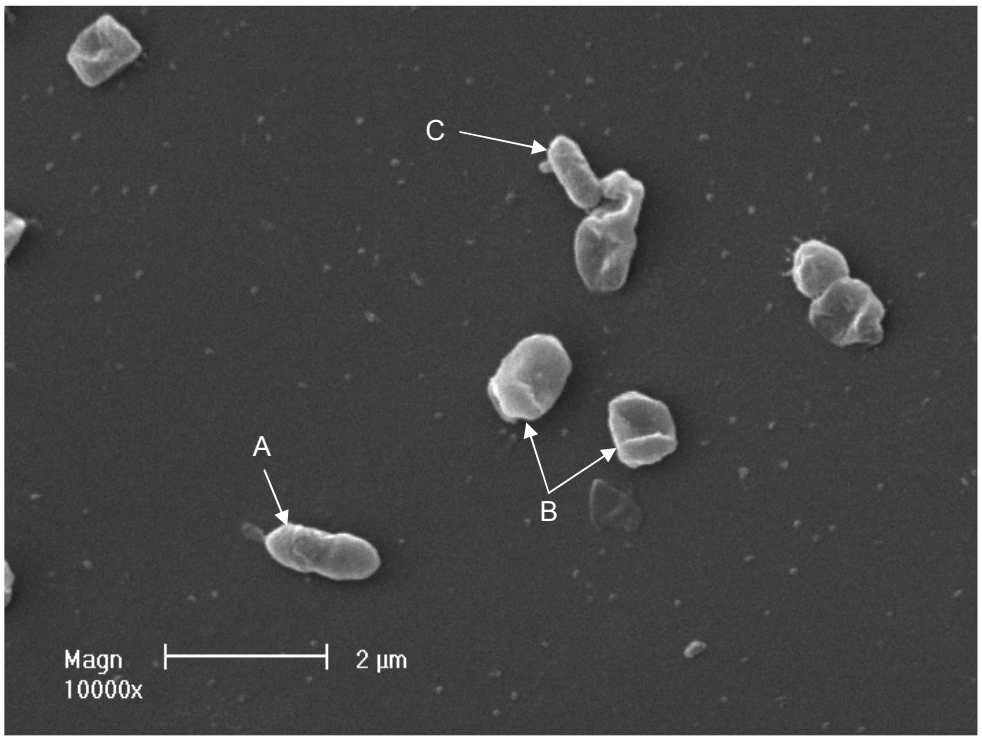


Figure 3.3: Scanning electron microscopy of *B. thetaiotaomicron* strain AR29 with partial encapsulation (A), heavy encapsulation (B) and unencapsulated (C) photographed at 10000 X magnification.

3.2.2: Detection of Int and Xis RNA transcripts in *E. coli* strain SCS110.

RT-PCR studies indicated that the *Int* and *Xis* genes on pBA are transcribed in *E. coli* strain SCS110 (Figure 3.4). RT-PCR products from the recombination genes migrated according to their predicted sizes of 375 and 182 bp. Transcripts of the *Bacteroides* genes, RepA and Clin^r, were present, in addition to transcripts from the *E. coli* Amp^r gene.

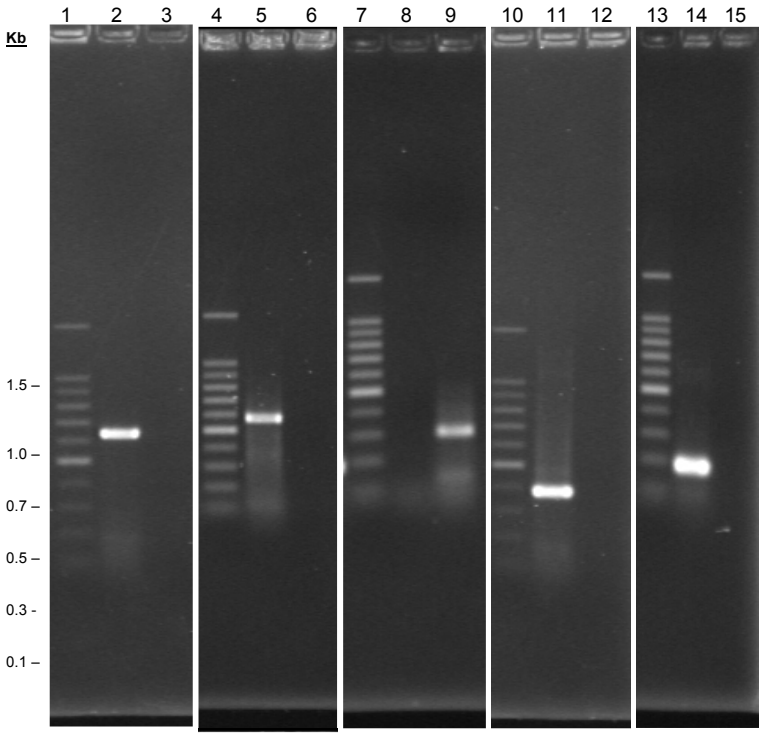


Figure 3.4: RT-PCR products from *E. coli* strain SCS110 containing plasmid pBA, to detect the presence of Amp^r, Clin^r, RepA, *Int* and *Xis* transcripts. Lanes 1, 4, 7, 10, and 13 : 100 bp ladder; Lane 2: RT-PCR using Amp^r primers; Lane 3: RT-PCR negative control using Amp^r primers; Lane 5: RT-PCR using Clin^r primers; Lane 6: RT-PCR negative control using Clin^r primers; Lane 8: RT-PCR negative control using Rep A primers; Lane 9: RT-PCR using RepA primers; Lane 11: RT-PCR using *Int* primers; Lane 12: RT-PCR negative control using *Int* primers; Lane 14: RT-PCR using *Xis* primers; Lane 15: RT-PCR negative control using *Xis* primers.

3.2.3:Detection of Int and Xis RNA transcripts in *B. uniformis* strain AR20.

Recombination genes, *Int* and *Xis*, were shown to be transcribed in *B. uniformis* strain AR20 (Figure 3.5), similarly to results from SCS110. AR20 was also shown to transcribe the RepA and Clin^r genes. Transcripts from the *E. coli* Amp^r gene were also detected from RT-PCR of RNA from AR20.

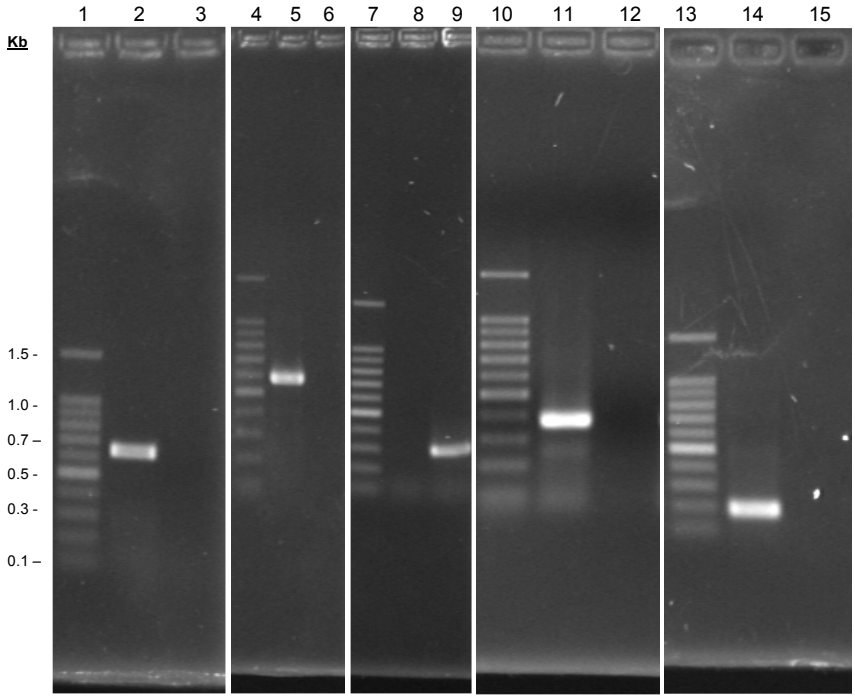


Figure 3.5: RT-PCR products from *B. uniformis* strain AR20 containing plasmid pBA, to detect the presence of Amp^r, Clin^r, RepA, Int and *Xis* transcripts. Lanes 1, 4, 7, 10, and 13: 100 bp ladder; Lane 2: RT-PCR using Amp^r primers; Lane 3: RT-PCR negative control using Amp^r primers; Lane 5: RT-PCR using Clin^r primers; Lane 6: RT-PCR negative control using Clin^r primers; Lane 8: RT-PCR negative control using Rep A primers; Lane 9: RT-PCR using RepA primers; Lane 11: RT-PCR using *Int* primers; Lane 12: RT-PCR negative control using *Int* primers; Lane 14: RT-PCR using *Xis* primers; Lane 15: RT-PCR negative control using *Xis* primers.

3.2.4: Detection of Int and Xis RNA transcripts in *B. thetaiotaomicron* strain AR29

The results of the RT-PCR studies on untransformed and transformed *B. thetaiotaomicron* AR29 are shown in Figure 3.6 and 3.7, respectively. The proposed *Int* and *Xis* genes were transcribed in *B. thetaiotaomicron* AR29, both in the presence and absence of the plasmid. The control genes (RepA, Clin^r and Amp^r) produced results that were predicted for pBA-transformed bacteria. Negative control experiments performed on untransformed AR29 generated indistinct or faint multiple products, suggesting that mispriming occurred when PCR was performed in the absence of specific primer target sites.

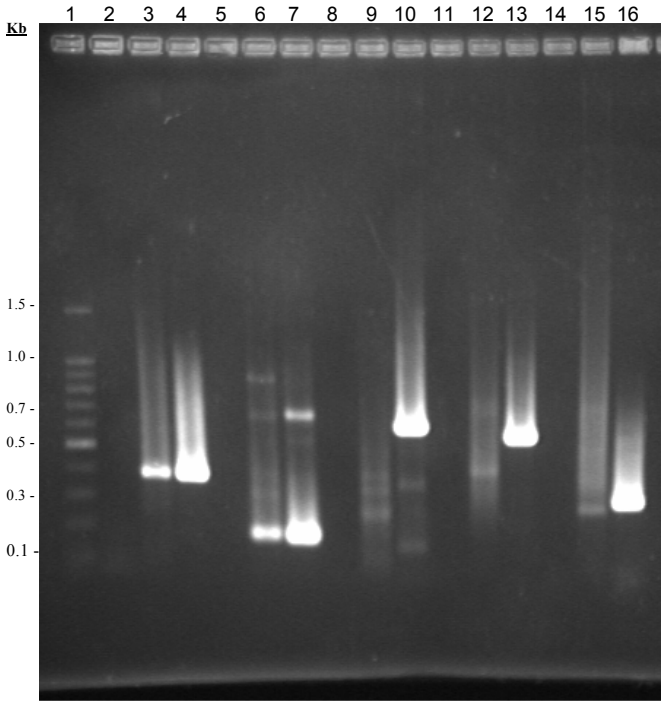


Figure 3.6: RT-PCR products from *B. thetaiotaomicron* strain AR29 without plasmid pBA, to detect the presence of *Int*, *Xis*, Amp^r, $Clin^r$ and RepA transcripts.
Lane 1: 100 bp ladder; Lane 2: RT-PCR negative control using *Int* primers; Lane 3: RT-PCR using *Int* primers; Lane 4: PCR positive control using *Int* primers; Lane 5: RT-PCR negatives control using *Xis* primers; Lane 6: RT-PCR using *Xis* primers; Lane 7: PCR positive control using *Xis* primers; Lane 8: RT-PCR negative control using Amp^r primers; Lane 9: RT-PCR using Amp^r primers; Lane 10: PCR positive control using Amp^r primers; Lane 11: RT-PCR negative control using $Clin^r$ primers; Lane 12: RT-PCR using $Clin^r$ primers; Lane 13: PCR positive control using $Clin^r$ primers; Lane 14: RT-PCR negative control using RepA primers; Lane 15: RT-PCR using RepA primers; Lane 16: PCR positive control using RepA primers.

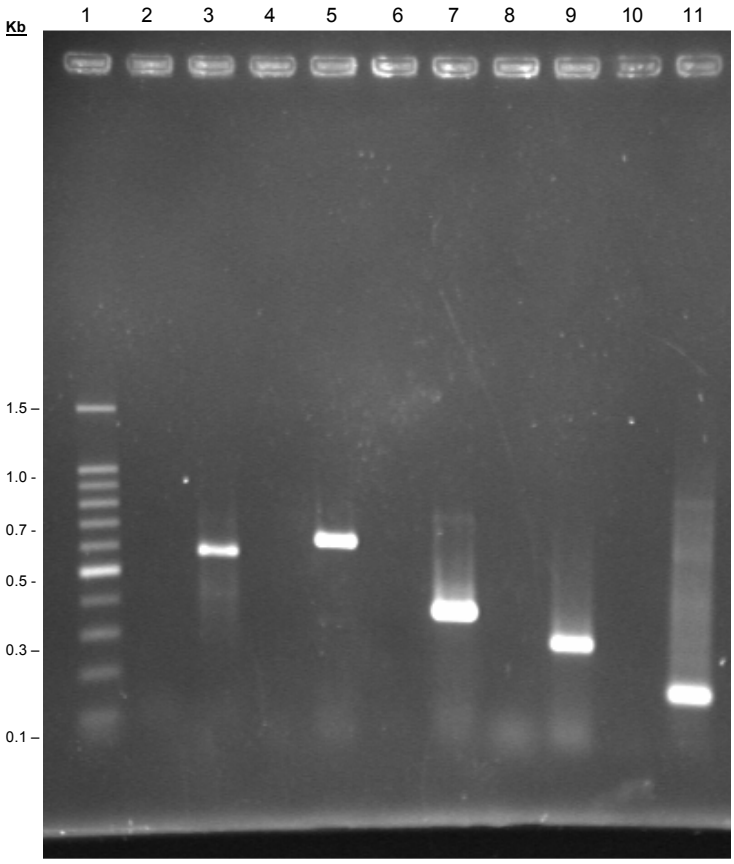


Figure 3.7: RT-PCR of *B. thetaiotaomicron*.strain AR29 transformed with plasmid pBA using primers to detect the presence of $Clin^r$, Amp^r, *Int*, RepA and *Xis* transcripts.
Lane 1: 100 bp ladder; Lane 2: RT-PCR negative control using $Clin^r$ primers; Lane 3: RT-PCR using $Clin^r$ primers; Lane 4: RT-PCR negative control using Amp^r primers; Lane 5: RT-PCR using Amp^r primers; Lane 6: RT-PCR negative control using *Int*; Lane 7: RT-PCR using *Int* primers; Lane 8: RT-PCR negative control using RepA primers; Lane 9: RT-PCR using RepA primers; 10: RT-PCR negative control using *Xis* primers; Lane 11: RT-PCR using *Xis* primers;

3.2.5: Sequence analysis of proposed Int and Xis

Nucleotide and *In silico* translated protein sequences of the proposed *Int* and *Xis* genes (Gregg *et al*., 1994) were used in a BLAST search of GenBank to determine whether these genes possessed structural similarities to other recombination-related genes and proteins. Results from nucleotide and protein amino acids sequence BLAST searches did not show any clear homology to the family of proteins to which these genes were thought to belong.

Protein sequence translations of the *Int* and *Xis* genes were aligned to 50 known integrases and 26 known excisionases, respectively (Appendix 1: Figure 3.1A and

3.2A). Alignment of *Int* indicated that position 195, 292, 336, 340, 365, 424, 448 and 547 share similar amino acids to other known phage *Int* proteins, whilst a consenus arginine is located at position 333. No consensus sequences were indicated in alignment of *Xis* protein sequences, with similarities found only at locations 100, 124, 134, and 157. Overall the alignment of these proteins to other known *Int* and *Xis* proteins showed poor homology to their hypothetical protein families.

3.3: Discussion

3.3.1 Encapsulation of AR20 and AR29

Some bacteria are known to form extracellular capsules, including *Pneumococci*, *Streptococci*, *Staphylococci*, *Meningococci*, *Salmonella* spp and *Bacteroides* spp (Kasper, 1986). Past studies on *Bacteroides fragilis* have shown that the capsule is composed of polysaccharides or lipopolysaccharides (Kasper *et al.,* 1983). Both *B. thetaiotaomicron* VPI-5482 and *B. fragilis* are capable of synthesizing at least 7 and 8 distinct capsular polysaccharides, respectively (Krinos *et al.*, 2001; Comstock and Coyne, 2003). Capsular polysaccharides have been shown to be an important virulence factor for *B. fragilis* (Pantosti *et al.,* 1991), with three (polysaccharide A (PS A), PS B and PS C) being shown to induced intra-abdominal abscess in rodents (Comstock *et al.,* 2000).

B. fragilis varies the expression of different capsular polysaccharides by reversible inversion of the promoters (Krinos *et al.*, 2001; Comstock and Coyne, 2003). The variability of *Bacteroides* surface architecture may affect the organism's interaction with the host or with other microbiota and allow competitive advantage for its survival. Encapsulation of *B. uniformis* AR20 and *B. thetaiotaomicron* AR29 provided a challenge during transformation, preventing the transfer of DNA across the cell membrane. Previous studies have demonstrated that encapsulation of bacterial cells provides a barrier to bacteriophage infection (Bernheimer and Tiraby, 1976; Burt *et al.*,

1978; Ramirez *et al.*, 1999; Gindreau *et al.*, 2000). The resistance of phage ϕBrb01 infection in capsulated *B. uniformis* (previously *ruminicola ss brevis*) strain AR20 has been observed by Klieve *et al.,* (1991). The inability of bacteriophage to infect capsulated bacteria may be due to inaccessibility of surface receptor site on the cells (Klieve *et al.*, 1991).

The development of capsules on ruminal bacteria has been attributed to the presence of excess nutrients such as carbohydrates (Russell, 1998). In the rumen, some bacteria convert excess carbohydrates to polysaccharide and store them as an intracellular reserve, an extracellular capsule or secrete them as "slime". The development of DF media was useful in the reduction of capsular material in AR20 and AR29, which enabled transformation and RT-PCR studies.

3.3.2 In silico DNA sequence analysis of ϕAR29 Int and Xis with other known recombinase gene in the GenBank

Analysis of the previously identified *Int* and *Xis* proteins using current data in GenBank was consistent with that previous report (Gregg *et al.*, 1994b). Recombination genes from different sources often show direct sequence homology and the weak protein sequence alignments suggest that these two open reading frames may not encode *Int* and *Xis* proteins as first proposed.

3.3.3 Detecting the transcripts of Int and Xis gene in SCS110, AR20 and AR29

Initial RNA preparations for RT-PCR studies produced false positive signals, which were discovered to be due to contaminating traces of DNA. A protocol proposed by Koo and Jaykus (2000) used primers with consecutively mismatched bases at positions 5 – 7 bases from the 3' end, to avoid false positive signals. Theoretically, the low temperature of the reverse transcription reaction allows the mismatched primer to

hybridize to target single stranded mRNA, but does not permit DNA denaturation and thus prevents DNA/primer association from occurring. Subsequently, in PCR with a high annealing temperature "mismatched" primers will have a high affinity for "mismatched" cDNA that was primed by them, but not for the original DNA template. As a result, products from RT-PCR are derived only from mRNA.

The recombination genes of temperate phages are usually located directly adjacent to the *attP* site (Cheong and Brooker, 1998). The ϕAR29 DNA fragment that was used to construct pBA had been concluded to carry the *attP* site and two open reading frames adjacent to the *attP*, which were proposed to be integrase and excisionase genes (Gregg *et al.*, 1994b). Expression studies on *Int* and *Xis* genes in pBA in the present study have shown that both genes are transcribed in *E. coli,* AR29 and AR20. Despite this, previous restriction and hybridisation studies did not show any evidence of plasmid pBA integrating into the genome of AR20 or AR29 (Wong *et al.,* 2003). Possible reasons for not observing integration of pBA include:

- matching attachment sites may not be present in pBA and AR20.
- essential factors such as host integration factor (HIF) and/or factor for inversion stimulation (FIS) which play a role in the bending of the double helix during the recombination process, may be missing. (Moreau *et al.,* 1999).
- the previously identified genes may not encode the true integrase and/or excisionase proteins.

However, the absence of proof for integration does not prove that integration did not occur. More information is required to allow confirmation of the integration of pBA into genomic DNA. This includes identifying the precise location of the attachment site in both the phage and the host genomes, and sequencing of the ϕAR29 genome to identify other open reading frames that could encode integration related genes.

Bacteriophage ϕAR29 has only been found as a lysogen in AR29. Therefore, the presence of *Int* and *Xis* transcripts in AR29 was not surprising, even in the absence of pBA transformation. Attempts at "curing" ϕAR29 from AR29 have been made previously, but without success (Klieve *et al.,* 1989).

Other interesting observations from the transcription studies were the detection by RT-PCR of clindamycin and RepA transcripts in *E. coli*, and ampicillin transcript in *Bacteroides.* The presence of these transcripts could indicate that promoters on pBA are recognized by both *E. coli* and *Bacteroides.* Alternatively, plasmid transcription might continue around the plasmid, from promoters that are specifically recognised in each of the host species.

Chapter 4: Characterisation of the Bacteriophage ϕAR29 Genome

4.0: Introduction

Bacteriophages are abundant in the environment, with around 10^{10} phage per litre of surface seawater and 10^{7} to 10^{9} per gram of sediment or topsoil (Rohwer and Edwards, 2002; Danovaro *et al.*, 2001). In the genomes of bacteria, prophages can constitute as much as 10-20% of the DNA (Casjens, 2003). Bacteriophages are also a major contributor to genetic exchange (Wagner and Waldor 2002; Canchaya *et al.,* 2003; Casjens 2003; Weinbauer and Rassoulzadegan 2004). Transduction rate by bacteriophage is estimated to occur at a rate of 10^{25}-10^{29} bp of DNA per year (Paul, 1999; Jiang and Paul, 1998; Wagner and Waldor, 2002; Casjens, 2003; Canchaya *et al.*, 2003; Weinbauer and Rassoulzadegan, 2004).

The presence of functional or defective prophages in a bacterial genome can also contribute important biological properties to the bacterial host (Wagner and Waldor, 2002; Casjens, 2003; Canchaya *et al.*, 2003; Weinbauer and Rassoulzadegan, 2004; Boyd *et al.*, 2001). During the lysogenic life cycle, genes involved in lytic growth are switched off. However, the maintenance of prophage in its quiescent state requires the expression of lysogenic conversion-genes. The expression of lysogenic genes may also alter the genetic properties of the host, which may include protection of the host against further phage infection, or increasing the virulence of a pathogenic host (Boyd *et al.*, 2001; Wagner and Waldor, 2002; Casjens, 2003). In many cases, genes from mutated, decaying prophage may remain functional. Examples of this include the recombination system of prophage Rac, cell lysis genes from prophage QIN, and virion-like particles from *Bacillus subtilis* prophage, PBSX (Okamoto *et al.*, 1968; Kaiser and Murray, 1979; Espion *et al.*, 1983). Furthermore, defective prophages are still able to participate in recombination events involving closely related phages (Casjens, 2003).

Bacteriophage research is currently undergoing a renaissance (Campbell, 2003; Nelson, 2004). The present focus of phage research is on studies of phage evolution and their role in natural ecosystems (Rohwer and Edwards, 2002). In addition, other areas that are receiving increased attention include:

1. the application of phage to treat bacterial diseases while antibiotic resistance is increasing among bacteria (Ho, 2001; Lakhno and Bordunovskii, 2001; Loeffler *et al.*, 2001; Nelson *et al.*, 2001; Smith *et al.*, 2001; Sulakvelidze *et al.*, 2001; Summers, 2001; Schuch *et al.*, 2002; Liu *et al.*, 2004)
2. the eradication of industrially destructive phage in the food industries (Brussow, 2001; Desiere *et al.*, 2001; Barrangou *et al.*, 2002)

(Campbell, 2003; Nelson, 2004).

The expansion of knowledge about bacteriophages has been aided by the development of genetic analysis technology and bioinformatics, allowing the sequencing of phage genomes and improved genomic annotation. In recent years a number of bacteriophage genomes have been sequenced completely. At the time of writing this thesis a total of 237 phage genome sequences are available on GenBank, 45 of which were sequenced in 2004. Despite this, there is little available on genomic sequences of bacteriophages derived from rumen.

4.0.1: Aim

The bacteriophage ϕAR29 integration cassette has been isolated and cloned (Gregg *et al.*, 1994b). The database available at the time of constructing integrative vector, pBA, suggested that the cloned integration fragment carries the phage *attP* site, *Int* and *Xis* gene (Gregg *et al.*, 1994b). Although studies described in Chapter 3, have shown that the genes cloned in pBA (integrase and excisionase) were transcribed in the

plasmid host, previous findings did not detect the integration of plasmid into the chromosome of AR29 or AR20. The aims of this study were:

- To obtain the complete genome sequence of bacteriophage ϕAR29.
- Characterisation of the genome, to establish the gene arrangement.
- Use the ϕAR29 genome sequence to help confirm the precise location of the *attP* site and identify the *attB* region in AR29.
- Explore the possibility that other candidate *Int* or *Xis* genes may be present in the phage, may be involved in the process of integration, and may be located outside the fragment cloned in plasmid pBA.

4.2: Material and Methods

4.2.1: Bacterial strains

B. thetaiotaomicron AR29 with prophage ϕAR29, used throughout the study, is described in Section 2.3. *Escherichia coli* strain PMC112 (Section 2.3) was used as a cloning host for bacteriophage DNA sequences.

4.2.2: Phage Genomic DNA Preparation

Phage genomic DNA was prepared by method as decribed by *Klieve et al.,* (1989), with some changes to the protocol. In brief, cultures of *B. thetaiotaomicron* AR29 were grown in 5 mL rumen fluid medium overnight at 39 °C. Ten microlitres of the overnight culture was inoculated into 5 mL of fresh rumen fluid medium and incubated at 39 °C until the medium turned slightly cloudy (approximately 5 hours), at which time mitomycin C was added to a final concentration of 2 μg/mL. After 24 hours growth at 39 °C, the lysate was centrifuged at 4 °C for 15 minutes at 9,000 × g. The supernatant was treated for 1 hour at 37 °C with DNase I (20 units/mL) and RNase (1 μg/mL) to remove bacterial DNA and RNA respectively. Phage particles in the

supernatant were pelleted by centrifugation at 50,000 × g at 4°C for 2 hours. The supernatant was discarded and the pellet was resuspended in 0.5 mL of TE buffer (pH 8) and treated with proteinase K (1mg/mL) for 1 hour at 50°C. Subsequently, 10 μL of 10% SDS was added and the sample was phenol:chloroform:isoamyl alcohol extracted. Phage genomic DNA was precipitated with ethanol and redissolved in water as described in section 2.6.2.

4.2.3: Testing for the Presence of Cohesive Termini (cos site)

Two procedures were used to test for the presence of a *cos* site in the ϕAR29 genome: (1) Restriction analysis using Southern blot detection, and (2) analysis by gel electrophoresis (Nauta, 1997).

1) Southern blot detection to test for the presence of a *cos* site was used to maximise the detection sensitivity for faint DNA bands. This involved:

1. Preparation of two separate Hind III restriction digest reactions of phage ϕAR29 genomic DNA.
2. DNA from the first Hind III reaction was heated at 60°C for 5 minutes to dissociate possible annealing at a *cos* site and was chilled immediately to 0°C. The digested samples were end-filled and labeled as described by Boehringer Mannheim Dig DNA Labeling and Detection kit (Cat. No. 1093657: Gebeyehu *et al.*, 1987). In brief: 3 μg of the DNA was labeled for 1 hour at 37 °C by the method shown in Table 4.1. The reaction was terminated by adding 2 μL of 0.2 M EDTA (pH 8.0) and the DNA was precipitated by adding 2.5 μL of 4 M LiCl and 75 μL of ethanol and chilling for 2 hours at -20 °C.
3. . The sample was then centrifuged at 20,800 × g for 15 minutes and the pellet was washed with 50 μL of ethanol. Precipitated DNA was re-centrifuged and the

ethanol discarded. After drying under vacuum, the pellet was re-dissolved in 20 μL of water.

Table 4.1: Composition of labeling reaction, as described by Boehringer Mannheim Dig DNA Labeling and Detection kit, that was used to test for the presence of a *cos* site in phage ϕAR29 genomic DNA

Component	Volume (μL)
Hexanucleotide mixed	2
dNTP mixture	2
Klenow enzyme	1
DNA template	0.5 – 3 μg
MilliQ water	to a volume total of 20

4. Both Hind III restriction digests were electrophoresed on a 1.2 % agrose gel in TAE buffer.
5. The electrophoresed DNA samples were transferred to hybridisation membranes by the method of Southern (1975) as described in *Short Protocols in Molecular Biology*. Briefly, the gel was:
    - washed with sterile distilled water then soaked in 10 gel volumes of 0.25 M HCl for 30 minutes
    - soaked in two changes of 10 volumes of denaturation solution for 20 minutes each.
    - rinsed in sterile distilled water and a soaked in 10 volumes of neutralisation solution.
    - blotted overnight to a neutral nylon membrane (Amersham) as described in *Short Protocols in Molecular Biology*.
6. Following the transfer, the Nylon membrane was rinsed with 2 × SSC and baked at 80°C for 1 hour or exposed to Ultraviolet (UV) light (254-nm wavelength) to fix the DNA permanently onto the membrane.

7. DIG DNA Labelling and Detection kit , and Roche DIG Wash and Block Buffer Set (Cat. No. 1585762) were used to visualise the membrane-bound DNA. In summary:
    - The membrane was washed twice in 2 × SSC/0.1% SDS at room temperature, followed by two washes in 0.1 × SSC/0.1% SDS at 68°C under constant agitation.
    - the membrane was rinsed briefly in DIG Washing buffer and was incubated in DIG Blocking solution for 1 hour at room temperature.
    - Anti-DIG-AP conjugate was added for 30 minutes, then the membrane was washed, equilibrated in detection buffer and transferred to a sealed bag containing colour-substrate solution (NBT/BCIP). The colour reaction was allowed to develop overnight in a dark location, without shaking.
8. Following colour development, the membrane was washed in distilled water and images of the bands were scanned using a UMAX ASTRA 610S scanner and analyzed using PHOTOSHOP 5.5 program.

2) The gel electrophoresis protocol used to test for the presence of a *cos* site involved:

1. A sample of phage ϕAR29 genomic DNA was treated with DNA ligase overnight.
2. The ligated sample and two unligated samples of phage ϕAR29 genomic DNA were digested with Hind III in three separate reactions.
3. The ligated sample and one of the two unligated Hind III digests were heated at 60 °C for 5 minutes to dissociate any annealed *cos* site prior to gel electrophesis.
4. The 3 restriction digests were compared by electrophoresis for differences that could indicate the presence of a *cos* site.

4.2.4: Sequencing of the ϕAR29 Genome

Two approaches were used to determine the genome sequence of ϕAR29.

1. The construction of a phage genomic library (Nauta, 1997; Loessner *et al.*, 2000; Vander Byl and Kropinski, 2000). DNA fragments from a Hind III digest of ϕAR29 genome were ligated to plasmid, pUK21 and cloned in *E. coli* strain PMC112. Cloned phage fragments were sequenced using M13 primers. To sequence the larger cloned fragments, SSC and 2SC primer sets (see Appendix 4; Table 4.1A) were designed for sequential primer walking, through the plasmid clones.
2. Primer walking using synthetic oligonucleotides designed from opposite ends of the phage integration module cloned in plasmid pBA(Loessner *et al.*, 2000). These primers are indicated by the prefix SSP (see Appendix 4; Table: 4.1A). .

After completing the initial sequencing, the genome sequence was confirmed by amplifying genome segments using forward and reverse 2SSP primers in PCR reactions (e.g. 2SSP1for and 2SSP1rev; see Appendix 4; Table 4.1A) and the resulting PCR products were sequenced from each strand, using the same primers.

All thermal cycling conditions and ethanol precipitation of sequencing reaction were as described in Chapter 2 section 2.6.17. Sequence contigs were assembled using Vector NTI Advance™.

4.2.5: Predicting ORFs

Open reading frames (ORF) were predicted using the GeneMark Heuristic model (Besemer and Borodovsky, 1999). ORFs with more than 30 codons were accepted as potential genes. A second ORF prediction program, FGENESV (Xu *et al.*, 2003b) was used to confirm the output generated by GeneMark Heuristic.

4.2.6: *B. thetaiotaomicron* AR29 Genomic DNA Preparation and Sequencing

Bacteria genomic DNA preparation was carried out as described by Woods *et al.*, (1989). In brief, *B. thetaiotaomicron* AR29 carrying the prophage was grown overnight in 10 mL of DF medium. Prior to genomic DNA preparation, the culture was incubated at 80°C for 15-20 mins to inactivate nucleases. Bacterial cells were collected by centrifugation at 20,800 × g for 5 min at 4 °C, washed with TE buffer (pH 8.0) and re-centrifuged as before. The cells were resuspended with 0.1 volume of TE (pH 8.0) and incubated with lysozyme (1 mg/ml) at 37°C for 30 mins. RNA and proteins were removed by incubating the preparation with 100 μL of 10% SDS and RNase (1 μg/mL) at 37°C for 30 mins, followed by overnight incubation with proteinase K (1mg/mL) at 55°C. An equal volume of phenol was added, mixed, and the sample was centrifuged at 20,800 × g for 10 mins. The aqueous layer was removed, extracted with an equal volume of chloroform, centrifuged at 20,800 × g for 5 mins, and the aqueous layer removed and retained. The DNA was ethanol precipitated as described in section 2.6.2, except that precipitation was at room temperature (approx 22°C).

Synthetic oligonucleotides complementary to the *attP* flanking sequences (Sbagatt1 for and rev) were used to sequence directly from the chromosomal DNA, to establish the precise location of the ϕAR29 *attB* site (Table 4.2). The first set of Sbgatt primers were designed from sequences that were proposed by Gregg *et al*., (1994) to contain the *attB*. Thermal cycling reactions and conditions for bacterial genomic DNA sequencing are shown in Table 4.3 and 4.4, respectively.

Table 4.2: Synthetic oligonucleotides used to sequence *attB* in AR29

Primer name	Sequence 5′→ 3′	Length (b)	Tm (° C)
Sbgatt1 for	ATGCGATGGGCGATAAAGTC	20	52
Sbgatt1 rev	AGTCGGTTATGGAGGCTTTCAC	22	55
Sbgatt2 for	CCTAGATGTTGTGCGAATGTTGTG	24	56
Sbgatt2 rev	TAGGTTTTTGCGTTCCGTTATGTC	24	54
Sbgatt3 For	GATGTTGTGCGAATGTTGT	19	47
Sbgatt3 rev	TAGCTATCATTTTACGGTTGTG	22	49

Table 4.3: The amount of template required per 40-µL sequencing reaction

Composition	**Volume (µL) per reaction**
Primer (3.2 pmole/µL)	3.5
Big Dye terminator version 3.1 mix	16
DNA (ng)	2000-3000
DMSO	2
Water	Volume to total of 40 µL

Table 4.4: Thermal cycle characteristics used during PCR

Cycle	1 cycle	45 cycles			1 cycle
Temperature (°C)	95	95	55-60	60	14
Duration (min or sec)	5 min	30 sec	20 sec	4 mins	indefinite

4.2.7: Predicting the Position of Regulatory Elements in ϕAR29 Genome

Promoters analysis was carried out *in silico* by PHIRE version 1.0 (Lavigne *et al.,* 2004) and Neural Network Promoter Prediction (NNPP:Reese, 2001). Bionagviator GCG Terminator software was used for transcription terminator prediction in the ϕAR29 genome (Brendel and Trifonov, 1984).

4.2.8: Predicting Translational Frameshifts in the ϕAR29 Genome

Frameshift analysis was carried out *in silico* using software "Programmed Frameshift Finder" (Xu *et al.,* 2004) and "FSFinder" (Moon *et al.,* 2004).

4.2.9: Identifying tRNA in ϕAR29 Genome

Online software, tRNAscan-SE Search Server, was used to search for any tRNA genes present in ϕAR29 genome (refer to Chapter 2, section 2.2).

4.2.10: Cloning and Examination of Lytic Activity of ϕAR29 Amidase Gene

Bacteriophage ϕAR29 amidase gene was cloned into the expression vector pTrcHis A (Invitrogen, Cat#V360-20: Egon *et al.*, 1983; Brosius *et al.*, 1985; Mulligan *et al.*, 1985) by:

1. Amplifying the genomic region containing the amidase gene with primers SABamHI and SAHindIII, which carry a *Bam*HI and *Hind*III restriction site, respectively.
2. Purification of the PCR product from a 1% TAE gel as described in section 2.6.0.
3. In separate reactions, the recombinant plasmid pTrcHis A and the PCR product were both digested with a combination of HindIII and BamHI.
4. The products from restriction digests were ligated together and
5. *E. coli* strain PMC112 was transformed with the ligation products

Table 4.5: Characteristics of primer SamidBam and SamidHind

Primer name	Sequence 5′→ 3′	Length (b)	Tm (°C)
SamidBam	GGCGACGGATCCAAGAGAGAAGATATAGA	29	61
SamidHind	TCACTAAGCTTTCATGGTCGGATCACTGT	29	60

To test for amidase activity, the turbidity of the culture was measured at intervals of 1 hour, over a total period of 10 hours, using a Shimadzu UV-1201 UV-VIS spectrophotometer. Five different reactions were performed and the composition of each reaction is shown

in Table 4.6. Expression of the cloned protein was induced with IPTG (1mM) at the 3rd hour of growth after samples were taken for absorbance readings.

Cells taken from Flask 2 and 3 at the 3rd and 5th hour were also used for Western blot analysis. These samples were harvested by centrifugation at 5,000 × *g* for 10 min at 4°C. The supernatant was discarded and each pellet was resuspended with 10 µl Ni-NTA denaturing lysis buffer (100 mM NaH_2PO_4, 10 mM Tris.HCl, pH 8.0, 8 M urea). After vortexing the tube for 1 min, the cellular debris was pelleted by centrifugation at 10,000 × *g* for 10 min at 4°C. The supernatant was transferred to a new tube and stored at -20°C until analysis.

Table 4.6: Composition of experimental cultures for amidase activity studies.

Flasks (note)	Volume of LB (mL)	Volume of 100 mg/mL of Ampcin (µL)	Volume of 1 M glucose (final concentration of 0.3%) (µL)	Volume of Lysozyme 10 mg/mL (final concentration of 0.2 mg/mL) (mL)	Volume of 0.5 M IPTG to a final concentration of 1 mM (µL)	Voume of overnight Inoculums (µL)	Water (mL)	Total Volume (mL)
pAmidase + Glucose - Lysozyme + addition of IPTG at 3rd hour (Suppressed pAmidase: -ve control)	48	50	100	0	100	200 µL of PM112 (pAmidase)	1.55	50
pAmidase – Glucose – Lysozyme – IPTG (uninduced pAmidase: +ve control)	48	50	0	0	0	200 µL of PM112 (pAmidase)	1.75	50
pAmidase – Glucose – Lysozyme + addition of IPTG at 3rd hour (Test)	48	50	0	0	100	200 µL of PM112 (pAmidase)	1.65	50
pAmidase – Glucose + Lysozyme – IPTG (Lysozyme activity: +ve control)	48	50	0	0	0	200 µL of PM112 (pAmidase)	1.75	50
pTrcHis – Glucose – Lysozyme + addition of IPTG at 3rd hour (pTrcHis A construct: -ve Control)	48	50	0	1	100	200 µL of PM112 (pTrcHis)	0.65	50

4.2.11: Western Blots

Samples harvested from Flask 2 and 3 were separated by SDS-PAGE, using a discontinuous Tris.glycine buffer system as described in *Short Protocols in Molecular Biology*. This involved treating 30 μL of protein sample with:

1. 10 μL of 4 × sample treatment buffer (250 mM Tris.HCl (pH 6.0), 8% SDS, 200 mM DTT, 40% glycerol and 0.04 % bromophenol blue).
2. Samples were boiled for 5 min immediately prior to loading 10 μl of the sample into wells in the gel.

The gel used for SDS-PAGE comprised a stacking gel (125 mM Tris.HCl pH 6.8, 4% acylamide, 0.15% bis-acrylamide and 0.1% SDS) and a separating gel (375 mM Tris.HCl ph 8.8, 12% acylamide, 0.31% bis-acrylamide and 0.1% SDS). Polymerisation of the gel was catalysed by addition of 0.1% TEMED and 0.05% freshly prepared ammonium sulphate solution and cast into the mini-Protean dual slab cell (Bio-Rad). Protein samples were electrophoresed at 150 V at room temperature until the bromophenol blue dye-front reached the bottom of the gel. Pre-stained molecular weight standards were electrophoresed in parallel with the samples. After electrophoresis, the gel was stained with Coomassie Brilliant Blue G250 (Biorad) or electro-transferred onto nitrocellulose membrane for Western blotting.

Electrophoretic transfer of separated proteins from the SDS-PAGE gel to nitrocellulose membrane was performed using the Towbin transfer buffer system, which involved:

1. Equilibrating the gel in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, pH 8.3) for 15 min.
2. The gel were transferred to nitrocellulose membrane (Protran) using the mini-Protean transblot apparatus (Bio-Rad). Electrophoretic transfer was performed at 30 V overnight at 4°C.

3. Nitrocellulose membrane containing the transferred proteins was blocked with 10 mL of Tris-buffered saline (TBS) containing 5% skim milk powder for 1 h at room temperature.
4. The membrane was washed with TBS containing 0.1% Tween 20 (TBST) and then incubated with 10 mL mouse anti-his antibody (Sigma; diluted 5,000-fold with TBST) for 1 h at room temperature.
5. The membrane was washed three times with TBST for 5 min before being incubated with 10 mL of goat anti-mouse IgG (whole molecule)-AP (Sigma; diluted 5,000-fold in TBST), for 1 h at room temperature.
6. Proteins on the membrane were detected using the Alkaline Phosphatase Substrate Kit (Biorad). The detection reaction was terminated by washing the membrane with distilled water and the membrane was dried and scanned.

4.3: Results:

4.3.1: Nucleotide Sequence of the ϕAR20 Genome

Nucleotide sequencing of ϕAR29 was completed on both DNA strands (see Appendix 2 for genomic sequence). The sequencing results revealed that ϕAR20 comprises a 35,558 bp genome, in close agreement with the size determined by restriction mapping (Figure 4.1). The GC content of the phage genome is 39.11%, which is slightly lower than in *B. thetaiotaomicron* VPI-5482, with 42.8%.

Southern blots were used to determine whether ϕAR29 possesses specific cohesive ends, but these studies gave inconclusive results (not shown). Hind III digestion and ligation studies were also applied to determine whether complementary termini were present. No changes occurred in the restriction pattern following ligation, which would be expected to link together the terminal fragments of a phage genome

with cohesive termini, thus suggesting the absence of cohesive ends (Figure 4.2) or possibly the absence of 5' terminal phosphate groups on the phage DNA. Migration of undigested phage DNA adjacent to the 23.13 kb band of Hind III digested λ DNA marker and the absence of DNA bands in or near the wells indicated that phage genomic DNA preparations contained only linear molecules. Furthermore, the ability to amplify genomic fragments across all regions of the circular genome map indicates that no specific part of the genome is represented at the termini. If the phage did possess cohesive 3' termini, at least one position within a circular genome map would be expected to be interrupted by the terminal break unless the genome was in a circular form.

Figure 4.1: (below) *Hin*dIII digestion of ϕAR29 genomic DNA. The position of undigested 35.5 kb ϕAR29 DNA suggests a linear molecule. Lanes 1: λ DNA digested with *Hin*d III; Lane 2: Undigested ϕAR29 DNA, Lane 3: ϕAR29 digested with *Bam*HI; Lane 4: ϕAR29 digested with *Cal* I; Lane 5: ϕAR29 digested with *Eco*RI; Lane 6: ϕAR29 digested with *Eco*RV; Lane 7: ϕAR29 digested with *Hin*d III; Lane 8: ϕAR29 digested with *Knp*I.

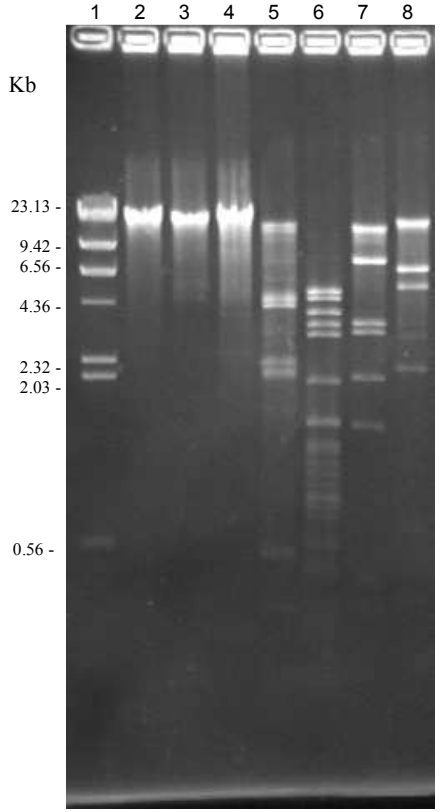
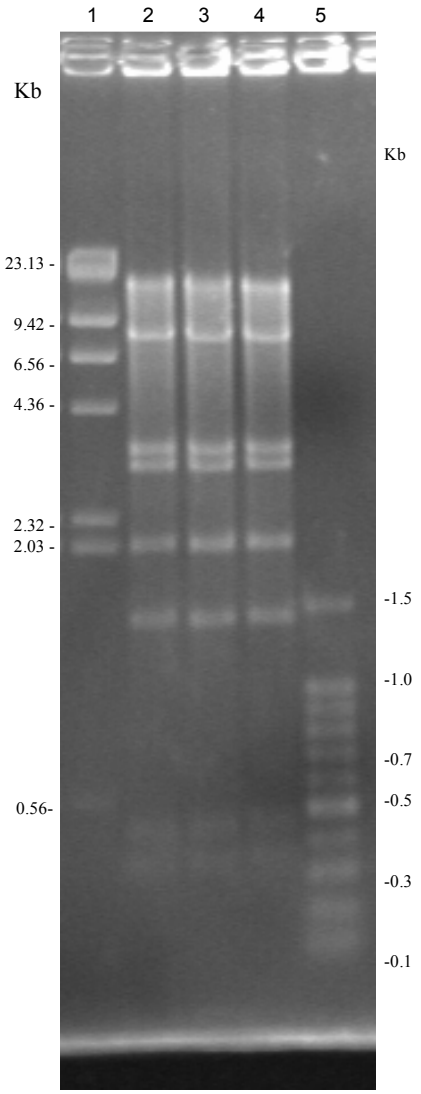


Figure 4.2: (Right) Ligation and *Hin*d III digestion of Bacteriophage ϕAR29 genome to characterise the termini of linear DNA molecules.

Lanes 1: λ DNA digested with Hind III; Lane 2: ϕAR29 DNA digested with Hind III, Lane 3: ϕAR29 DNA digested with *Hin*d III and heat treated for 5 min at 60°C prior to electrophoresis; Lane 4: ϕAR29 treated with T4 DNA ligase before digestion with *Hin*d III and heated for 5 min at 60°C prior to electrophoresis; Lane 5: 100 bp ladder.



4.3.2: Identification of ϕAR29 Open Reading Frames, Functional Prediction of Genome Organization

The open reading frame prediction programs Fgenes (Softberry) and Heuristic predicted 53 protein-coding regions within the ϕAR29 genome. The ORFs are summarized in Table 4.7. The nucleotide sequences and the deduced amino acids (aa) sequences of each ORF were subjected to BLAST analysis against the GenBank in an attempt to identify their function and to allow the ϕAR29 gene arrangement to be compared with those of other bacteriophages. The ORFs can be represented as three groups:

1. Identifiable ORFs
2. Significantly matched ORFs, and
3. Poorly matched ORFs.

Table 4.7: AR29 predicted ORFs with their position, deduced amino acids sequence length, protein mass, putative function and BLASTP analysis results.

ORF	Position (start and terminating condon)	+/- strand	Size (aa)	kDa	Function	Best match Genbank Accession No.: Protein(s)	BLASTP E Value (Score) Identity %
1	742(AUG)-308(UGA)	-	145	16.04	Unidentified	CAA56475:replicase [Pseudomonas phage PP7]. (552 aa) (Olsthoorn *et al.*, 1995)	0.59 (32) 30
2	1145(AUG)-735(UAA)	-	137	15.22	Unidentified	ZP_00144218: Glutamate-1-semialdehyde 2,1-aminomutase [Fusobacterium nucleatum subsp. vincentii ATCC 49256].(434 aa) (Kapatral *et al.*, 2003)))	0.16 (41) 31
3	3847(AUG)-1142(UGA)	-	902	99.83	Unidentified	ZP_00145627: Glutamate dehydrogenase/leucine dehydrogenase [Psychrobacter sp. 273-4].(448 aa)	0.19 (41) 31
4	4491(AUG)-	-	214	24.7	Possible small	ZP_00213073: UDP-glucose	0.25

	3850(UAA)				subunit Terminase ??	pyrophosphorylase [Burkholderia cepacia R18194].(295 aa)	(37) 28
5	5650(AUG)-4472(UAA)	-	393	44.66	Large subunit Terminase	AAL82258: large terminase [Staphylococcus aureus phage phi 11].(447 aa) (Iandolo *et al.*, 2002)	8e-37 (153) 32
6	6180(AUG)-5650(UGA)	-	178	20.03	Unidentified	NP_815687: hypothetical protein EF2018 [Enterococcus faecalis V583].(153 aa) (Paulsen *et al.*, 2003)	9e-03 (42) 26
7	6780(AUG)-6238(UGA)	-	181	21.34	Unidentified	glutamine-dependent carbamyl phosphate synthetase [Bacteroides fragilis YCH46].(1073 aa) (Kuwahara *et al.*, 2004)	4.8 (33) 25
8	7187(AUG)-7068(UAG)	-	40	4.5	Unidentified		
9	8016(AUG)-7180(UAA)	-	280	31.76	Unidentified		
10	8572(AUG)-8003(UGA)	-	190	21.47	Unidentified	ZP_00314704: Histone acetyltransferase HPA2 and related acetyltransferases [Microbulbifer degradans 2-40]. (320 aa)	3e-03 (43) 25
11	9175(AUG)-8849(UAA)	-	110	7.00	Unidentified	AAW70513: gp42 [Salmonella typhimurium bacteriophage ES18]. (121 aa) (Casjens *et al.*, 2005)	8e-05 (44) 42
12	9311(AUG)-9565(UAA)	+	86	9.28	Unidentified		
13	9974(AUG)-9789(UAA))	-	62	7.08	Unidentified		
14	10293(AUG)-9994(UAA)	-	100	11.68	Unidentified		
15	11157(AUG)-10306(UAA)	-	284	32.60	Unidentified		
16	11983(AUG)-11264(UGA)	-	227	27.94	DnaC	ZP_00307502: Predicted ATPase [Cytophaga hutchinsonii].(212 aa)	8e-09 (62) 26
17	12644(AUG)-11901(UAA)	-	248	28.00	Possible DnaD??	YP_099609:hypothetical protein BF2328 [Bacteroides fragilis YCH46].(295 aa)	1e-48 (194) 69%

						(Kuwahara *et al.*, 2004)	
18	13225(AUG)-12818(UAG)	-	136	15.76	Unidentified	AAQ14164: PmgM-morphorgenetic function [Enterobacteria phage P1]. (192 aa)(Lobocka *et al.*, 2004)	3e-16 (86) 40
19	14205(AUG) -13768(UAA)	-	146	16.94	Unidentified		
20	14507(AUG)-14208(UAA)	-	100	11.07	Unidentified		
21	14992(AUG)-14513(UAA)	-	160	18.18			
22	15648(AUG)-14989(UGA)	-	220	25.04	Unidentified	YP_164283:hypothetical protein F116p19 [Pseudomonas aeruginosa phage F116]. (251 aa)	4e-05 (50) 27
23	16081(AUG)-15785(UAA)	-	99	11.24	Possible Excisionase??	YP_133702:hypothetical protein NBU1_08 (Orf2X)[Bacteroides uniformis]. (104 aa) (Shoemaker *et al.*, 2000)	6e-09 (62) 37
24	16271(AUG)-16092(UAG)	-	60	6.75	Unidentified		
25	16675(AUG)-16295 (UAG)	-	127	14.69	HTH-LUXR	NP_947645.1: Bacterial regulatory protein, LuxR family [Rhodopseudomonas palustris CGA009]. (217 aa) (Larimer *et al.*, 2004)	7e-04 (45) 46
26	16887(AUG)-16675(UGA)	-	71	8.24	Unidentified		
27	16964(AUG)-17248 (UAA)	+	95	10.81	Unidentified		
28	17385(AUG)-17143(UAG)	-	81	9.24	Unidentified		
29	17864(AUG)-17691(UAG)	-	58	6.74	Unidentified		
30	18121(AUG)-17864(UAA)	-	86	10.19	Unidentified		
31	18322(AUG)-19020(UAA)	+	233	26.48	cI repressor	P14819: Phage Phi 80 cI repressor. (236 aa) (Ogawa *et al.*, 1988)	4e-03 (40) 29
32	19067(AUG)-19234(UAG)	+	56	6.35	Unidentified		
33	19246(AUG)-20007(UAA)	+	254	29.17	Unidentified		
34	20019(AUG)-20549(UAA)	+	177	20.03	Unidentified		
35	20870(AUG)-22249(UAA)	+	460	54.22	Integrase	YP_133699: IntN1 [Bacteroides uniformis]. (445 aa) (Shoemaker	1e-31 (139) 26

						et al., 2000)	
36	22522(AUG)-227010(UAA)	+	63	6.75	Unidentified		
37	23103(AUG)-22793(UAG)	-	104	12.18	Unidentified		
38	23136(AUG)-23234(UAA)	+	33	3.73	Unidentified		
39	23746(AUG)-23231(UAG)	-	172	19.95	Unidentified		
40	24243(AUG)-23743(UGA)	-	167	18.42	Amidase	*Bacteroides thetaiotaomicron* NP_810940.1: N-acetylmuramoyl-L-alanine amidase [Bacteroides thetaiotaomicron VPI-5482]. (145 aa) (Xu *et al.*, 2003a)	3e-48 (192) 62
41	24476(AUG)-24255(UAG)	-	74	8.09	Unidentified		
42	24735(AUG)-24454(UAA)	-	94	10.39	Unidentified		
43	25287(AUG)-24748(UAA)	-	180	20.31	Unidentified		
44	26431(AUG)-25280(UAG)	-	384	43.12	Unidentified	YP_040054: aldo/keto reductase family protein [Staphylococcus aureus subsp. aureus MRSA252]. (312 aa) (Holden *et al.*, 2004)	0.05 (41) 31
45	26975(AUG)-26451(UAA)	-	175	19.90	Unidentified		
46	27508(AUG)-26978(UAA)	-	177	18.46	Unidentified		
47	28035(AUG)-27511(UAA)	-	175	18.33	Unidentified		
48	28570(AUG)-28040(UAA)	-	174	17.99	Unidentified		
49	29533(AUG)-28589(UAA)	-	315	34.8	Unidentified		
50	30553(AUG)-29564(UAA)	-	330	36.46	Unidentified		
51	33209(AUG)-30660(UAA)	-	850	92.96	Possible Tail Tape measure??	AAN28244: tail length determinator [Bacteriophage WPhi]. (815 aa)	8e-03 (42) 21
52	34896(AUG)-33214(UAA)	-	561	63.10	Unidentified		
53	311(AUG)-34898(UAA)	-	323	36.16	Unidentified		

4.3.3: Identifiable ORFs

4.3.3.1: ORF5 (Terminase Large subunit)

ORF5 showed a conserved domain similar to a known phage large terminase subunit (Gual *et al.*, 2000; Chai *et al.*, 1992; Marchler-Bauer *et al.*, 2003). The function of terminase is in the recognition of phage DNA, ATP-dependent cleavage of DNA concatemer, and packaging of the terminally redundant molecule into a capsid. BLAST searches and multiple alignment revealed that ORF5 is closely related to terminases found in bacteriophages ϕ11, Lj928, Lj771, 315.6, SSP1 and prophage Lj965 (Figure 4.8) with amino-acid identity of 32 %, 31 %, 31%, 30 %, 24 % and 22%, respectively. Their E-values were 9e-36, 1e-29, 1e-29 , 8e-29, 4e-07 and 0.008.

Table 4.8: BLAST result of Bacteriophage Terminase and the accession number of the sequence to which the ϕAR29 ORF aligned.

Terminase of other phages	Accession No.: Length of protein (aa)	Similarity % (Identity %) E-value	Reference
ϕ11	NP_803283.1: (447)	50 (32) 9e-36	(Bolotin *et al.*, 2001)
Lj928	NP_958532.1: (422)	46 (31) 1e-29	(Ventura *et al.*, 2004)
Lj771	AAK27930.1: (423)	46 (31) 1e-29	(Desiere *et al.*, 2000)
315.6	NP_665239.1: (425)	48 (30) 8e-29	(Beres *et al.*, 2002)
Ssp1	NP_690654.1: (422)	38 (24) 4e-07	(Tavares *et al.*, 1992)
Lj965	AAS08294.1: (424)	39 (22) 0.008	(Pridmore *et al.*, 2004)

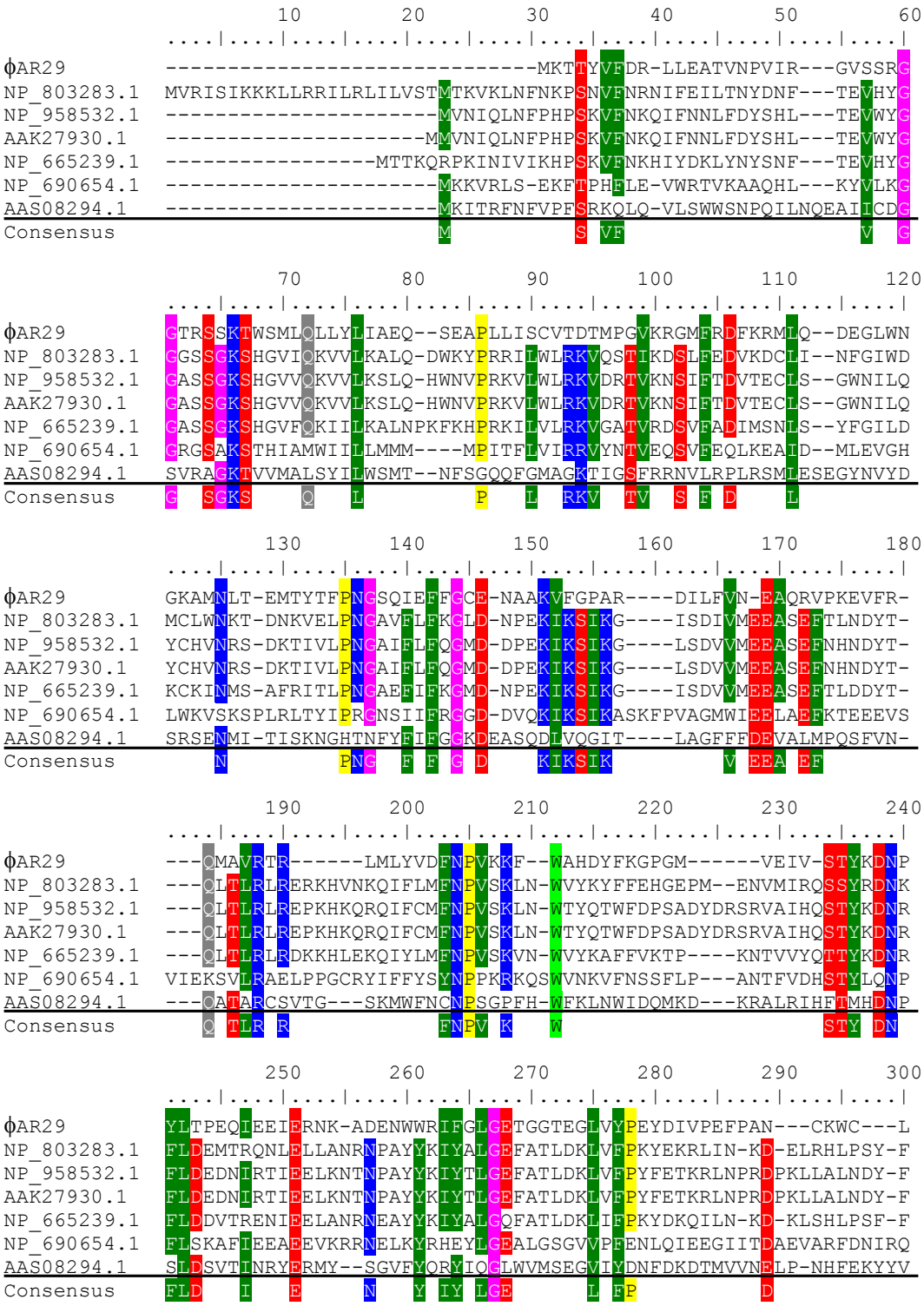


Figure 4.3: Multiple alignment of ϕAR29 terminase with other known bacteriophage terminases. Residue with coloured background indicates identicalmatch.

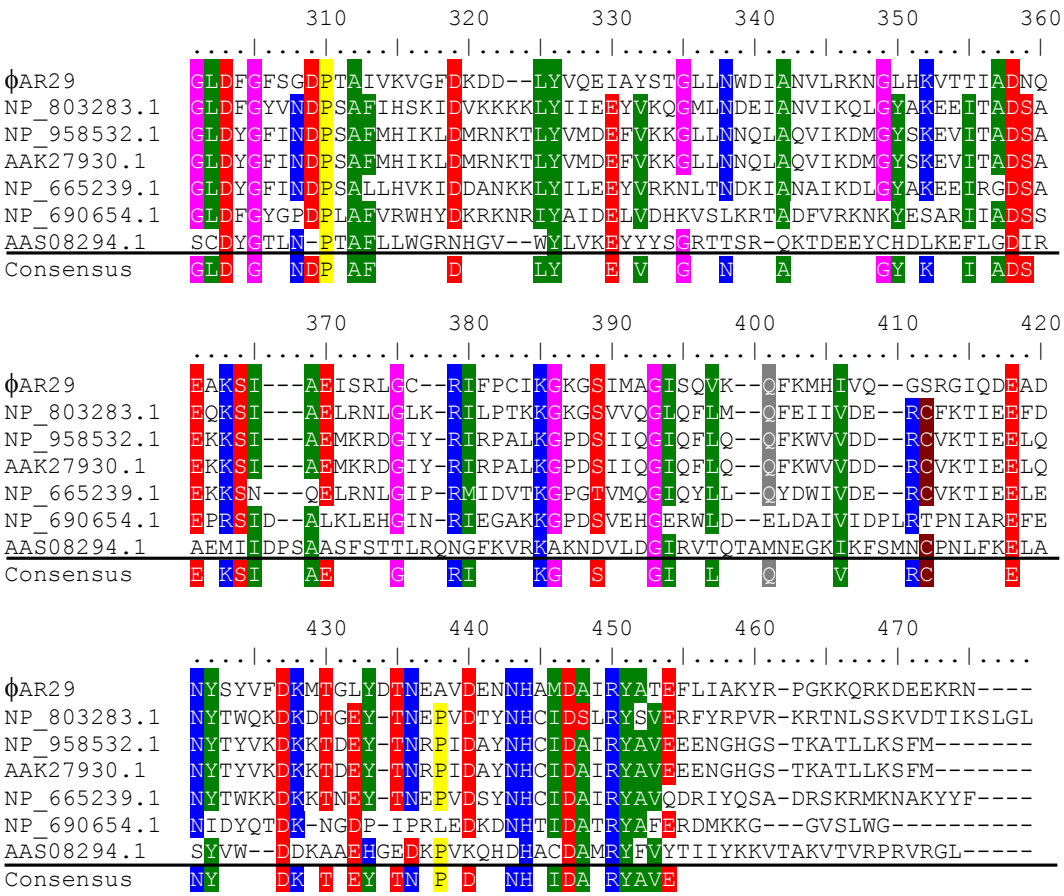


Figure 4.3 *(continued)*: Multiple alignment of ϕAR29 terminase with other known bacteriophage terminases.

4.3.3.2: ORF16 (DnaC protein)

The analysis programs NCBI, InterPro and Motif Scan were able to identify AAA-ATPase domains within ORF16 (Figure 4.4). The major role of AAA family proteins is to perform chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes (Confalonieri and Duguet, 1995). Furthermore, NCBI also showed homologies of ORF16 to domains in DNA replication protein, dnaC. DnaC is required for the formation of a dnaB-dnaC complex. The role of this protein is to help deliver the helicase enzyme to the DNA template. NCBI BLAST showed 26%, 27%, 29% and 28% identity respectively to replication proteins from *Cytophaga hutchinsonii*, *Fusobacterium nucleatum subsp. Vincentii*, *Clostridium thermocellum* and

Leptospira interrogans serovar Copenhageni,. Multiple alignment of these proteins and dnaC (or dnaC-like) proteins of *Bacillus subtilis*, *E.coli* and prophage pi1 showed five conserved motifs at the C-terminus of the protein at positions 20, 95, 125,140 and 165 ORF16 (Figure 4.5).

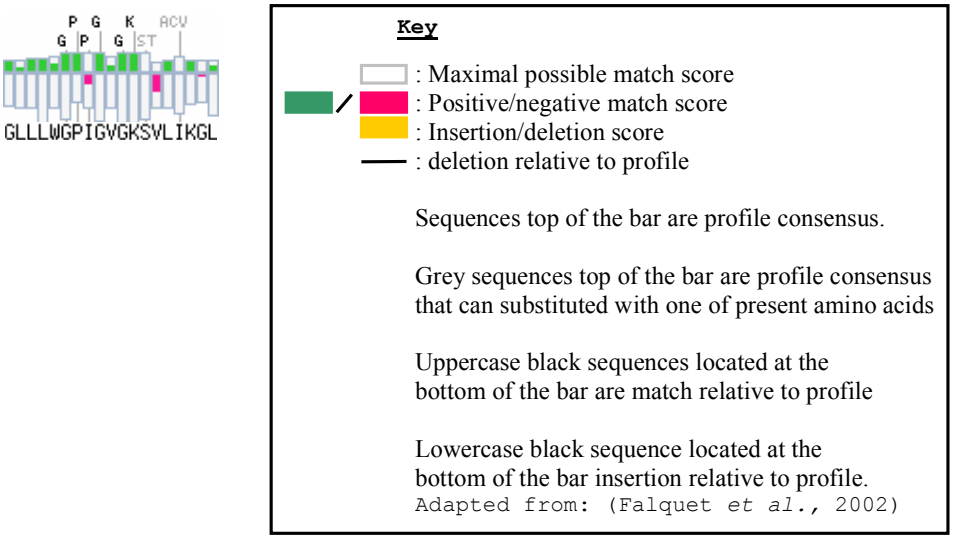


Figure 4.4: Motifs identified in ORF16 by MotifScan (Falquet *et al.*, 2002).

Table 4.9: DnaC Proteins included in the alignment with ORF16.

Organism and protein	Access No. : Length of protein (aa)	References
Cytophaga hutchinsonii Predicted ATPase	ZP_00307502.1: (212)	Unpublished
Fusobacterium nucleatum subsp. Vincentii Replicative DNA helicase	ZP_00144305.1: (266)	(Kapatral *et al.*, 2003)
Clostridium thermocellum DNA replication protein	ZP_00311701: (241)	(Copeland *et al.*, Unpublished)
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130 DNA replication protein DnaC	YP_003073.1: (285)	(Nascimento *et al.*, 2004)
Bacillus subtilis Phage-like element PBSX protein xkdC	P39782: (266)	(McDonnell *et al.*, 1994)
E. coli DNA replication protein	NP_290977.1: (245)	(Perna *et al.*, 2001)
Prophage pi1 replication protein	AAK04547: (291)	(Bolotin *et al.*, 2001)

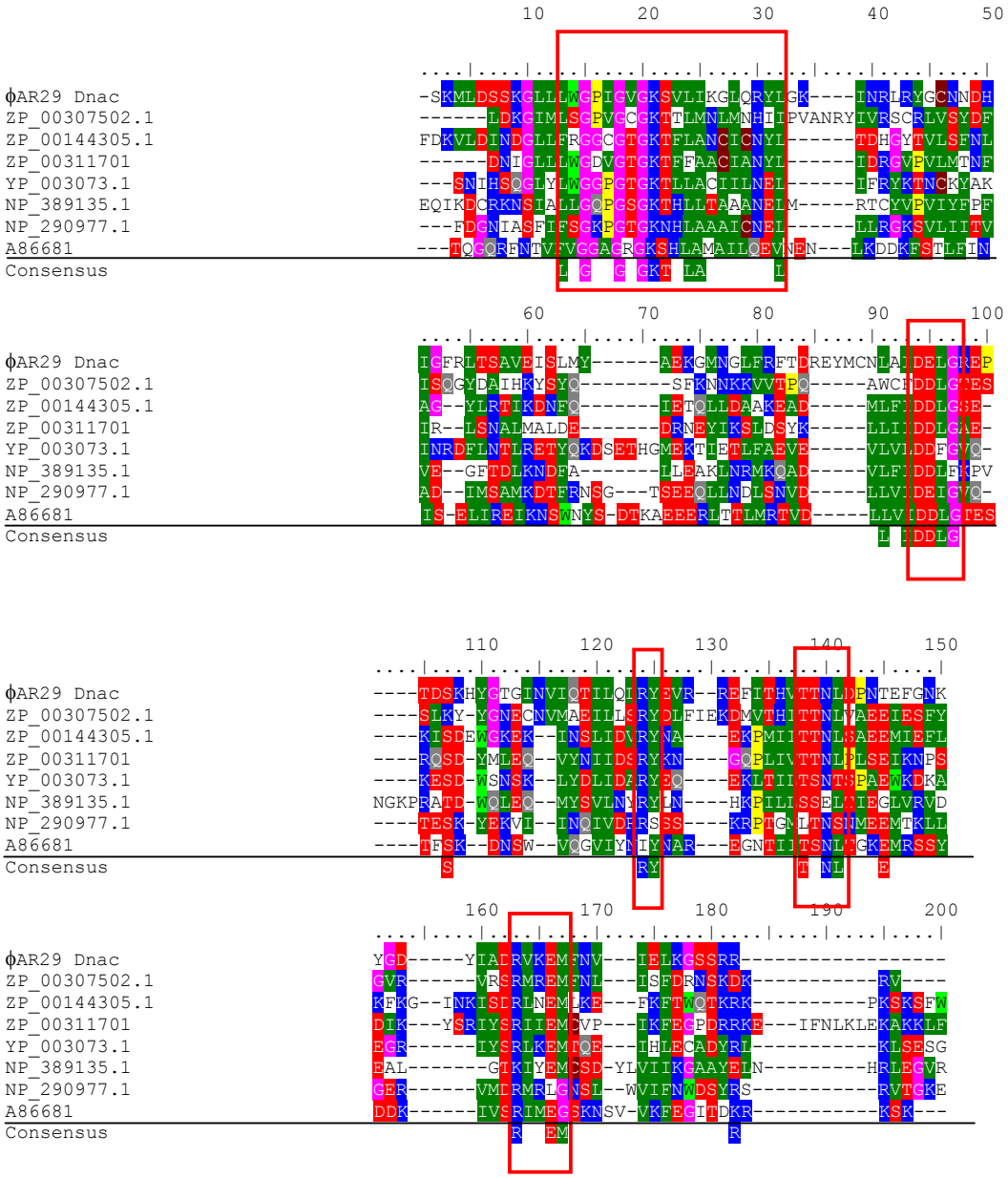


Figure 4.5: Multiple alignment of ORF16 with dnaC-like proteins from *C. hutchinsonii, F. nucleatum subsp. Vincentii, C. thermocellum, L. interrogans serovar Copenhageni. B. subtilis, E. coli* and prophage pi1. Red Boxed areas indicate conserved motifs. Red box indicates conserved region.

4.3.3.3: ORF 25 (Transcription response regulator protein)

Amino acid BLAST analysis of ORF25 indicated that 6 transcription response regulator conserved domains were located at the N-terminus of the protein. These domains were of helix-turn-helix Lux regulon (HTH LuxR), GerE, CitB, CsgD, MalT and TtrR (Marchler-Bauer *et al.*, 2005). Their E-values were 2e-06, 3e-05, 7e-08, 5e-07, 3e04 and

0.006, respectively. Due to the position of ORF25 in relation to ORF 31, it may function like a *Cro* protein that governs the genetic switch of phage ϕAR29 between lysogenic and lytic life cycle.

Table 4.10: Results of conserved domain search on NCBI.

Proteins.	E-values	Accession number
HTH LuxR	2e-06	gnl|CDD|25324
GerE	3e-05	gnl|CDD|25436
CitB	7e-08	gnl|CDD|11904
CsgD	5e-07	gnl|CDD|12157
MalT	3e04	gnl|CDD|12258
TtrR	0.006	gnl|CDD|13729

4.3.3.4: ORF 31 (cI protein)

BLAST analysis of amino acid sequence deduced from ORF31 showed that the encoded protein has a transcriptional regulator protein domain (Accession:COG2932, with an E-value of 1e-07). Further studies using InterPro database identified ORF31 as a protein (Accession:IPRO10982) similar to the λ repressor, *cI*. Bacteriophage λ *cI* repressor is part of the phage lysogenic/lytic growth switch and is essential for maintaining lysogeny of phage. ORF31 also has strong similarity to *cI* repressors of phages Phi 80 (Ogawa *et al.,* 1988), ST104 (Tanaka *et al.,* 2004) and D3 (Kropinski, 2000) (46 % over 111 aa; 45 % over 160 aa and 41 % 194 aa, respectively). Alignment of ORF 31 with 13 other bacteriophage *cI*-like proteins revealed that the C-termini consist of 6 motifs at positions 4, 18, 45, 73, 81 and 100 (Figure 4.6). ORF 31 also has strong homology to a predicted transcriptional regulator found in the bacterium *Cytophaga hutchinsoni* with an E-vale of 1e-11 and an identity value of 27% over 259 aa (Altschul *et al.*, 1997).

Table 4.11: Results of BLAST analysis that shows the similarity and identity value (%) between ϕAR29 ORF31 and other bacteriophage *cI* repressors.

Phage	*cI* protein Access No. : Length of protein (aa)	Similiarty % (identity %) E-value	References
phi-80	P14819: (236)	46 (29) 0.004	(Ogawa *et al.*, 1988)
ST104	YP_006379.1: (229)	45 (26) 0.24	(Tanaka *et al.*, 2004)
D3	NP_061565.1: (223)	41 (25) 0.32	(Kropinski, 2000)

Table 4.12: Repressor cI proteins of other bacteriophages used in the alignment with ORF31and their accession number.

Protein	Access No.	References
Lambda	AAA96581	(Weigel *et al.*, 1973)
phi-80	P14819	See Table 4.11
ST104	YP_006379.1	See Table 4.11
D3	NP_061565.1	See Table 4.11
BPP-1	NP_958701.1	(Liu *et al.*, 2004)
EJ-1	NP_945243.1	(Romero *et al.*, 2004)
D3112	NP_938208.1	(Wang *et al.*, Unpublished)
HK620	NP_112053.1	(Clark *et al.*, 2001)
CP-933H	NP_285963	(Perna *et al.*, 2001)
HK97	NP_037735.1	(Juhala *et al.*, 2000)
434	S32822	(Nikolnikov *et al.*, 1984)
VT2-SA	BAA89781.1	BAA89781.1
V	NP_599066.1	NP_599066.1

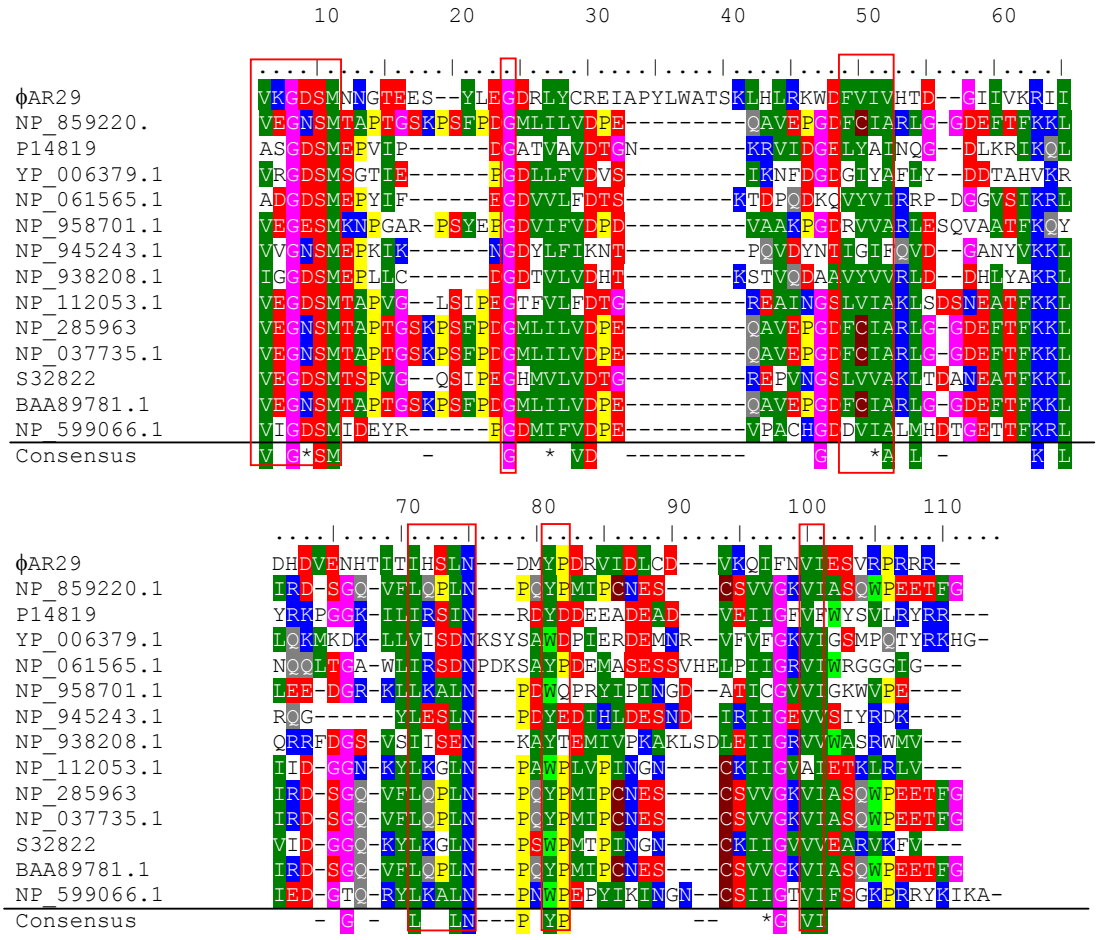


Figure 4.6: Multiple C-terminal alignment of OR31 with known cI repressors. Red boxed regions indicate conserved motifs.

4.3.3.5: ORF 35 (Integrase)

The deduced protein was identified as a phage integrase, which is probably responsible for the integration and excision of ϕAR29. Interpro Scan search revealed that ORF35 is similar to the Cre recombinase, which belongs to the Lambda-like (tyrosine) recombinase family (Kwon *et al.*, 1997; Guo *et al.*, 1997). Online software, MotifScan, found a motif region located between positions 255 and 436 (Figure 4.7). The database identified this motif region was identified as characteristic of the Phage Integration Family (Accession No.: PF00589.9) with an N-score and E-value of 10.849 and 0.0003, respectively (Falquet *et al.,* 2002). A domain search on NCBI also showed similarity to transponson Tn4399 motifs, with an E vaule of 7e-10 (Marchler-Bauer and

Bryant, 2004). Protein Tn4399 and related integrases include various bacterial integrases, including cLV25, a *Bacteroides fragilis* chromosomal transfer factor integrase similar to the *Bacteroides* mobilizable transposon, Tn4399 (Marchler-Bauer *et al.*, 2005).

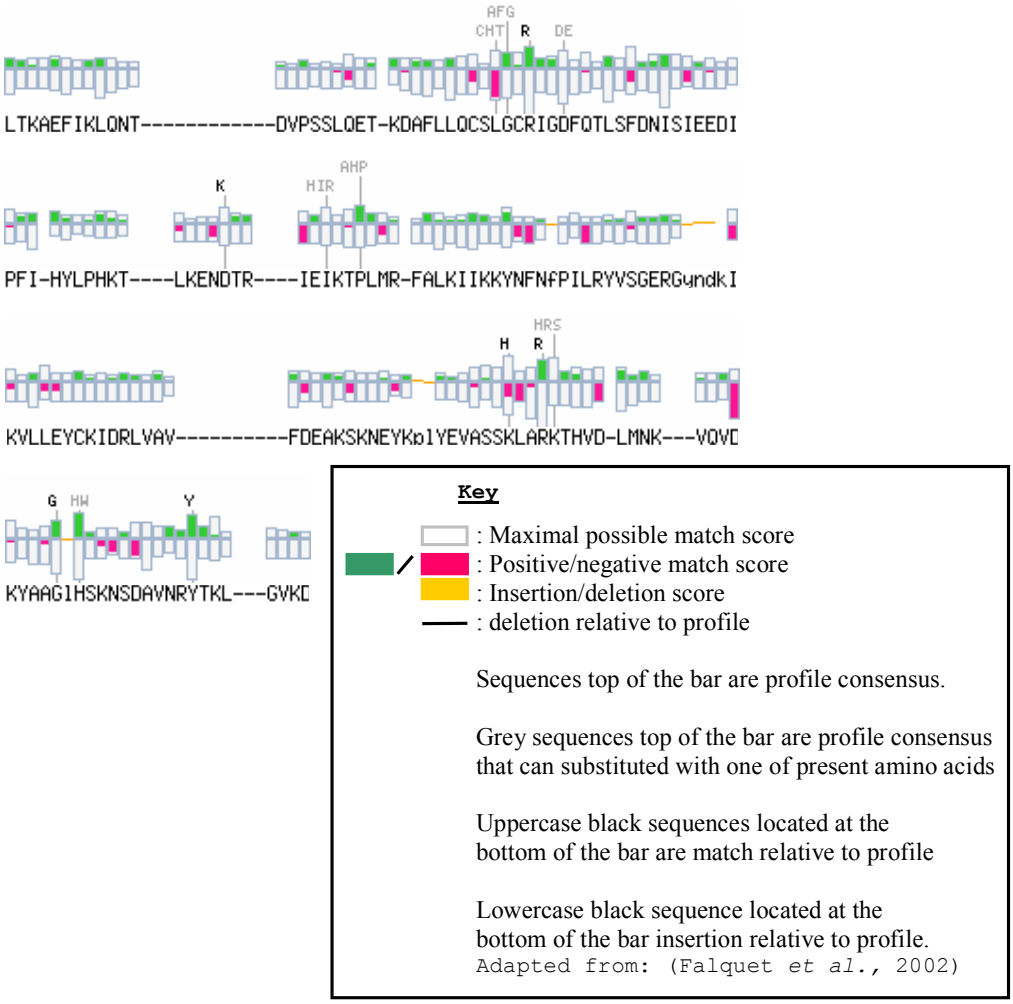


Figure 4.7: Motifs in ORF35 identified by MotifScan (Falquet *et al.*, 2002).

Multiple alignment was constructed, aligning the C-terminus of the ϕAR29 integrase to other integrase, recombinase, or transposase enzymes found in *Bacteroides uniformis*, *B. thetaiotaomicron* VPI-5482, *Porphyromonas gingivalis* W83, *Clostridium acetobutylicum* ATCC 824, *Lactobacillus gasseri*, Lambda and Cre recombinase. The

consensus of the tyrosine recombinase family Arg(212)-Lys(235)-His(308)-Arg(311)-His(333)-Tyr(342) in phage λ aligned closely with the ϕAR29 integrase, with the exception of a conservative substitution of His with Lys (Figure 4.8).

Table 4.13: Integrase and Transposases from bacteria and phage that were used in the alignment of ORF35.

Protein	Access No.	Reference:
Bacteroides uniformis (IntN1)	AAF74437.1	(Shoemaker *et al.*, 1996)
Bacteroides thetaiotaomicron VPI-5482 (Int)	NP_812390	(Xu *et al.*, 2003)
Bacteroides thetaiotaomicron VPI-5482 (Int)	NP_811180	(Xu *et al.*, 2003)
Bacteroides thetaiotaomicron VPI-5482 (Tn)	NP_813650	(Xu *et al.*, 2003)
Porphyromonas gingivalis W83 (Int)	NP_905090	(Nelson *et al.*, 2003)
Bacteroides thetaiotaomicron VPI-5482 (Tn)	NP_808989	(Xu *et al.*, 2003)
Clustridium acetobutylicum ATCC824 (Int)	NP_347802	(Nolling *et al.*, 2001)
Lactobacillus gasseri(Int)	ZP_00046094	Unpublished
Bacteroides thetaiotaomicron VPI-5482 (Tn)	NP_811190	(Xu *et al.*, 2003)
Bacteroides thetaiotaomicron VPI-5482 (Int)	CAC47923	(Xu *et al.*, 2003)
Lambda (Int)	P03700	(Sanger *et al.*, 1982)
Cre Enterobacteria	YP_006472.1	(Lobocka *et al.*, 2004)

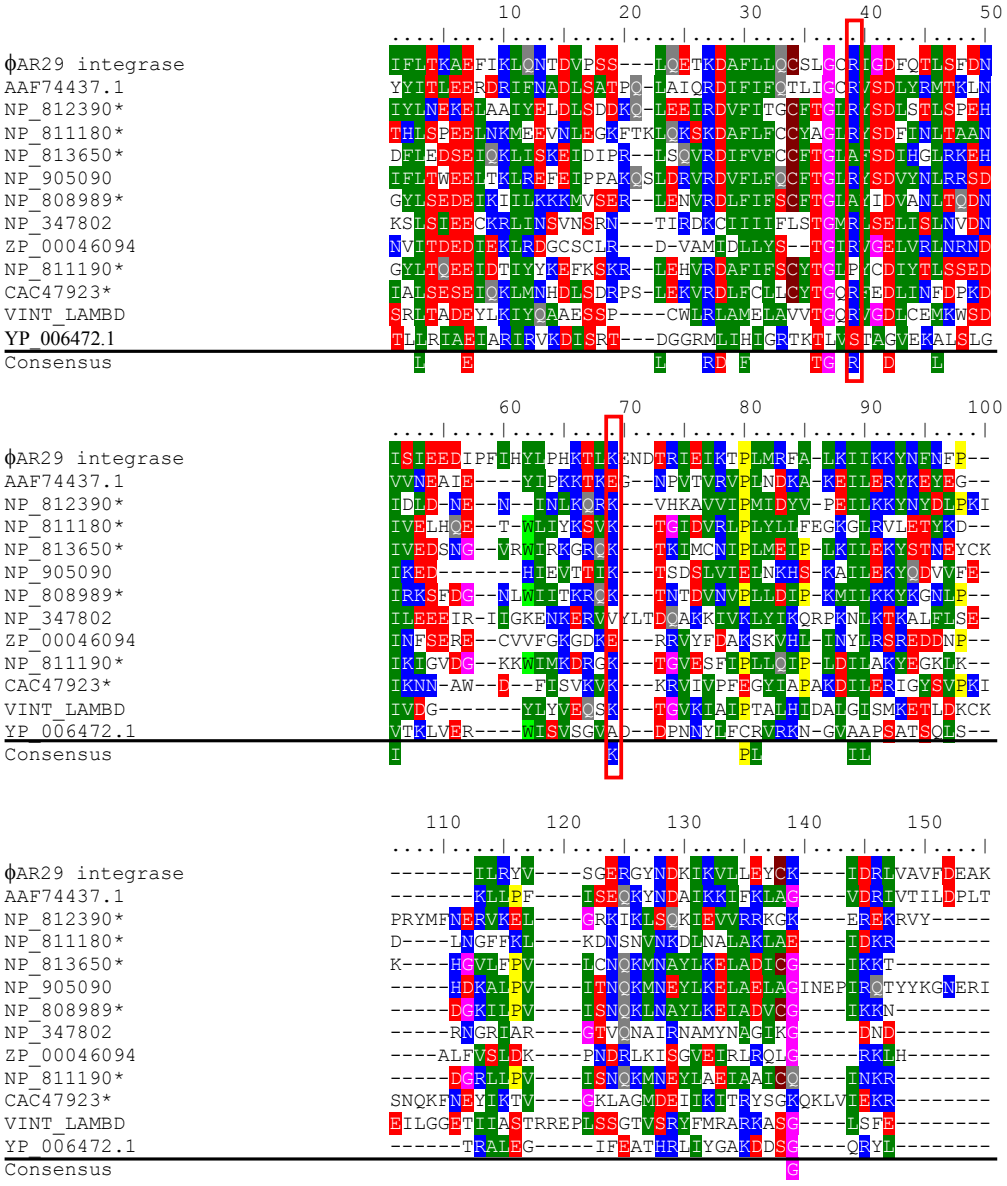


Figure 4.8: Alignment of ϕAR29 integrase to other integrease, recombinase and transposase enzymes. The red boxed residues are the R-K-H-R-H-Y consensus sequence found at the C terminus of the tyrosine integrase family.

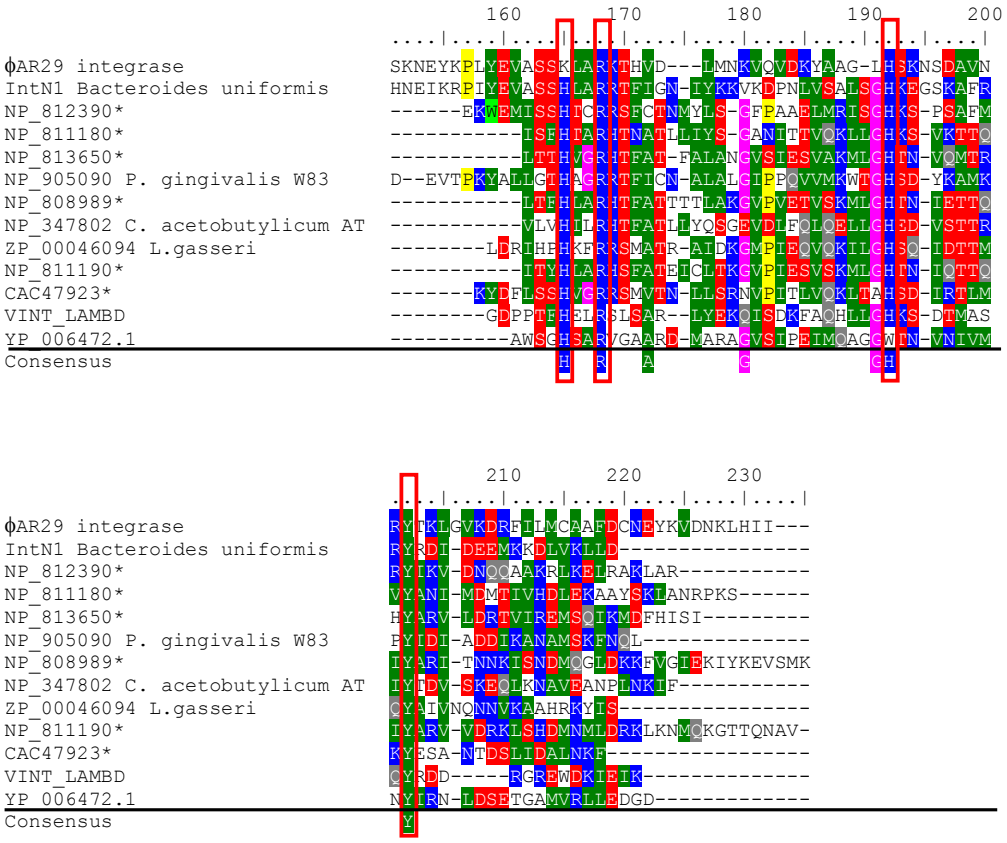


Figure 4.8 *(continued)***:** Alignment of ϕAR29 integrase to other integrease, recombinase and transposase enzymes. The red boxed residues are the R-K-H-R-H-Y consensus sequence found at the C terminus of the tyrosine integrase family.

4.3.3.6: ORF40 (N-acetylmuramoyl-L-alanine amidase)

The notional product of ORF40 contains an N-acetylmuramoyl-L-alanine amidase domain (Cheng *et al.*, 1994; Potvin *et al.*, 1988; Wang *et al.*, 1991). The conserved motifs were found distributed at amino acid positions 2 to 161 (Figure 4.9). Amidases are proteins that are capable of breaking down cell walls and are used by phage to release progeny during the late phase of the lytic cycle. Alignment of this deduced amino acid sequence to the lysozymes and amidases of T7 (Dunn and Studier, 1983), Phi A1122 (Garcia *et al.*, 2003), gh-1 (Kovalyova and Kropinski, 2003), phiYeO3-12 (Pajunen *et al.*, 2001), and T3 (Pajunen *et al.*, 2002), revealed 36% - 37% identical matches. When ORF40 was aligned to amidases found in *B. thetaiotaomicron* VPI-5482 (Xu *et al.*, 2003), the percentage of identical matches increased to a range

from 42% - 62% (Table 4.14). Multiple alignment of phage and *Bacteroides* amidases showed that ϕAR29 contains 6 motifs that are unique to bacteriophages, but also contains 16 motifs that are found in *Bacteroides* (Figure 4.10). Phylogenetic tree created from the multiple alignments clear illustrates ϕAR29 amidase is closely related to the amidase derived from the *Bacteroides* (Figure 4.11).

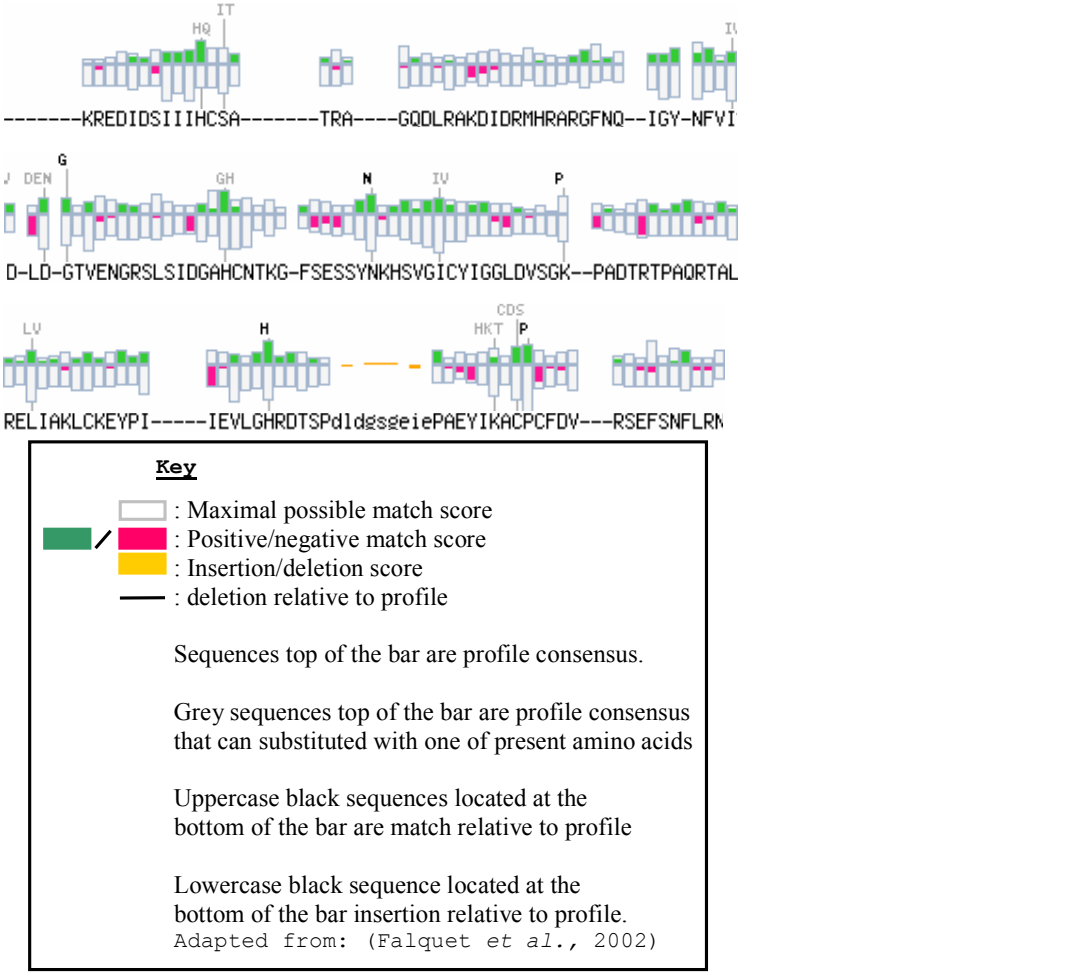


Figure 4.9: Motifs identified in ORF40 by MotifScan (Falquet *et al.*, 2002).

Table 4.14: The p-distance, standard errors (S.E) and pairwise alignment of ϕAR29 amidase with closely related N-acetylmuramoyl-L-alanine amidases found in phages (red) and bacteria (black).

Organism	p-Distance	S.E.	Identity(%)	Similarity(%)	Gaps(%)	E-value
T7 (P00806: Dunn and Studier, 1983)*	0.607	0.042	36	55	13	8e-22
Phi A1122 (NP_848277.1: Garcia *et al.*, 2003)*	0.600	0.042	37	54	13	7e-22
gh-1 (NP_813758.1: Kovalyova and Kropinski, 2003)*	0.607	0.042	36	55	13	1e-21

phiYeO3-12 (NP_052084.1: Pajunen *et al.*, 2001)*	0.600	0.042	37	54	14	2e-21
T3 (NP_523313.1: Pajunen *et al.*, 2002)*	0.615	0.042	36	54	14	7e-21
B. thetaiotaomicron VPI5482 (NP_810940: Xu *et al.*, 2003)	0.533	0.043	62	76	5	1e-48
B. thetaiotaomicron VPI5482 (NP_812112: Xu *et al.*, 2003) Putative amidase	0.533	0.043	52	65	4	2e-38
B. thetaiotaomicron VPI5482 (NP_810081: Xu *et al.*, 2003)	0.533	0.043	50	66	4	8e-37
B. thetaiotaomicron VPI5482 (NP_810904: Xu *et al.*, 2003)	0.533	0.043	54	67	4	2e-36
B. thetaiotaomicron VPI5482 (NP_810432: Xu *et al.*, 2003)	0.533	0.043	50	65	4	4e-35
B. thetaiotaomicron VPI5482 (NP_813647: Xu *et al.*, 2003)	0.533	0.043	49	62	4	3e-33
B. thetaiotaomicron VPI5482 (NP_812942: Xu *et al.*, 2003)	0.533	0.043	44	63	4	4e-32
B. thetaiotaomicron VPI5482 (NP_809619: Xu *et al.*, 2003) Putative amidase	0.533	0.043	53	65	7	2e-31
B. thetaiotaomicron VPI5482 (NP_810619: Xu *et al.*, 2003) Putative amidase	0.533	0.043	43	54	13	2e-23
B. thetaiotaomicron VPI5482 (NP_810554: Xu *et al.*, 2003)	0.533	0.043	42	54	13	4e-23

Bacteriophage lysozyme/amidase is indicated by "*".

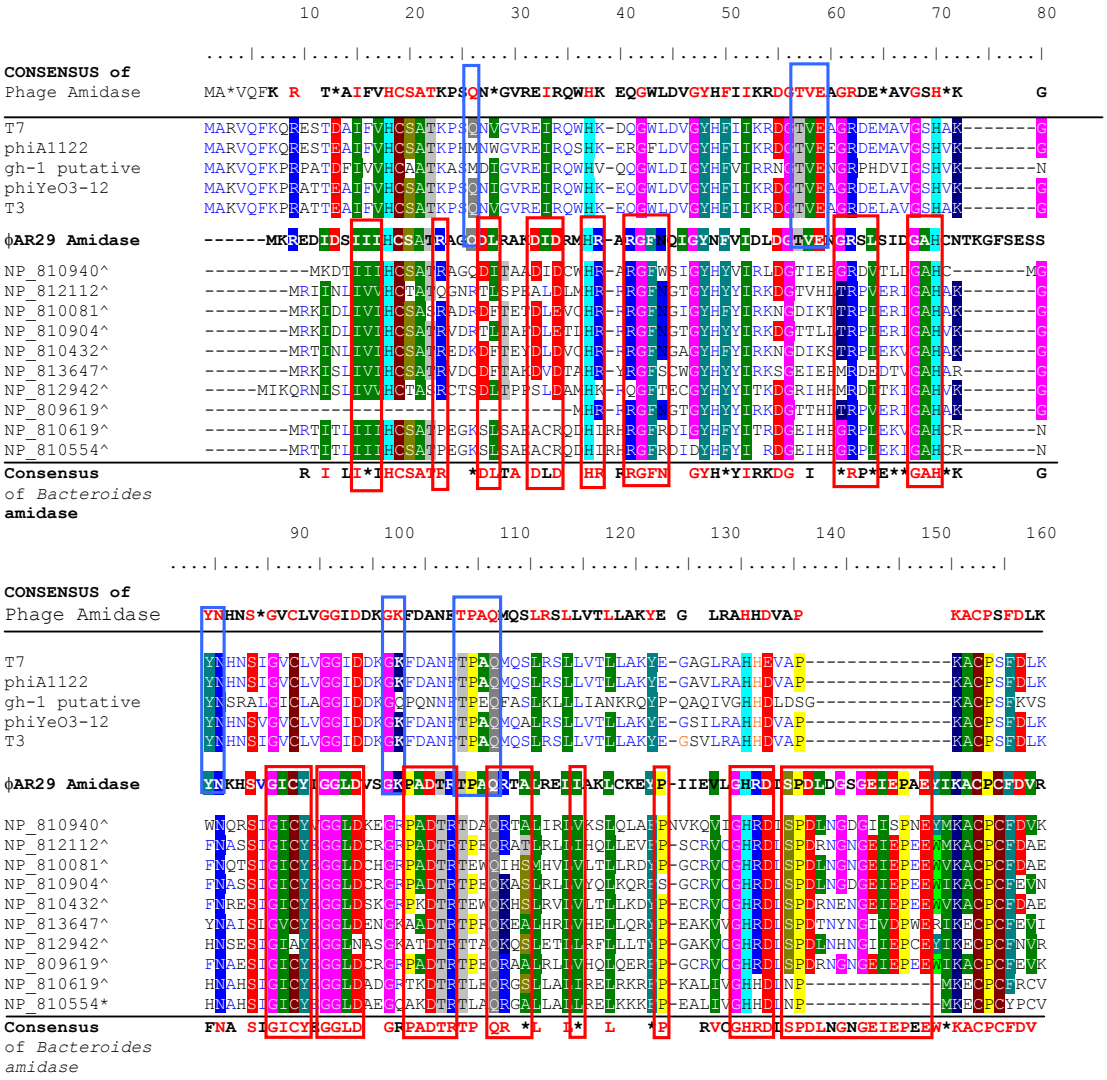


Figure 4.10: Alignment of the putative amidase from ϕAR29 with amidases from other bacteriophages and *B. thetaiotaomicron* VPI-5482. Blue sequences in the alignment indicate identical matches to consensus sequence. Residues with coloured background reveal homology or similarity matches between phage or VPI-5482 amidase to that of ϕAR29. Red amino acid letters in the consensus indicate conservation of sequence between the consensus and ϕAR29 amidase. Substitution of an amino acid is represented as "*" in the consensus. Blue and red boxed regions indicate conserved motifs in bacteriophage and *Bacteroides* integrases respectively. "^" indicates that the amidase is from *B. thetaiotaomicron* VPI-5482.

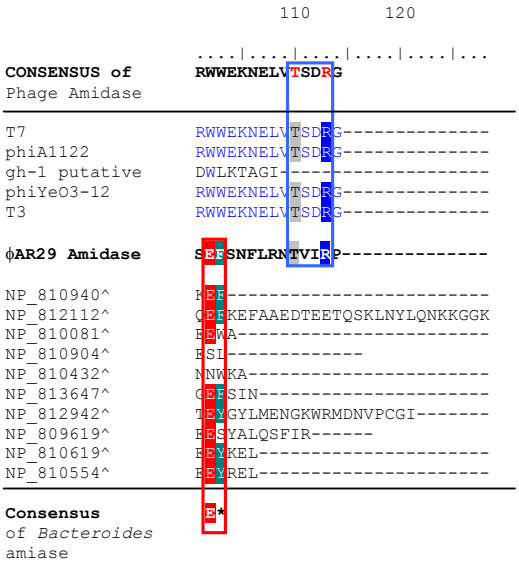


Figure 4.10 *(continued)*: Alignment of Amidase from ϕAR29 with amidases from other bacteriophages and *B. thetaiotaomicron* VPI-5482. Blue sequences in the alignment indicate identical matches to consensus sequence. Residues with coloured background reveal homology or similarity matches between phage or VPI-5482 amidase to that of ϕAR29. Red amino acid letters in the consensus indicate conservation of sequence between the consensus and ϕAR29 amidase. Substitution of an amino acid is represented as "*" in the consensus. Blue and red boxed regions indicate conserved motifs in bacteriophage and *Bacteroides* integrases respectively. "^" indicates that the amidase is from *B. thetaiotaomicron* VPI-5482.

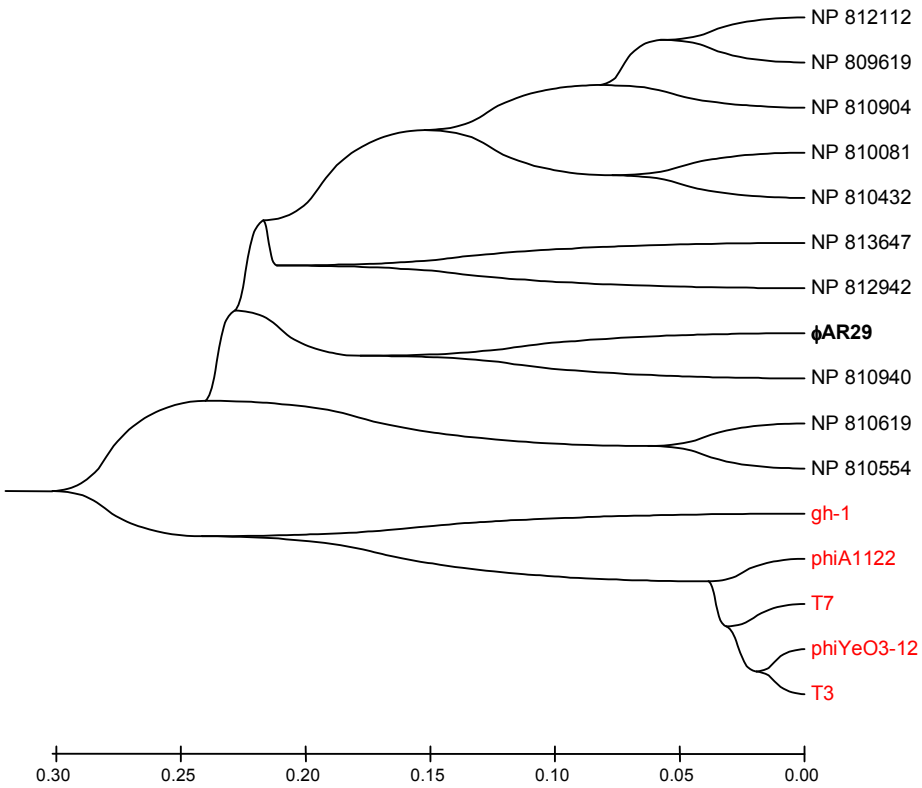


Figure 4.11: Tree showing the relationship of ϕAR29 amidase to those from *Bacteroides* (in black) and those from bacteriophage (red).

4.3.4: ORFs with significant matches to known genes and proteins

4.3.4.1: ORF4

Interpro scan on EBI revealed that ORF 4 contains a putative DNA binding domain (Accessioin:IPR009061) with a conserved structure that consists of a three-helical fold. DNA binding domains can be found in several different protein families, including exision repair factor (XPA) (Buchko *et al.,* 1999) and the gpNU1 subunit of lambda terminase (de Beer *et al.,* 2002). The 5'end of ORF4 overlaps the 3' of the terminase (ORF 5) by 17 bases (Figure 4.12). The presence of conserved domains and the proximity to ORF5 suggest ORF4 may encode the small subunit of the terminase.

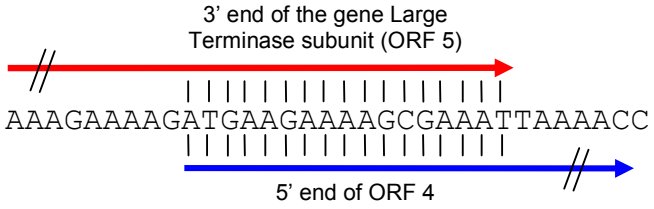


Figure 4.12: the 3′ end of Terminase subunit (red arrow) overlaps the 5′ end of ORF 4 (blue arrow) by 17 bases (vertical bars between sequence and arrow).

4.3.4.2: ORF6

The product deduced from ORF 6 is very simlar to two proteins of unknown function from *Enterococcus facecalis* V583 (NP_815687.1) and bacteriophage Tuc2009 (NP_108707.7), with and E-value of 3e-05 and 0.012. Their respective identities are 26 % (over 123 aa) and 26% (over 153 aa).

4.3.4.3: ORF17

Results from a BLAST analysis of ORF 17 showed similarities to the RecB protein from *Clostridium thermocellum* ATCC 27405 (ZP_00311700.1). The two proteins share 58 % identity at the C-terminus with an E-value of 3e-20. The search also

showed a 44 % identity over 107 aa and 48 % identity over of 62 aa to prophage Lp2 protein 20 (Kleerebezem *et al.*, 2003) and bacteriophage T5 D14 protein (Kaliman *et al.*, 1988b; Kaliman *et al.*, 1988a), respectively . More interestingly, the BLAST analysis resulted in pairwise alignment to the dnaD protein from *Enterococcus faecalis* (Paulsen *et al.*, 2003). The dnaD protein was similar in size and alignment showed 20 % identity over 238 aa. In *Bacillus subtilis* dnaD is thought to be involved in helicase (dnaB) loading during the process of restarting stalled replication forks (Bruand *et al.,* 2005).

4.3.4.4: ORF18

ORF18 has 40% identity over 115 aa to a hypothetical protein derived from *Polaromonas* sp. JS666 (ZP_00363673.1) with an E-value of 2e-15. Protein pmgM of phage P1 (Lobocka *et al.*, 2004) and an unknown gene from phage T1 (Roberts *et al.*, 2004) were also nominated as possible matches by the BLAST analysis. The alignment produced respective identities of 31 % (over 104 aa, E=9e-07) and 28 % (over 120 aa, E=2e-04). The GenBank has listed a putative morphorgenic function for pmgM protein from phage P1.

4.3.4.5: ORF22

ORF22 showed 33% identity over 153 aa to a hypothetical protein from *Mesorhizobium* sp.BNC1 MBNC02003433. Interpro scan on EBI revealed no conserved domains. However, MEBNC02003433, was found to contain a high mobility group (HMG) box domain at its C-terminus. Proteins possessing an HMG-box belong to a large, diverse family that participate in the regulation of DNA-dependent processes that require the bending and unwinding of chromatin, such as transcription, replication, and strand repair (Zdobnov and Apweiler, 2001; Bustin, 1999). Many of these proteins are regulators of gene expression (Bustin, 1999).

4.3.4.6: ORF 23

A domain search on ORF 23 found that the C-terminus of the encoded protein aligned to a conserved domain found at the N-terminus of a site-specific integrase-resolvase (Score of 41 and E-value of 4e-05: Marchler-Bauer and Bryant, 2004). However, BLAST analysis of the protein revealed alignment with many unidentified hypothetical proteins of similar size from the genomic sequence of *Bacteroides thetaiotaomicron* VPI-5482 (Xu *et al.*, 2003). The best match from the study was to a protein tagged Orf2x. Pairwise alignment of ORF 23 with Orf2x showed them to share 37% identity from amino-acid positions 4 to 97 with an E-value of 2e-09. Orf2x is known to be involved in the excision process of non-replicating *Bacteroides* unit 1 (NBU 1:Shoemaker *et al.*, 2000:Nadja Shoemaker, personal comm.). Interestingly, the BLAST studies revealed an alignment to a similar length portion of a putative DNA binding protein belonging to the excisionase family, which was found in the genomic sequence of *Porphyromonas gingivalis* W83 (length 102 aa, 25 % identity, E-value of 0.19: Nelson *et al.,* 2003). Similarities to excision related proteins suggest that the protein product of ORF23 may be required for the excision process of ϕAR29.

4.3.4.7:ORF51

BLAST analysis of ORF 51 suggested that the gene may be a tail tape measure protein. Identities of 22% (over 135 aa) and 20% (over 143 aa) were found in tail length tape measure proteins of phages WPhi and phi-13 (Iandolo *et al.*, 2002), respectively.

4.3.5: Poorly Matched ORFs

Forty open reading frames were defined as "poorly matched" i.e. they have an E-value of greater than 1 and their putative function was not identifiable from the BLAST searches**.**

4.3.6: tRNA Gene Prediction

tRNAscan-SE was able to identify a leucine tRNA (anti-codon: AAG) on the negative strand of the ϕAR29 genome (Figure 4.13). The program also noted the tRNA as a pseudo gene. Interestingly, this gene contains an intron of 41 bp, commencing at position 16948 (Table 4.15).

Table 4.15: Characteristic of a pseudo-tRNA gene identified within ϕAR29 genome

Strand	tRNA	AntiCondon	Start	End	Intron start	Intron end	tRNA type
-	Leu	AAG	17028	16914	16989	16948	Pseudo

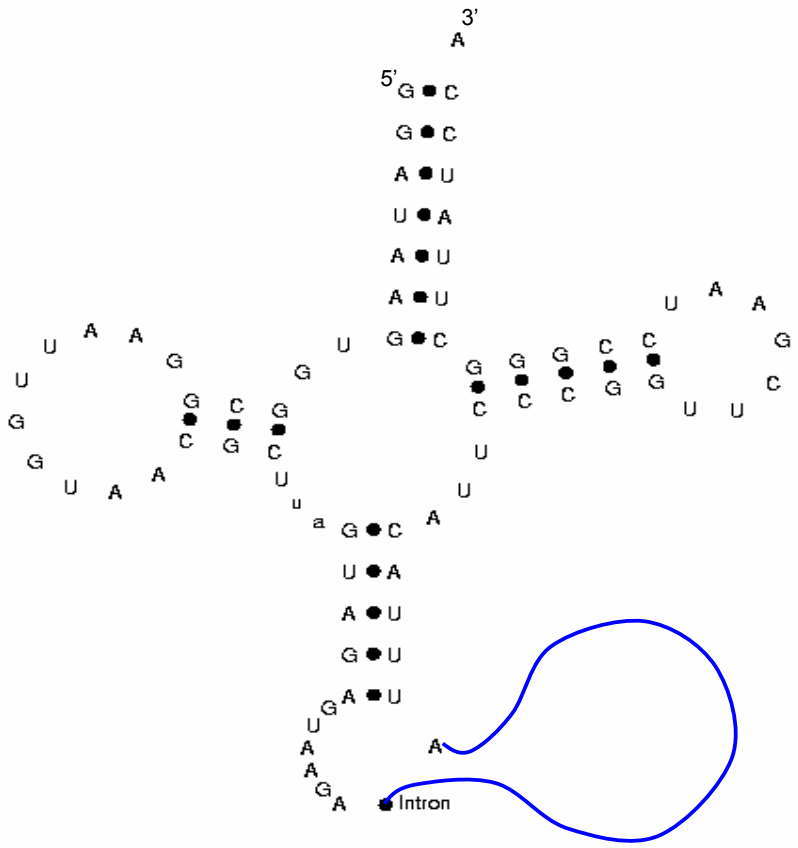


Figure 4.13: 2D structure of pseudo-leucine tRNA (AAG) present in the ϕAR29 genome.

4.3.7: Frameshift Analysis

The programme "FSFinder" did not identify any -1 or +1 frameshifts within ϕAR29 genome. However, when the genome was analyzed using "Programmed Frameshift Finder" software, two -1 frameshifts were found (Table: 4.16). Removal of the -1 frame "slippery" region and re-analysing did not result in combining of the adjacent reading frames by the GeneMark Heuristic model.

Table 4.16: The result of frameshift analysis using "Programmed Frameshift Finder".

Location	+/- Strand	Position	slippery sequence	Start of the first ORF	End of the first ORF	End of the second ORF	A number of Overlap Nucleotides (bases)
Crossover junction between ORF 6 and 5	-	5674	GGGGGAA	(ORF6) 6180	5650	(ORF5) 4475	22
Crossover junction between ORF 46 and 45	-	26993	TTTAAAA	(ORF46) 27508	26981	(ORF45) 26454	5

4.3.8: Regulatory Elements in ϕAR29 Genome.

The computer program PHIRE version 1.0 was applied, to determine the presence of regulatory elements within the ϕAR29 genome. A summary of the output is presented in Table 4.17. The majority of the sequences identified by the program did not have assigned functions. The absence of palindromes among the sequences made it unlikely that they included any transcription terminators. Interestingly, a potential promoter at position 18232-18213 has exactly the same as sequence as the opposite strand at 18215 -18234 (Figure 4.14A).

Software and visual analysis predicted 4 potential promoter regions (Table 4.18). Conserved *Bacteroides* motifs, -7 (TAnnTTTG) and -33 (TTTG), were found in these regions. Sequence from 18193-18316 contains two -7 and -33 promoter domains, apparently initiating transcription in opposite directions. Regulatory sequences

predicted by PHIRE were found in or in close proximity to the potential promoter regions predicted by other means.

Bionavigator GCG Terminator software predicted 7 rho-independent transcription terminators (Table 4.17). The majority were found at the proposed ends of the phage cycle regions (e.g lytic, lysogenic and replication). Terminators 4 and 5 are complementary sequences and were found at the end of the lytic and lysogenic part of the genome (Figure 4.14B).

Table 4.17: PHIRE prediction for potential regulator sites (promoters, ori or rho-independent terminators) and Bionavigator terminator prediction results in the bacteriophage ϕAR29 genome. Sequences with unknown function requires future work to clarify their role.

	Location	Sequence	Strand (+/-)	Function
	18245-18264	**TACATAATAATTTATGTAT**T	+	Potential promoter
	18215-18234	**TACATAATAT**TTTATGTAGA	+	Potential promoter
	18298-18317	CA**CATAATAATTTATGTATA**	+	Potential promoter
	22224-22243	**TA**G**ATAATAA**G**TTA**CA**TAT**T	+	Potential promoter
	20766-20747	**TA**TC**T**G**ATAATT**A**ATGTAT**T	-	Unknown
	18232-18213	**TACATAA**A**A**TA**TTATGTAT**C	-	Potential promoter
Consensus		**TA**c**ATAATA**at**TTATGTAT**n		
	17760-17741	**AGAAACGAAAGAAGAAATTA**	-	Unknown
	05696-05677	**AGAAACG**G**AAGA**T**GAAA**GG**A**	-	Unknown
	05900-05881	**AGAAA**GT**AAAGAAGAAA**A**TA**	-	Unknown
	15986-15967	**AAAAAC**A**AAAGAAGAA**G**TT**G	-	Unknown
	24727-24708	G**GAAACGA**G**AGA**C**GA**T**ATTA**	-	Unknown
Consensus		**AGAAAC**g**AAAGA**a**GAAA**t**TA**		
	30016-29997	**AAAG**A**GGAAA**T**CGT**A**GAAAC**	-	Unknown
	00706-00687	**AAA**AA**GGAAA**T**CTT**C**GAA**TC	-	Unknown
	12194-12175	**A**TA**GG**G**GAAA**C**CGT**AA**AAC**	-	Unknown
	32888-32869	G**AAG**C**GGAAA**G**CGT**G**GAAAC**	-	Unknown
Consensus		**AA**a**G**n**GGAAA**n**CGT**n**GAAAC**		
	34927-34908	**AAAACA**G**CAACATTAAAA**T**T**	-	Unknown
	08251-08232	**AAAACA**A**CAACA**G**TAAAA**CC	-	Unknown
	08464-08445	**AAAA**ATG**CAA**A**AT**C**AAAA**T**T**	-	Unknown
	09677-09658	**AAAACA**A**CAACATT**G**AAA**AT	-	Unknown
Consensus		**AAAACA**n**CAACATT**n**AAA**n**T**		
				Type of terminator
Terminators 1	8748-8730	**AT**A**ATGT**CATCA**ACAT**A**AT**	-	I-shaped
2	11260-11241	**TC**GC**ATAG**TGGT**CTAT**CA**GA**	-	L-shaped
3	20555-20585	**ATAA**T**AGCCCGT**CTAAAAA**AC GGGCT**T**TTAT**	+	I-shaped
4	22721-22747	*__AAGAGGTAGCT__TATTC__AGCT ACCTCTT__	+	X-shaped
5	22747-22721	*__AAGAGGTAGCT__GAATA__AGCT ACCTCTT__	-	X-shaped
6	22778-22793	**AG**A**AAG**TAAA**CTT**C**CT**	+	L-shaped
7	22959-22943	**AGA**A**TA**GGTAC**TA**A**TCT**	-	I-shaped

Note: "*"indicates complementary strands of the same sequence.

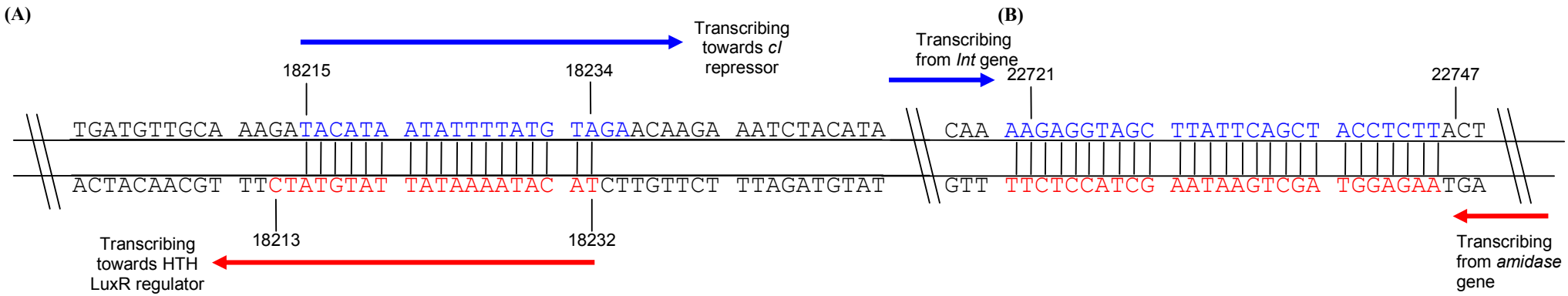


Figure 4.14: Overlapping regulator elements (A) and terminators (B) determined by PHIRE and Terminator GCG, respectively. Blue and red nucleotide sequences present sense and anti-sense strands, respectively. Blue and red arrows indicate the direction of transcription on sense and anti-sense strands, respectively. Bar lines in between nucleotide sequences reveal overlapping regions.

Table 4.18: Promoter regions predicted by Neural Network Promoter Prediction software (NNPP) and presence of *Bacteroides* -33 and -7 motifs within the predicted regions.

promoter region predicted by programme NNPP	Potential *Bacteroides* promoter -33(TTTG) and -7 (TAnnTTTG) consensus in bold font.	Comments
8604-8819	**TTTG**AAAAAAAACAGTCTTAAATT**AGTATTTG**TA CAATTAATAATTATATATTTGCAAACGTTTTTAA GCCCTAACAAGGCAACTAAAAGAAAT	Positioned at 8729-8573 the sequence is located directly upstream of ORF 10
18193-18316	**TTTG**CAACATCAAACAACATCCAACACTGCAAAG GTGC**GAAGTTTG**AGTGAGAAAACCAAATATTTTA CATAACTAAAAATAGGTAAGACA	Positioned at 18214-18122 the sequence is located directly upstream of ORF30
	TTTGGTTTTCTCACTCAAACTTCG**CACCTTTG**CA GTGTTGGATGTTGTTTGATGTTGCAAAGATACAT AATATTTTATGTAGAACAAGAAATCTACATAATA ATTTATGTATTTAACTTTTATTTTCAATTAAACG CCGGATAAATCACATAATAATTTATGTATATATA	Positioned at 18152-18321 the sequence is located directly upstream of the cI gene (ORF 31).
20613-20833	**TGTTG**CATTTTAGTTGTGCAATTAAG**CATTTTCG** TTGTGCAAAAACAGTATATTTGCACAACCGTAAA ATGATAGCTAT	Positioned at 20791-20869 the sequence is located directly upstream of the integrase gene (ORF 35)

4.3.9: Comparison of ϕAR29 gene arrangement with other bacteriophages

The genomic map of ϕAR29 is illustrated in Figure 4.15. The assignment of probable functions to ORFs and the identification of likely promoter regions allows the phage genome to be classified into three sections: lysogenic, replicative, and lytic. The genomic arrangement was compared to six other phages. A summary of genome arrangements seen in other phages is shown in Figure 4.16. The genome of ϕAR29 has similar arrangement to those of phages Sfi21, U136 and SM1. Lambda and P22 phages differ in that their lysis genes are located upstream of the terminase. In ϕAR29 these genes are located downstream of the terminase.

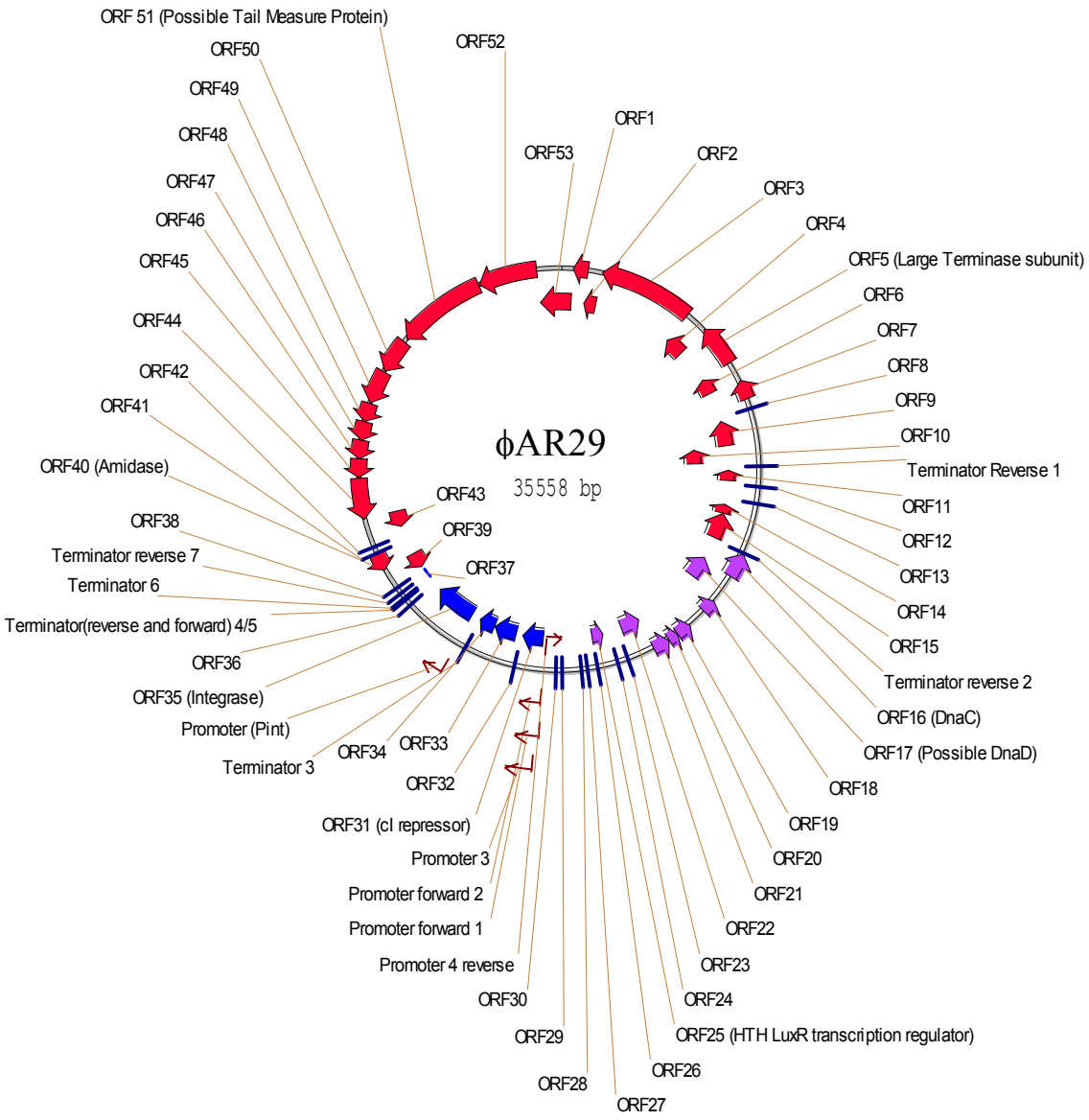


Figure 4.15: ϕAR29 genomic map. Blue, purple and red arrows indicate ORFs involved in lysogenic, replication and lytic cycle respectively. Lytic ORFs may include structural, packaging and lytic proteins.

Bacteriophage	Genome Organization				
Lambda	Integrase	Replication	Lysis	Terminase	Tail protein
P22	Integrase	Replication	Lysis	Terminase	tail protein
Sfi21	Integrase	Replication	Terminase	tail protein	Lysis
Ul36	Integrase	Replication	Terminase	tail protein	Lysis
SM1	Integrase	Replication	Terminase	tail protein	Lysis
ϕAR29	Integrase	Replication	Terminase	tail protein	Lysis

Figure 4.16: Comparison of genomic arrangement of ϕAR29 with other bacteriophages. The arrows indicate the direction of transcription.

4.3.10: Locating and Identifying the Integration Site.

The *attP* site of ϕAR29 was located by sequencing the bacterial genome, using primer Sbgatt2, and marking the change from phage sequence to bacterial genome sequence. The result was confirmed with forward and reverse primers Sbgatt3. The site on the bacterial genome at which ϕAR29 integrates into its host was identified by BLAST analysis of the attB region against the GenBank. The search resulted in a 100% match to the arginine-tRNA molecule found in *B. thetaiotaomicron* VPI-5482 (AE016934).

Alignment of phage sequence with left and right flanking sequences of the *att* site (*attL and attR*) revealed a 16 bases *att* core region (ATGTGACCCCGGTGCG) where genetic material is exchanged (Figure 4.17). The *att* core is located 162 bases upstream of the start codon of the integrase gene on the phage genome and at the tail of the Arg-tRNA on the bacterial host genome (Figure 4.18).

Surrounding the *attP* core region, sequence motif imperfect repeats were found to occur 7 times on the sense strand with 3 imperfect copies of the same repeat on the anti-sense strand. The consensus sequence of these repeats is ANGTTGTGCAA (Figure

4.19). A 4-base inverted repeat was present at the *att* core site. These repeated sequences were not present in the *attB* region.

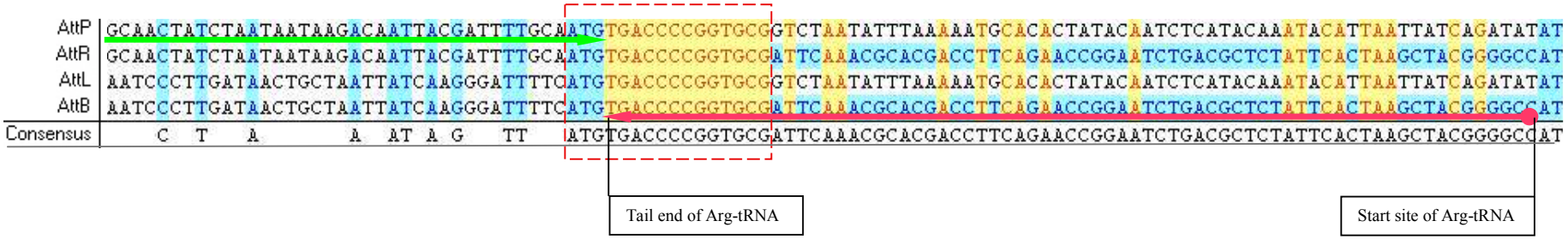


Figure 4.17: Alignment of *AttP*, *AttR*, *AttL*, and *AttB* against the arginine tRNA gene. The green and red arrows indicate the direction of transcription at the *Att*P site and of the arginine tRNA, respectively. The red broken box surrounds the 16 bases core attachment sequence.

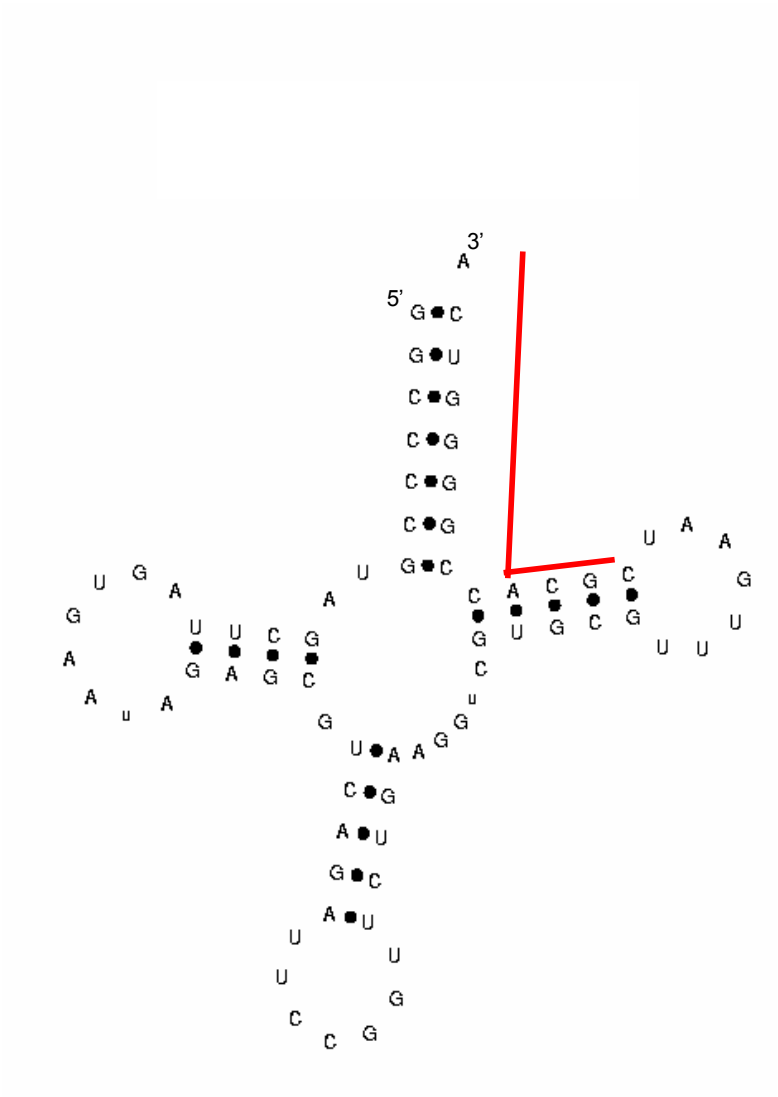
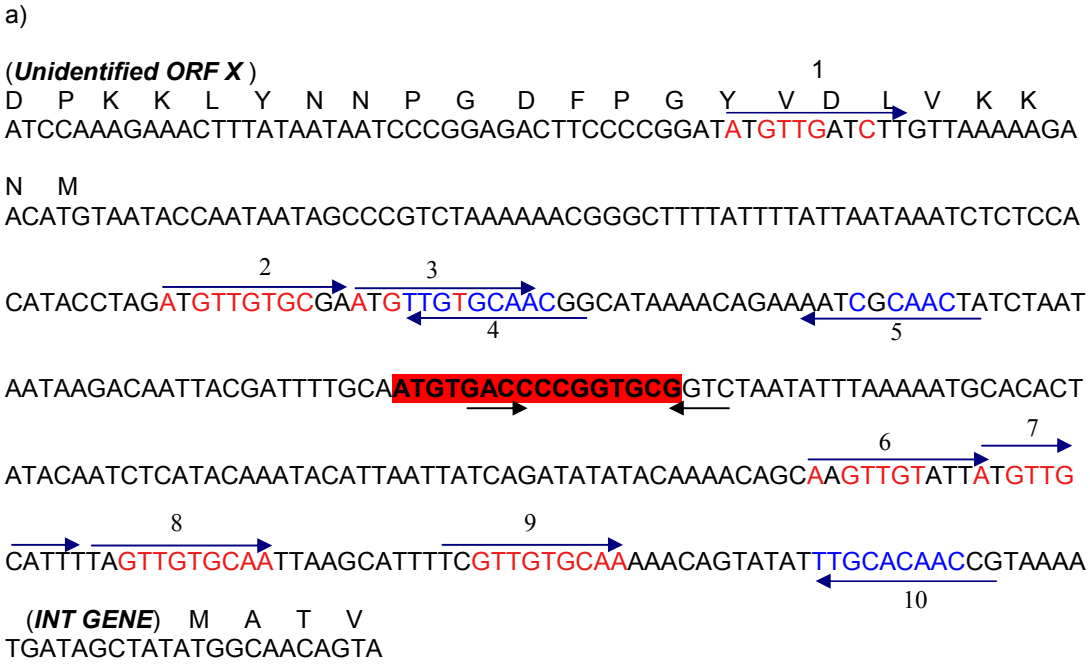


Figure 4.18: ϕAR29 integrates into the 3' end of $tRNA^{arg}$, as indicated by the red line.



(b)

```
1    ATGTTGATCTT
2    ATGTTGTGCGA
3    ATGTTGTGCAA
4    CCGTTGCACAA
5    TAGTTGCGATT
6    AAGTTGTATTA
7    ATGTTGCATTT
8    TAGTTGTGCAA
9    TCGTTGTGCAA
10   CGGTTGTGCAA
Consensus:  ANGTTGTGCAA
```

Figure 4.19: Characterization of the ϕAR29 *att* site. (a) DNA sequence of *att*P. The amino acid sequence indicates the 3′ end of an unidentified ORF and 5′ end of the integrase gene. Highlighted red sequence is the *att* core. The red and blue sequences are direct and inverted repeats, respectively. Arrows indicate the direction of the repeat. (b) Alignment of all the perfect and imperfect repeats from *attP* created a consensus of ANGTTGTCAA. The cyan and yellow highlighted sequences indicate partially and completely conserved regions, respectively.

4.3.11: Identifying the Site and Orientation of Integration of ϕAR29 into *B. thetaiotaomicron* AR29

To provide an understanding of the region of bacterial chromosome surrounding the arg-tRNA gene, a BLAST search was performed for the *attB* sequence. The search revealed 96% homology to sequences from *B. thetaiotaomicron* VPI-5482, with an E-value of 0. Nucleotide sequences 63 bases downstream and 220 upstream of the tRNA gene were identified as encoding conserved hypothetical kinase-like protein (AAO77235.1) and aspartate ammonia ligase (AAO77236.1), respectively.

Sequence analysis also determined that, following recombination, the ϕAR29 genome is placed downstream of the arg-tRNA gene with the ϕAR29 *cI* repressor located nearest to the *attR* site (Figure 4.20)**.** Genomic sequence of *B. thetaiotaomicron* VPI-5482 from the GenBank did not list any promoter sequence between the kinase like gene and arg-tRNA. However, the putative promoter in front of the *Int* gene of ϕAR29 is located 84 bases downstream of the core region found in *attL*. Therefore, transcription of *Int* from the integrated prophage may be regulated by its own promoter.

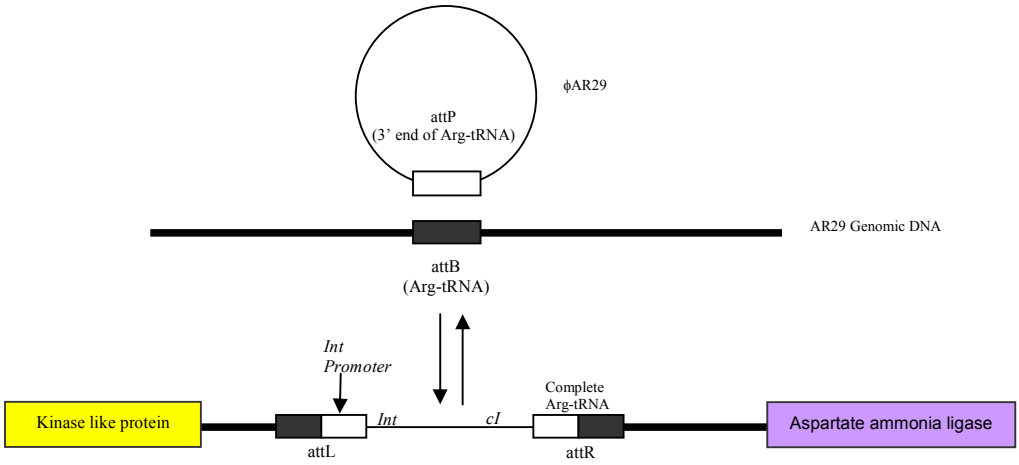


Figure 4.20: Diagrammatic illustration of the integration of ϕAR29 into AR29.

4.3.12: Amidase Activity Studies

No significant changes to PMC 112 growth were observed when the cloned ϕAR29 amidase-like gene was induced with IPTG (Figure 4.21). Following inoculation, all cultures, except those incubated in the presence of glucose (Flask 1), grew to similar optical densities in the first 3 hours. Flask 1 cultures with glucose produced the most rapid increase in cell density until reaching stationary phase at 6 hours. The lysozyme treated culture showed a lower absorbance reading at each time point in comparison to the other cultures. Cells that were induced with IPTG (Flask 3) did not show significant difference in cell density when compared to uninduced cultures (Flask 2) or cultures transformed with only the vector, pTrcHis A (Flask 5).

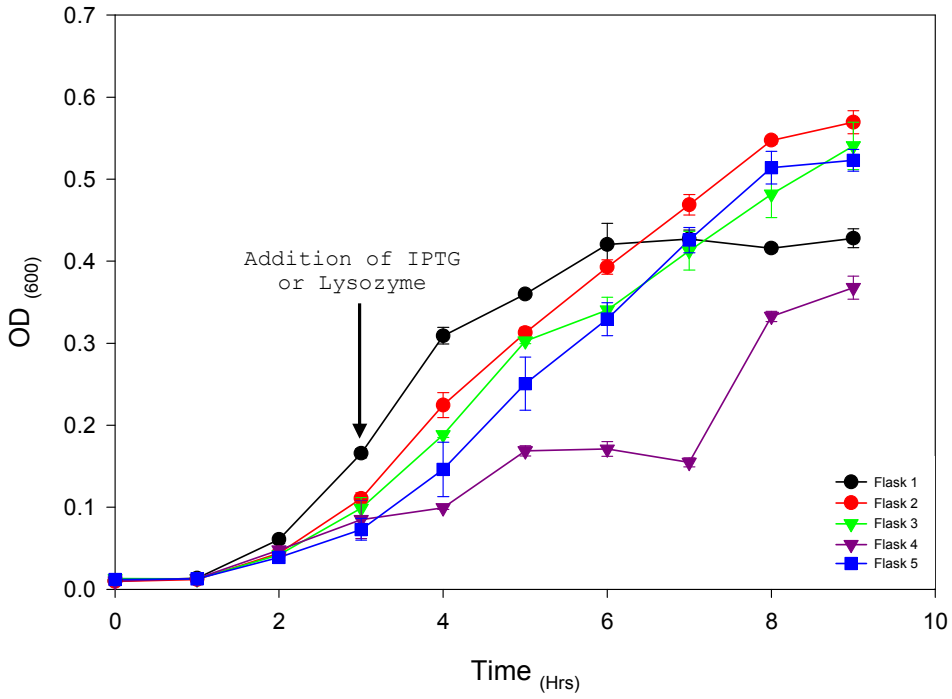


Figure 4.21: The effect of ϕAR29 amidase-like gene on *E.coli* PMC112 cell density over 9 hours.
Flask 1: pAmidase + Glucose – Lysozyme + addition of IPTG at 3 hours (Suppressed pAmidase: -ve control)
Flask 2: pAmidase – Glucose – Lysozyme – IPTG (uninduced pAmidase: +ve control)
Flask 3: pAmidase – Glucose – Lysozyme + addition of IPTG at 3 hours (Test)
Flask 4: pAmidase – Glucose + Lysozyme – IPTG (Lysozyme activity: +ve control)
Flask 5: pTrcHis – Glucose – Lysozyme + addition of IPTG at 3 hours (pTrcHis A construct: -ve Control)

Cells taken from the uninduced (Flask 2) and IPTG induced (Flask 3) cultures at 3 and 5 hours were used for Western blot analysis. The Western blot showed bands of approximately 22 kDa from all samples (Figure 4.22). This is in agreement with the theortical size of the amidase protein with a His tag attached. At 3 hours, both flask 2 and flask 3 cultures were predicted to produce protein bands of similar intensity, since neither was induced at this stage. Unexpectedly, at 3 hours, the flask 3 sample produced less protein than from Flask 2. At 5 hours, however, Flask 3 produced a markedly more intense protein band, confirming that the culture produced more protein following induction with IPTG.

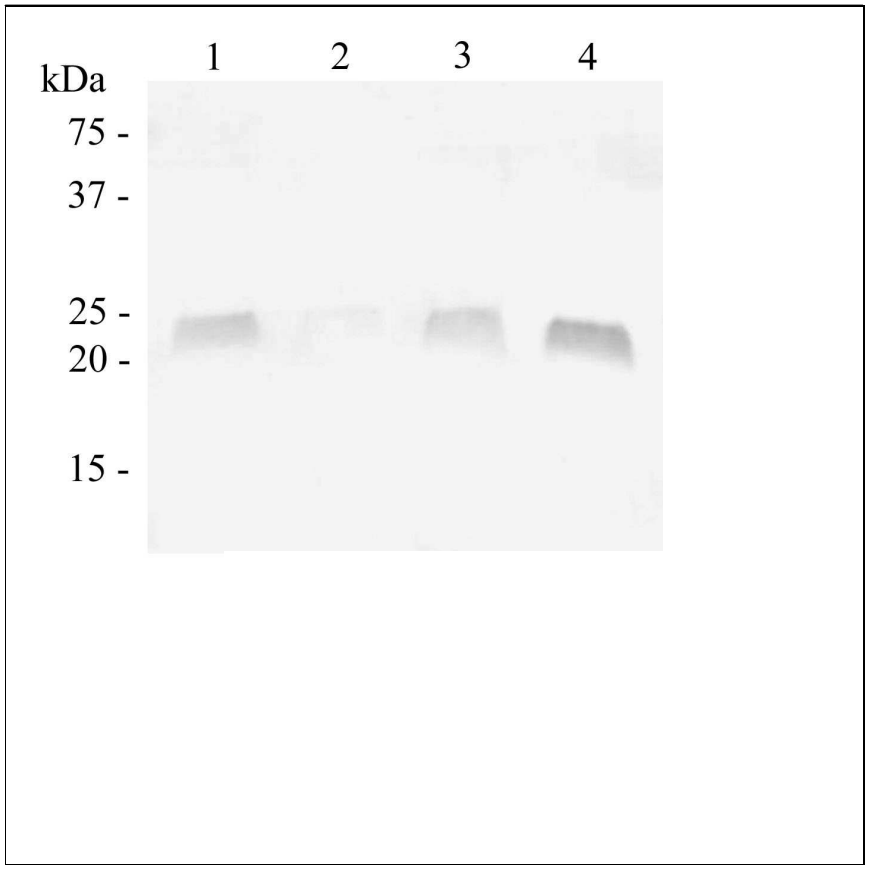


Figure 4.22: Western Blot of amidase protein produced by culture from Flask 2 and Flask 3 at 3 and 5 hours.

Lane 1: protein from Flask 2 (uninduced pAmidase: +ve control) at 3 hours.
Lane 2: protein from Flask 3 (test) at 3 hours, prior to induction.
Lane 3 protein from Flask 2 at 5 hours.
Lane 4 protein from Flask 3 at 5 hours, 2 hours post-induction.

4.4: Discussion

The complete genome sequence of bacteriophage ϕAR29 has been obtained in this study. Major difficulty was encountered in inducing lytic release of ϕAR29 from AR29, for the extraction of phage DNA. While the stability of the lysogenic ϕAR29 has been documented (Klieve *et al.,* 1989), another factor that may have contributed to the low yield may be the heavy encapsulation of AR29, which was apparently induced by the rich rumen fluid medium. Encapsulation of bacterial cells have been shown to block phage infection (Bernheimer and Tiraby, 1976; Burt *et al.*, 1978; Ramirez *et al.*, 1999; Gindreau *et al.*, 2000) and may prevents complete lysis of bacteria cell. The surrounding capsule may provide structural support that inhibits lysis of cellular membrane. Although the DF media developed in Chapter 3 helped reduce encapsulation, growth of AR29 in the medium was always poor in comparison to growth of AR20. The DNA yield was improved by increasing the culture volume used for extractions from 5 to 10 mL.

4.4.1: DNA Sequence of ϕAR29

At the time of writing, ϕAR29 is the only *Bacteroides* bacteriophage for which the complete nucleotide sequence has been determined. The analysis of the guanine and cytosine (GC) value can reveal codon usage patterns that are typical for that genome. The GC content of ϕAR29 could not be compared to its host's genome as there are no complete sequences of *B. thetaotaiomicron* AR29. However, analysis of ϕAR29 suggests that its GC content is comparable to a more thoroughly characterised strain: *B. thetaotaiomicron* VPI-5482 (Xu *et al.*, 2003)

. The structural nature of ϕAR29 is incompletely understood. The presence of complementary termini (a *cos* site) could not be demonstrated in ϕAR29. Southern blot analysis was inclusive, and restriction analysis showed no indication that ϕAR29

possesses cohesive termini. The migration of undigested phage DNA adjacent to the 23.13 kb size-marker indicated that ϕAR29 genomic DNA was extracted as a linear molecule, since relaxed circular or supercoiled DNA of this size migrates at a slower rate than the linear form (Qiagen News, January 2000).

4.4.2: Characterisation of the ϕAR29 Genome

The 35,558 bp genome of ϕAR29 contains 53 coding regions with two major clusters of ORFs transcribing in opposite directions. Like many temperature phages, it appears that the ϕAR29 genome is organized into life-cycle-specific gene clusters. The diverging transcriptional initiation sites are located between ORF30 and ORF31. These two diverging ORF clusters may be responsible for the lytic and lysogenic development of the phage, respectively.

Of the 53 deduced proteins from ϕAR29, only 6 have conserved domains with strong similarity to known proteins, which allows the assignment of probable functions (Terminase, dnaC, HTH LUX response regulator, *cI* repressor, integrase and amidase). Probable roles for the majority of ORFs remain unidentified, although 7 can be guessed at from their limited homologies.

The product of ORF4 could possibly be a terminase small subunit, based on the fact that the gene overlaps the ORF encoding a terminase large subunit. In addition, the presence of a gpNU1 domain further indicates this function (Yang *et al.*, 1999). However, the position of ORF4, downstream of the large terminase subunit, is inconsistent with most previous findings. In a majority of cases, phage small terminase sub-unit genes are located upstream, overlapping the large terminase unit (Ventura *et al.*, 2002; Loessner *et al.*, 2000; Siboo *et al.*, 2003). Bacteriophage P2 is one exception, where its small terminase unit is located some distance downstream of the large subunit and the two genes are transcribed in opposite directions (Linderoth *et al.,* 1991).

Significant alignment of the ORF17 translation product to a *C. thermocellum* RecB protein suggested that it could play a role in the recombinational repair of double-strand DNA breaks. The RecB protein has homology to known helicases such as Rep and uvrD proteins. Douglas *et al.* (1998) proved experimentally that the N-terminal of RecB protein is an ATP-dependent helicase and the C-terminal is associated with nuclease activity of RecBCD. BLAST results from ORF17 revealed similarities to proteins 20 and D14, from Bacteriophage Lp2 and T5 respectively, which were found to comprise a DnaD domain. The presence of DnaC upstream of ORF17, the location of the gene in the early part of lytic ORF clusters, along with the evidence above, indicates that ORF17 is likely to be involved in the replication of the ϕAR29 genome.

The predicted gene product from ORF23 has strong homology to the Orf2X protein, which has recently been proposed to be involved in the excision of a mobile element, NBU1 [non-replicating Bacteroides unit] (Nadja Shoemaker, pers. Communication). If ORF23 is an excisionase (*Xis*) its position 4789 bp from the *Int* gene in ϕAR29 is inconsistent with most findings among lambdoid phages (Echols and Guarneros, 1983; Baker *et al.*, 1991; Vander Byl and Kropinski, 2000; Juhala *et al.*, 2000). The coding regions for *Int* and *Xis* are generally located near the *att* site, within close proximity of each other.

Translation product of ORF52 was shown to have similarities with the tail tape measure protein (Tmp), which determines the length of the tail shaft of long tailed phages by a ruler-mechanism (Casjens, 2003; Zimmer *et al.*, 2003). The length of the phage tail shaft is directly proportal to the size of the Tmp protein. Phage A118, with a tail length of 300 nm, has a Tmp consisting of 1794 aa: twice the size of the λ Tmp of 853 aa, which correspnds with a 150 nm tail. The ϕAR29 tail measured 120 nm in length and its Tmp consists of 850 aa, which is similar to that of λ Tmp (Klieve *et al.,* 1989; Loessner *et al.,* 2000).

Due to the lack of homology with known genes or proteins, most of the ORFs in the ϕAR29 genome remain undefined. Despite this, the likely function of genes can be tentatively proposed by comparing the position of putative genes and terminators with details from other, published phage genomes. The limited analysis possible from available information has indicated that ϕAR29 has a similar genome arrangement to phages Sfi21, ul36 and SM1. By arranging genes from these phage under the headings Lysogenic; Replication; Transcription regulation; Packaging, head and tail protein; and Lysis, possible function can be assigned to undefined ORFs in ϕAR29 (Figure 4.23).

Phage	Genome Organization				
Sfi21	Lysogenic	Replication	Transcription Regulation	Packaging Head and Tail protein	Lysis
SM1	Lysogenic	Replication	Transcription Regulation	Packaging Head and Tail protein	Lysis
Ul36	Lysogenic	Replication	?	Packaging Head and Tail protein	Lysis
ϕAR29	Lysogenic (ORF31-ORF36) (+ strand)	Replication (ORF30-ORF16) (-strand)	? (ORF15-ORF11) (-strand)	Packaging Head and Tail protein (ORF10-ORF41) (-strand)	Lysis (ORF40-ORF37) (-strand)

Figure 4.23: Comparison of genomic arrangement of ϕAR29 with other phages, to determine the likely ORF functions. The arrows indicate the direction of transcription.

Bioinformatic studies of ϕAR29 ORFs have also indicated that some of its genes are similar to genes found in *Bacteroides* genomes. A BLAST search of *Int* and N-acetylmuramoyl-L-alanine amidase genes showed most matches were to genes of similar function, found in *B. thetaiotaomicron* VPI-5428. Identification of *Int* in ϕAR29 was through its homology to *B. uniformis* NBU1 INT1 genes and to the presence of a phage-like integrase domain. The presence of consensus sequence R-K-H-R-H-Y (Groth and Calos, 2004) has allowed the classification of ϕAR29 *Int* gene into the tyrosine integrase protein family. Although there was substitution of a histidine residue for a lysine, the change is conservative, as both amino acids are basic, polar molecules. The H residue is represented in >90% of the tyrosine members, but is not absolutely

conserved (Groth and Calos, 2004). The newly identified *Int* gene is located 1624 bases downstream of the originally proposed *Int* (Gregg *et al.*, 1994b)

Multiple alignment of amidase from phage and bacterial sources identified the ϕAR29 amidase to be closely related to those found in the close relative of its host, *B. thetaiotaomicron* VPI-5428. This is not surprising, as ϕAR29 would be expected to require mechanisms that mimic its host amidase for successful degradation of the bacterial membrane prior to lysis (Burt *et al.,* 1978; Kasper *et al.,* 1983; Coyne *et al.,* 2000). The similarity of phage and host amidase has been report in other phages such as pneumococcal bacteriophage Cp-1 (Garcia *et al.,* 1988; Diaz *et al.,* 1990), SPβ (Regamey and Karamata, 1998), ϕB6 (Romero *et al.*, 2004a), and ϕHer (Romero *et al.*, 2004a). The strong similarity of ϕAR29 amidase to the host also may suggest that the gene could have been acquired from the host chromosome through a transduction event, or from closely related phages that have not been sequenced.

4.4.3: Prediction of a Pseudo-tRNA (Leu)

The only tRNA gene identified in ϕAR29 genome was categorised as a leucine pseudo-tRNA. This gene is located between the promoter though to be responsible for expression of lytic genes, at 18214 – 18122, and HTH LuxR regulator (16675 – 16295). The presence of an intron in the tRNA was a surprise finding. Previous understanding was that bacterial genes do not contain introns. However it has recently been realised that bacterial genomes contain introns from two of the four major classes of introns: group I and group II (Haugen *et al.,* 2005). Unlike eukaryotic spliceosomal introns, bacterial introns have the ability to undergo self-splicing and appear able to transpose to uncoccupied target sites (Eickbush, 1999).

Group I introns are widely distributed in protists, but occur infrequently in bacteria (Cannone *et al.*, 2002; Ko *et al.*, 2002; Nesbo and Doolittle, 2003; Haugen *et*

al., 2005) and bacteriophages (Edgell *et al.*, 2000; Sandegren and Sjoberg, 2004; Haugen *et al.*, 2005). The process by which these introns are spliced and transposed to an unoccupied site is termed homing (Eickbush, 1999). Homing is a gene conversion process that involves intron site-specific endonuclease, which makes a double-strand break in an intronless allele (Lambowitz and Belfort, 1993; Belfort and Perlman, 1995; Parker *et al.*, 1999). The mechanism of homologous recombination then uses the "intron-present" allele to repair the recipient site, thus resulting in inheritance of the intron (Parker *et al.*, 1999; Martinez-Abarca and Toro, 2000). The first of such introns to be identified from a phage was in the thymidylate synthase gene of bacteriophage T4 (Chu *et al.*, 1984; Martinez-Abarca and Toro, 2000).

Group II introns are large self-splicing RNA molecules that are believed to have been ancestors of nuclear pre-mRNA introns. Introns in this group were initially identified in organelle genomes of lower eukaryotes and plants (Martinez-Abarca and Toro, 2000; Haugen *et al.*, 2005). In 1993, group II introns were first identified in *Calothrix species* (Ferat and Michel, 1993) and since then, species from numerous bacterial genera have been found to contain these mobile elements, including *Bacillus* (Huang *et al.,* 1999; Okinaka *et al.,* 1999), *Clostridium* (Mullany *et al.,* 1996), *Escherichia* (Tobe *et al.*, 1999; Makino *et al.*, 1998), *Lactococcus* (Mills *et al.,* 1996; Belhocine *et al.,* 2005), *Pseudomonas* (Yeo *et al.*, 2001; Kholodii *et al.*, 2000), *Serratia* (Kulaeva *et al.,* 1998), *Sinorhizbium* (Martinez-Abarca *et al.,* 1998), *Sphingomonas* (Romine *et al.,* 1999), *Shigella* (Rajakumar *et al.,* 1997) and *Streptococcus* (Coffey *et al.,* 1998).

In bacteria, group II introns generally encode a reverse transcriptase (RT), and are retrotransposable elements (Dai and Zimmerly, 2003; Dai and Zimmerly, 2002). In brief, the splicing and mobility of group II introns occurs when intron encoded RT proteins binds to pre-mRNA transcripts containing an intron and induce an RNA-

catalysed splicing reaction (Figure 4.24:(Matsuura *et al.*, 1997; Cousineau *et al.*, 1998; Martinez-Abarca and Toro, 2000; Mohr *et al.*, 2000; Zimmerly *et al.*, 2001). The process produces a complex ribonucleoprotein that contains the RT and the lariat intron. Retrohoming (mobility of group II introns) is initiated when the intron reverse splices either partially or completely into an intronless allele site. Subsequently, the zinc domain of the RT induces a nick in the antisense strand of the intron insertion site, 9 or 10 bp downstream of the exon junction. The cleaved DNA is then used as a primer for the RT to reverse transcribe a cDNA of the intron. Finally, a host RecA-independent repair mechanism is used to replicate the sense-strand and to join the nicked insertion site with the newly synthesized intron (Matsuura *et al.*, 1997; Cousineau *et al.*, 1998; Martinez-Abarca and Toro, 2000; Mohr *et al.*, 2000; Zimmerly *et al.*, 2001).

Group I introns have been identified in other bacterial tRNA genes. In cyanobacteria these elements are found to interrupt the tRNA-Leu(UAA) gene, or the tRNA-fMet gene, or both (Biniszkiewicz *et al.,* 1994; Paquin *et al.,* 1997), whilst 3 divergent species of α-proteobacteria (Reinhold-Hurek and Shub, 1992; Paquin *et al.*, 1999) and one species of β-proteobacterium (Reinhold-Hurek and Shub, 1992) were found to have introns inserted into tRNA-Arg(CCU) and tRNA-Ile(CAU), respectively. All of these introns are found in the tRNA gene anti-codon loop (Kuhsel *et al.*, 1990; Xu *et al.*, 1990; Biniszkiewicz *et al.*, 1994; RajBhandary, 1994; Paquin *et al.*, 1997; Paquin *et al.*, 1999; Rudi and Jakobsen, 1999; Vepritskiy *et al.*, 2002), in the case of tRNA-Leu, at the position: U-intron-AA. (Kuhsel *et al.,* 1990; Xu *et al.,* 1990; Paquin *et al.,* 1997; Besendahl *et al.,* 2000). ϕAR29 tRNA has a Leucine anticodon (CCG), which is different from those of cyanobacteria. The intron also differs in that it is inserted one base position down stream of the anti-codon rather than within the anti-codon, like the intron of tRNA-Leu(UAA).

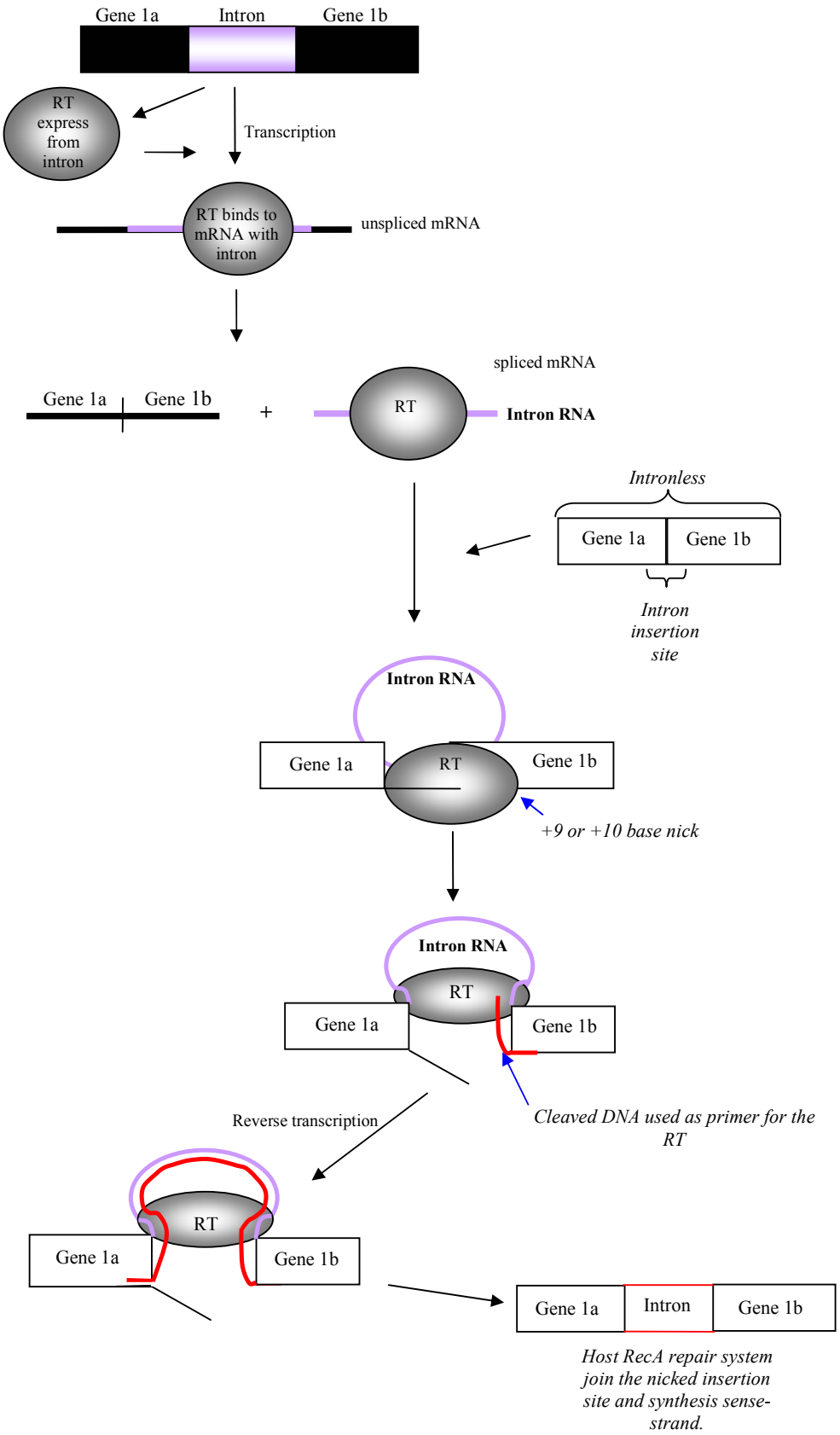


Figure 4.24: Flow diagram, illustrating the mobility of group II introns. (Eickbush, 1999; Martinez-Abarca and Toro, 2000)

4.4.4: Coding Frameshifts

Analysis software identified the presence of two possible -1 frameshifts in ORF6/ORF5 and ORF46/ORF45 within the ϕAR29 genome. However, removal of the frameshift slippery sequence did not alter the size of the ORF in concern. Frameshifting is induced by specific signals in the mRNA that cause the ribosome to switch to an alternative coding frame at a specific site (Farabaugh, 1996; Du and Hoffman, 1997). The two frameshift types are the -1 and +1 frameshift (Deogun *et al.,* 2004; Moon *et al.,* 2004). The common -1 frameshift consist of a "slippery" site (where ribosomes change reading frames) and a stimulatory RNA structure such as a pseudoknot or a stem-loop located downstream (Deogun *et al.,* 2004; Moon *et al.,* 2004; Xu *et al.,* 2004). In viruses frameshifts play an important role in a variety of biological process, including morphogenesis (Du and Hoffman, 1997), for example, bacteriophage λ tail assembly proteins gpG and gpGT. Although both proteins are required for the assembly of phage tail, neither protein is incorporated into the mature tail structure. The protein gpGT is encoded as a result of a -1 translational frameshift at the 3' end of the gpG gene (Levin *et al.*, 1993). Studies by Xu (2001) have revealed that production of both gpG and gpGT, in the correct ratio, is crucial for the production of biologically active tails (Xu, 2001; Xu *et al.*, 2004).

Identifying positions for programmed frameshifts is a difficult process, largely due to their diverse nature (Moon *et al.,* 2004). Although there are computational models available, the output predicts many false positives or requires reference protein sequences together with DNA sequence from similar organisms to prime the model (Moon *et al.,* 2004).

Programmed frameshifts have been characterized in several phages that included λ (Christie *et al.*, 1996; Levin *et al.*, 1993; Hayes and Bull, 1999), P2 (Christie *et al.,* 2002), T4 (Brown *et al.*, 1993; Ripley and Clark, 1986; Streisinger and Owen, 1985;

Ripley and Shoemaker, 1983), T7 (Pierce and Masker, 1992), and A2 (Garcia *et al.,* 2004). Recently, Xu et al., (2004) described a strongly conserved -1 frameshift in a region that lies between the major tail protein gene and the tape measure protein of dsDNA phages. The presence of a frameshift was detected even when the sequences at that location differed (Xu *et al.,* 2004). ϕAR29 did not show an apparent frameshift at this region.

4.4.5: Regulatory Elements

The B*acteroides* -7 and -33 promoter consensus sequence has been characterised (Bayley *et al.*, 2000; Thorson, 2003) and it has been shown that disruption of these sequences has abolished or sharply reduced promoter activity (Bayley *et al.,* 2000). *Bacteroides* promoter sequences were present in all three of the predicted promoter regions of ϕAR29.

Region 18193-18316 contains 2 potential promoter sequences that are transcribed divergently and appear to constitute a bidirectional promoter between two ϕAR29 gene clusters. Interestingly, the parts of each promoter that contain the putative –33 regions, overlap each other (Figure 4.25).

The presence of HTH LuxR regulator gene (ORF25), near the *cI* repressor gene (ORF31) and the promoter region 18193-18316, may be the governing switch between lytic and lysogenic life cycles. The HTH LuxR regulator (ORF 25) may function like a λ *Cro* protein, which prevents the phage from entering lysogency but allows induction of the lytic cycle . *Cro* proteins of bacteriophages are known to have HTH motifs (Fromknecht *et al.*, 2003; LeFevre and Cordes, 2003; Nilsson and Widersten, 2004).

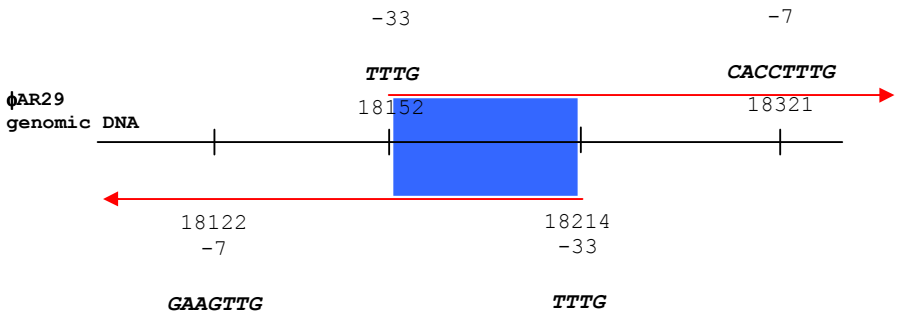


Figure 4.25: Two diverging promoter regions identified within phage ϕAR29. The red arrows and blue region indicate the direction of transcription and the overlapping region between the two promoters, respectively.

Transcriptional terminators can be classified into 5 categories: L-shaped, I-shaped, V-shaped, U-shaped and X-shaped, dependent on the type of stem loop they form (Unniraman *et al.,* 2002). The majority of stem structures found in ϕAR29 were either L-shaped or I-shaped. The terminators identified here have either short poly-U tails, or no poly-U tail. Past studies have shown that terminators can have varying efficiencies, despite the absence of a poly-U tail, such as those found in bacteriophage ϕC31 (Unniraman *et al.*, 2002; Ingham *et al.*, 1995). One unusual X-shaped terminator was identified between ORFs 36 and 37: the last ORFs transcribed from each of the two diverging clusters. It appears likely, therefore, that ORFs 36 and 37 may share the same, bidirectional terminator.

4.4.6: *Att* Sites

4.4.6.1: AttP

In previous studies the precise location of phage ϕAR29 *attP* site was not established (Gregg *et al.*, 1994b). Sequencing of ϕAR29 and the bacterial DNA flanking the integrated phage has allowed precise identification of the phage *attP* region. Although the majority of the *attP* sites found in other phages, including the well characterised λ-like phages, are located downstream of their respective integrase gene,

the *attP* of ϕAR29 is positioned 162 bp upstream of the newly identified *Int* gene. There have been reports that phages Epsilon 15 (Kropinski pers comm.); L5 (Lee *et al.*, 1991b); D29 (Ribeiro *et al.,* 1997); Ms6 (Freitas-Vieira *et al.,* 1998); HK620 (Clark *et al.,* 2001); ϕBT1 (Gregory *et al.,* 2003) and Sf6 (Casjens *et al.,* 2004) share the same *attP-Int* configuration as ϕAR29. In the case of *Myxococcus xanthus* phages, Mx8 and Mx9, the *attP* lies within the 3' end of the *Int* gene (Magrini *et al.,* 1999).

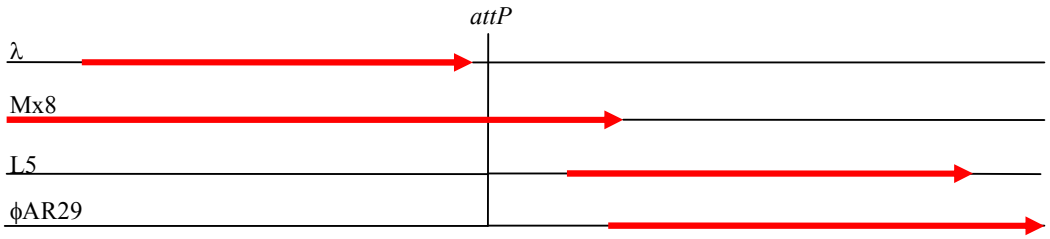


Figure 4.23: Comparative organisation of the integration modules in bacteriophages. The red arrows represent *Integrase* genes.

4.4.6.2 AttB site

As stated in Chapter 1, bacteriophages often target the 3' terminus of tRNA genes as their *attB* site. Some of the reasons proposed for this preference are:

1. The use of the symmetrical sequence in the tRNA gene as a recognition site.
2. Stability. tRNA genes are generally more stable than protein coding regions. In bacteria, divergence rate per base pair for tRNA genes are 4 – 9-fold lower than for protein coding regions.
3. Small size. The small size of tRNA gene also reduces the amount of sequence in the *attP* site that is required for capturing and restoring host sequences during the integration procedure.

(Williams, 2002; Yang *et al.*, 2002)

Another reason for phage to use the 3' terminus of a tRNA gene, instead of the 5' end, may be the presence of regulatory signals, such as promoter sites, that are generally located at the 5' side of the tRNA gene. Restoring such sites during integration would be more complex than for sites at the 3' end of tRNA genes. Although the 3' terminus may possess a terminator sequence, they are easily replaced and have a broader host range than promoter signals (Zhao and Williams, 2002). In general, insertion of the phage genome into a tRNA gene can be performed in such a way as to avoid the disruption of the gene, which in other gene types may be lethal (Yang *et al.,* 2002).

The complete genome sequence of *B. thetaiotaomicron* VPI-5428 (Xu *et al.*, 2003) allowed the identification of the ϕAR29 arg-tRNA *attB* site. The position of the *att* core at the 3' end of a tRNA gene is a common occurrence among phages of gram-negative bacteria such as 16-3, 186, HP1, P4, P22 and ϕCTX, and in actinomycete integrative plasmids such as SLP1, pSAM2 and pMEA100 (Reiter *et al.*, 1989; Yang *et al.*, 2002; Campbell, 1992). Bacteriophage ϕAR29 *attB* fulfills two of the 4 general rules for integration into tRNA genes (section 1.4.4.1). These are:

1. The core sequence is the 3' terminus of a tRNA gene, and
2. Integration restores the tRNA gene.

(Gabriel *et al.,* 1995)

The other two rules are:

- The anticodon is part of the core region,
- The target tRNA site must be a functional gene,

(Gabriel *et al.,* 1995)

ϕAR29 integration at the 3' end of the tRNA did not satisfy the first point, whilst the latter rule remains undetermined, as there is no experiment data supporting the tRNA functionality.

The location of *attB* within the 3' end of the tRNA, allows the site to be listed as class III (see Chapter 1). Numerous phages require a class III *attB,* including ϕR73 (Sun *et al.,* 1991), P4 (Pierson and Kahn, 1987), A2 (Proux *et al.,* 2002), 933I (Allison *et al.,* 2002), T12 (McShan *et al.,* 1997), mv4 (Auvray *et al.,* 1999), TPW22 (Petersen *et al.,* 1999) and VPIϕ (Petersen *et al.,* 1999). Interestingly, *Bacteroides* mobile elements NBU1 and NBU2 also belong to this group (Williams, 2002), with these mobile elements reported to insert at the 3' end of Leu-tRNA and Ser-tRNA genes, respectively.

4.4.6.3: Core attachment site

Past studies have revealed that phages integrating into tRNA genes contain 17 – 182 bp of DNA that is homologous between *att*B and *att*P. The ϕAR29 16 bp *att* core sequence was at the lower end of this scale. The re-defined position of *att*P is 142 bp from the previously proposed site, placing it somewhat further upstream of the newly identified *Int* gene (Gregg *et al.*, 1994b).

Several imperfect direct and inverted repeats were observed surrounding *attP*. In more intensively studied phage systems, these repeats are thought to represent binding or recognition sites for *Int* and IHF (integrative host factor) proteins (Bruttin *et al.*, 1997; Groth and Calos, 2004). The consensus of these repeats is very similar to those determined by Gregg *et al.* (1994).

4.4.7: Determining the Function of ORF40 (amidase)

Lytic enzymes used for breaking down cell walls during cell division and host lysis can be found in the genomes of bacteria and phage, respectively (Diaz *et al.*, 1990; Romero *et al.*, 2004a). In bacteriophage this gene is sometimes referred to as endolysin or lysin (Diaz *et al.,* 1990). These lytic enzymes can be classified into one of two general groups:

1. Those that attack the glycosidic bond in the cell wall (i.e lysozymes and transglycosylases), or
2. Those that break the peptide bonds (i.e. amidases)

(Diaz *et al.,* 1990)

The lytic enzyme identified in bacteriophage ϕAR29 belongs to the latter group. Amidases act by hydrolysing the link between *N*-acetylmuramoyl residues and L-amino acid residues, preferentially the D-lactyl-L-Ala link, in certain cell-wall glycopeptides (Diaz *et al.,* 1990; Rigden *et al.,* 2003) . Due to time limitations, investigation into the activity of ϕAR29 amidase was incomplete. However, attempts were made at cloning and confirming the function of the putative amidase gene in *E. coli*. Characterisation of this gene was thought to be beneficial, since the enzyme might be used as a tool for lysing cells. This could result in improved methods for extraction of phage DNA and of plasmid from encapsulated bacteria such as AR29. Although the amidase gene was successfully cloned into pTric vector, the data from the expression of the lytic gene in *E. coli* did not indicate any significant cell lysis effect, when comparing the test culture (Flask 3) and control cultures (Flask 2 and Flask 5). It was concluded that lytic activity was not induced in these experiments. Reasons for a lack of induction may include the inability of the protein to recognise substrates in *E. coli* that are normally present in *B. thetaiotaomicron* AR29. However, further study is required to determine whether active

protein was generated from the cloned gene. Sequence alignment (Figure 4.10) has revealed that the ϕAR29 amidase shares many domains with amidases found in *Bacteroides*. Interestingly, a portion of the C-terminus of the ϕAR29 amidase (amino-acids 130 to 150), which is involved in cell wall recognition and binding (Diaz *et al.*, 1990; Regamey and Karamata, 1998; Rigden *et al.*, 2003; Romero *et al.*, 2004a), is absent from amidases of other phages (Figure 4.11). A likely explanation for not observing lytic activity in *E. coli* cultures is the absence of a phage lytic protein, holin, which is a membrane-spanning protein that allows the phage amidase to access the peptidoglycan moiety by forming holes in the cytoplasmic membrane (Gaeng *et al.,* 2000; Kashige *et al.,* 2000). In most lytic phage systems studied, the activity of endolysins is dependent on the presence of holin in the host cell (Gaeng *et al.,* 2000). Examination of ϕAR29 sequences did not reveal the presence of a holin like gene.

To fully determine the lytic function of ORF40, the protein needs to be purified and tested for activity on cell-wall components from AR29. Furthermore, confirmation of cell lysis may be improved by microscopy. Measurement of culture turbidity alone may be an insensitive procedure for confirming cell lysis, as two different type of lytic events can occur:

1. by total cellular degradation, whereby the lytic reaction leaves featureless dedris, or
2. a lytic reaction that leaves a nonrefractile host shell

4.5: Conclusion

The bacteriophage ϕAR29 genomic sequence has been completed and probable functions have been assigned to 6 of the 53 ORFs. This is believed to be the first *Bacteroides* phage, and possibly the first ruminal bacteriophage to be fully sequenced. ϕAR29 has illustrated mosaicism in the genome and, overall, is not closely similar to

other known phage genomes. The work carried out here has provided a substantial basis for future work, which will be required for a more complete understanding of the phage. Important questions that remain unanswered include:

- the functions of proteins encoded by the undefined ORFs, and
- mechanisms for the regulation of ϕAR29 lysogenic and lytic life-cycle.

At this stage, it is not clear whether ϕAR29 is entirely functional as a lytic phage, since a lytic host has not been identified (Klieve *et al.*, 1989). As an alternative to the identification of a different host for lytic infection, purification of an unlysogenised strain of AR29 would greatly facilitate future work. In addition, the application of DNA footprint techniques would allow the location of binding sites of regulatory and recombination proteins. Despite the limitations of the scope of this work, sequencing of the ϕAR29 genome has led to a better understanding of the phage recombination mechanism, which has helped explain previous work, and will be valuable in subsequent investigations.

Chapter 5: Chromosomal integration of plasmid pBA in *B. thetaiotaomicron* AR29

5.0: Introduction

One of the prerequisites for the practical application of recombinant DNA technology is to maintain the stability of introduced genes in genetically modified organisms. To achieve this, many plasmid vectors and integration vectors have been developed. Unlike plasmid vectors, integration vectors insert DNA into the chromosome of the intended host (Groth and Calos, 2004). For reasons of stability and safety, selection marker genes and undesirable vector-derived sequences that are not required in the final construct may be removed from the integration vector.

Integration mechanisms for vectors have used one of two methods:

1. endogenous homologous recombination into a certain DNA sequence on the chromosome, or
2. site-specific recombination into a highly specific target site

(Shimizu-Kadota, 2001; Rossignol *et al.*, 2002)

The latter method, which is used by bacteriophage, involves a phage encoded integrase, and the phage attachment site (*attP*) (Bruttin *et al.,* 1997). The advantages of bacteriophage integration mechanisms include:

1. the integration process is likely to be harmless to the host cell, as the phage attachment site is a naturally evolved integration site.
2. The stability of transgenic integrants in a non-selective environment is generally greater than those generated by homologous recombination.
3. Generally, only a single copy of the sequence is integrated and is not subject to amplification.

(Shimizu-Kadota, 2001; Yang *et al.*, 2002)

5.0.1: Phage Integrases and Attachment (*att*) sites

The integrases of bacteriophages recognize distinct DNA sequences that may be 30 bp or longer. Based on amino acid sequence homology and catalytic residues, these site specific recombinases can be classified into either the serine or tyrosine family of integrases. Both families accomplish recombination by covalent attachment to DNA, with integrase monomers bound to each *att* site (*attB and attP*). Neither recombinase family requires DNA synthesis or energy substrate cofactors to achieve integration. Despite these similarities there are differences between serine and tyrosine integrases (Table 5.1).

Table 5.1: The general characteristics of Tyrosine and Serine integrase families. (Groth and Calos, 2004)

	Tyrosine integrases	**Serine integrases**
Catalytic residue	Tyrosine	Serine
Mechanism of integration	Formation and resolution of Holliday junction	2-bp staggered cleavage, 180° rotation and ligation.
attB	Short with overlap region flanked by short imperfect inverted repeats.	Short with overlap region occasionally flanked by short imperfect inverted repeats.
attP	Extended with multiple binding sites for *Int*, IHF and cofactors.	Short; overlap region occasionally flanked by short imperfect inverted repeats.
Overlap region (bp)	6 – 8	3 – 12
Requirement of host factors	Yes (e.g. IHF, FIS)	No

5.0.1.1: Serine Integrases

Serine integrases, belonging to the serine recombinase family, possess a serine catalytic residue (Smith and Thorpe, 2002; Groth and Calos, 2004). These proteins, which include enzymes referred to as transposases, have been proposed to be a sub-family or separate, closely related family to other serine recombinases, i.e resolvase and invertase (Thorpe *et al.,* 2000). Amino-acid sequence alignment of these large serine integrases

revealed an N-terminal catalytic domain, a region of conserved cysteine residues and a conserved C-terminal region of unknown function (Figure 5.1; Smith and Thorpe, 2002).

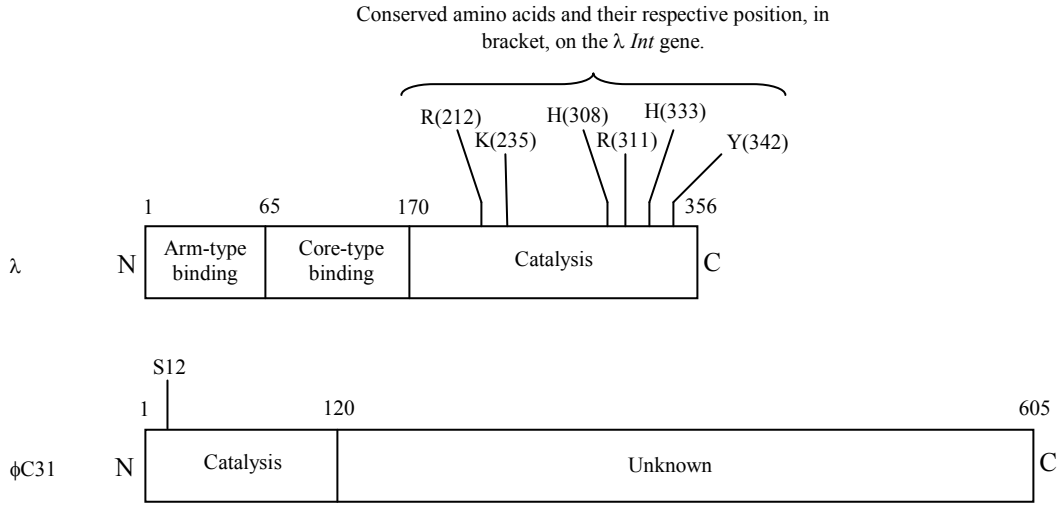


Figure 5.1: The protein domain structure of tyrosine and serine recombinase families, represented by the most intensively studied phage λ and ϕC31 integrase respectively. (Groth and Calos, 2004)

Serine family integrases cleave the DNA backbone using the hydroxyl group of the catalytic serine (Figure 5.2 A). The enzyme makes a double strand break staggered by 2 bp at the middle of the *att* core (Figure 5.3 A) to produce recessed 5' ends and 3'-OH overhangs. Each of the 5' ends of the cleaved DNA strands is linked to one of the recombinase molecules through the conserved serine reside, which is presumed to provide the primary nucleophilic hydroxyl group in the cleavage reaction. Following this, the cleaved DNA is rotated 180° prior to ligation (Smith and Thorpe, 2002; Stark *et al.*, 1992; Hallet and Sherratt, 1997; Smith *et al.*, 2004).

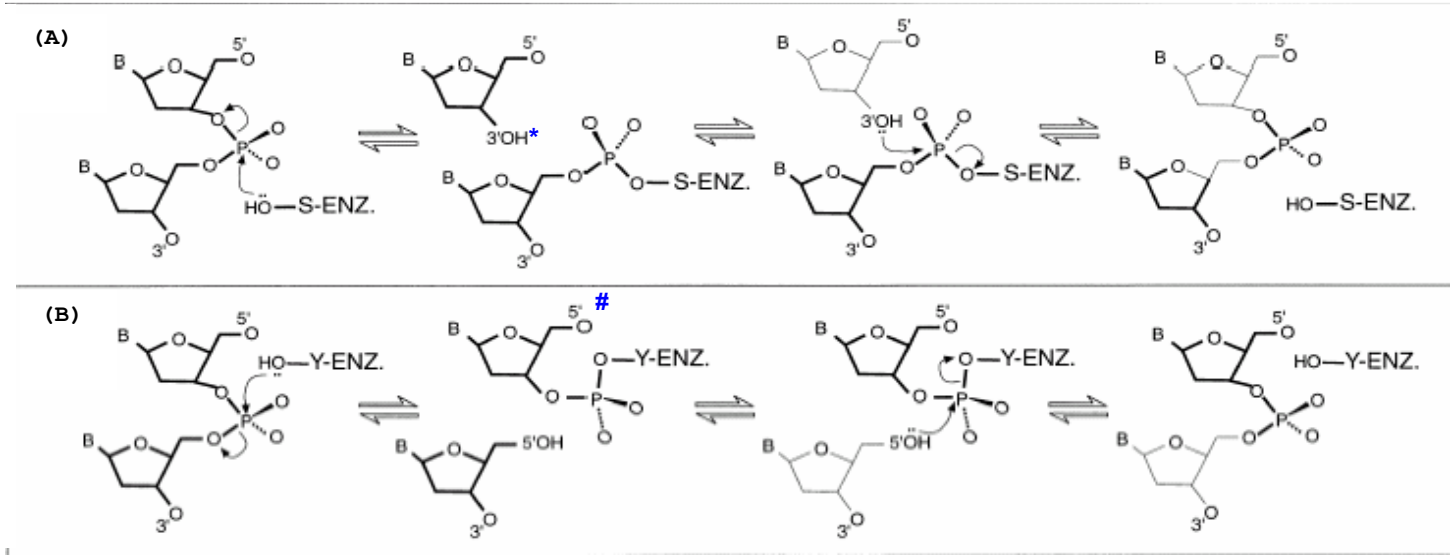


Figure 5.2: Chemistry of site-specific recombination by serine family, "ENZ-S-", (A) and tyrosine family of integrases, "ENZ-Y-" (B). Serine integrase use a serine residue for cleaving DNA, resulting in a 3'-OH group (*) for strand exchange, whilst in tyrosine integrase the catalytic residue is tyrosine and the 5'-OH group (#) is catalytic in the recombination process. Donor and recipient phosphate backones are drawn in thick and thin lines, respectively. (Hallet and Sherratt, 1997)

(A) Diagram illustrating the serine-based recombination.

(B) Diagram illustrating the tyrosine-based recombination.

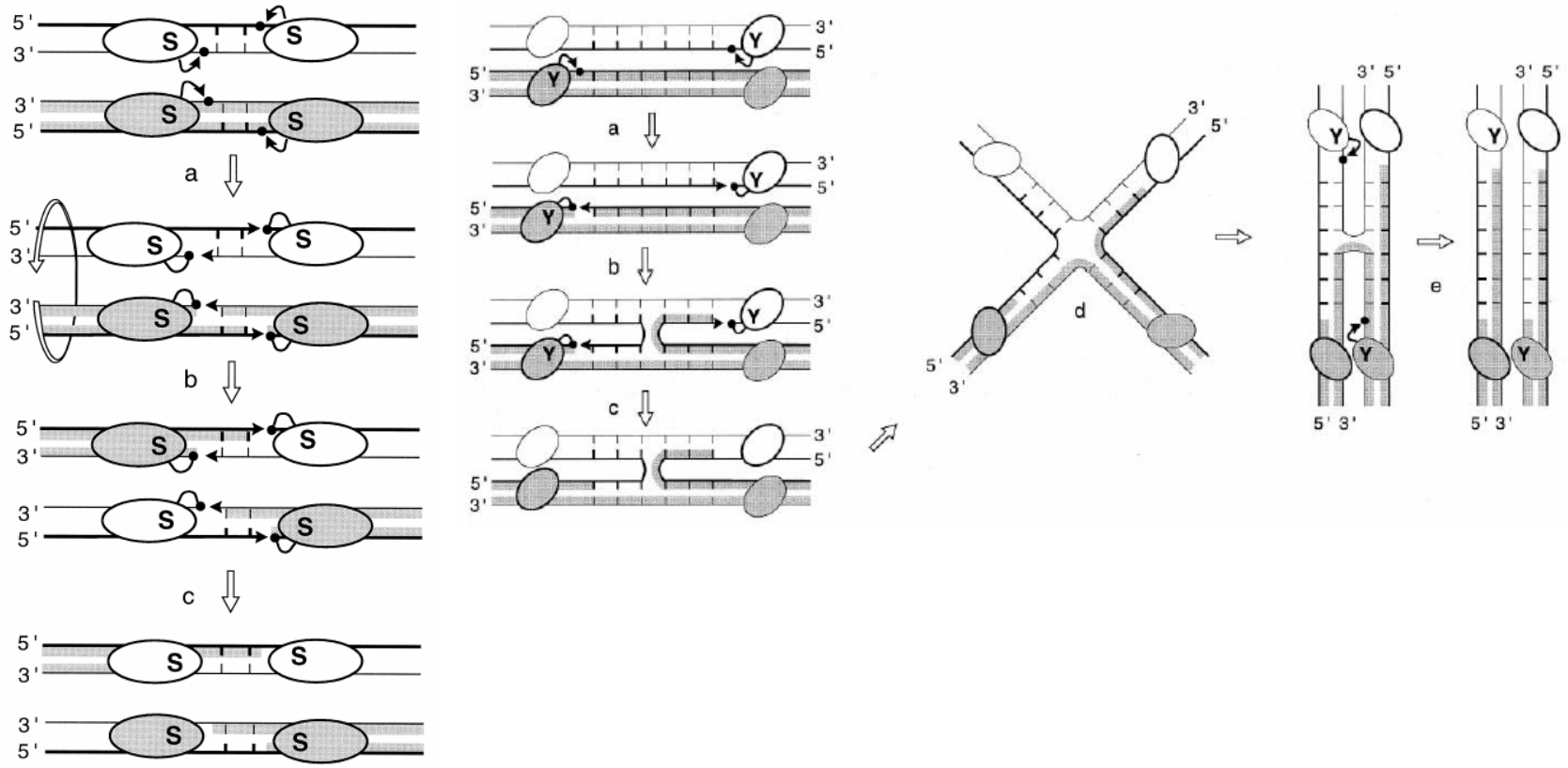


Figure5.3: Diagrams illustrating the recombination process of serine family (A) and tyrosine family of integrase (B). In serine integrase, serine subunits (oval shape with letter S) bind to the core region and induce a cut on the sense (thick line) and anti-sense (thin line) DNA. The nucleophilic cleavage (a) is achieved by the serine residue (small arrow attached to serine subunit) attacking at the phosphate backbone (black dot) groups. Subsequently, the strand exchange occurs by 180° rotation (b) prior to re-ligation (c). With tyrosine integrase, the tyrosine residues (oval shape with letter Y) induce cleavage on the sense strand (thick line: a), prior to strand exchange (b) and ligation of the swapped strands (c). After ligation, the DNA forms a Holliday structure (d). Following the dissociation of the Holliday structure, the integrase tyrosine residues carry out the same cleavage, strand exchange and ligation, from step a-c on the anti-sense strand.(Hallet and Sherratt, 1997)

To date there is no evidence that integrases from the serine recombinase family require any cofactor for their integrative function. Past studies have shown purified integrases from phages ϕC31, TP901-1 and R4 were able to function *in vitro* and in mammalian cells, in the absence of phage or bacterial-encoded proteins (Thyagarajan *et al.*, 2001; Thyagarajan *et al.*, 2000; Stoll *et al.*, 2002; Olivares *et al.*, 2001; Groth *et al.*, 2000). As a result of this, they have been targeted for use in gene replacement technology in mammals (Groth and Calos, 2004).

5.0.1.2: Serine Integrase Attachment

Integrase *att* sites consist of an overlap region where the sequence of *attB* is identical to that of *attP* and where crossing over occurs. Serine integrase *att* sites consist of a short overlap region of ≥ 3 bp and are usually flanked by two imperfect inverted repeats (Groth and Calos, 2004). Studies have shown that in phages TP901-1 and ϕC31 the 2 bp overhang resulting from *Int* cleaving the DNA must be complimentary to the opposite strand for efficient recombination (Combes *et al.*, 2002; Breuner *et al.*, 2001).

5.0.1.3: Tyrosine Integrases

Sometimes referred to as the Integrase family, the tyrosine family of recombinases utilise a tyrosine catalytic residue (Y342 in λ) to cleave the DNA backbone (Nunes-Duby *et al.,* 1998). In the well characterized 356 amino acid integrase from phage λ, the N- and C-terminal domains are involved in protein-protein interaction between integrase monomers (Campbell *et al.,* 2002). Amino acid positions from 1 – 64 are responsible for binding the arm-type sites of *attP*, which are the direct repeats flanking the attachment core site (Figure 5.1). Amino acid residues 65 – 169 participate in core-type binding (Campbell *et al.*, 2002;

Cho *et al.*, 2002a; Subramaniam *et al.*, 2003; Groth and Calos, 2004). The C-terminal of the tyrosine recombinase family consists of a domain with five highly conserved amino acids. These are the R, K, H, R, and H residues which occur in λ at positions 212, 235, 308, 311 and 333, respectively (Figure 5.1). The H residue at position 308 is present in >90% of the recombinases in this family, however it is not absolutely conserved. In addition, the H amino acid positioned at 333 can be substituted with a tryptophan residue (Groth and Calos, 2004; Nunes-Duby *et al.*, 1998).

The recombination processes of λ-like integrases involve the catalytic tyrosine residue, which is involved in covalent binding of DNA to the integrase molecule. The hydroxyl group of the tyrosine participates in DNA strand cleavage by nucleophilically attacking the phosphate backbone of DNA (Figure 5.2 B(Groth and Calos, 2004). Unlike serine-based integrases, tyrosine recombination enzymes covalently attach to the 3' end of one DNA strand, while the free 5' hydroxyl group interacts with the protein-DNA bond of the opposite *att* site, resulting in a Holliday junction (Figure 5.3 B;(Sadowski, 1986; Stark *et al.*, 1992; Hallet and Sherratt, 1997)

Unlike the serine recombinase family, tyrosine recombinases require cofactors or proteins, such as integration host factors (IHF), to aid the recombination process. Richet *et al.* (1988) were able to illustrate that IHF and integrase form a complex, prior to binding at *attP*, followed by attachment at *attB*. Although it is not fully understood, IHF is believed to participate by binding to sites in the extended *attP* and induce bending of the DNA to form a structure called an intasome (Robertson and Nash, 1988; Weisberg and Landy, 1983; Nash, 1990; Hwang and Scocca, 1990; Dorgai *et al.*, 1998).

5.0.1.4: Tyrosine Integrase Att site

Tyrosine integrases require a simple *attB* and a complex *attP*. The size of the *attB* overlap region ranges from 6 to 8 bp and i*s* generally flanked by two inverted repeats called core-type binding sites (Figure 5.3; Sarkar *et al.*, 2001; Cho *et al.*, 2002a). These sites also direct the integrase to perform the recombination reaction at a specific location within the DNA squence. In contrast, the *attP* consists of direct repeats called arm-type binding sites, which flank the core-type site (Hakimi and Scocca, 1994; Pena *et al.*, 1997). Furthermore, *attP* also contains binding sites for IHF (Craig and Nash, 1983; Craig and Nash, 1984; Groth and Calos, 2004).

Studies based on λ integrase have shown the enzyme to bind with a high affinity to arm regions of the *attP* site and with low affinity to the core site (Nash, 1990). The presence of IHF subsequently induces the folding of DNA, bringing the strongly bound *Int* protein into close proximity with the lower affinity central core-type site (Nash, 1990; Swalla *et al.*, 2003). In addition, experiments have revealed that *Int* binds weakly to the P2 arm-type binding site and it has been found that P2 is required only for excisive recombination (Bauer *et al.,* 1986; Numrych *et al.,* 1990).

5.0.2: Plasmid pBA

Integration vector pBA was constructed for the transformation of two *Bacteroides* strains and of *E.coli* (see Chapter 1; Wong *et al*, 2003). However, hybridization studies and experiments described in Chapter 3 did not show any evidence of plasmid pBA integrating into the genome of AR20, AR29 or *E. coli* strain SCS110. Nevertheless, the absence of proof for integration does not dictate that integration did not occur. The lack of information available during the work described in Chapter 3, such as the precise location of attachment and the sequence surrounding the *attB* site, limited the sensitivity of methods for detecting

integration. Sequencing of the ϕAR29 genome (Chapter 4) has identified an integrase gene that may be responsible for phage integration, suggesting that the proposed integrase in pBA may not function in that role.

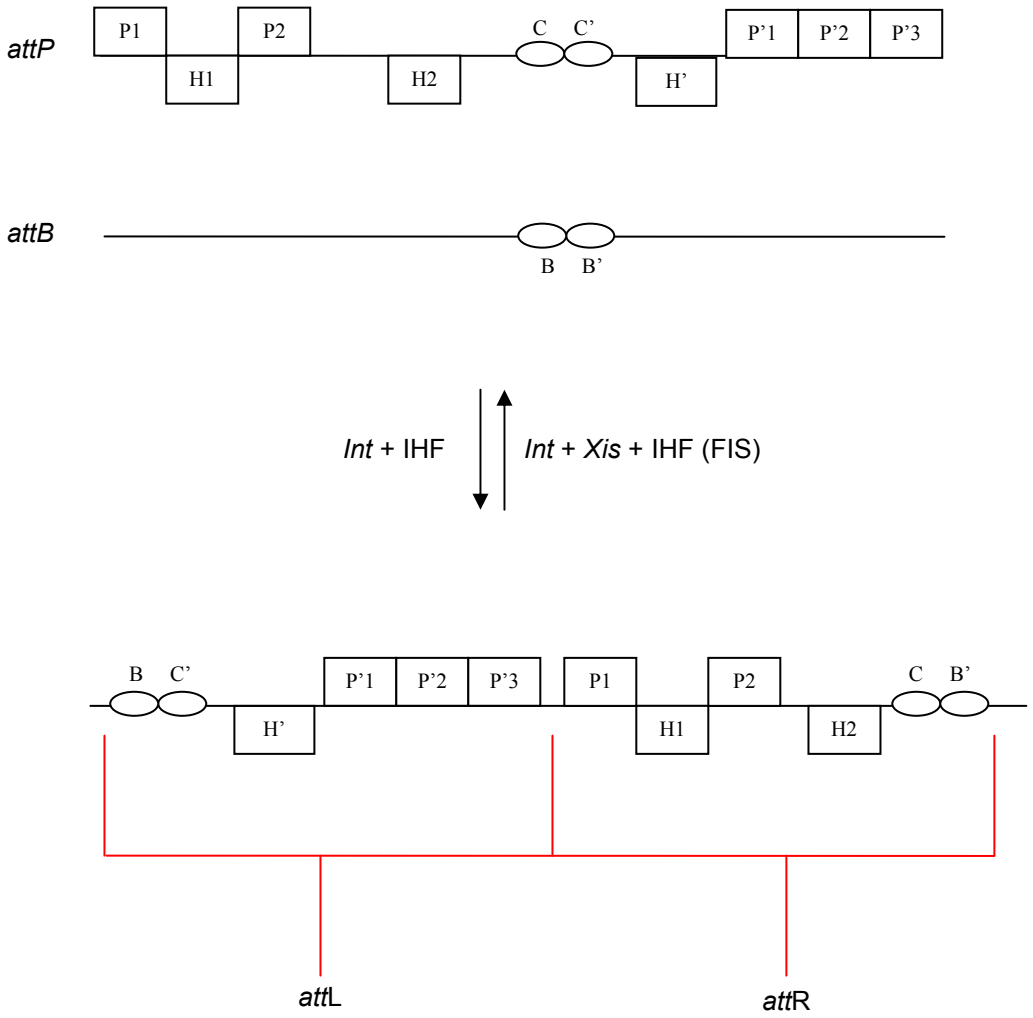


Figure 5.3: The position of DNA and protein binding sites in λ *att* site prior to, and after, recombination. The *attP* consist of five arm-type binding sites (P), three IHF attachment site (H) and a core-type binding region (C and C'). Within *attB*, only the core type sites B and B' are used during integration.(Goodman *et al.*, 1999; Swalla *et al.*, 2003)

5.0.3 Aim:

The aim of work described in this chapter was to

- confirm that pBA contains the true *attP* site,
- determine whether pBA contains a functional *Int* gene, and
- observe whether plasmid pBA was able to integrate into the genome of AR29.

5.1: Materials and Methods

5.1.1: Bacterial strains and Plasmids, Sequence Analysis and RNA Isolation

The transformation of *B. uniformis* strain AR20, and *B. thetaiotaomicron* AR29 with plasmid pBA, RNA isolation procedure and Sequence analysis are described in Chapter 2 sections 2.6.9.2, 2.6.14 and 2.6.17, respectively.

5.1.2: Testing for the Expression of Prophage *Int* Genes from *B. thetaiotaomicron* AR29

RT-PCR was performed as described in Chapter 3, section 3.1.3. The primers used for the detection of *Int* sequences in AR29 RNA are listed in Table 5.2.

Table 5.2: Primers used for RT-PCR studies. The red bases indicate the alteration made to create a hybridisation mismatch to the original sequence (blue).

Primers name	Sequence 5'→ 3'	Length (bases)	Tm (° C)	Product (bases)
Srec for	AAAGCCTCCATAACCGACTCAACT	25	72.3	535
Srec rev	TCTCCAATTCGACAACCTAAACTACA CCT	21		

*5.*1.3: Testing for pBA Integration

5.1.3.1: Preparation of samples for PCR analysis

Bacteroides strains AR20 and AR29 were transformed with plasmid pBA, grown overnight, and each strain was re-inoculated into 5 ml of fresh, defined, non-rumen-fluid broth medium. After 24 hours of growth at 39°C, the culture was chilled on ice and a small sample of the culture was tested by PCR, using a tooth pick swab. This involved:

1. touching the culture with a sterile toothpick and transferring the cells to the PCR tube by smearing it onto the inner wall.
2. PCR mixture and conditions used were as shown in Table 2.5 and 2.6 (2.6.4).

5.1.3.2: Tests for the presence of integrated plasmid pBA

Tests to determine whether pBA plasmid integrated into the genomes of AR20 and AR29 used primer combinations listed in table 5.3. Primer sets 1 and 2 were designed to amplify a fragment of 2.8 Kbp and 1.4 Kbp respectively, if the plasmid was integrated in one orientation relative to the tRNA gene that contains the *attB* site. Primer sets 3 and 4 were designed to amplify a fragment of 1.4 Kbp and 2.7 Kbp respectively, if the plasmid was integrated in the opposite orientation (Table 5.4).

5.1.3.3: Detection of free, non-integrated plasmid pBA in AR20 and AR29.

The presence of non-integrated plasmid was tested using primer set 5 (Table 5.4).

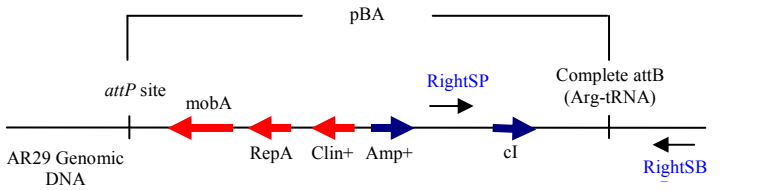
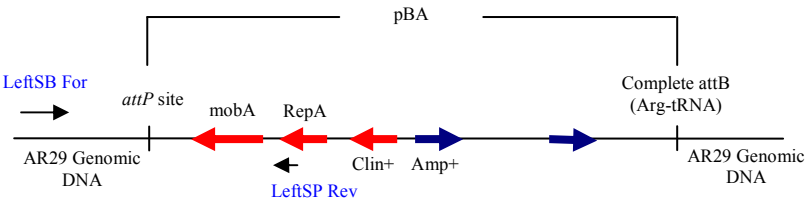
5.1.3.4: Tests for plasmid pBA integration adjacent to prophage in B. thetaiotaomicron AR29.

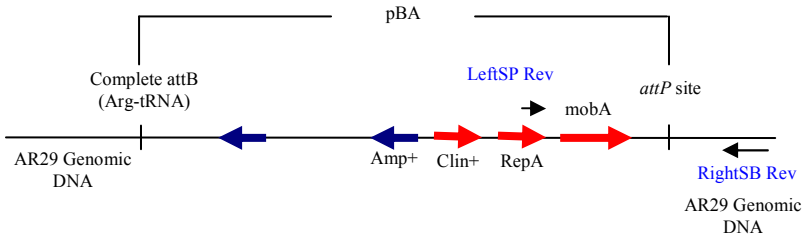
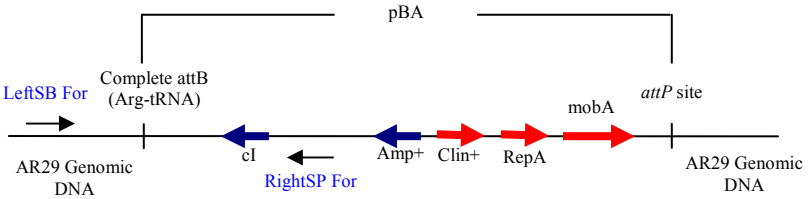
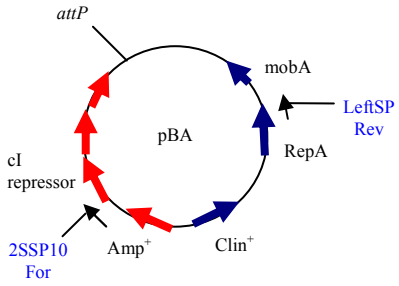
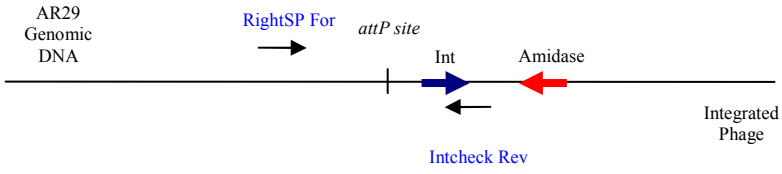
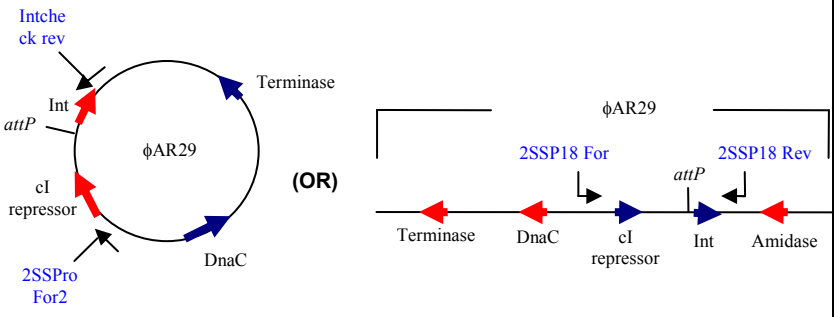
The possible integration of pBA into the bacterial chromosome adjacent to the phage genome, at the reformed *attB* site, was investigated using primer sets 10 and 11 to test for the integration of pBA downstream from the *cI* gene or upstream from the *Int* gene of ϕAR29, respectively (Table 5.4).

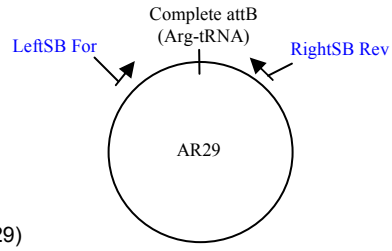
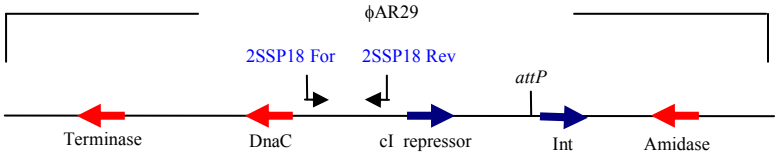
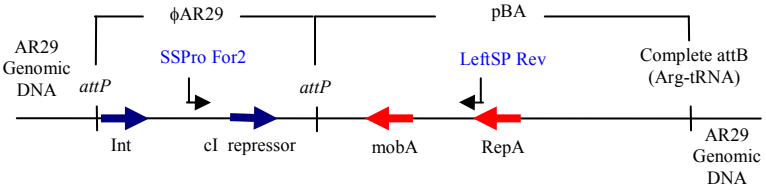
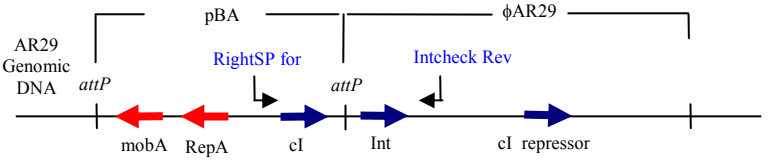
Table 5.3: A list of primers used to test for plasmid pBA integration into the AR29 genome.

Name of Primers	Primer Sequence
RightSP For	ATGTTCTTTCCTGCGTTATC
RightSB Rev	ATGTTAAGGGACGGTAGTGA
LeftSP Rev	CAACGCAAGGACAACCAGTA
LeftSB For	CCGATAAAGGATTGCAGGTA
2SSP10 For	TCCGAGCGAAAATCACTAATA
Intcheck Rev	CCTCGTTCTCCTGATACATAGCG
SSPro F2	GCAGTGTTGGATGTTGTTTGAT
2SSP18 For	CGTAGATGGTCGTTTCCTTTC
2SSP18 Rev	GATGGCTGCTGATGTGTATTG

Table 5.4: Sets of Primer combinations (from Table 5.3) used to test for the integration of pBA into AR29.

Primer set	Primer 1	Primer 2	Product Size (bases)	The position of primers (black arrow) to determine the possible orientation of pBA integration.
1	RightSB Rev	RightSP For	2,780	 (Integrated pBA located downstream of Arg-tRNA)
2	LeftSB For	LeftSP Rev	1,409	 (Integrated pBA located downstream of Arg-tRNA)

3	RightSB Rev	LeftSP Rev	1,417	 (Integrated pBA located upstream of Arg-tRNA)
4	LeftSB For	RightSP For	2,711	 (Integrated pBA located upstream of Arg-tRNA)
5	2SSP10 for	LeftSP Rev	2,283	 (Unintegrated plasmid)
6	LeftSB for	Intcheck rev	1,335	 (Integrated phage)
7	SSPro for2	Intcheck rev	3,754	 (Free phage with intact *attP*)

8	LeftSB For	RightSB Rev	264	 (Non recombine AR29)
9	2SSP18F For	2SSP18 Rev	880	 (Presence of Phage)
10	SSPro For2	LeftSP Rev	3,826	 (pBA integrates right of lysogenic phage genome)
11	RightSP for	Intcheck rev	3,850	 (pBA integrates left of lysogenic genome)

5.1.4: PCR Detection of Unrecombined *B. thetaiotaomicron* AR29 *attB* Site.

B. thetaiotaomicron AR29 cultures were tested using primer set 8 (Table 5.4), which were predicted to amplify a DNA fragment of 264 bp from the bacterial genome if the attB site was free of integrated phage ϕAR29 or plasmid pBA.

5.1.5: Approach to Curing Lysogenic Phage ϕAR29 from *B. thetaiotaomicron* AR29

5.1.5.1: Monitoring the removal of ϕAR29 by determining the presence of intact attP, and integrated phage ϕAR29 in AR29 by using PCR.

Two separate sets of cultures of *B. thetaiotaomicron* AR29 were grown in DF medium, at 39 °C and 44 °C respectively. Each culture was inoculated into fresh broth (0.05 mL into 5 mL), followed by 24 hours of growth, and the procedure was repeated the next for 56 days. Tests for the presence of bacteria with uninterrupted *attB* sites were performed every day prior to reinoculation. PCR with primer sets 6 and 7 was used to detect the presence of integrated or circularized phage ϕAR29, respectively (Table 5.4). A positive control culture was prepared in the same way as the test, except that clindamycin was included in the broth. Primer set 9 was used to detect the presence of phage, independent of its DNA form.

5.2: Results:

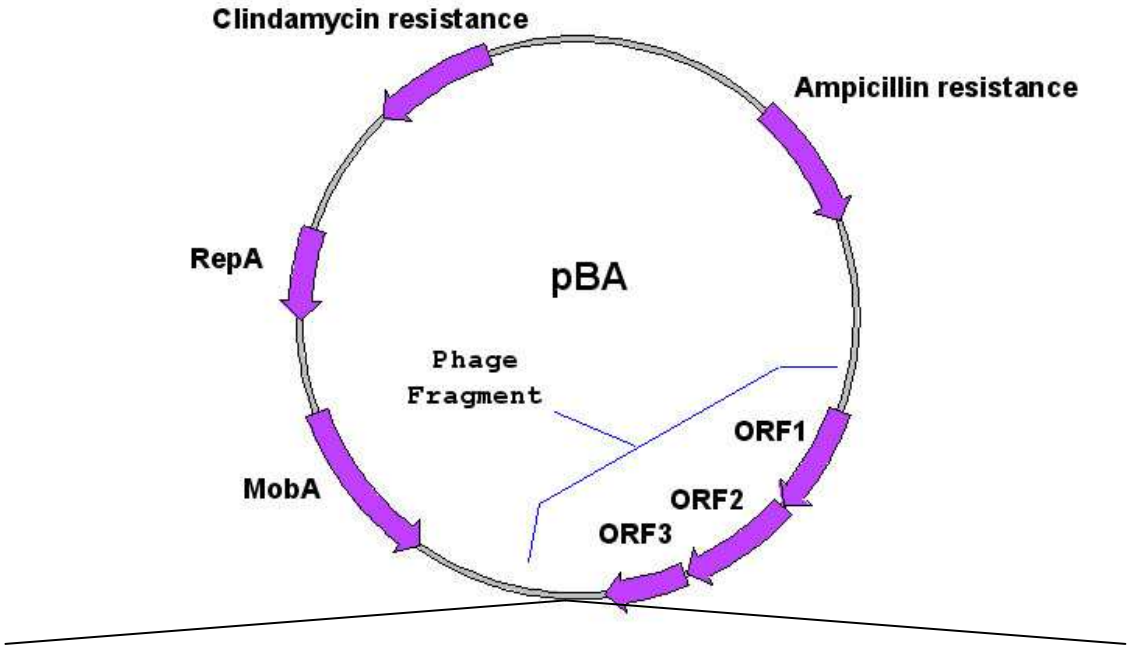
5.2.1: Identification of *Int* and *att* site in pBA

A review of pBA sequence revealed that the first 435 bases of the newly identified ϕAR29 *Int* gene were included in the phage DNA in pBA (Figure 5.5). In addition, the *attP* site was confirmed to be present in the plasmid, but at a location 117 bp from that previously hypothesised.

5.2.2:Expression of *Int* from the ϕAR29 Prophage

The results of RT-PCR studies on untransformed *B. thetaiotaomicron* AR29, for the detection of *Int* transcript, are presented in Figure 5.6 and show that mRNA transcribed

from the *Int* gene was present, as demonstrated by the presence of RT-PCR product that migrated according to the predicted size of 535 bp.



atgtgaccccggtgcggtctaatatttaaaaatgcacactatacaatctcatacaaatacatta
attatcagatatatacaaaacagcaagttgtattatgttgcattttagttgtgcaattaagcat
tttcgttgtgcaaaaacagtatatttgcacaaccgtaaaatgatagctat*atggcaacagtaaa
agcatttattcgatcaagcaaaaaggataactttgttaatatacgatttcgtttatccgatggt
aggaatatacaattatttcacacttccgagattttagtacaaccttcaatttgggatgagaaaa
aggaacaatacaaatccaaagtacttataccaatccattgcaaaactagagaagagttttataa
agacataacggaacgcaaaaacctaattctacgtttgtacaccgatcacaagatagaaaccagt
gaacaactaaatcaatatatagatgaatatattaatccggagaagtatattaacaaggaatcaa
gtgaaagcctccataaccgactcaacttatacatagagcaatgttataaggatggaatatttgg
cgaaggcagaaagaaacattatgatgtattattacg*tcaggctcattttatactcatatcaaca
acttccagtgatg

Figure 5.5: Nucleotide sequence of phage *Int* fragment cloned in pBA. Blue and black sequences indicate DNA derived from phage ϕAR29 and vector respectively. Yellow highlighted bases illustrate the *attP* site. The start codon for the 435 bases of *Int* gene that were cloned in pBA and the direction of transcription are represented by the asterisk and the black arrow, respectively.

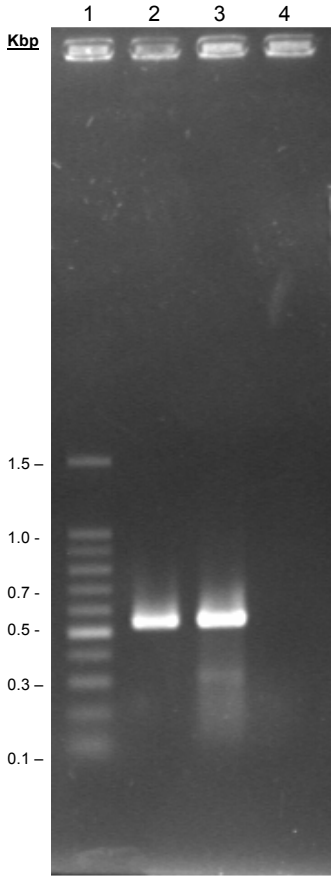


Figure 5.6: Agarose gel showing 535 bp product from RT-PCR of *B. thetaiotaomicron*.AR29 RNA, using primers Srec for and Srec rev to detect the presence of integrase mRNA.
Lane 1: 100 bp ladder; **Lane 2:** RT-PCR positive control using Int2 primers; **Lane 3:** RT-PCR using Int2 primers; **Lane 4:** PCR negative control using Int2 primers.

5.2.3: Test for Integration of pBA into AR20 and AR29 Genomes

The results of PCR using primer sets 1 – 4 to detect chromosomal integration of pBA are shown in Figure 5.7. No PCR products were detected from AR20. In contrast, reactions using AR29 DNA as a source of template produced bands with primer sets 1 and 2, but not with primer sets 3 or 4. The respective size of these bands corresponded to the predicted values of 2.78 kbp and 1.41 kbp.

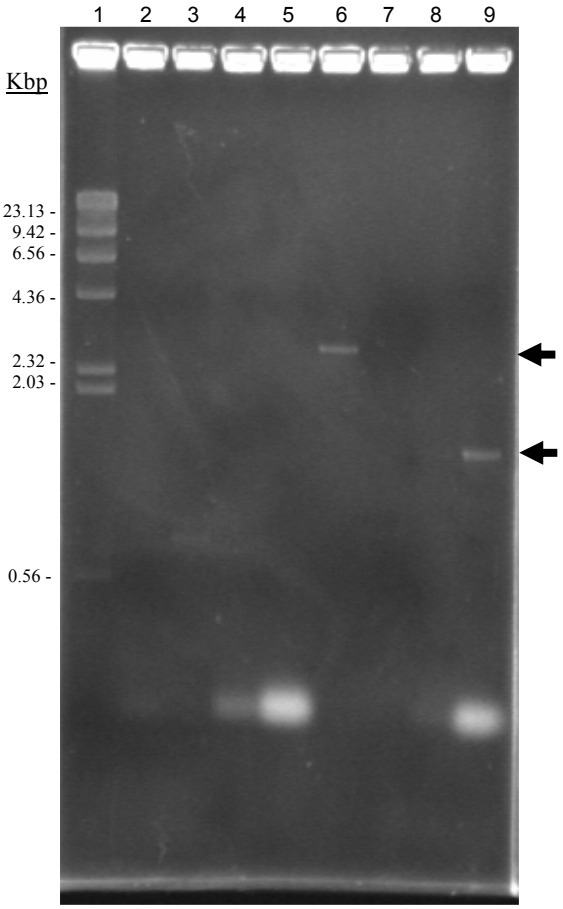


Figure 5.7: Agarose gel showing PCR tests for the detection of integrated plasmid pBA in AR20 and AR29
Lane 1: λ DNA digested with Hind III; **Lane 2:** PCR using primer set 1 on AR20; **Lane 3:** PCR using primer set 3 on AR20; **Lane 4:** PCR using primer set 4 on AR20; **Lane 5:** PCR using primer set 2 on AR20; **Lane 6:** PCR using primer set 1 on AR29; **Lane 7:** PCR using primer set 3 on AR29; **Lane 8:** PCR using primer set 4 on AR29. ; **Lane 9:** PCR using primer set 2 on AR29.

The smaller of the two PCR products from AR29 was sequenced (Figure 5.8), to confirm the integration of pBA into AR29 (Figure 5.9 A). To ensure that the sequencing result was derived from the *attP* site of the phage DNA fragment cloned in pBA, rather than from ϕAR29, primer LeftSP Rev was used to sequence the 1.41 kb PCR product to confirm the presence of the pBA RepA gene (Figure 5.9 B). The sequence from primer LeftSB For clearly illustrates that the bacterial *attB* region was replaced during integration by the phage core site and sequence upstream from the truncated *Int* gene in pBA. Results

from Left SP Rev confirmed the integration of pBA by showing the presence of 3' terminal sequences from the RepA gene of pBA within the PCR product (Figure 5.9 B).

5.2.4: Characterisation of pBA Integration and Excision of Lysogenic Phage ϕAR29

PCR tests were conducted to determine whether the lysogenic phage and integrated plasmid were integrated in tandem in the genome of AR29. Concurrently, attempts were also made to cure AR29 of phage ϕAR29.

The results of PCR analysis, following 24 hours of culture growth, are shown in Figure 5.10. The results indicated that the integrated forms of both pBA and ϕAR29 were present in a single AR29 culture. DNA products were also detected from PCR that was designed to detect bacterial *att* sequence that were not interrupted by integration of exogenous DNA. i.e. showing the presence of *attB* without plasmid or prophage integration. In addition, the presence of non-integrated plasmid was identified in all cultures. However, PCR products were not generated from primers designed to detect the integration of pBA adjacent to ϕAR29. The presence of integrated phage was confirmed by the predicted PCR products, from cultures grown with and without clindamycin at 39°C and 44°C. Interestingly, non-integrated phage (intact *attP*) was detected only in cultures grown at 39°C with or without clindamycin.

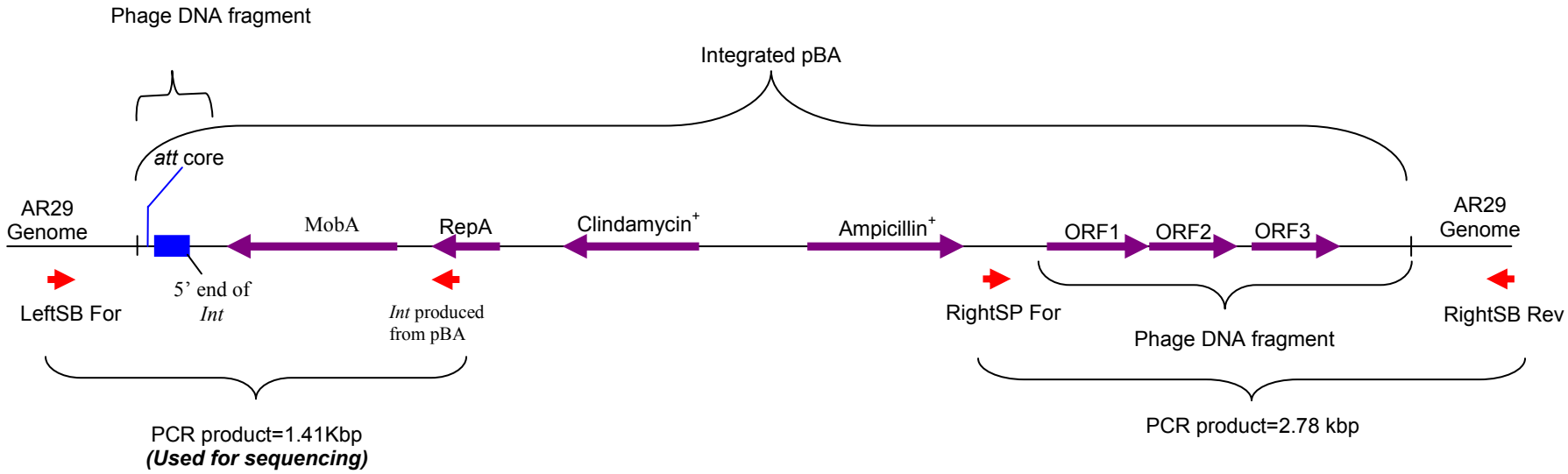


Figure 5.8: A diagram illustrating the position of primers (red arrows) used to detect the integration of pBA into the AR29 genome and their respective PCR product size. The 1.41Kbp product was sequenced to confirm integration of pBA.

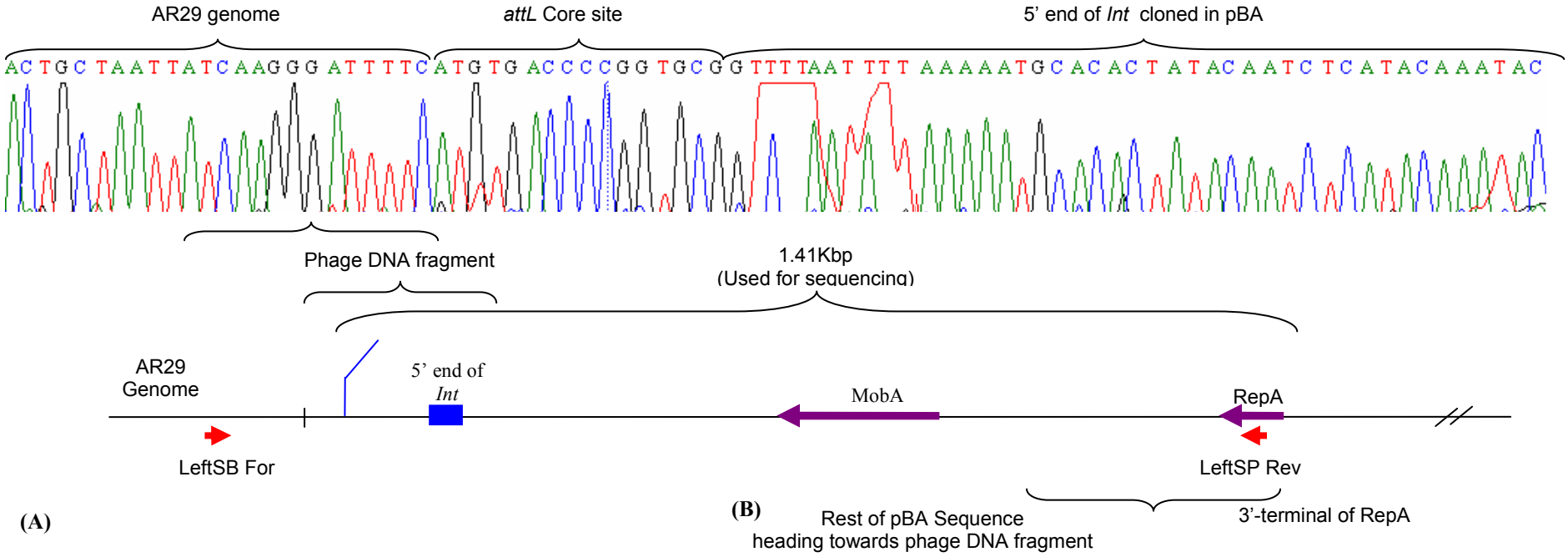


Figure 5.9: DNA Sequencing results confirming the integration of plasmid pBA into the AR29 chromosome. The arrows indicate the position of primers LeftSB For and LeftSP Rev used to amplify the 1.41 PCR product for sequencing. (A) The *attL* region containing the bacterial genome, core att site, and sequence upstream of the integrase gene cloned in pBA. (B) The PCR product was also sequenced using primer LeftSPRev, which shows the 3' terminal sequence of the RepA gene from pBA.

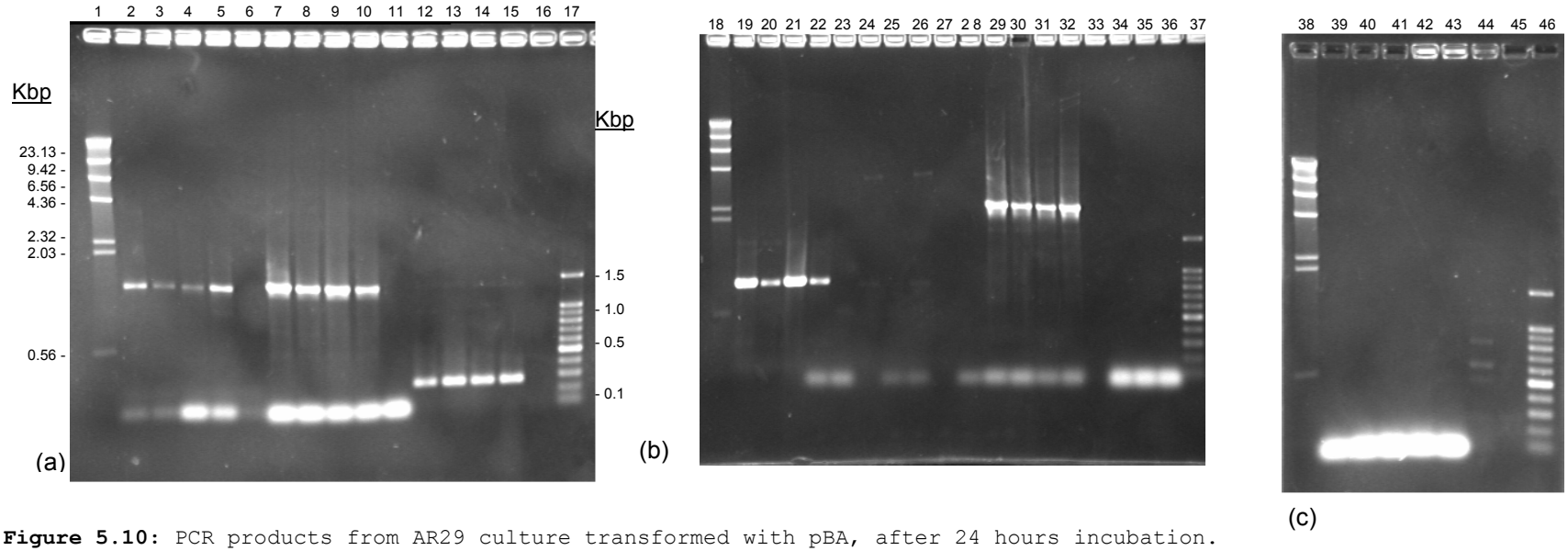


Figure 5.10: PCR products from AR29 culture transformed with pBA, after 24 hours incubation.
Lane 1, 18 and 38: λ DNA digested with Hind III; **Lane 17, 37 and 46:** 100 bp DNA Ladder;
Lanes 2 – 6: PCR to detect integrated phage in AR29 grown at: lane 2: 39°C, Lane 3: 44°C, Lane 4: 39°C with clindamycin, Lane 5: 44°C with clindamycin, Lane 6: negative control;
Lanes 7 – 11: PCR to detect integrated pBA in AR29 grown at: Lane 7: 39°C, Lane 8: 44°C, Lane 9: 39°C with clindamycin, Lane 10: 44°C with clindamycin, Lane 11: negative control;
Lanes 12 – 16: PCR to detect AR29 without integrants when grown at: Lane 12: 39°C, Lane 13: 44°C, Lane 14: 39°C with clindamycin, Lane 15: 44°C with clindaymin, Lane 16: negative control;
Lanes 19 – 23: PCR to detect the presence of phage in AR29(both as integrated and non-integrated) grown at: Lane 19: 39°C, Lane 20: 44°C, Lane 21: 39°C with clindamycin, Lane 22: 44°C with clindamycin, Lane 23: negative control;
Lanes 24 – 28: PCR to detect re-circularized phage in AR29 grown at: Lane 24: 39°C, Lane 25: 44°C, Lane 26: 39°C with clindamycin, Lane 27: 44°C with clindamycin, Lane 28: negative control;
Lanes 29 – 33: PCR to detect non-integrated plasmid in AR29 grown at: Lane 29: 39°C, Lane 30: 44°C, Lane 31: 39°C with clindamycin, Lane 32: 44°C with clindamycin, Lane 33: negative control;
Lanes 34 – 36, 39, 40: PCR to detect integration of pBA downstream of phage genome in AR29 at: Lane 34: 39°C, Lane 35: 44°C, Lane 36: 39°C with clindamycin, Lane 39: 44°C with clindamycin, Lane 40: negative control;
Lanes 41 – 45: PCR to detect integration of pBA upstream of phage genome in AR29 at: Lane 41: 39°C, Lane 42: 44°C, Lane 43: 39°C with clindamycin, Lane 44: 44°C with clindamycin, Lane 45: negative control.

After two months of sub-culturing, PCR tests failed to detect the presence of integrated phage in all cultures grown at 39°C or 44°C, with and without clindamycin (Figure 5.11: Lanes 2 – 5). Furthermore, results from primer sets 2 and 3 showed that these cultures contained integrated forms of pBA, but also contained AR29 with an uninterrupted *attB* site (Figure 5.11: Lanes 7 – 16). Nevertheless, although PCR did not detect integrated phage, primer set 9 was able to identify the presence of apparently non-integrated phage in cultures grown at 39°C with and without clindamycin and at 44°C grown without clindamycin. The presence of phage and excised phage with intact *attP* was not detected in AR29 grown at 44°C with clindamycin. This cured A29 culture was cryopreserved for future used.

5.2.5: Identification of Potential *attB* Site in Other Bacterial Genomes

In addition to Arg-tRNA(CGG) previously identified in *B. thetaiotaomicron* VPI-5482 in Chapter 4, BLAST analysis searching for sequences matching the AR29 *attB* site revealed similar arg-tRNA (CGG) genes in *Porphyromonas gingivalis* W83 and *Synechococcus elongatus* PCC 6301. Pairwise alignment of the tRNA genes, with that from AR29, found 85.7% and 76.7% identity, respectively. Although, BLAST analysis did not match AR29 *attB* to any sequence in *E. coli*, pairwise alignment of ϕAR29 *attB* with the *E. coli* Arg-tRNA (CGG) gene showed 58.4% identity.

Multiple alignment of tRNA genes, showed that the 16 bp *att* core of AR29 is conserved in VPI-5482, while W83 showed four base-substitutions in positions 1, 10, 13 and 15. Both PCC 6301 and *E. coli* have only 44% of the core sequence in common with ϕAR29 (Figure 5.12).

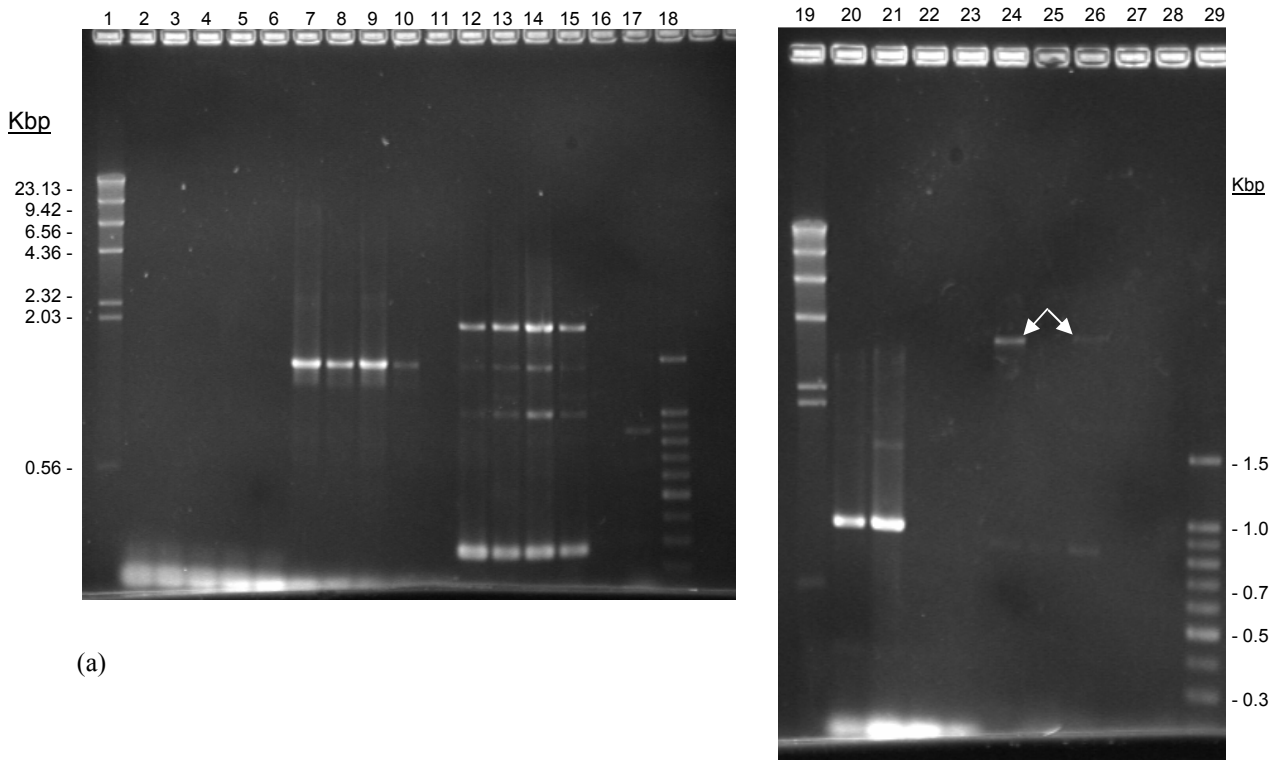


Figure 5.11: PCR to detect integrated pBA and prophage ϕAR29 in pBA-transformed AR29 cultures, following 2 months of sub-culturing.

Lane 1, 18 and 38: λ DNA digested with Hind III;
Lanes 2 – 6: PCR to detect integrated phage in AR29 grown at: lane 2: 39°C, Lane 3: 44°C, Lane 4: 39°C with clindamycin, Lane 5: 44°C with clindamycin, Lane 6: negative control;
Lanes 7 – 11: PCR to detect integrated pBA in AR29 grown at: Lane 7: 39°C, Lane 8: 44°C, Lane 9: 39°C with clindamycin, Lane 10: 44°C with clindamycin, Lane 11: negative control;
Lanes 12 – 16: PCR to detect AR29 without integrants when grown at: Lane 12: 39°C, Lane 13: 44°C, Lane 14: 39°C with clindamycin, Lane 15: 44°C with clindaymin, Lane 16: negative control;
Lane 18 and 29: 100 bp DNA Ladder;
Lanes 17, 20 – 23: PCR to detect the presence of phage in AR29 grown at: Lane 17: 39°C, Lane 20: 44°C, Lane 21: 39°C with clindamycin, Lane 22: 44°C with clindamycin, Lane 23: negative control;
Lanes 24 – 28: PCR to detect re-circularized phage in AR29 cultures grown at: Lane 24: 39°C, Lane 25: 44°C, Lane 26: 39°C with clindamycin, Lane 27: 44°C with clindamycin, Lane 28: negative control;

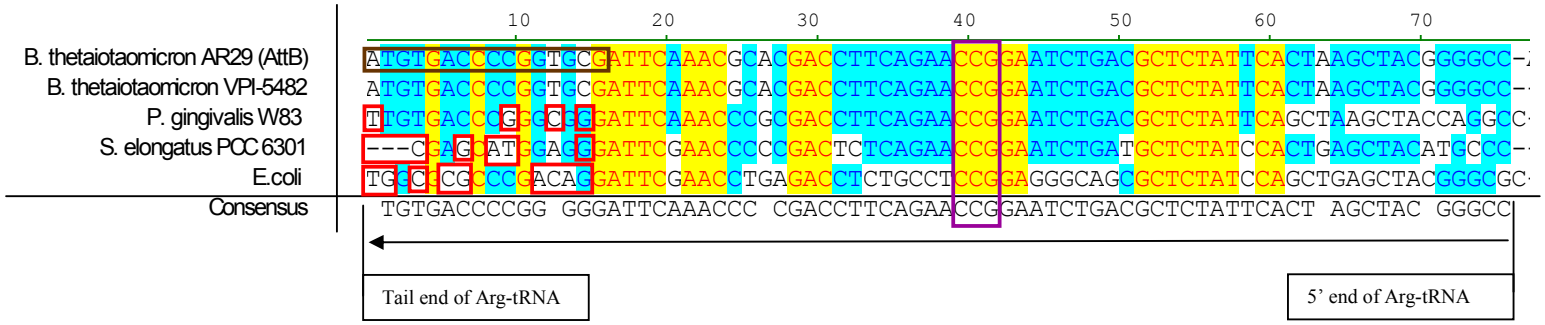


Figure 5.12: Multiple Alignment of ϕAR29 *att*B Arg-tRNA(CGG) gene with similar genes found in *B. thetaiotaomicron* VPI-5482, *P. gingivalis* W83, *S. elongatus* PCC 6301 and *E. coli*. The black arrows indicate the direction of transcription of arginine tRNA. The brown box surrounds the 16-base core attachment sequence, while the pink boxed region indicates the anticodon region of the tRNA. The *att* core nucleotides surrounded by red boxes indicate base differences from the core region of AR29.

5.3: Discussion

In the previous chapter, ϕAR29 *Int* was described as belonging to the tyrosine family of recombinases that participate in the formation of Holliday junctions during the integration process. Although only 436 bp of the 5'-terminal of the ϕAR29 integrase gene was cloned in pBA, experiments have shown the integration of plasmid pBA into the genome of AR29. In contrast, the same PCR process did not to detect the integration of pBA into the chromosome of *B.uniformis* AR20. Four possible reasons for the failure to detect integration into the AR20 genome are:

1) the absence of a suitable *attB* site in AR20,
2) oligonucleotide priming on the AR20 genome was too weak for PCR to occur, due to dissimilarities in the genomic sequences of AR29 and AR20,
3) integration may have occurred at a different location from that in AR29, leaving no functional priming site on the flanking genomic DNA, or
4) Absence of a functional integrase protein.

Due to a lack of knowledge about the AR20 genome there is no evidence to support or disprove the first three proposals. These three possibilities could be tested by hybridisation of restriction-digested pBA-transformed AR20 genomic DNA, with probes from the appropriate regions of the ϕAR29 genome. The absence of ϕAR29 prophage from AR20 and the incomplete cloning of the confirmed *Int* gene in pBA, suggest that the fourth reason remains a possible explanation.

5.3.1: Presence of *Int* in AR29 Untransformed with pBA

The removal of ϕAR29 from the bacterial genome and the integration of pBA both require the presence of *Int* protein. RT-PCR analysis on untransformed AR29 containing the

prophage has shown that the newly identified *Int* gene (Chapter 4) was transcribed in the host. It has not confirmed experimentally whether the *Int* mRNA was translated in those culture, but in prokaryotes transcription and translation are generally coupled processes (Pastushok and Kennell, 1974; Hondel *et al.*, 1975; Danchin *et al.*, 1981; Chen and Zubay, 1983; Aksoy *et al.*, 1984; Gowrishankar and Harinarayanan, 2004). If ϕAR29 gene expression is a coupled process, then translation to produce integrase in untransformed AR29 cultures can be assumed. However, it remains uncertain whether the detected transcript was derived from prophage or possibly from free phage released by induction, that could be re-infecting cure cells. Since, PCR analyses were able to detect presence of integrated phage, intact *attB* and intact *attP*, within the population as a whole, both mechanisms could occur.

5.3.2: Absence of Co-Integrated pBA and ϕAR29

Amplification experiments using primer sets 1 and 2 showed that the integrated pBA plasmid was located downstream of the tRNA gene, as predicted from the integration point of ϕAR29. The results also revealed that integrated phage and integrated pBA were able to co-exist in a culture. In theory, the presence of integrated phage and integrated vector might be explained by the insertion of pBA adjacent to ϕAR29 at the re-formed tRNA gene, but this was not observed from the PCR results. Studies on the integration mechanism of mycobacteriophage L5 have shown that the insertion of *attP* into the *attB* site disrupts the re-formed attachment site (Pena *et al.*, 1999; Pena *et al.*, 2000; Saviola and Bishai, 2004): Dr Beatrice Saviola pers. comm.). Disruption of *attB* is brought about by the inheritance of an integrase arm-type binding site (P) that normally flanks the core region of the *attP* site (Figure 5.12 A). As with phage λ, L5 "P" sites interact with the *Int* protein and

IHF to induce DNA folding that allows the formation of an integrative intasome (Pena *et al.,* 1997). However, when these binding sites are present at a reformed *attB*, folding of bacterial genomic DNA blocks the interaction of *attP* sequences with the reformed *attB* site. In phage L5 P sites, P4 and P5, interact with the integrase and IHF to form intramolecular bridges with the core of *attP* (Figure 5.13B; Pena *et al.,* 1999; Pena *et al.,* 2000). When integrated, P4 and P5 were found to create the same folding in the re-formed bacterial attachment site, thus preventing subsequent integration of additional L5 genomes (Figure 5.13 C).

The failure of PCR to detect the presence of integrated pBA adjacent to ϕAR29 genome despite of their co-existence, may suggested that the integrated page and integrated pBA are presence in different cells within a population. If timer permits, culture should be plated out on agar medium and individual colonies should be tested for the presence of integrated pBA and AR29. This may also provide the frequencies of occurrence of the different possible integration events that is happening within the culture.

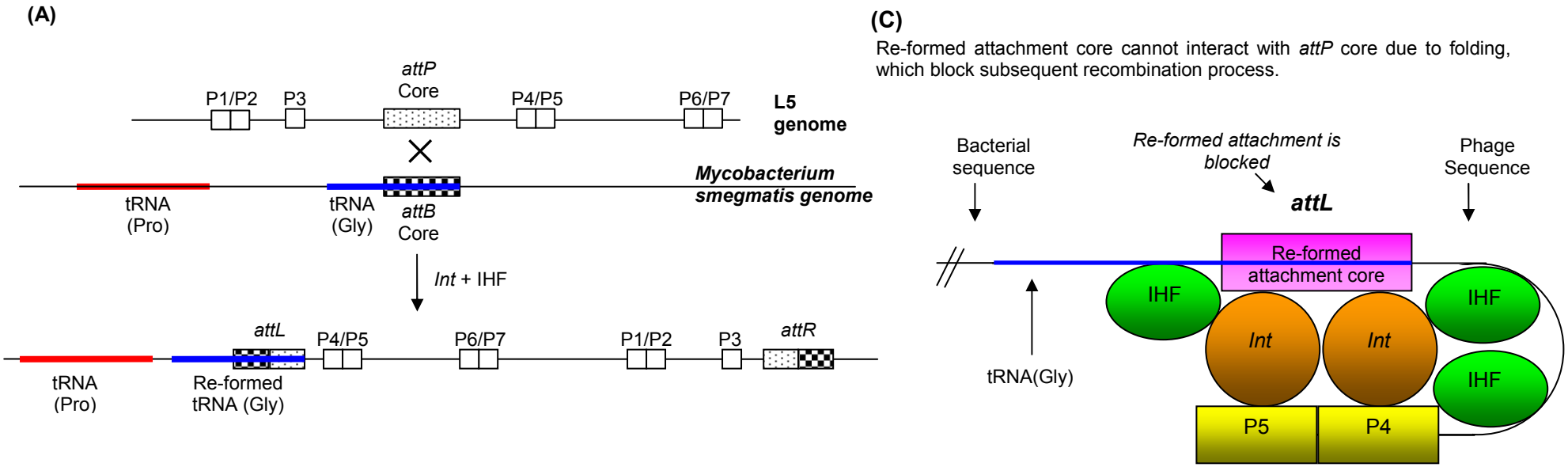
5.3.3: Detection of intact *attB* and *attP*

Bacteria that appeared free of integrated phage or plasmid, were detected within all cultures of pBA-transformed AR29 for the two-month duration of the experiment. In contrast to previous findings (Figure 5.14; Klieve *et al.,* 1989), the results obtained suggests that not all AR29 cells in a culture necessarily possess the lysogenic ϕAR29. It appears unlikely that PCR products of apparently non-lysogenised bacteria are derived from DNA remnants of lysed AR29, since:

1. the high endonuclease level of AR29 would be expected to cause rapid DNA degradation following the phage lytic cycle.

2. the presence of integrated plasmid in some cells requires that viable cells with an intact *attB* site were present in the culture at some stage.

The detection of intact *attB* after 24 hours of growth indicated unlysogenised cells were present at very early stages. Although the PCR analysis shows existence of intact *attB*, it is not certain whether this represents cells completely cured of phage (i.e. no phage in the cell) or from cured cells that contain replicating or packaged free phage. It is possible that unlysogenised cells may exist continually as a small percentage of the population isolate. Time permitting, plating out the mixed cultures and examining individual colonies by PCR should clarify whether some cells within the population are cured of phage.



(B) The formation of L5 integrative intasome at *attP* core prior integration into *attB*..

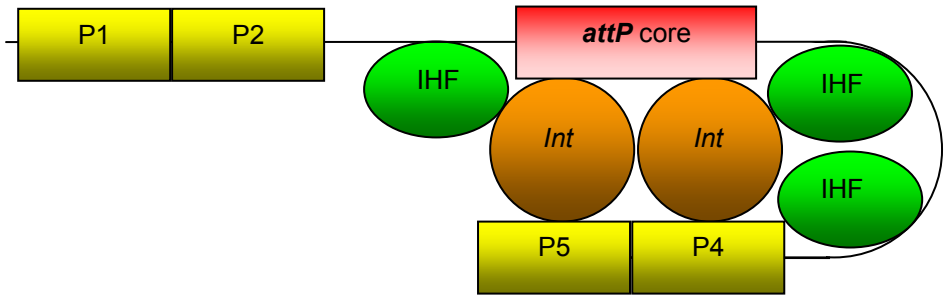


Figure 5.13: Integration mechanism of mycobacteriophage L5. L5 contains arm-type binding sites (P1-P7) flanking the *attP* region. L5 integrates into the 3' end of a Gly-tRNA gene (A). Analysis of *attP* revealed that two arm-type binding sites (P4-P5) interact with the attachment core sequence during the integration process, which induces the folding of the *attP* region to form the intasome complex (B). However, following integration these arm-type binding sites are located downstream, near the re-formed attachment site. Due to their ability to interact with core sequence, the integrated form of the P4 and P5 will bind to the core region in the re-formed attachment site, thus preventing subsequent integration by a 2nd copy of the L5 genome (C). (Pena *et al.*, 1997; Pena *et al.*, 1999; Pena *et al.*, 2000)

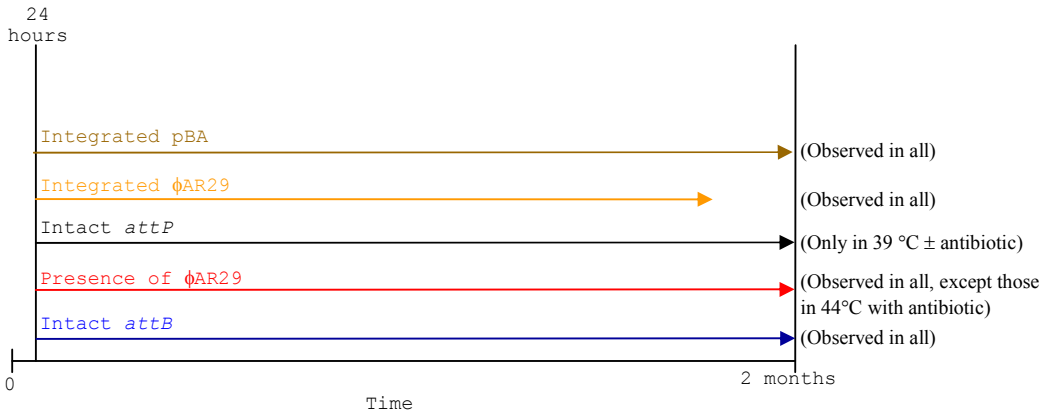


Figure 5.14: A diagram illustrating the extent to which PCR detected integrated plasmid pBA, integrated ϕAR29, presence of phage and intact *attP* and *attB* sites.

PCR analysis also identified free phage within the culture. The presence of intact *attP* sequences could be derived from:

1. excised phage ϕAR29 forming a circular genome
2. circularized or linear genome of infecting ϕAR29, or
3. remnants of DNA from circularized or linear ϕAR29 genome. This is regarded as unlikely because of the high levels of bacterial nucleases.

Although all cultures, except the cured sample, showed the presence of phage DNA, phage with an intact *attP* site were found only in cultures grown at 39 °C (+/-) antibiotic. If ϕAR29 is indeed produced as terminally redundant DNA molecules, then PCR across the *attP* site should have amplified that region from any culture containing intact, free phage. It is unclear why intact *attP* was not observed in cultures grown at 44°C throughout the duration of the experiment.

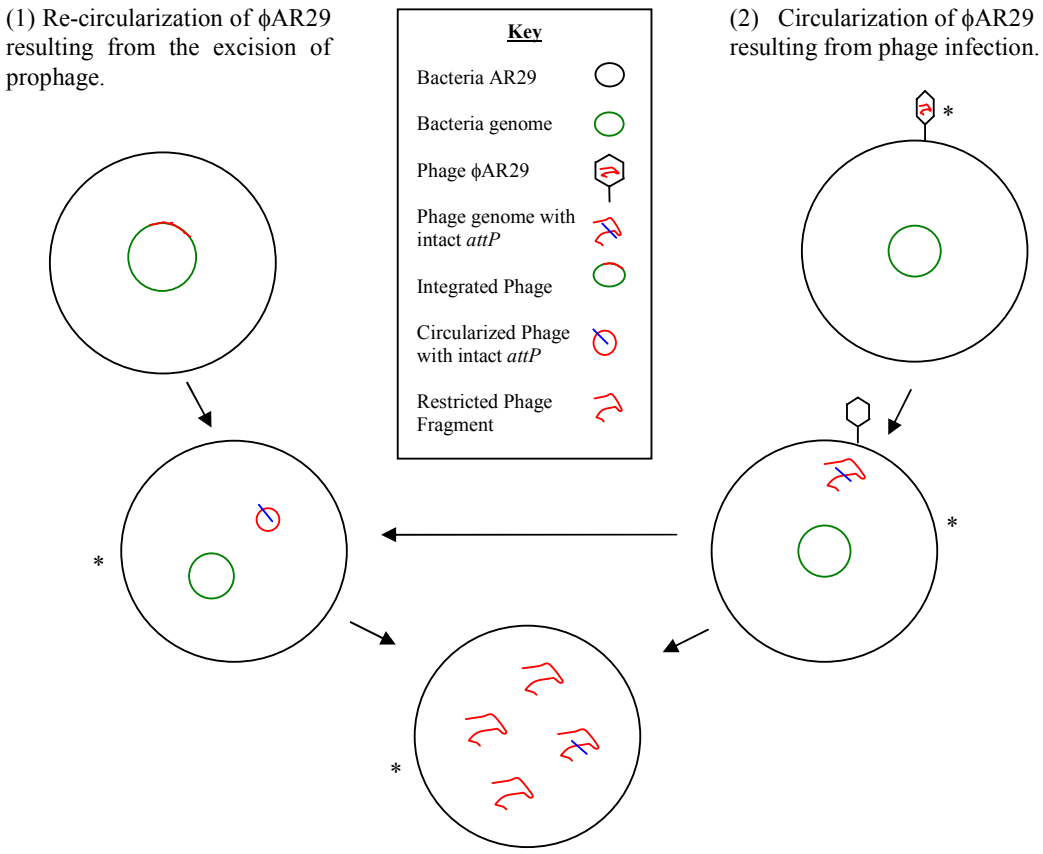


Figure 5.15: Possible pathways from which phage with an intact *attP* could be detected (*).

5.3.4: Detection of integrated phage and integrated pBA

Prophage ϕAR29 was present throughout most of the analyses, but PCR failed to detect integrated phage a few weeks prior to the termination of monitoring. In cultures grown at 44 °C the prophage was completely excised only a week earlier than those cultures grown at 39 °C. The detection of PCR product from intact phage but not from prophage, at the end of the two months suggests that the phage could have been packaged, circularised or existing as a linear free entity.

The presence of integrated pBA after just 24 hours of growth indicated that integration occurred relatively early in its occupation of AR29. It is proposed that

integration of pBA may occur within the first 4 culture cycles prior to the first PCR analysis. This could have occurred within the stages:

1. growth of transformed cells on plates for 48 hours
2. culture of transformants from the plates in overnight cultures, which were cryopreserved
3. Inoculation of cryopreserved cells into new medium and growth overnight (24 hours)
4. Inoculation from overnight culture into fresh broth and growth for another 24 hours prior to first PCR analysis.

5.3.5: Mechanisms for Excision of ϕAR29, Integration of pBA, and Curing of AR29

In the past, the use of acridine orange, UV irradiation, and mitomycin C induction, failed to cure AR29 of the ϕAR29 prophage (Klieve *et al.,* 1989). The studies described here showed that daily subculturing, for two months after transformation with pBA, produced cultures with no detectable integrated form of ϕAR29. However, in most cultures, non-integrated phage sequences were detectable throughout the experiment. Complete elimination of phage ϕAR29 from strain AR29 within the two-month period available, was achieved only in cultures grown at 44°C in the presence of pBA and clindamycin. Due to the lack of further experimental data, the mechanisms of excision of ϕAR29 and integration of pBA are not fully understood. Nevertheless, proposals can be made on possible mechanisms for the integration and excision processes.

Three models that might explain the excision of prophage and integration of pBA are:

1. Both *Xis* and *Int* could be expressed from prophage
2. *Xis* could be expressed from pBA and *Int* expressed from prophage
3. Both *Xis* and *Int* could be expressed from pBA

Model 1:

One possible mechanism for the excision of prophage and replacement with pBA, is for the phage to re-enter the lytic cycle (Figure 5.16 C). During this state, prophage should be removed from the host, leaving unlysogenised AR29 with an intact *attB* site for plasmid integration. Once the lysogen was removed, the observed integration of plasmid pBA into AR29 would be possible through the presence of an *attP* site in pBA. However, excisionase produced the from the prophage generally initiates an irreversible lytic cycle (Voet and Voet, 1990; Little *et al.*, 1999). As described in section 1.4.3.2.1, the lytic conversion of a λ prophage involves the host SOS response, which produces *RecA* protein that cleaves the lysogenic maintenance protein, *cI* repressor. This allows the transcription of *Xis* and other proteins, such as structure and packaging proteins, required for the completion of the lytic cycle. Since ϕAR29 possesses a gene that is homologous with known *cI* protein genes, the phage ϕAR29 repressor could be cleaved by a similar SOS response in AR29. Therefore, like λ, excised prophage ϕAR29 would most probably be packaged and lyse its host, rather than be re-integrated.

Despite this, packaged phage might not fatally lyse an encapsulated host cell, in which the capsule could maintain cellular integrity, and the bacterium may possibly survive the lytic reaction, as suggested in (section 5.3.3). In conditions such as low nutrient availability, reduction of capsule thickness can occur. Subsequent re-infection of these cells with ϕAR29, or of cell previously unlysogenised, could promote the integration of one of the more numerous pBA molecules into the genome. Free plasmid was shown to have copy numbers in the range 4 – 16 copies/cell (Wong *et al.*, 2003) while the phage is most likely to be present as a single DNA molecule. With the production of integrase enzyme, the relative copy numbers may increase the probability

of pBA integrating, rather than a phage genome. Although the number of integrated plasmids in each cell was not measured, the changes in *attB* during the integration process make multiple integration unlikely.

Model 2:

This model is based on the assumption that *Xis* and *Int* proteins are produced from pBA and prophage, respectively. The basis for this model is derived from the different observations from Klieve *et. al.,* (1989) who found no evidence for phage curing, and from this study. A major difference between the cultures used in these two studies is that pBA transformed AR29 was used in the present study. This suggests that the presence of pBA within the cells may be a contributing factor for the excision of ϕAR29 and integration of plasmid. In the absence of a clear alternative, the ORF previously identified as encoding *Xis* remains the most likely source of the enzyme responsible for excisive recombination.

In the presence of integrase expressed from prophage for maintaining lysogeny, expression of *Xis* from pBA, could provide the conditions for excision of the prophage, without induction of cellular SOS systems. The same balance would prevent the re-insertion of viral DNA into the chromosome of the host (Figure 5.16 B). Similarly, such a situation would probably prevent integration of pBA due to continued expression of the *Xis* protein.

In this model, since prophage was apparently excised without induction of *RecA*, and therefore without induction of the lytic cycle, it is possible that excised prophage may remain in an intermediate stage between lysogenic and lytic cycles, through the inhibition of gene expression by *cI* repressor protein. Until induction of a lytic cycle is prompted, by expression of *RecA* to cleave *cI*, the excised prophage might continue to synthesise *Int* protein. Under these conditions, a reduction in *Xis* production from pBA,

perhaps through a decline in plasmid copy number, could shift the balance towards integration of either pBA or reintegration of ϕAR29. The origin of transcription of *Xis* mRNA in pBA remains unknown. When integrated into the AR29 genome, the expression of *Xis* would be expected to trigger excision of the plasmid. Experiments to isolate cultures with integrated plasmid, without the presence of phage, would allow expression studies on the integrated plasmid-borne genes to be conducted.

To determine whether the presence of pBA in AR29 does indeed have an effect on the excision of ϕAR29, a similar PCR analysis experiment could be performed on untransformed cultures of the lysogenised AR29. In such a study, the demonstration of intact *attB* sequences, with progressive loss of prophage from the genome, would indicate that unlysogenised host cells can persist, within a culture. This would also imply that pBA may not contribute to the excision of prophage. Alternatively, the absence of intact *attB* sequences, and failure of prophage to be displaced from the genome would confirm that pBA is a likely contributing factor in the excision of prophage. To eliminate the possibility of results being confused by persistence of DNA fragments from lysed cells, future experiments could be performed on cells that have been washed in saline and DNase treated, to remove DNA remnants from the outside of the cells.

Model 3:

The *Int* and *Xis* proteins may both be expressed from pBA. Although the sequence of the newly identified *Int* gene shows strong similarity to other known integrases, confirmation that the protein functions as predicted will require functional studies. This, together with the apparent mosaic structure of the phage genome, dictates that the previously proposed *Int* gene must also be considered as a possible source of integrase activity. The production of *Int* and *Xis* from pBA would be likely to result in

excision of both phage and plasmid. The excised prophage may persist as a dormant form, as proposed in model 2. If the previously proposed integrase protein can catalyse the insertion of pBA, the total levels of integrase would be raised since integrase activity could be transcribed and translated from both the phage and the plasmid. This would alter the *Xis/Int* ratio and promote a favourable integrative reaction. Both excised prophage and free plasmid would have the potential for integration into the host. On the other hand, if prophage was stimulated to re-enter a lytic cycle, this will favour an excisive reaction, since phage and pBA would both be producing the *Xis* protein.

5.3.6: The Identification of Potential *attB* Sites in Other Organisms

BLAST analysis and sequence alignment of Arg-tRNA(CGG) genes from other sources has identified a potential *attB* site in *B. thetaiotaomicron* VPI-5482, for integrative vectors derived from the ϕAR29 recombination mechanism. Although, *P. gingivalis* W83 shared strong homology with the Arg-tRNA found in AR29, only 75% of the 16 base sequence in the core region was conserved. Experiments using an integrative vector with ϕAR29 *attP* and *Int* protein would be required to determine whether the tRNA genes in VPI-5482 or W83 could be used as *attB*. Other factors that may prevent plasmid integrating into these bacteria are the absence of integrase (arm-type) binding sites in the sequence surrounding the *attB* core, and the absence of the appropriate IHF.

The tRNA gene from *S. elongatus* PCC 6301 and *E. coli* revealed a low percentage of identity (44%) to the core region. This would be likely to prevent the ϕAR29 integration mechanism from recognising the tRNA as an *attB* site and is concluded to be a key reason for not observing pBA integration in *E. coli*.

5.4: Conclusions

Removal of the ϕAR29 prophage and the integration of pBA into the AR29 genome was demonstrated in these experiments. This has provided confirmation of the precise identity and location of *attP* and *attB* sites. Sequence analysis has also revealed potential *attB* sites in other bacteria, in which pBA might be used. A culture of AR29 containing cells with an intact, unrecombined *attB* site and cells with integrated pBA, has been isolated. This should allow separation of the two forms, which can be used for future studies. The strain with integrated pBA will be useful for future excision studies, while the fully cured strain will provide a host for future integration vector studies and bacteriophage life-cycle studies. Despite the large amount of data obtained here, fundamental understanding on how pBA and ϕAR29 are integrated and excised remains elusive. In further investigations on the lysogenic and lytic cycles of ϕAR29, functional identification of the excisionase gene and the original hypothetical integrase gene will be essential.

Chapter 6: General Conclusion and Closing Remarks

The development of transformation vectors based on bacteriophage lysogenic mechanisms has been reported in many publications (Kuhstoss *et al.*, 1989; Lee *et al.*, 1991a; Lee *et al.*, 1991b; Wuenscher *et al.*, 1991; Raya *et al.*, 1992; Lee and Hatfull, 1993; Dupont *et al.*, 1995; Auvray *et al.*, 1997; Alvarez *et al.*, 1998; Freitas-Vieira *et al.*, 1998; Van Mellaert *et al.*, 1998; Kaatz *et al.*, 1999; Moreau *et al.*, 1999; Shimizu-Kadota *et al.*, 2000; Wisniewski *et al.*, 2000; Springer *et al.*, 2001; Combes *et al.*, 2002; Lauer *et al.*, 2002; Gregory *et al.*, 2003; Li *et al.*, 2003; Saviola and Bishai, 2004). The initial aim of the project was to provide a better understanding of the lysogenic mechanism of ϕAR29 to address the lack of developments in site specific integration vectors for rumen bacteria. This initial plan included:

- determining whether the *Int* and *Xis* ORFs cloned in pBA were transcribed in various hosts,
- examining the similarity of the proposed *Int* and *Xis* ORFs to integrase and excisionase genes reported in gene/protein databanks,
- confirm the location of the *attP* site and investigate the integration process in more detail.

It became apparent during the early phase of the project that considerable benefit would result from sequencing the ϕAR29 genome, which would allow the genetic components of the phage to be examined and could identify other possible recombination genes.

6.0: Characterisation of the integration module of bacteriophage ϕAR29, cloned in pBA

Success in developing integration vectors depends heavily on molecular knowledge of the virus and host concerned, including genome structure, genes responsible for the recombination process and the location and nature of *att* sites. In this study, the sequencing of bacteriophage ϕAR29 genome has allowed confident identification of an integrase gene and the attachment sites, *attP* and *attB*, that are involved in the phage integrative recombination process. There was no experimental confirmation that the ORFs identified within the previously cloned fragment (Gregg *et al.*, 1994a) were the *Int* and *Xis* gene that mediate integration and excision processes (Gregg *et al.*, 1994b; Wong *et al.*, 2003). In addition, previous investigations had revealed no evidence of pBA integrating into the genome of AR29. Therefore an essential aim of the project was to determine whether the ϕAR29 ORFs present in pBA were transcribed in AR29 or *E. coli* and to clarify the identification of the proposed *Int* and *Xis* genes.

As with previous findings (Gregg *et al.*, 1994a), only poor homology was detected when the proposed *Int* and *Xis* ORFs in pBA were aligned against the now far more numerous *Int* and *Xis* gene and protein sequences. This is despite the fact that in the intervening years, the numbers of sequences available have risen from 5 excisionases and 8 integrases to a current 26 excisionases and 50 integrases.

The possibility that the cloned phage fragment may not contain the complete integration module was emphasised when an *Int* gene with considerable similarity to other *Int* genes was identified close to the proposed *attP* site. Despite uncertainty about integrase activity in pBA, the plasmid was integrated into AR29 genome, although the precise mechanism remains unresolved. One possible explanation is that integrase activity was

provided from another source, such as non-integrated ϕAR29. Although the newly identified *Int* is viewed as the most likely candidate for the integrative process, the involvement of the previously proposed *Int* gene in pBA cannot be ruled out until functional studies are conducted. However, DNA sequence from the integration borders of pBA confirmed the precise location of the *attP*/*attB* sites and the orientation of plasmid integration.

6.1: Complexity of the Phage genomic sequence

Information gathered from the genome sequence of ϕAR29 has help provided a better understanding to the integrative mechanism of the phage. In addition, the complete genomic sequence has revealed much about the gene arrangement in the phage and has identified potential regulatory elements controlling the ϕAR29 lytic and lysogenic cycles.

The majority of the 53 ORFs have little or no resemblance to other reported genes. However, this is not unexpected. The complete sequence of 10 new mycobacteriophage genomes by Pedulla *et al.,* (Pedulla *et al.*, 2003b; Pedulla *et al.*, 2003a) has revealed that over 50% of the ORFs in these genomes are unrelated to any sequences in the GenBank. Recent work on sequencing of phage from sea water produced a similar outcome (Rohwer, 2003). It has been suggested that less than 0.0002% of the global phage metagenome has been sampled (Rohwer, 2003) which must contribute to the difficulties of identifying ORFs in phage genomes.

6.1.1: Mosaic of the phage genome

Bioinformatic analysis of ϕAR29 has demonstrated that small portions of the genome resemble specific functional regions of bacteriophages from *E. coli*,

Staphylococcus, *Lactobacillus* and *Pseudomonas*, whilst some ϕAR29 genes appear more closely related to genes from its host. Data presented here clearly support the proposed model that phage genomes are a mosaic built from components called genetic modules (Wang *et al.*, 2004; Campbell, 2003). These designated modules can be functional segments of varying size, ranging from whole genomic segments to single genes or gene domains. Such modules are thought to be available from a large common genetic pool, in which access to the pool is not uniform among phages (Campbell, 2003). Through a variety of mechanisms, including interchanging of modules by homologous recombination, transduction and horizontal gene transfer, bacteriophages are able to acquire new genetic information, resulting in unrelated bacteriophages possessing homologous genes(Loessner *et al.,* 2000).

The lack of similarity of many bacteriophage ϕAR29 genes to those of other phages, and the mosaicism of phage genomes, has made it difficult to suggest how ϕAR29 might have evolved. This is further complicated by the lack, in phage genomes, of any taxonomically convenient markers, such as the 16S ribosomal genes that have served as a taxonomic tool of cellular microorganisms (Weinbauer and Rassoulzadegan, 2004; Rohwer and Edwards, 2002; Nelson, 2004). The 3 decades old International Committee on Taxonomy of Viruses (ICTV) phage taxonomic classification scheme has been inadequate in classifying phage (Nelson, 2004). Of the total phage genome presence in the GenBank at the end of 2004, approximately 40% of phage are unclassified beyond the level of family and 10% are not assigned to an order, using the ICTV taxonomic scheme (Nelson, 2004).

6.2: Current approach of bacteriophage Integrative vectors in mammalian cells as tools for gene therapy.

Both tyrosine and serine integrase families have been targeted as tools for gene therapy in mammalian cells.

6.2.1: Tyrosine integrases

Generally tyrosine-based integrases, like those of lambdoid phage, have not been seen as candidates for gene therapy research in mammalian cells and development in this area is still in its infancy. This is largely due to their requirements for supercoiled DNA, host factors and the large size of the *attP* site (Groth and Calos, 2004). Nevertheless, the tyrosine family does include integrases that do not require host factors, such as Cre and FLP (O'Gorman *et al.*, 1991; Sauer and Henderson, 1988; Sauer, 1994). As a result they are able to carry out deletion and translocation-type recombinations in higher eukaryotic cells (O'Gorman *et al.*, 1991; Sauer and Henderson, 1988; Sauer, 1994).

It has been reported that the wild-type integrase of phage HK022 can perform integrative recombination in mouse NIH3T3 cells (Kolot *et al.,* 1999), COS1 monkey cells (Kolot and Yagil, 2003) and human embryonic kidney cell line 293T (Kolot *et al.,* 2003). Interestingly, HK022 integrase was able to carry out excision of integrated plasmid without the presence of IHF and excisionase protein in mammalian cells (Kolot *et al.,* 1999; Kolot *et al.,* 2003).

Recently two λ phage integrase mutants, Int-h and Int-h/218, were developed to function without the need for IHF or supercolied DNA (Lorbach *et al.,* 2000). Both these enzymes contain a E174K mutation, but the Int-h/218 mutant has an additional E218K mutation. Int-h and Int-h/218 have been shown to catalyse integration into human BL60 and HeLa cells at a frequency of 16% and 6-30%, respectively (Lorbach *et al.,* 2000).

6.2.2: Serine integrases

The requirement by serine-based integrases for a simple attachment site and the ability to function autonomously without IHF have made them a more obvious choice for gene therapy. Serine integrases also have the advantage that they insert into a limited number of sites compared to the randomly integrating vectors derived from retroviruses. There are several ways in which enzyme such ϕC31, R4 and TP901-1 could be used:

1. The creation of an attachment site by insertion of an *attB* or *attP* site into the genome of an organism.

   Although this is an imprecise and low frequency event, the establishment of an *attB* or *attP* site in the eukaryotic chromosome allows plasmids carrying *attP* or *attB* and an integrase gene to be integrated at high efficiency. Both ϕC31 and R4 integrases have been used by this method in both human and mouse cells (Olivares *et al.,* 2001; Thyagarajan *et al.,* 2001; Quenneville *et al.,* 2004).

2. By cassette exchange.

   This is achieved by creating an integrative cassette that contains an insert flanked with *attB* sites which will integrate into a site on the mammalian chromosome that is flanked by *attP sites* (Figure 6.1). The application of this method was successfully achieved in mouse ES cells (Belteki *et al.,* 2003).

3. Naturally occurring pseudo *attP* sites in mammalian cells can be used as the integration site for plasmid carrying an integrase gene and an *attB* site.

Using phage site-specific recombination mechanism as a tool for gene therapy Ortiz-Urda *et al.* (2002) were able to restore the function of human Type VII collagen (COL7A1) protein in skin cells from recessive dystrophic epidermolysis bullosa (EB) patients. This was achieved by integrating a plasmid carrying ϕC31 integrase, a COL7A1 gene and *attB* site into a pseudo-*attP* site, HpsA, in human skin cells that were grafted onto immune-deficient mice (Ortiz-Urda *et al.*, 2002; Ortiz-Urda *et al.*, 2003b; Ortiz-Urda *et al.*, 2003a). Similar results were also obtained in studies that involved the correction of human skin cells and fibroblasts lacking laminin B3 protein and COL7A1 proteins, respectively (Ortiz-Urda *et al.*, 2002; Ortiz-Urda *et al.*, 2003b; Ortiz-Urda *et al.*, 2003a). Despite success in animal and *in vitro* models, the approach failed to proceed beyond Phase I study on EB patients (Quenneville *et al.*, 2004; Bauer and Laimer, 2004).

Studies by Olivare *et al.* (2002) have demonstrated that the application of gene therapy using ϕC31 integrase can create transgenic mammalian cells that are able to secrete therapeutic proteins. The studies showed that the human coagulation factor IX (hFIX) gene was integrated into a pseudo-*attP* site at MpsL1 and MpsL2 sites in liver cells of mice. The hFIX gene was capable of producing and secreting therapeutic levels of proteins (4 μg/mL) throughout the duration of the 8-month experiment.

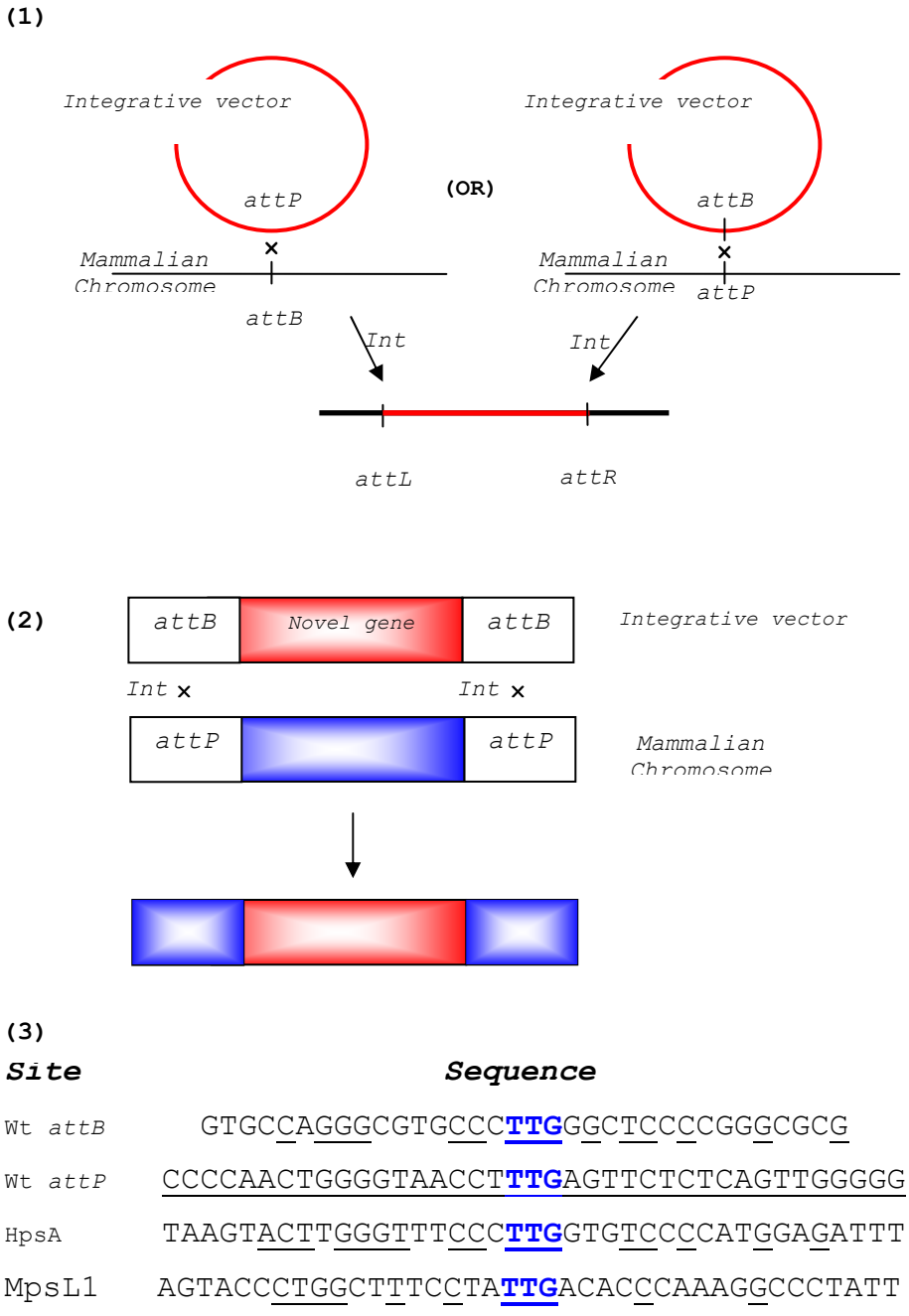


Figure 6.1: Diagram illustrating the 3 possible mechanisms by which serine based integrative vectors can be inserted into mammalian chromosome (Groth and Calos, 2004).

6.3: Future Work

Due to the lack of similarities to other phage genes on GenBank, the sequence did not reveal as much detail on the bacteriophage genome as first expected. On the other hand,

information gathered from the characterisation of ϕAR29 has contributed to the understanding of the phage integrative mechanism and provides a sound base for subsequent work to further the understanding of ϕAR29 and aid the refinement of pBA as an integrative vector.

6.3.1: Genomic Analysis

Characterisation of the ϕAR29 genome will be a continuous process that will depend either on the growth of information in genetic databases, or on a detailed study of phage gene function. As more phages are sequenced, particularly those infecting *Bacteroides* species, the accumulated information should help identify the currently unknown ORFs in ϕAR29.

For those ORFs identified as sites of translational frameshifting, the function of those genes should be confirmed experimentally

6.3.2: Refinement of pBA

Despite results showing integration of pBA into the AR29 genome, considerable improvement can be made. Some of the aims of future work to improve the efficiency of pBA as an integrative vector include:

- Confirming the function of the newly identified *Int* gene and its incorporation into a redesigned plasmid. The predicted promoter sequence location upstream of the *Int* gene should also be investigated. Confirming its promoter role may be central in maximising the efficiency of pBA integration.
- Improve pBA as a stable, single-copy integrative plasmid. Due to the strict regulation on the release of genetically modified organisms (GMO), scientists are

looking into the development of vectors capable of single copy integration and high stability upon release.

Despite attempts to meet government regulations on GMO, most of the procedures discussed are only practical in controlled environments such as bioreactors and laboratories. In non-controlled surroundings, like the rumen, bacteria will be subjected to phage infections and possibly the subsequent release of recombinant DNA into the environment by lysis of bacterial cells carrying integrated plasmid. A possible mechanism to stabilize pBA for release into an open environment is to include the *cI* gene and its promoter in the vector construction. The presence of *cI* in the cell could provide construct stability in the event of future infection, by repressing transcription of the ϕAR29 genome, thus preventing lysis.

The redesigned pBA could contain parts of ϕAR29 in the order: *attP*, ϕAR29 promoter governing the lytic cycle of ϕAR29, *Int* gene, and *cI* gene (Figure 6.2). In this arrangement, the *Int* protein could stimulate pBA integration and the *cI* protein could help to maintain the integrated state through repressing the expression of both the *Int* gene and the *cI* gene itself, from the cloned lytic promoter (Figure 6.1). However, if *cI* dissociated from its binding site due to *RecA* cleavage, induced by the host SOS system, this would lead to the subsequent expression of *Int* which could ensure that pBA remains integrated.

Future constructs based on pBA may benefit from using the homologous recombination procedure as described by Shimizu-Kadota (2001) to remove unnecessary or undesirable sequences from the plasmid. These include selective marker genes and replicons required for independent replication in bacterial hosts.

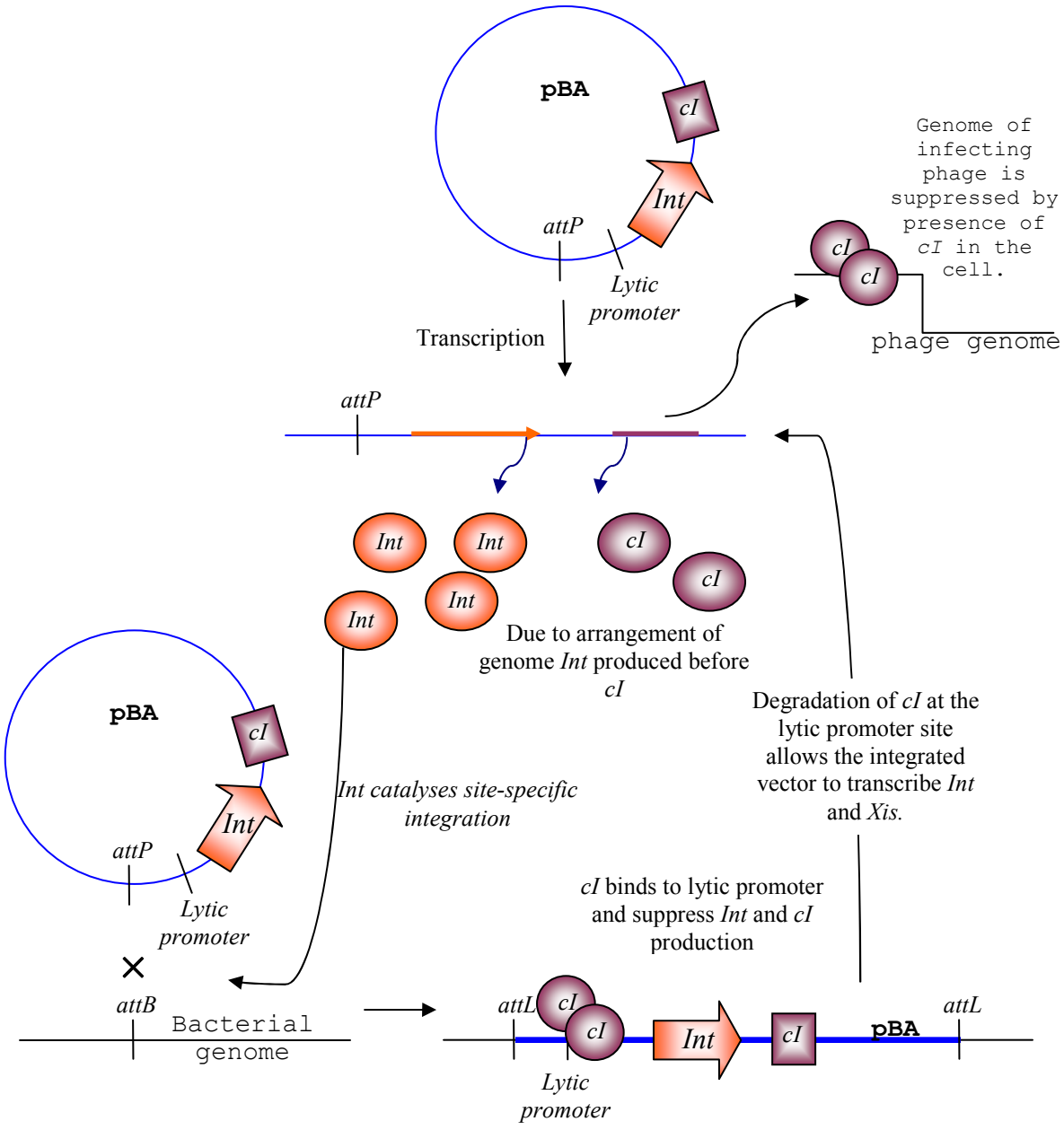


Figure: 6.3: A flow diagram illustrating a mechanism to stabilize pBA for release into an open environment. This involves cloning parts of ϕAR29 DNA in the order: *attP*, ϕAR29 promoter governing the lytic cycle of ϕAR29, *Int* gene, and *cI* gene, into vector pBA.

6.3.3: Application of pBA in Other Bacteria

The application of pBA as a molecular tool for transformation studies and replacement technology may extend to other bacterial hosts. BLAST analysis of the ϕAR29 *attB* site showed that *B. thetaiotaomicron* VPI-5482 has the same tRNA sequence. VPI-5482 is a symbiotic colonic bacterium that is a dominant member of the intestinal microbiota of humans and other mammals. Previously studies have shown the human strains *B. thetaiotaomicron* have a role in inducing host fucosylation, stimulating angiogenesis and inducing an innate immune response in the gut (Hooper *et al.*, 1999; Stappenbeck *et al.*, 2002; Hooper *et al.*, 2003). Vector pBA could be useful in the genetic engineering of VPI-5482 to further improve the ability of the organism to provide immune responses to the host.

Another bacterium, *Porphyromonas gingivalis* W83, was found to carry a Arg-tRNA gene that is similar in sequence to the AR29 gene. *P.* gingivalis is a pathogen that is associated with adult periodontal disease. Despite the 4 base substitutions, it may be possible for pBA to integrate into the genome of W83. If so, pBA may be useful in controlling the pathogenicity of W83.

Interestingly, studies have found that AR29 can colonise the large intestine of monogastric animals (Gregg, 2003). This may extend the use of pBA transformed AR29 in animals other the ruminants.

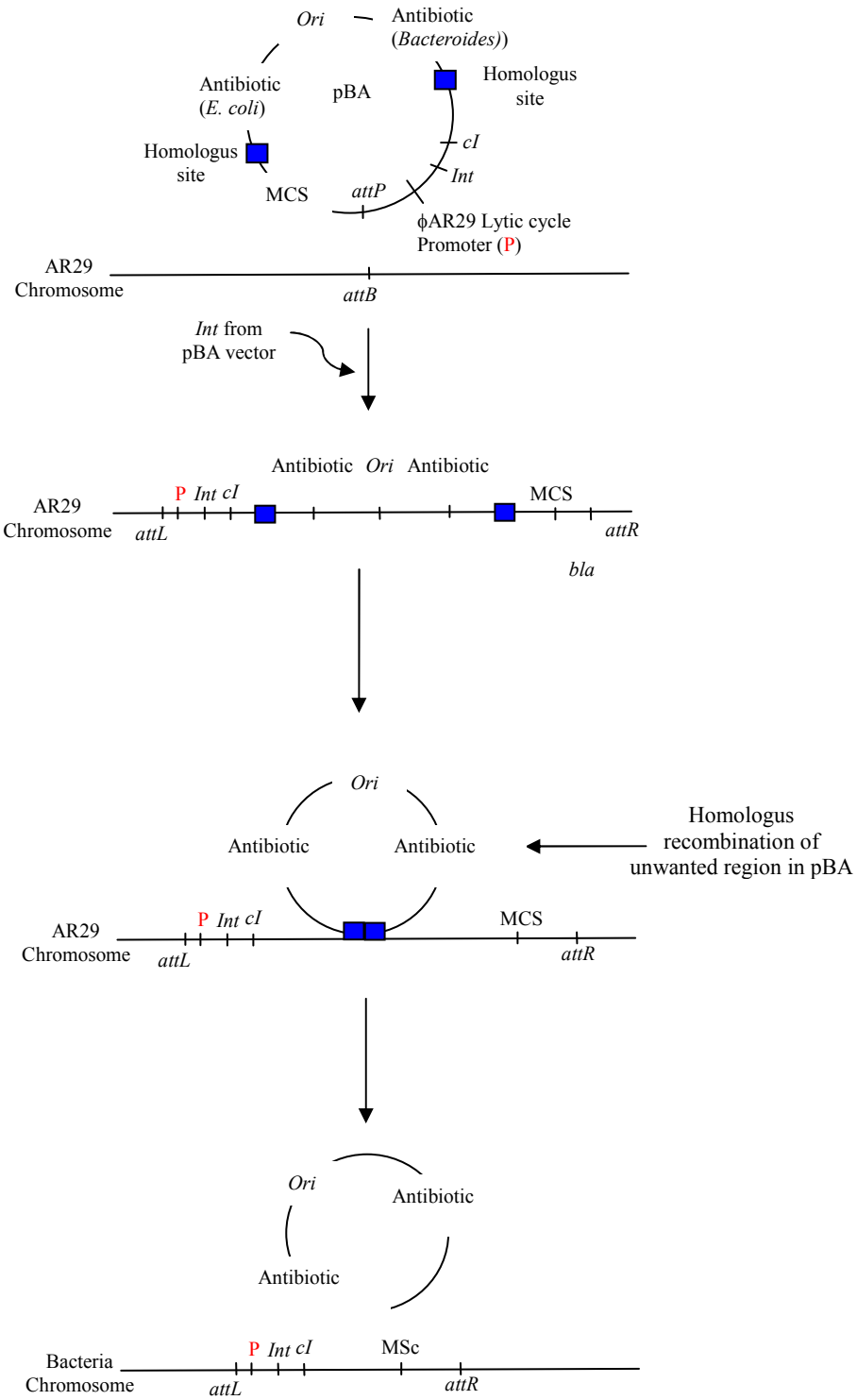


Figure 6.3: A flow diagram illustrating the use of homologous sites in an integrative plasmid such as pBA, to remove unwanted sequences from the integrated plasmid (e.g. *ori* and antibiotic resistance genes).

6.4: Closing remarks

The major aim of this project has been achieved. Newly acquired information from ϕAR29 has led to some understanding of the phage's life cycle. The detailed examination of phage genomic sequence and *attB* flanking regions has also redefined the *att* sites and aided the identification of a possible *Int* gene that was incompletely cloned into pBA and remained undiscovered in earlier work (Gregg *et al*. 1994).

More importantly, the investigation has shown that pBA is capable of integrating into the genome of a ruminal bacterium and appears to achieve this through prompting the excision of the prophage. A source of integrase that allowed the integrative recombination of pBA may be from phage particles infecting cured cells of AR29, or possibly from phage that was excised through the action of *Xis* encoded by pBA. Information gained from this study has provided a sound basis for future investigation and refinement of pBA as a manipulative and analytical tool.

References

Abraham, J., Mascarenhas, D., Fischer, R., Benedik, M., Campbell, A. and Echols, H. (1980). DNA sequence of regulatory region for integration gene of bacteriophage lambda. *Proc Natl Acad Sci U S A*, **77**(5): 2477-81.

Adams, J. C., Gazaway, J. A., Brailsford, M. D., Hartman, P. A. and Jacobson, N. L. (1966). Isolation of bacteriophages from the boive rumen. *Experientia*, **22**: 717-718.

Aksoy, S., Squires, C. L. and Squires, C. (1984). Translational coupling of the trpB and trpA genes in the Escherichia coli tryptophan operon. *J Bacteriol*, **157**(2): 363-7.

Alexander, D. C., Devlin, D. J., Hewitt, D. D., Horan, A. C. and Hosted, T. J. (2003). Development of the Micromonospora carbonacea var. africana ATCC 39149 bacteriophage pMLP1 integrase for site-specific integration in Micromonospora spp. *Microbiology*, **149**(Pt 9): 2443-53.

Allison, G. E., Angeles, D., Tran-Dinh, N. and Verma, N. K. (2002). Complete genomic sequence of SfV, a serotype-converting temperate bacteriophage of Shigella flexneri. *J Bacteriol*, **184**(7): 1974-87.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, **215**(3): 403-10.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**(17): 3389-402.

Alvarez, M. A., Herrero, M. and Suarez, J. E. (1998). The site-specific recombination system of the Lactobacillus species bacteriophage A2 integrates in gram-positive and gram-negative bacteria. *Virology*, **250**(1): 185-93.

Ambrozic, J., Ferme, D., Grabnar, M., Ravnikar, M. and Avgustin, G. (2001). The bacteriophages of ruminal prevotellas. *Folia Microbiol (Praha)*, **46**(1): 37-9.

Asmundson, R. V. and Kelly, W. J. (1987). Isolation and characterization of plasmid DNA from Ruminococcus. *Current Microbiol*, **16**: 97-100.

Attwood, G. T. and Brooker, J. D. (1992). Complete nucleotide sequence of a Selenomonas ruminantium plasmid and definition of a region necessary for its replication in Escherichia coli. *Plasmid*, **28**(2): 123-9.

Auvray, F., Coddeville, M., Ordonez, R. C. and Ritzenthaler, P. (1999). Unusual structure of the attB site of the site-specific recombination system of Lactobacillus delbrueckii bacteriophage mv4. *J Bacteriol*, **181**(23): 7385-9.

Auvray, F., Coddeville, M., Ritzenthaler, P. and Dupont, L. (1997). Plasmid integration in a wide range of bacteria mediated by the integrase of Lactobacillus delbrueckii bacteriophage mv4. *J Bacteriol*, **179**(6): 1837-45.

Avgustin, G., Wright, F. and Flint, H. J. (1994). Genetic Diversity and Phylogenetic Relationships Among Strains of Prevotella (Bacteroides) Ruminicola From the Rumen. *International Journal of Systematic Bacteriology*, **47**(2): 284-288.

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

Barrangou, R., Yoon, S. S., Breidt Jr, F., Jr., Fleming, H. P. and Klaenhammer, T. R. (2002). Characterization of six Leuconostoc fallax bacteriophages isolated from an industrial sauerkraut fermentation. *Appl Environ Microbiol*, **68**(11): 5452-8.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. and Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Res*, **32**(Database issue): D138-41.

Bauer, C. E., Hesse, S. D., Gumport, R. I. and Gardner, J. F. (1986). Mutational analysis of integrase arm-type binding sites of bacteriophage lambda. Integration and excision involve distinct interactions of integrase with arm-type sites. *J Mol Biol*, **192**(3): 513-27.

Bauer, J. W. and Laimer, M. (2004). Gene therapy of epidermolysis bullosa. *Expert Opin Biol Ther*, **4**(9): 1435-43.

Bayley, D. P., Rocha, E. R. and Smith, C. J. (2000). Analysis of cepA and other Bacteroides fragilis genes reveals a unique promoter structure. *FEMS Microbiol Lett*, **193**(1): 149-54.

Beard, C. E., Gregg, K., Kalmokoff, M. and Teather, R. M. (2000). Construction of a promoter-rescue plasmid for Butyrivibrio fibrisolvens and its use in characterization of a flagellin promoter. *Curr Microbiol*, **40**(3): 164-8.

Beard, C. E., Hefford, M. A., Forster, R. J., Sontakke, S., Teather, R. M. and Gregg, K. (1995). A stable and efficient transformation system for Butyrivibrio fibrisolvens OB156. *Curr Microbiol*, **30**(2): 105-9.

Bechet, M., Pheulpin, P., Flint, H. J., Martin, J. and Dubourguier, H. C. (1993). Transfer of hybrid plasmids based on the replicon pRRI7 from Escherichia coli to Bacteroides and Prevotella strains. *J Appl Bacteriol*, **74**(5): 542-8.

Becker, W. M., Reece, J. B. and Poenie, M. F. (1996) *The World of The Cell,* The Benjamin/ Cummings Publication Company, Inc., Californa.

Belfort, M. and Perlman, P. S. (1995). Mechanisms of intron mobility. *J Biol Chem*, **270**(51): 30237-40.

Belhocine, K., Yam, K. K. and Cousineau, B. (2005). Conjugative transfer of the Lactococcus lactis chromosomal sex factor promotes dissemination of the Ll.LtrB group II intron. *J Bacteriol*, **187**(3): 930-9.

Belteki, G., Gertsenstein, M., Ow, D. W. and Nagy, A. (2003). Site-specific cassette exchange and germline transmission with mouse ES cells expressing phiC31 integrase. *Nat Biotechnol*, **21**(3): 321-4.

Benzer, S. (1957) In *The Chemical Basis of Heredity*(Eds, McElroy, W. D. and Glass, B.) The Johns Hopkins Press., Baltimore, pp. 70-93.

Beres, S. B., Sylva, G. L., Barbian, K. D., Lei, B., Hoff, J. S., Mammarella, N. D., Liu, M. Y., Smoot, J. C., Porcella, S. F., Parkins, L. D., Campbell, D. S., Smith, T. M., McCormick, J. K., Leung, D. Y., Schlievert, P. M. and Musser, J. M. (2002). Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc Natl Acad Sci U S A*, **99**(15): 10078-83.

Berger, E., Jones, W. A., Jones, D. T. and Woods, D. R. (1989). Cloning and sequencing of an endoglucanase (end1) gene from Butyrivibrio fibrisolvens H17c. *Mol Gen Genet*, **219**(1-2): 193-8.

Bernheimer, H. P. and Tiraby, J. G. (1976). Inhibition of phage infection by pneumococcus capsule. *Virology*, **73**(1): 308-9.

Besemer, J. and Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res*, **27**(19): 3911-20.

Besendahl, A., Qiu, Y. L., Lee, J., Palmer, J. D. and Bhattacharya, D. (2000). The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. *Curr Genet*, **37**(1): 12-23.

Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N. and Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene*, **116**(1): 43-9.

Bigger, C. H., Murray, K. and Murray, N. E. (1973). Recognition sequence of a restriction enzyme. *Nat New Biol*, **244**(131): 7-10.

Biniszkiewicz, D., Cesnaviciene, E. and Shub, D. A. (1994). Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. *Embo J*, **13**(19): 4629-35.

Birnboim, H. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids. Res.*, **7**(6): 1513-1523.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science*, **277**(5331): 1453-74.

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D. and Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. *Genome Res*, **11**(5): 731-53.

Boyd, E. F., Davis, B. M. and Hochhut, B. (2001). Bacteriophage-bacteriophage interactions in the evolution of pathogenic bacteria. *Trends Microbiol*, **9**(3): 137-44.

Brendel, V. and Trifonov, E. N. (1984). A computer algorithm for testing potential prokaryotic terminators. *Nucleic Acids Res*, **12**(10): 4411-27.

Breuner, A., Brondsted, L. and Hammer, K. (2001). Resolvase-like recombination performed by the TP901-1 integrase. *Microbiology*, **147**(Pt 8): 2051-63.

Brondsted, L. and Hammer, K. (1999). Use of the integration elements encoded by the temperate lactococcal bacteriophage TP901-1 to obtain chromosomal single-copy transcriptional fusions in Lactococcus lactis. *Appl Environ Microbiol*, **65**(2): 752-8.

Brondsted, L., Ostergaard, S., Pedersen, M., Hammer, K. and Vogensen, F. K. (2001). Analysis of the complete DNA sequence of the temperate bacteriophage TP901-1: evolution, structure, and genome organization of lactococcal bacteriophages. *Virology*, **283**(1): 93-109.

Brooker, J. D. and McCarthy, J. M. (1997). Gene knockout of the intracellular amylase gene by homologous recombination in Streptococcus bovis. *Curr Microbiol*, **35**(3): 133-8.

Brown, M. D., Ripley, L. S. and Hall, D. H. (1993). A proflavin-induced frameshift hotspot in the thymidylate synthase gene of bacteriophage T4. *Mutat Res*, **286**(2): 189-97.

Bruand, C., Velten, M., McGovern, S., Marsin, S., Serena, C., Ehrlich, S. D. and Polard, P. (2005). Functional interplay between the Bacillus subtilis DnaD and DnaB proteins essential for initiation and re-initiation of DNA replication. *Mol Microbiol*, **55**(4): 1138-50.

Brussow, H. (2001). Phages of dairy bacteria. *Annu Rev Microbiol*, **55**: 283-303.

Bruttin, A., Foley, S. and Brussow, H. (1997). The site-specific integration system of the temperate Streptococcus thermophilus bacteriophage phiSfi21. *Virology*, **237**(1): 148-58.

Buchko, G. W., Daughdrill, G. W., de Lorimier, R., Rao, B. K., Isern, N. G., Lingbeck, J. M., Taylor, J. S., Wold, M. S., Gochin, M., Spicer, L. D., Lowry, D. F. and Kennedy, M. A. (1999). Interactions of human nucleotide excision repair protein XPA with DNA and RPA70 Delta C327: chemical shift mapping and 15N NMR relaxation studies. *Biochemistry*, **38**(46): 15116-28.

Burt, S., Meldrum, S., Woods, D. R. and Jones, D. T. (1978). Colonial variation, capsule formation, and bacteriophage resistance in Bacteroides thetaiotaomicron. *Appl Environ Microbiol*, **35**(2): 439-43.

Bustin, M. (1999). Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol*, **19**(8): 5237-46.

Campbell, A. (1962). Episomes. *Adv. Genet.*, **11**: 101-145.

Campbell, A. (2003). The future of bacteriophage biology. *Nat Rev Genet*, **4**(6): 471-7.

Campbell, A., del-Campillo-Campbell, A. and Ginsberg, M. L. (2002). Specificity in DNA recognition by phage integrases. *Gene*, **300**(1-2): 13-8.

Campbell, A. M. (1992). Chromosomal insertion sites for phages and plasmids. *J Bacteriol*, **174**(23): 7495-9.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A. and Brussow, H. (2003). Prophage genomics. *Microbiol Mol Biol Rev*, **67**(2): 238-76, table of contents.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N. and Gutell, R. R. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**(1): 2.

Carroll, D., Kehoe, M. A., Cavanagh, D. and Coleman, D. C. (1995). Novel organization of the site-specific integration and excision recombination functions of the Staphylococcus aureus serotype F virulence-converting phages phi 13 and phi 42. *Mol Microbiol*, **16**(5): 877-93.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol*, **49**(2): 277-300.

Casjens, S., Winn-Stapley, D. A., Gilcrease, E. B., Morona, R., Kuhlewein, C., Chua, J. E., Manning, P. A., Inwood, W. and Clark, A. J. (2004). The chromosome of Shigella flexneri bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J Mol Biol*, **339**(2): 379-94.

Casjens, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schicklmaier, P., Schmieger, H., Pedulla, M. L., Ford, M. E., Houtz, J. M., Hatfull, G. F. and Hendrix, R. W. (2005). The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol*, **187**(3): 1091-104.

Chai, S., Bravo, A., Luder, G., Nedlin, A., Trautner, T. A. and Alonso, J. C. (1992). Molecular analysis of the Bacillus subtilis bacteriophage SPP1 region encompassing genes 1 to 6. The products of gene 1 and gene 2 are required for pac cleavage. *J Mol Biol*, **224**(1): 87-102.

Chen, H. Z. and Zubay, G. (1983). Prokaryotic coupled transcription-translation. *Methods Enzymol*, **101**: 674-90.

Cheng, X., Zhang, X., Pflugrath, J. W. and Studier, F. W. (1994). The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc Natl Acad Sci U S A*, **91**(9): 4034-8.

Cheong, J. P. and Brooker, J. D. (1998). Lysogenic bacteriophage M1 from Selenomonas ruminantium: isolation, characterization and DNA sequence analysis of the integration site. *Microbiology*, **144 (Pt 8)**: 2195-202.

Cho, E. H., Alcaraz, R., Jr., Gumport, R. I. and Gardner, J. F. (2002a). Characterization of bacteriophage lambda excisionase mutants defective in DNA binding. *J Bacteriol*, **182**(20): 5807-12.

Cho, E. H., Gumport, R. I. and Gardner, J. F. (2002b). Interactions between Integrase and Excisionase in the Phage Lambda Excisive Nucleoprotein Complex. *J Bacteriol*, **184**(18): 5200-5203.

Chopin, A., Bolotin, A., Sorokin, A., Ehrlich, S. D. and Chopin, M. (2001). Analysis of six prophages in Lactococcus lactis IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res*, **29**(3): 644-51.

Christie, G. E., Cale, S. B., Isaksson, L. A., Jin, D. J., Xu, M., Sauer, B. and Calendar, R. (1996). Escherichia coli rpoC397 encodes a temperature-sensitive C-terminal frameshift in the beta' subunit of RNA polymerase that blocks growth of bacteriophage P2. *J Bacteriol*, **178**(23): 6991-3.

Christie, G. E., Temple, L. M., Bartlett, B. A. and Goodwin, T. S. (2002). Programmed translational frameshift in the bacteriophage P2 FETUD tail gene operon. *J Bacteriol*, **184**(23): 6522-31.

Chu, F. K., Maley, G. F., Maley, F. and Belfort, M. (1984). Intervening sequence in the thymidylate synthase gene of bacteriophage T4. *Proc Natl Acad Sci U S A*, **81**(10): 3049-53.

Church, D. C. (1993) *The Ruminant Animal : Digestive Physiology and Nutrition,* Englewood Cliffs, New Jersey.

Clark, A. J., Inwood, W., Cloutier, T. and Dhillon, T. S. (2001). Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J Mol Biol*, **311**(4): 657-79.

Clark, C. A., Beltrame, J. and Manning, P. A. (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**(1): 43-52.

Clark, R. G., Cheng, K. J., Selinger, L. B. and Hynes, M. F. (1994). A conjugative transfer system for the rumen bacterium, Butyrivibrio fibrisolvens, based on Tn916-mediated transfer of the Staphylococcus aureus plasmid pUB110. *Plasmid*, **32**(3): 295-305.

Cocconcelli, P. S., Ferrari, E., Rossi, F. and Bottazzi, V. (1992). Plasmid transformation of Ruminococcus albus by means of high-voltage electroporation. *FEMS Microbiol Lett*, **73**(3): 203-7.

Coffey, T. J., Enright, M. C., Daniels, M., Morona, J. K., Morona, R., Hryniewicz, W., Paton, J. C. and Spratt, B. G. (1998). Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae. *Mol Microbiol*, **27**(1): 73-83.

Combes, P., Till, R., Bee, S. and Smith, M. C. (2002). The streptomyces genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system. *J Bacteriol*, **184**(20): 5746-52.

Comstock, L. E. and Coyne, M. J. (2003). Bacteroides thetaiotaomicron: a dynamic, niche-adapted human symbiont. *Bioessays*, **25**(10): 926-9.

Comstock, L. E., Pantosti, A. and Kasper, D. L. (2000). Genetic diversity of the capsular polysaccharide C biosynthesis region of Bacteroides fragilis. *Infect Immun*, **68**(11): 6182-8.

Confalonieri, F. and Duguet, M. (1995). A 200-amino acid ATPase module in search of a basic function. *Bioessays*, **17**(7): 639-50.

Cotta, M. A., Whitehead, T. R. and Rasmussen, M. A. (1997). Survival of the recombinant Bacteroides thetaiotaomicron strain BTX in in vitro rumen incubations. *J Appl Microbiol*, **82**(6): 743-50.

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J. E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A. M. and Belfort, M. (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*, **94**(4): 451-62.

Coyne, M. J., Kalka-Moll, W., Tzianabos, A. O., Kasper, D. L. and Comstock, L. E. (2000). Bacteroides fragilis NCTC9343 produces at least three distinct capsular polysaccharides: cloning, characterization, and reassignment of polysaccharide B and C biosynthesis loci. *Infect Immun*, **68**(11): 6176-81.

Craig, N. L. and Nash, H. A. (1983). The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell*, **35**(3 Pt 2): 795-803.

Craig, N. L. and Nash, H. A. (1984). E. coli integration host factor binds to specific sites in DNA. *Cell*, **39**(3 Pt 2): 707-16.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, **227**(5258): 561-3.

Crick, F. H., Barnett, L., Brenner, S. and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, **192**: 1227-32.

Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. and Cozzarelli, N. R. (1999). The topological mechanism of phage lambda integrase. *J Mol Biol*, **289**(4): 747-75.

Culvenor, C. C. J. (1987) In *Temperate Pastures: Their Production, Use and Management*CSIOR, Australia, pp. 435-459.

Cybinski, D. H., Layton, I., Lowry, J. B. and Dalrymple, B. P. (1999). An acetylxylan esterase and a xylanase expressed from genes cloned from the ruminal fungus Neocallimastix patriciarum act synergistically to degrade acetylated xylans. *Appl Microbiol Biotechnol*, **52**(2): 221-5.

Dai, L. and Zimmerly, S. (2002). Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res*, **30**(5): 1091-102.

Dai, L. and Zimmerly, S. (2003). ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *Rna*, **9**(1): 14-9.

Dalrymple, B. P. and Swadling, Y. (1997). Expression of a Butyrivibrio fibrisolvens E14 gene (cinB) encoding an enzyme with cinnamoyl ester hydrolase activity is negatively regulated by the product of an adjacent gene (cinR). *Microbiology*, **143 (Pt 4)**: 1203-10.

Dalrymple, B. P., Swadling, Y., Cybinski, D. H. and Xue, G. P. (1996). Cloning of a gene encoding cinnamoyl ester hydrolase from the ruminal bacterium Butyrivibrio fibrisolvens E14 by a novel method. *FEMS Microbiol Lett*, **143**(2-3): 115-20.

Danchin, A., Dondon, L., Joseph, E. and Ullmann, A. (1981). Transcription-translation coupling and polarity: a possible role of cyclic AMP. *Biochimie*, **63**(5): 419-24.

Daniel, A. S., Martin, J., Vanat, I., Whitehead, T. R. and Flint, H. J. (1995). Expression of a cloned cellulase/xylanase gene from Prevotella ruminicola in Bacteroides vulgatus, Bacteroides uniformis and Prevotella ruminicola. *J Appl Bacteriol*, **79**(4): 417-24.

Danovaro, R., Dell'Anno, A., Trucco, A., Serresi, M. and Vanucci, S. (2001). Determination of virus abundance in marine sediments. *Appl Environ Microbiol*, **67**(3): 1384-7.

Davies, R. W. (1980). DNA sequence of the int-xis-Pi region of the bacteriophage lambda; overlap of the int and xis genes. *Nucleic Acids Res*, **8**(8): 1765-82.

de Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M. and Catalano, C. E. (2002). Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Mol Cell*, **9**(5): 981-91.

Dean, R. G., Martin, S. A. and Carver, C. (1989). Isolation of plasmid DNA from the ruminal bacterium *Selenomonas ruminantium* HD4. *Letters in Applied Microbiology*, **8**: 45-8.

Deogun, J. S., Donis, R., Komina, O. and Ma, F. (2004) In *Proceedings of the second conference on Asia-Pacific bioinformatics - Volume 29*, Vol. 29 (Ed, Che, Y.-P. P. n.) New Zealand, pp. 239 - 246.

Desiere, F., McShan, W. M., van Sinderen, D., Ferretti, J. J. and Brussow, H. (2001). Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic Streptococci: evolutionary implications for prophage-host interactions. *Virology*, **288**(2): 325-41.

Desiere, F., Pridmore, R. D. and Brussow, H. (2000). Comparative genomics of the late gene cluster from Lactobacillus phages. *Virology*, **275**(2): 294-305.

Diaz, E., Lopez, R. and Garcia, J. L. (1990). Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. *Proc Natl Acad Sci U S A*, **87**(20): 8125-9.

Dobrindt, U., Blum-Oehler, G., Nagy, G., Schneider, G., Johann, A., Gottschalk, G. and Hacker, J. (2002). Genetic structure and distribution of four pathogenicity islands (PAI I(536) to PAI IV(536)) of uropathogenic Escherichia coli strain 536. *Infect Immun*, **70**(11): 6365-72.

Dorgai, L., Papp, I., Papp, P., Kalman, M. and Orosz, L. (1993). Nucleotide sequences of the sites involved in the integration of phage 16-3 of Rhizobium meliloti 41. *Nucleic Acids Res*, **21**(7): 1671.

Dorgai, L., Sloan, S. and Weisberg, R. A. (1998). Recognition of core binding sites by bacteriophage integrases. *J. Mol. Biol.*, **277**: 1059-1070.

Doyle, K. (1996) *Promega protocols and applications guide,* Promega Corporation, U.S.A.

Du, Z. and Hoffman, D. W. (1997). An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots. *Nucleic Acids Res*, **25**(6): 1130-5.

Dupont, L., Boizet-Bonhoure, B., Coddeville, M., Auvray, F. and Ritzenthaler, P. (1995). Characterization of genetic elements required for site-specific integration of Lactobacillus delbrueckii subsp. bulgaricus bacteriophage mv4 and construction of an integration-proficient vector for Lactobacillus plantarum. *J Bacteriol*, **177**(3): 586-95.

Echols, H. (1975). Constitutive integrative recombination by bacteriophage lambda. *Virology*, **64**(2): 557-9.

Echols, H. and Guarneros, G. (1983) In *Lamb II*(Eds, Hendrix, R. W., Roberts, J. W., Stahl, F. W. and Weisberg, R. A.) Cold Spring Harbor Laboratory, USA, pp. 75-92.

Edgell, D. R., Belfort, M. and Shub, D. A. (2000). Barriers to intron promiscuity in bacteria. *J Bacteriol*, **182**(19): 5281-9.

Eickbush, T. H. (1999). Mobile introns: retrohoming by complete reverse splicing. *Curr Biol*, **9**(1): R11-4.

Ellis, D. E., Whitman, P. A. and Marshall, R. T. (1973). Effects of homologous bacteriophage on growth of Pseudomonas fragi WY in milk. *Appl Microbiol*, **25**(1): 24-5.

Elo, P., Semsey, S., Kereszt, A., Nagy, T., Papp, P. and Orosz, L. (1998). Integrative promoter cloning plasmid vectors for Rhizobium meliloti. *FEMS Microbiol Lett*, **159**(1): 7-13.

Espion, D., Kaiser, K. and Dambly-Chaudiere, C. (1983). A third defective lambdoid prophage of Escherichia coli K12 defined by the lambda derivative, lambdaqin111. *J Mol Biol*, **170**(3): 611-33.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. and Bairoch, A. (2002). The PROSITE database, its status in 2002. *Nucleic Acids Res*, **30**(1): 235-8.

Farabaugh, P. J. (1996). Programmed translational frameshifting. *Microbiol Rev*, **60**(1): 103-34.

Ferat, J. L. and Michel, F. (1993). Group II self-splicing introns in bacteria. *Nature*, **364**(6435): 358-61.

Ferenczi, S., Ganyu, A., Blaha, B., Semsey, S., Nagy, T., Csiszovszki, Z., Orosz, L. and Papp, P. P. (2004). Integrative plasmid vector for constructing single-copy reporter systems to study gene regulation in Rhizobium meliloti and related species. *Plasmid*, **52**(1): 57-62.

Firkins, J. L., Weiss, W. P. and Piwonka, E. J. (1992). Quantification of intraruminal recycling of microbial nitrogen using nitrogen-15. *J Anim Sci*, **70**(10): 3223-33.

Fliegerova, K., Pazoutova, S., Pristas, P. and Flint, H. J. (2000). Highly conserved DNA sequence present in small plasmids from Selenomonas ruminantium. *Plasmid*, **44**(1): 94-9.

Flint, H. J., Martin, J., McPherson, C. A., Daniel, A. S. and Zhang, J. X. (1993). A bifunctional enzyme, with separate xylanase and beta(1,3-1,4)-glucanase domains, encoded by the xynD gene of Ruminococcus flavefaciens. *J Bacteriol*, **175**(10): 2943-51.

Flint, H. J., Thomson, A. M. and Bisset, J. (1988). Plasmid-associated transfer of tetracycline resistance in Bacteroides ruminicola. *Appl Environ Microbiol*, **54**(4): 855-60.

Forsberg, C. F., Cheng, K.-J. and White, B. A. (1997) In *Gastrointestinal Microbiology*, Vol. 1 (Eds, Mackie, R. I. and White, B. A.) Chapman and Hall, New York, pp. 319-379.

Freitas-Vieira, A., Anes, E. and Moniz-Pereira, J. (1998). The site-specific recombination locus of mycobacteriophage Ms6 determines DNA integration at the tRNA(Ala) gene of Mycobacterium spp. *Microbiology*, **144 (Pt 12)**: 3397-406.

Friedman, D. I. and Gottesman, M. (1983) In *Lambda II*(Eds, Hendrix, R. W., Roberts, J. W., Stahl, F. W. and Weisberg, R. A.) Cold Spring Harbor Laboratory, USA, pp. 21-52.

Fromknecht, K., Vogel, P. D. and Wise, J. G. (2003). Combinatorial redesign of the DNA binding specificity of a prokaryotic helix-turn-helix repressor. *J Bacteriol*, **185**(2): 475-81.

Gabriel, K., Schmid, H., Schmidt, U. and Rausch, H. (1995). The actinophage RP3 DNA integrates site-specifically into the putative tRNA(Arg)(AGG) gene of Streptomyces rimosus. *Nucleic Acids Res*, **23**(1): 58-63.

Gaeng, S., Scherer, S., Neve, H. and Loessner, M. J. (2000). Gene cloning and expression and secretion of Listeria monocytogenes bacteriophage-lytic enzymes in Lactococcus lactis. *Appl Environ Microbiol*, **66**(7): 2951-8.

Garcia, E., Garcia, J. L., Garcia, P., Arraras, A., Sanchez-Puelles, J. M. and Lopez, R. (1988). Molecular evolution of lytic enzymes of Streptococcus pneumoniae and its bacteriophages. *Proc Natl Acad Sci U S A*, **85**(3): 914-8.

Garcia, P., Rodriguez, I. and Suarez, J. E. (2004). A -1 ribosomal frameshift in the transcript that encodes the major head protein of bacteriophage A2 mediates biosynthesis of a second essential component of the capsid. *J Bacteriol*, **186**(6): 1714-9.

Gardner, R. G., Russell, J. B., Wilson, D. B., Wang, G. R. and Shoemaker, N. B. (1996). Use of a modified Bacteroides-Prevotella shuttle vector to transfer a reconstructed beta-1,4-D-endoglucanase gene into Bacteroides uniformis and Prevotella ruminicola B(1)4. *Appl Environ Microbiol*, **62**(1): 196-202.

Gellert, M. (1967). Formation of covalent circles of lambda DNA by E. coli extracts. *Proc Natl Acad Sci U S A*, **57**(1): 148-55.

Georgopoulos, C. P., Hendrix, R. W., Casjens, S. R. and Kaiser, A. D. (1973). Host participation in bacteriophage lambda head assembly. *J Mol Biol*, **76**(1): 45-60.

Gibson, T. J. (1984) Cambridge University, England.

Gindreau, E., Lopez, R. and Garcia, P. (2000). MM1, a temperate bacteriophage of the type 23F Spanish/USA multiresistant epidemic clone of Streptococcus pneumoniae: structural analysis of the site-specific integration system. *J Virol*, **74**(17): 7803-13.

Gish, W. and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat Genet*, **3**(3): 266-72.

Gobius, K. S., Xue, G. P., Aylward, J. H., Dalrymple, B. P., Swadling, Y. J., McSweeney, C. S. and Krause, D. O. (2002). Transformation and expression of an anaerobic fungal xylanase in several strains of the rumen bacterium Butyrivibrio fibrisolvens. *J Appl Microbiol*, **93**(1): 122-33.

Goodman, H. J. and Woods, D. R. (1993). Cloning and nucleotide sequence of the Butyrivibrio fibrisolvens gene encoding a type III glutamine synthetase. *J Gen Microbiol*, **139 (Pt 7)**: 1487-93.

Goodman, S. D. and Scocca, J. J. (1989). Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of Haemophilus influenzae. *J Bacteriol*, **171**(8): 4232-40.

Goodman, S. D., Velten, N. J., Gao, Q., Robinson, S. and Segall, A. M. (1999). In vitro selection of integration host factor binding sites. *J Bacteriol*, **181**(10): 3246-55.

Gottesman, S. (1981). Lambda site-specific recombination: the att site. *Cell*, **25**(3): 585-6.

Gowrishankar, J. and Harinarayanan, R. (2004). Why is transcription coupled to translation in bacteria? *Mol Microbiol*, **54**(3): 598-603.

Gregg, K. (1992). Development in rumen bacterial molecular biology. *AgBiotech News and Information*, **4**(12): 375-380.

Gregg, K., Allen, G. and Beard, C. (1996). Genetic manipulation of rumen bacteria from potential to reality. *Aust. J. Agric. Res.*, **47**: 247-256.

Gregg, K., Cooper, C. L., Schafer, D. J., Sharpe, H., Beard, C. E., Allen, G. and Xu, J. W. (1994a). Detoxification of the Plant Toxin Fluoroacetate By a Genetically Modified Rumen Bacterium. *Bio-Technology*, **12**(13): 1361-1365.

Gregg, K., Hamdorf, B., Henderson, K., Kopecny, J. and Wong, C. M. (1998). Genetically modified ruminal bacteria protect sheep from fluoroacetate poisoning. *Appl Environ Microbiol.*, **64**(9): 3496-8.

Gregg, K., Kennedy, B. G. and Klieve, A. V. (1994b). Cloning and DNA sequence analysis of the region containing attP of the temperate phage phi AR29 of Prevotella ruminicola AR29. *Microbiology*, **140 (Pt 8)**: 2109-14.

Gregg, K. and Sharpe, H. (1991) In *Physiological Aspects of Digestioni and Metabolism in Ruminants: Proceedings of the Seventh International Symposium on Ruminant Physiology*Academic Press, Inc., pp. 719-735.

Gregory, M. A., Till, R. and Smith, M. C. (2003). Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. *J Bacteriol*, **185**(17): 5320-3.

Groth, A. C. and Calos, M. P. (2004). Phage integrases: biology and applications. *J Mol Biol*, **335**(3): 667-78.

Groth, A. C., Olivares, E. C., Thyagarajan, B. and Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proc Natl Acad Sci U S A*, **97**(11): 5995-6000.

Gual, A., Camacho, A. G. and Alonso, J. C. (2000). Functional analysis of the terminase large subunit, G2P, of Bacillus subtilis bacteriophage SPP1. *J Biol Chem*, **275**(45): 35311-9.

Guarneros, G., Montanez, C., Hernandez, T. and Court, D. (1982). Posttranscriptional control of bacteriophage lambda gene expression from a site distal to the gene. *Proc Natl Acad Sci U S A*, **79**(2): 238-42.

Guo, F., Gopaul, D. N. and van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**(6646): 40-6.

Gussin, G., Johnson, A., Pabo, C. and Sauer, R. (1983) In *Lamb II*(Eds, Hendrix, R. W., Roberts, J. W., Stahl, F. W. and Weisberg, R. A.) Cold Spring Harbor Laboratory, USA, pp. 93-122.

Hakimi, J. M. and Scocca, J. J. (1994). Binding sites for bacteriophage HP1 integrase on its DNA substrates. *J Biol Chem*, **269**(33): 21340-5.

Haldimann, A. and Wanner, B. L. (2001). Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol*, **183**(21): 6384-93.

Hall, T. A. (1999) In *Nucleic Acids Research Symposium Series 41*, pp. 95-98.

Hallet, B. and Sherratt, D. J. (1997). Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol Rev*, **21**(2): 157-78.

Hamdorf, B. J. (1998) In *Division of Molecular and Cellular Biology*University of New England, Armidale, pp. 216.

Hammond, A. C. (1995). Leucaena toxicosis and its control in ruminants. *J Anim Sci*, **73**(5): 1487-92.

Hart, M. G. and Ellison, J. (1970). Ultraviolet reactivation in bacteriophage lambda. *J Gen Virol*, **8**(3): 197-208.

Haugen, P., Simon, D. M. and Bhattacharya, D. (2005). The natural history of group I introns. *Trends Genet*, **21**(2): 111-9.

Hayes, S. and Bull, H. J. (1999). Translational frameshift sites within bacteriophage lambda genes rexA and cI. *Acta Biochim Pol*, **46**(4): 879-84.

Hedgpeth, J., Goodman, H. M. and Boyer, H. W. (1972). DNA nucleotide sequence restricted by the RI endonuclease. *Proc Natl Acad Sci U S A*, **69**(11): 3448-52.

Hefford, M. A., Teather, R. M. and Forster, R. J. (1993). The complete nucleotide sequence of a small cryptic plasmid from a rumen bacterium of the genus Butyrivibrio. *Plasmid*, **29**(1): 63-9.

Hendrix, W. R. (1983) In *LambdaII*(Eds, Hendrix, R. W., Roberts, J. W., Stahi, F. W. and Weisberg, R. A.) Cold Spring Harbor Laboratory, USA, pp. 75-92.

Hermanova, A., Pristas, P., Molnarova, V., Fliegerova, K. and Javorsky, P. (2001). Plasmids of Selenomonas ruminantium and development of host-vector system. *Folia Microbiol (Praha)*, **46**(4): 289-91.

Hermesz, E., Olasz, F., Dorgai, L. and Orosz, L. (1992). Stable incorporation of genetic material into the chromosome of Rhizobium meliloti 41: construction of an integrative vector system. *Gene*, **119**(1): 9-15.

Hershey, A. D., and M. Chase. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.*, **36**: 39-56.

Hespell, R. B. and Whitehead, T. R. (1991a). Conjugal transfer of Tn916, Tn916*DE* and pAMB1 from *Enterococcus faecalis* to *Butyrivibrio fibrisolvens* strains. *Appl Environ Microbiol*, **57**: 2703-2709.

Hespell, R. B. and Whitehead, T. R. (1991b). Introduction of Tn916 and pAMB1 into *Streptococcus bovis* JB1 by conjugation. *Appl Environ Microbiol*, **57**: 2710-2713.

Ho, K. (2001). Bacteriophage therapy for bacterial infections. Rekindling a memory from the pre-antibiotics era. *Perspect Biol Med*, **44**(1): 1-16.

Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J. and Schweizer, H. P. (1998). A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa mutants. *Gene*, **212**(1): 77-86.

Hoang, T. T., Kutchma, A. J., Becher, A. and Schweizer, H. P. (2000). Integration-proficient plasmids for Pseudomonas aeruginosa: site-specific integration and use for engineering of reporter and expression strains. *Plasmid*, **43**(1): 59-72.

Holden, M. T., Feil, E. J., Lindsay, J. A., Peacock, S. J., Day, N. P., Enright, M. C., Foster, T. J., Moore, C. E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S. D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K. D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B. G., Spratt, B. G. and Parkhill, J. (2004). Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci U S A*, **101**(26): 9786-91.

Hondel, C. A., Konings, R. N. and Schoenmakers, J. G. (1975). Regulation of gene activity in bacteriophage M13 DNA: Coupled transcription and translation of purified genes and gene-fragments. *Virology*, **67**(2): 487-97.

Hooper, L., Stappenbeck, T., Hong, C. and Gordon, J. I. (2003). Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol*, **4**: 269-273.

Hooper, L., Xu, J., Falk, P., Midtvedt, T. and Gordon, J. I. (1999). A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A*, **96**: 9833-9838.

Hsu, P. L. and Landy, A. (1984). Resolution of synthetic att-site Holliday structures by the integrase protein of bacteriophage lambda. *Nature*, **311**(5988): 721-6.

Huang, C. C., Narita, M., Yamagata, T., Itoh, Y. and Endo, G. (1999). Structure analysis of a class II transposon encoding the mercury resistance of the Gram-positive Bacterium bacillus megaterium MB1, a strain isolated from minamata bay, Japan. *Gene*, **234**(2): 361-9.

Hudman, J. F. and Gregg, K. (1989). Genetic diversity among strains of bacteria from the
rumen. *Current Microbiology*, **19**: 313-318.

Hulo, N., Sigrist, C. J., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P. and Bairoch, A. (2004). Recent improvements to the PROSITE database. *Nucleic Acids Res*, **32**(Database issue): D134-7.

Hungate, R. E. (1966) *The Rumen and Its Microbes,* Academic Press, New York.

Hwang, E. S. and Scocca, J. J. (1990). Interaction of integration host factor from Escherichia coli with the integration region of the Haemophilus influenzae bacteriophage HP1. *J Bacteriol*, **172**(9): 4852-60.

Iandolo, J. J., Worrell, V., Groicher, K. H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P., Qi, S., Zhu, H. and Roe, B. A. (2002). Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of Staphylococcus aureus 8325. *Gene*, **289**(1-2): 109-18.

Ingham, C. J., Hunter, I. S. and Smith, M. C. (1995). Rho-independent terminators without 3' poly-U tails from the early region of actinophage oC31. *Nucleic Acids Res*, **23**(3): 370-6.

Iverson, W. G. and Millis, N. F. (1976a). Characterization of Streptococcus bovis bacteriophages. *Can J Microbiol*, **22**(6): 847-52.

Iverson, W. G. and Millis, N. F. (1976b). Lysogeny in Streptococcus bovis. *Can J Microbiol*, **22**(6): 853-7.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, **3**: 318-56.

Jiang, S. C. and Paul, J. H. (1998). Gene transfer by transduction in the marine environment. *Appl Environ Microbiol*, **64**(8): 2780-7.

Jiang, W. H., Patterson, J. A. and Steenson, L. R. (1995). Isolation and characterization of a temperate bacteriophage from a ruminal acetogen. *Curr Microbiol*, **31**(6): 336-9.

Jones, R. J. and Lowry, J. B. (1984). Australian goats detoxify the goitrogen 3-hydroxy-4(1H) pyridone (DHP) after rumen infusion from an Indonesian goat. *Experientia*, **40**(12): 1435-6.

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F. and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J Mol Biol*, **299**(1): 27-51.

Kaatz, G. W., Seo, S. M. and Foster, T. J. (1999). Introduction of a norA promoter region mutation into the chromosome of a fluoroquinolone-susceptible strain of Staphylococcus aureus using plasmid integration. *Antimicrob Agents Chemother*, **43**(9): 2222-4.

Kaiser, K. and Murray, N. E. (1979). Physical characterisation of the "Rac prophage" in E. coli K12. *Mol Gen Genet*, **175**(2): 159-74.

Kaneko, J., Kimura, T., Narita, S., Tomita, T. and Kamio, Y. (1998). Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage phiPVL carrying Panton-Valentine leukocidin genes. *Gene*, **215**(1): 57-67.

Kapatral, V., Ivanova, N., Anderson, I., Reznik, G., Bhattacharyya, A., Gardner, W. L., Mikhailova, N., Lapidus, A., Larsen, N., D'Souza, M., Walunas, T., Haselkorn, R., Overbeek, R. and Kyrpides, N. (2003). Genome analysis of F. nucleatum sub spp vincentii and its comparison with the genome of F. nucleatum ATCC 25586. *Genome Res*, **13**(6A): 1180-9.

Kapfhammer, D., Blass, J., Evers, S. and Reidl, J. (2002). Vibrio cholerae phage K139: complete genome sequence and comparative genomics of related phages. *J Bacteriol*, **184**(23): 6592-601.

Karzai, A. W., Roche, E. D. and Sauer, R. T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol*, **7**(6): 449-55.

Kashige, N., Nakashima, Y., Miake, F. and Watanabe, K. (2000). Cloning, sequence analysis, and expression of Lactobacillus casei phage PL-1 lysis genes. *Arch Virol*, **145**(8): 1521-34.

Kasper, D. L. (1986). Bacterial capsule--old dogmas and new tricks. *J Infect Dis*, **153**(3): 407-15.

Kasper, D. L., Weintraub, A., Lindberg, A. A. and Lonngren, J. (1983). Capsular polysaccharides and lipopolysaccharides from two Bacteroides fragilis reference strains: chemical and immunochemical characterization. *J Bacteriol*, **153**(2): 991-7.

Katz, E., Thompson, C. J. and Hopwood, D. A. (1983). Cloning and expression of the tyrosinase gene from Streptomyces antibioticus in Streptomyces lividans. *J Gen Microbiol*, **129**(9): 2703-14.

Kholodii, G., Mindlin, S. Z., Gorlenko Zh, M., Bass, I. A., Kaliaeva, E. S. and Nikiforov, V. G. (2000). [Molecular genetic analysis of the Tn5041 transposition system]. *Genetika*, **36**(4): 459-69.

Klieve, A. V., Bain, P. A., Yokoyama, M. T., Ouwerkerk, D., Forster, R. J. and Turner, A. F. (2004). Bacteriophages that infect the cellulolytic ruminal bacterium Ruminococcus albus AR67. *Lett Appl Microbiol*, **38**(4): 333-8.

Klieve, A. V. and Bauchop, T. (1988). Morphological diversity of ruminal bacteriophages from sheep and cattle. *Appl Environ Microbiol*, **54**(6): 1637-41.

Klieve, A. V., Gregg, K. and Bauchop, T. (1991). Isolation and characterization of lytic phages from Bacteroides ruminicola ss brevis. *Current Microbiology*: 183-7.

Klieve, A. V., Hudman, J. F. and Bauchop, T. (1989). Inducible bacteriophages from ruminal bacteria. *Appl Environ Microbiol*, **55**(6): 1630-4.

Klieve, A. V. and Swain, R. A. (1993). Estimation of ruminal bacteriophage numbers by pulsed-field gel electrophoresis and laser densitometry. *Appl Environ Microbiol*, **59**(7): 2299-303.

Ko, M., Choi, H. and Park, C. (2002). Group I self-splicing intron in the recA gene of Bacillus anthracis. *J Bacteriol*, **184**(14): 3917-22.

Koo, K. and Jaykus, L. A. (2000). Selective amplification of bacterial RNA: use of a DNA primer containing mismatched bases near its 3' terminus to reduce false-positive signals. *Lett Appl Microbiol*, **31**(3): 187-92.

Kobayashi, Y., Forster, R. J., Hefford, M. A., Teather, R. M., Wakita, M., Ohmiya, K. and Hoshino, S. (1995). Analysis of the sequence of a new cryptic plasmid, pRJF2, from a rumen bacterium of the genus Butyrivibrio: comparison with other Butyrivibrio plasmids and application in the development of cloning vector. *FEMS Microbiol Lett*, **130**(2-3): 137-43.

Kobayashi, Y., Okuda, N., Matsumoto, M., Inoue, K., Wakita, M. and Hoshino, S. (1998). Constitutive expression of a heterologous Eubacterium ruminantium xylanase gene (xynA) in Butyrivibrio fibrisolvens. *FEMS Microbiol Lett*, **163**(1): 11-7.

Kodaira, K. I., Oki, M., Kakikawa, M., Watanabe, N., Hirakawa, M., Yamada, K. and Taketo, A. (1997). Genome structure of the Lactobacillus temperate phage phi g1e: the whole genome sequence and the putative promoter/repressor system. *Gene*, **187**(1): 45-53.

Kolot, M., Meroz, A. and Yagil, E. (2003). Site-specific recombination in human cells catalyzed by the wild-type integrase protein of coliphage HK022. *Biotechnol Bioeng*, **84**(1): 56-60.

Kolot, M., Silberstein, N. and Yagil, E. (1999). Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022. *Mol Biol Rep*, **26**(3): 207-13.

Kolot, M. and Yagil, E. (2003). Determinants that target the integrase of phage HK022 into the mammalian nucleus. *J Mol Biol*, **325**(4): 629-35.

Krinos, C. M., Coyne, M. J., Weinacht, K. G., Tzianabos, A. O., Kasper, D. L. and Comstock, L. E. (2001). Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature*, **414**(6863): 555-8.

Kropinski, A. M. (2000). Sequence of the genome of the temperate, serotype-converting, Pseudomonas aeruginosa bacteriophage D3. *J Bacteriol*, **182**(21): 6066-74.

Kuhsel, M. G., Strickland, R. and Palmer, J. D. (1990). An ancient group I intron shared by eubacteria and chloroplasts. *Science*, **250**(4987): 1570-3.

Kuhstoss, S., Richardson, M. A. and Rao, R. N. (1989). Site-specific integration in Streptomyces ambofaciens: localization of integration functions in S. ambofaciens plasmid pSAM2. *J Bacteriol*, **171**(1): 16-23.

Kuhstoss, S., Richardson, M. A. and Rao, R. N. (1991). Plasmid cloning vectors taht integrates site-specifically in *Streptomyces* spp. *Gene*, **97**: 143-146.

Kulaeva, O. I., Koonin, E. V., Wootton, J. C., Levine, A. S. and Woodgate, R. (1998). Unusual insertion element polymorphisms in the promoter and terminator regions of the mucAB-like genes of R471a and R446b. *Mutat Res*, **397**(2): 247-62.

Kumar, S., Tamura, K. and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, **5**(2): 150-63.

Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., Kuhara, S., Hattori, M., Hayashi, T. and Ohnishi, Y. (2004). Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. *Proc Natl Acad Sci U S A*, **101**(41): 14919-24.

Kwon, H. J., Tirumalai, R., Landy, A. and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science*, **276**(5309): 126-31.

Lakhno, V. M. and Bordunovskii, V. N. (2001). [Use of bacteriophage therapy in surgical practice]. *Vestn Khir Im I I Grek*, **160**(1): 122-5.

Lambowitz, A. M. and Belfort, M. (1993). Introns as mobile genetic elements. *Annu Rev Biochem*, **62**: 587-622.

Larimer, F. W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M. L., Pelletier, D. A., Beatty, J. T., Lang, A. S., Tabita, F. R., Gibson, J. L., Hanson, T. E., Bobst, C., Torres, J. L., Peres, C., Harrison, F. H., Gibson, J. and Harwood, C. S. (2004). Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nat Biotechnol*, **22**(1): 55-61.

Lauer, P., Chow, M. Y., Loessner, M. J., Portnoy, D. A. and Calendar, R. (2002). Construction, characterization, and use of two Listeria monocytogenes site-specific phage integration vectors. *J Bacteriol*, **184**(15): 4177-86.

Lavigne, R., Sun, W. D. and Volckaert, G. (2004). PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics*, **20**(5): 629-35.

Lazarevic, V., Dusterhoft, A., Soldo, B., Hilbert, H., Mauel, C. and Karamata, D. (1999). Nucleotide sequence of the Bacillus subtilis temperate bacteriophage SPbetac2. *Microbiology*, **145 (Pt 5)**: 1055-67.

Lee, C. Y., Buranen, S. L. and Ye, Z. H. (1991a). Construction of single-copy integration vectors for Staphylococcus aureus. *Gene*, **103**(1): 101-5.

Lee, M. H. and Hatfull, G. F. (1993). Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. *J Bacteriol*, **175**(21): 6836-41.

Lee, M. H., Pascopella, L., Jacobs, W. R., Jr. and Hatfull, G. F. (1991b). Site-specific integration of mycobacteriophage L5: integration-proficient vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin. *Proc Natl Acad Sci U S A*, **88**(8): 3111-5.

LeFevre, K. R. and Cordes, M. H. (2003). Retroevolution of lambda Cro toward a stable monomer. *Proc Natl Acad Sci U S A*, **100**(5): 2345-50.

Leong, J. M., Nunes-Duby, S. E., Oser, A. B., Lesser, C. F., Youderian, P., Susskind, M. M. and Landy, A. (1986). Structural and regulatory divergence among site-specific recombination genes of lambdoid phage. *J Mol Biol*, **189**(4): 603-16.

Levin, M. E., Hendrix, R. W. and Casjens, S. R. (1993). A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *J Mol Biol*, **234**(1): 124-39.

Lewis, J. A. and Hatfull, G. F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res*, **29**(11): 2205-16.

Li, X., Zhou, X. and Deng, Z. (2003). Vector systems allowing efficient autonomous or integrative gene cloning in Micromonospora sp. strain 40027. *Appl Environ Microbiol*, **69**(6): 3144-51.

Lillehaug, D. and Birkeland, N. K. (1993). Characterization of genetic elements required for site-specific integration of the temperate lactococcal bacteriophage phi LC3 and construction of integration-negative phi LC3 mutants. *J Bacteriol*, **175**(6): 1745-55.

Linderoth, N. A., Ziermann, R., Haggard-Ljungquist, E., Christie, G. E. and Calendar, R. (1991). Nucleotide sequence of the DNA packaging and capsid synthesis genes of bacteriophage P2. *Nucleic Acids Res*, **19**(25): 7207-14.

Lindsey, D. F., Mullin, D. A. and Walker, J. R. (1989). Characterization of the cryptic lambdoid prophage DLP12 of Escherichia coli and overlap of the DLP12 integrase gene with the tRNA gene argU. *J Bacteriol*, **171**(11): 6197-205.

Little, J. W., Shepley, D. P. and Wert, D. W. (1999). Robustness of a gene regulatory circuit. *Embo J*, **18**(15): 4299-4307.

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T., McCarty, J., Srikumar, R., Williams, D., Wu, J. J., Gros, P., Pelletier, J. and DuBow, M. (2004). Antimicrobial drug discovery through bacteriophage genomics. *Nat Biotechnol*, **22**(2): 185-91.

Lobocka, M. B., Rose, D. J., Plunkett, G., 3rd, Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M. B. and Blattner, F. R. (2004). Genome of bacteriophage P1. *J Bacteriol*, **186**(21): 7032-68.

Lockington, R. A., Attwood, G. T. and Brooker, J. D. (1988). Isolation and characterization of a temperate bacteriophage from the ruminal anaerobe Selenomonas ruminantium. *Appl Environ Microbiol*, **54**(6): 1575-80.

Loeffler, J. M., Nelson, D. and Fischetti, V. A. (2001). Rapid killing of Streptococcus pneumoniae with a bacteriophage cell wall hydrolase. *Science*, **294**(5549): 2170-2.

Loessner, M. J., Inman, R. B., Lauer, P. and Calendar, R. (2000). Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of Listeria monocytogenes: implications for phage evolution. *Mol Microbiol*, **35**(2): 324-40.

Lorbach, E., Christ, N., Schwikardi, M. and Droge, P. (2000). Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. *J Mol Biol*, **296**(5): 1175-81.

Lowe, T. M. and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*, **25**(5): 955-64.

Lu, G. and Moriyama, E. N. (2004). Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform*, **5**(4): 378-88.

Luria, S. E., Delbruck, M. and Schlichting, C. (1943). Mutations of bacteria from vrius sensitivity to virus resistance. *Genetics*, **28**: 491.

Macrina, F. L., Tobian, J. A., Jones, K. R., Evans, R. P. and Clewell, D. B. (1982). A cloning vector able to replicate in Escherichia coli and Streptococcus sanguis. *Gene*, **19**(3): 345-53.

Madden, T. L., Tatusov, R. L. and Zhang, J. (1996). Applications of network BLAST server. *Methods Enzymol*, **266**: 131-41.

Magrini, V., Creighton, C. and Youderian, P. (1999). Site-specific recombination of temperate Myxococcus xanthus phage Mx8: genetic elements required for integration. *J Bacteriol*, **181**(13): 4050-61.

Makino, K., Ishii, K., Yasunaga, T., Hattori, M., Yokoyama, K., Yutsudo, C. H., Kubota, Y., Yamaichi, Y., Iida, T., Yamamoto, K., Honda, T., Han, C. G., Ohtsubo, E., Kasamatsu, M., Hayashi, T., Kuhara, S. and Shinagawa, H. (1998). Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic Escherichia coli O157:H7 derived from Sakai outbreak. *DNA Res*, **5**(1): 1-9.

Mann, S. P., Hazlewood, G. P. and Orpin, C. G. (1986). Characterization of a cryptic plasmid (pOM1) in Butyrivibrio fibrisolvens by restriction endonuclease analysis and its cloning in Eschersichia coli. *Current Microbiol*, **13**: 17-22.

Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S., Hurwitz, D. I., Jackson, J. D., Jacobs, A. R., Lanczycki, C. J., Liebert, C. A., Liu, C., Madej, T., Marchler, G. H., Mazumder, R., Nikolskaya, A. N., Panchenko, A. R., Rao, B. S., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Vasudevan, S., Wang, Y., Yamashita, R. A., Yin, J. J. and Bryant, S. H. (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res*, **31**(1): 383-7.

Marsh, J. L., Erfle, M. and Wykes, E. J. (1984). The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. *Gene*, **32**(3): 481-5.

Martin, S. A. and Dean, R. G. (1989). Characterization of a plasmid from the ruminal bacterium Selenomonas ruminantium. *Appl Environ Microbiol*, **55**(12): 3035-8.

Martinez-Abarca, F. and Toro, N. (2000). Group II introns in the bacterial world. *Mol Microbiol*, **38**(5): 917-26.

Martinez-Abarca, F., Zekri, S. and Toro, N. (1998). Characterization and splicing in vivo of a Sinorhizobium meliloti group II intron associated with particular insertion sequences of the IS630-Tc1/IS3 retroposon superfamily. *Mol Microbiol*, **28**(6): 1295-306.

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., Cavanagh, S., Dunny, G. M., Belfort, M. and Lambowitz, A. M. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev*, **11**(21): 2910-24.

Matteuzzi, D. and Sozzi, T. (1971). Bifidobacterium bacteriophage from calf rumen. *Z Allg Mikrobiol*, **11**(1): 57-8.

Matzura, O. and Wennborg, A. (1996). RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Comput Appl Biosci*, **12**(3): 247-9.

May, T., Kocherginskaya, S. A., Mackie, R. I., Vercoe, P. E. and White, B. A. (1996). Complete nucleotide sequence of a cryptic plasmid, pBAW301, from the ruminal anaerobe Ruminococcus flavefaciens R13e2. *FEMS Microbiol Lett*, **144**(2-3): 221-7.

McDonnell, G. E., Wood, H., Devine, K. M. and McConnell, D. J. (1994). Genetic control of bacterial suicide: regulation of the induction of PBSX in Bacillus subtilis. *J Bacteriol*, **176**(18): 5820-30.

McShan, W. M., McLaughlin, R. E., Nordstrand, A. and Ferretti, J. J. (1998). Vectors containing streptococcal bacteriophage integrases for site-specific gene insertion. *Methods in Cell Science*, **20**: 51-57.

McShan, W. M., Tang, Y. F. and Ferretti, J. J. (1997). Bacteriophage T12 of Streptococcus pyogenes integrates into the gene encoding a serine tRNA. *Mol Microbiol*, **23**(4): 719-28.

McSweeney, C. S., Dalrymple, B. P., Gobius, K. S., Kennedy, P. M., Krause, D. O., Mackie, R. I. and Xue, G. P. (1999). The application of rumen biotechnology to improve the nutritive value of fibrous feedstuffs: pre- and post-ingestion. *Livestock Production Science*, **59**(2-3): 265-283.

Mercer, D. K., Patel, S. and Flint, H. J. (2001). Sequence analysis of the plasmid pRRI2 from the rumen bacterium Prevotella ruminicola 223/M2/7 and the use of pRRI2 in Prevotella/Bacteroides Shuttle Vectors. *Plasmid*, **45**(3): 227-32.

Mertz, J. E. and Davis, R. W. (1972). Cleavage of DNA by R 1 restriction endonuclease generates cohesive ends. *Proc Natl Acad Sci U S A*, **69**(11): 3370-4.

Mills, D. A., McKay, L. L. and Dunny, G. M. (1996). Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J Bacteriol*, **178**(12): 3531-8.

Miyamoto, H., Nakai, W., Yajima, N., Fujibayashi, A., Higuchi, T., Sato, K. and Matsushiro, A. (1999). Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions. *DNA Res*, **6**(4): 235-40.

Mohr, G., Smith, D., Belfort, M. and Lambowitz, A. M. (2000). Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev*, **14**(5): 559-73.

Moitoso de Vargas, L. and Landy, A. (1991). A switch in the formation of alternative DNA loops modulates lambda site-specific recombination. *Proc Natl Acad Sci U S A*, **88**(2): 588-92.

Moon, S., Byun, Y., Kim, H. J., Jeong, S. and Han, K. (2004). Predicting genes expressed via -1 and +1 frameshifts. *Nucleic Acids Res*, **32**(16): 4884-92.

Moreau, S., Blanco, C. and Trautwetter, A. (1999). Site-specific integration of corynephage phi16: construction of an integration vector. *Microbiology*, **145 (Pt 3)**: 539-48.

Morrison, M., Mackie, R. I. and White, B. A. (1992). Partial purification and characterization of Ral8I, a class-IIS restriction endonuclease from Ruminococcus albus 8 which recognizes 5'-GGATC. *Gene*, **111**(1): 105-8.

Morrison, M., Mackie, R. I. and White, B. A. (1994). The restriction endonuclease RflFII, isolated from Ruminococcus flavefaciens FD-1, recognizes the sequence 5'-AGTACT-3', and is inhibited by site-specific adenine methylation. *FEMS Microbiol Lett*, **122**(1-2): 181-5.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silventoinen, V., Studholme, D. J., Vaughan, R. and Wu, C. H. (2005). InterPro, progress and status in 2005. *Nucleic Acids Res*, **33 Database Issue**: D201-5.

Mullany, P., Pallen, M., Wilks, M., Stephen, J. R. and Tabaqchali, S. (1996). A group II intron in a conjugative transposon from the gram-positive bacterium, Clostridium difficile. *Gene*, **174**(1): 145-50.

Mulligan, M. E., Brosius, J. and McClure, W. R. (1985). Characterization in vitro of the effect of spacer length on the activity of Escherichia coli RNA polymerase at the TAC promoter. *J Biol Chem*, **260**(6): 3529-38.

Nakamura, M., Nagamine, T., Harada, C., Tajima, K., Matsui, H. and Benno, Y. (2003). Expression of Ruminococcus albus xylanase gene (xynA) in Streptococcus bovis 12-U-1. *Curr Microbiol*, **47**(1): 71-4.

Nakamura, M., Nagamine, T., Ogata, K., Tajima, K., Aminov, R. I. and Benno, Y. (1999). Sequence analysis of small cryptic plasmids isolated from Selenomonas ruminantium S20. *Curr Microbiol*, **38**(2): 107-12.

Nakamura, M., Ogata, K., Nagamine, T., Tajima, K., Matsui, H. and Benno, Y. (2000). Characterization of the cryptic plasmid pSBO2 isolated from Streptococcus bovis JB1 and construction of a new shuttle vector. *Curr Microbiol*, **41**(1): 27-32.

Nakamura, M., Ogata, K., Nagamine, T., Tajima, K., Matsui, H. and Benno, Y. (2001). The replicon of the cryptic Plasmid pSBO1 isolated from Streptococcus bovis JB1. *Curr Microbiol*, **43**(1): 11-6.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y. and Hayashi, T. (1999). The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of Pseudomonas aeruginosa: implications for phage evolution and horizontal gene transfer via bacteriophages. *Mol Microbiol*, **31**(2): 399-419.

Nascimento, A. L., Ko, A. I., Martins, E. A., Monteiro-Vitorello, C. B., Ho, P. L., Haake, D. A., Verjovski-Almeida, S., Hartskeerl, R. A., Marques, M. V., Oliveira, M. C., Menck, C. F., Leite, L. C., Carrer, H., Coutinho, L. L., Degrave, W. M., Dellagostin, O. A., El-Dorry, H., Ferro, E. S., Ferro, M. I., Furlan, L. R., Gamberini, M., Giglioti, E. A., Goes-Neto, A., Goldman, G. H., Goldman, M. H., Harakava, R., Jeronimo, S. M., Junqueira-de-Azevedo, I. L., Kimura, E. T., Kuramae, E. E., Lemos, E. G., Lemos, M. V., Marino, C. L., Nunes, L. R., de Oliveira, R. C., Pereira, G. G., Reis, M. S., Schriefer, A., Siqueira, W. J., Sommer, P., Tsai, S. M., Simpson, A. J., Ferro, J. A., Camargo, L. E., Kitajima, J. P., Setubal, J. C. and Van Sluys, M. A. (2004). Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. *J Bacteriol*, **186**(7): 2164-72.

Nash, H. A. (1990). Bending and supercoiling of DNA at the attachment site of bacteriophage lambda. *Trends Biochem Sci*, **15**(6): 222-7.

Nauta, A. (1997) In *Genetics*University of Groningen, Netherlands.

Nelson, D. (2004). Phage taxonomy: we agree to disagree. *J Bacteriol*, **186**(21): 7029-31.

Nelson, D., Loomis, L. and Fischetti, V. A. (2001). Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proc Natl Acad Sci U S A*, **98**(7): 4107-12.

Nelson, K. E., Fleischmann, R. D., DeBoy, R. T., Paulsen, I. T., Fouts, D. E., Eisen, J. A., Daugherty, S. C., Dodson, R. J., Durkin, A. S., Gwinn, M., Haft, D. H., Kolonay, J. F., Nelson, W. C., Mason, T., Tallon, L., Gray, J., Granger, D., Tettelin, H., Dong, H., Galvin, J. L., Duncan, M. J., Dewhirst, F. E. and Fraser, C. M. (2003). Complete genome sequence of the oral pathogenic Bacterium porphyromonas gingivalis strain W83. *J Bacteriol*, **185**(18): 5591-601.

Nesbo, C. L. and Doolittle, W. F. (2003). Active self-splicing group I introns in 23S rRNA genes of hyperthermophilic bacteria, derived from introns in eukaryotic organelles. *Proc Natl Acad Sci U S A*, **100**(19): 10806-11.

Nilsson, M. T. and Widersten, M. (2004). Repertoire selection of variant single-chain Cro: toward directed DNA-binding specificity of helix-turn-helix proteins. *Biochemistry*, **43**(38): 12038-47.

Nolan, J. V. and Stachiw, S. (1979). Fermentation and nitrogen dynamics in Merino sheep given a low-quality-roughage diet. *Br J Nutr*, **42**(1): 63-80.

Numrych, T. E., Gumport, R. I. and Gardner, J. F. (1990). A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophage Lambda. *Nucl. Acids. Res.*, **18**: 3953-3959.

Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. and Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res*, **26**(2): 391-406.

Obregon, V., Garcia, J. L., Garcia, E., Lopez, R. and Garcia, P. (2003). Genome organization and molecular analysis of the temperate bacteriophage MM1 of Streptococcus pneumoniae. *J Bacteriol*, **185**(7): 2362-8.

Ogata, K., Aminov, R. I., Nagamine, T., Benno, Y., Sekizaki, T., Mitsumori, M., Minato, H. and Itabashi, H. (1996). Structural organization of pRAM4, a cryptic plasmid from Prevotella ruminicola. *Plasmid*, **35**(2): 91-7.

Ogawa, T., Ogawa, H. and Tomizawa, J. (1988). Organization of the early region of bacteriophage phi 80. Genes and proteins. *J Mol Biol*, **202**(3): 537-50.

O'Gorman, S., Fox, D. T. and Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science*, **251**(4999): 1351-5.

Ohara, H., Miyagi, T., Kaneichi, K., Karita, S., Kobayashi, Y., Kimura, T., Sakka, K. and Ohmiya, K. (1998). Structural analysis of a new cryptic plasmid pAR67 isolated from Ruminococcus albus AR67. *Plasmid*, **39**(1): 84-8.

Ohmiya, K., Hoshino, C. and Shimuzu, S. (1989). Cellulose-dependent and penicillin-resistance plasmids isolated from Rumincoccus albus. *Asian Journal of Animal Science 2, 501-5*, **2**: 501-5.

Okamoto, K., Mudd, J. A., Mangan, J., Huang, W. M., Subbaiah, T. V. and Marmur, J. (1968). Properties of the defective phage of Bacillus subtilis. *J Mol Biol*, **34**(3): 413-28.

Okazaki, T. (1969). [Refining method of DNA polymerase]. *Tanpakushitsu Kakusan Koso*, **14**(12): 1078-85.

Okazaki, T. and Okazaki, R. (1969). Mechanism of DNA chain growth. IV. Direction of synthesis of T4 short DNA chains as revealed by exonucleolytic degradation. *Proc Natl Acad Sci U S A*, **64**(4): 1242-8.

Okinaka, R. T., Cloud, K., Hampton, O., Hoffmaster, A. R., Hill, K. K., Keim, P., Koehler, T. M., Lamke, G., Kumano, S., Mahillon, J., Manter, D., Martinez, Y., Ricke, D., Svensson, R. and Jackson, P.

J. (1999). Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes. *J Bacteriol*, **181**(20): 6509-15.

Olivares, E. C., Hollis, R. P. and Calos, M. P. (2001). Phage R4 integrase mediates site-specific integration in human cells. *Gene*, **278**(1-2): 167-76.

Olsthoorn, R. C., Garde, G., Dayhuff, T., Atkins, J. F. and Van Duin, J. (1995). Nucleotide sequence of a single-stranded RNA phage from Pseudomonas aeruginosa: kinship to coliphages and conservation of regulatory RNA structures. *Virology*, **206**(1): 611-25.

Orpin, C. G. and Munn, E. A. (1974). The occurrence of bacteriophages in the rumen and their influence on rumen bacterial populations. *Experientia*, **30**(9): 1018-20.

Ortiz-Urda, S., Lin, Q., Green, C. L., Keene, D. R., Marinkovich, M. P. and Khavari, P. A. (2003a). Injection of genetically engineered fibroblasts corrects regenerated human epidermolysis bullosa skin tissue. *J Clin Invest*, **111**(2): 251-5.

Ortiz-Urda, S., Thyagarajan, B., Keene, D. R., Lin, Q., Calos, M. P. and Khavari, P. A. (2003b). PhiC31 integrase-mediated nonviral genetic correction of junctional epidermolysis bullosa. *Hum Gene Ther*, **14**(9): 923-8.

Ortiz-Urda, S., Thyagarajan, B., Keene, D. R., Lin, Q., Fang, M., Calos, M. P. and Khavari, P. A. (2002). Stable nonviral genetic correction of inherited human skin disease. *Nat Med*, **8**(10): 1166-70.

Pantosti, A., Tzianabos, A. O., Onderdonk, A. B. and Kasper, D. L. (1991). Immunochemical characterization of two surface polysaccharides of Bacteroides fragilis. *Infect Immun*, **59**(6): 2075-82.

Papp, I., Dorgai, L., Papp, P., Jonas, E., Olasz, F. and Orosz, L. (1993). The bacterial attachment site of the temperate Rhizobium phage 16-3 overlaps the 3' end of a putative proline tRNA gene. *Mol Gen Genet*, **240**(2): 258-64.

Paquin, B., Heinfling, A. and Shub, D. A. (1999). Sporadic distribution of tRNA(Arg)CCU introns among alpha-purple bacteria: evidence for horizontal transmission and transposition of a group I intron. *J Bacteriol*, **181**(3): 1049-53.

Paquin, B., Kathe, S. D., Nierzwicki-Bauer, S. A. and Shub, D. A. (1997). Origin and evolution of group I introns in cyanobacterial tRNA genes. *J Bacteriol*, **179**(21): 6798-806.

Paradis, F. W., Zhu, H., Krell, P. J., Phillips, J. P. and Forsberg, C. W. (1993). The xynC gene from Fibrobacter succinogenes S85 codes for a xylanase with two similar catalytic domains. *J Bacteriol*, **175**(23): 7666-72.

Parker, M. M., Belisle, M. and Belfort, M. (1999). Intron homing with limited exon homology. Illegitimate double-strand-break repair in intron acquisition by phage t4. *Genetics*, **153**(4): 1513-23.

Pastushok, C. and Kennell, D. (1974). Residual polarity and transcription-translation coupling during recovery from chloramphenicol or fusidic acid. *J Bacteriol*, **117**(2): 631-40.

Paul, J. H. (1999). Microbial gene transfer: an ecological perspective. *J Mol Microbiol Biotechnol*, **1**(1): 45-50.

Paulsen, I. T., Banerjei, L., Myers, G. S., Nelson, K. E., Seshadri, R., Read, T. D., Fouts, D. E., Eisen, J. A., Gill, S. R., Heidelberg, J. F., Tettelin, H., Dodson, R. J., Umayam, L., Brinkac, L., Beanan, M., Daugherty, S., DeBoy, R. T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K. A., Dougherty, B. A. and Fraser, C. M. (2003). Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis. *Science*, **299**(5615): 2071-4.

Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J. G., Jacobs, W. R., Jr., Hendrix, R. W. and Hatfull, G. F. (2003a). Origins of highly mosaic mycobacteriophage genomes. *Cell*, **113**(2): 171-82.

Pedulla, M. L., Ford, M. E., Karthikeyan, T., Houtz, J. M., Hendrix, R. W., Hatfull, G. F., Poteete, A. R., Gilcrease, E. B., Winn-Stapley, D. A. and Casjens, S. R. (2003b). Corrected sequence of the bacteriophage p22 genome. *J Bacteriol*, **185**(4): 1475-7.

Pena, C. E., Kahlenberg, J. M. and Hatfull, G. F. (1999). Protein-DNA complexes in mycobacteriophage L5 integrative recombination. *J Bacteriol*, **181**(2): 454-61.

Pena, C. E., Kahlenberg, J. M. and Hatfull, G. F. (2000). Assembly and activation of site-specific recombination complexes. *Proc Natl Acad Sci U S A*, **97**(14): 7760-5.

Pena, C. E., Lee, M. H., Pedulla, M. L. and Hatfull, G. F. (1997). Characterization of the mycobacteriophage L5 attachment site, attP. *J Mol Biol*, **266**(1): 76-92.

Perna, N. T., Plunkett, G., 3rd, Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A. and Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature*, **409**(6819): 529-33.

Petersen, A., Josephsen, J. and Johnsen, M. G. (1999). TPW22, a lactococcal temperate phage with a site-specific integrase closely related to Streptococcus thermophilus phage integrases. *J Bacteriol*, **181**(22): 7034-42.

Pierce, J. C. and Masker, W. (1992). Frameshift mutagenesis in bacteriophage T7. *Mutat Res*, **281**(2): 81-7.

Pierson, L. S., 3rd and Kahn, M. L. (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**(3): 487-96.

Plunkett, G., 3rd, Rose, D. J., Durfee, T. J. and Blattner, F. R. (1999). Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product. *J Bacteriol*, **181**(6): 1767-78.

Potvin, C., Leclerc, D., Tremblay, G., Asselin, A. and Bellemare, G. (1988). Cloning, sequencing and expression of a Bacillus bacteriolytic enzyme in Escherichia coli. *Mol Gen Genet*, **214**(2): 241-8.

Prescott, L., M.Harley, J. P. and Klein, D. A. (1996) *Microbiology,* Wm. C. Brown Publishers, Boston.

Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A. C., Zwahlen, M. C., Rouvet, M., Altermann, E., Barrangou, R., Mollet, B., Mercenier, A., Klaenhammer, T., Arigoni, F. and Schell, M. A. (2004). The genome sequence of the probiotic intestinal bacterium Lactobacillus johnsonii NCC 533. *Proc Natl Acad Sci U S A*, **101**(8): 2512-7.

Proux, C., van Sinderen, D., Suarez, J., Garcia, P., Ladero, V., Fitzgerald, G. F., Desiere, F. and Brussow, H. (2002). The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria. *J Bacteriol*, **184**(21): 6026-36.

Ptashne, M. (1967). Specific binding of the lambda phage repressor to lambda DNA. *Nature*, **214**(85): 232-4.

Quenneville, S. P., Chapdelaine, P., Rousseau, J., Beaulieu, J., Caron, N. J., Skuk, D., Mills, P., Olivares, E. C., Calos, M. P. and Tremblay, J. P. (2004). Nucleofection of muscle-derived stem cells and myoblasts with phiC31 integrase: stable expression of a full-length-dystrophin fusion gene by human myoblasts. *Mol Ther*, **10**(4): 679-87.

Rajakumar, K., Sasakawa, C. and Adler, B. (1997). Use of a novel approach, termed island probing, identifies the Shigella flexneri she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect Immun*, **65**(11): 4606-14.

RajBhandary, U. L. (1994). Initiator transfer RNAs. *J Bacteriol*, **176**(3): 547-52.

Ramirez, M., Severina, E. and Tomasz, A. (1999). A high incidence of prophage carriage among natural isolates of Streptococcus pneumoniae. *J Bacteriol*, **181**(12): 3618-25.

Raya, R. R., Fremaux, C., De Antoni, G. L. and Klaenhammer, T. R. (1992). Site-specific integration of the temperate bacteriophage phi adh into the Lactobacillus gasseri chromosome and molecular characterization of the phage (attP) and bacterial (attB) attachment sites. *J Bacteriol*, **174**(17): 5584-92.

Recktenwald, J. and Schmidt, H. (2002). The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. *Infect Immun*, **70**(4): 1896-908.

Reese, M. G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. *Comput Chem*, **26**(1): 51-6.

Regamey, A. and Karamata, D. (1998). The N-acetylmuramoyl-L-alanine amidase encoded by the Bacillus subtilis 168 prophage SP beta. *Microbiology*, **144 (Pt 4)**: 885-93.

Reinhold-Hurek, B. and Shub, D. A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature*, **357**(6374): 173-6.

Reiter, W. D., Palm, P. and Yeats, S. (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res*, **17**(5): 1907-14.

Ribeiro, G., Viveiros, M., David, H. L. and Costa, J. V. (1997). Mycobacteriophage D29 contains an integration system similar to that of the temperate mycobacteriophage L5. *Microbiology*, **143 (Pt 8)**: 2701-8.

Rigden, D. J., Jedrzejas, M. J. and Galperin, M. Y. (2003). Amidase domains from bacterial and phage autolysins define a family of gamma-D,L-glutamate-specific amidohydrolases. *Trends Biochem Sci*, **28**(5): 230-4.

Ripley, L. S. and Clark, A. (1986). Frameshift mutations produced by proflavin in bacteriophage T4: specificity within a hotspot. *Proc Natl Acad Sci U S A*, **83**(18): 6954-8.

Ripley, L. S. and Shoemaker, N. B. (1983). A major role for bacteriophage T4 DNA polymerase in frameshift mutagenesis. *Genetics*, **103**(3): 353-66.

Robertson, C. A. and Nash, H. A. (1988). Bending of the bacteriophage lambda attachment site by Escherichia coli integration host factor. *J Biol Chem*, **263**(8): 3554-7.

Rohwer, F. (2003). Global phage diversity. *Cell*, **113**(2): 141.

Rohwer, F. and Edwards, R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol*, **184**(16): 4529-35.

Romero, P., Lopez, R. and Garcia, E. (2004a). Characterization of LytA-like N-acetylmuramoyl-L-alanine amidases from two new Streptococcus mitis bacteriophages provides insights into the properties of the major pneumococcal autolysin. *J Bacteriol*, **186**(24): 8229-39.

Romero, P., Lopez, R. and Garcia, E. (2004b). Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus. *Virology*, **322**(2): 239-52.

Romine, M. F., Stillwell, L. C., Wong, K. K., Thurston, S. J., Sisk, E. C., Sensen, C., Gaasterland, T., Fredrickson, J. K. and Saffer, J. D. (1999). Complete sequence of a 184-kilobase catabolic plasmid from Sphingomonas aromaticivorans F199. *J Bacteriol*, **181**(5): 1585-602.

Rossignol, M., Moulin, L. and Boccard, F. (2002). Phage HK022-based integrative vectors for the insertion of genes in the chromosome of multiply marked Escherichia coli strains. *FEMS Microbiol Lett*, **213**(1): 45-9.

Rudi, K. and Jakobsen, K. S. (1999). Complex evolutionary patterns of tRNA Leu(UAA) group I introns in the cyanobacterial radiation [corrected]. *J Bacteriol*, **181**(11): 3445-51.

Russell, J. B. (1998). Strategies that ruminal bacteria use to handle excess carbohydrate. *J Anim Sci*, **76**(7): 1955-63.

Sadowski, P. (1986). Site-specific recombinases: changing partners and doing the twist. *J Bacteriol*, **165**(2): 341-7.

Sam, M. D., Papagiannis, C. V., Connolly, K. M., Corselli, L., Iwahara, J., Lee, J., Phillips, M., Wojciak, J. M., Johnson, R. C. and Clubb, R. T. (2002). Regulation of directionality in bacteriophage lambda site-specific recombination: structure of the Xis protein. *J Mol Biol*, **324**(4): 791-805.

Sandegren, L. and Sjoberg, B. M. (2004). Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J Biol Chem*, **279**(21): 22218-27.

Sarkar, D., Radman-Livaja, M. and Landy, A. (2001). The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. *Embo J*, **20**(5): 1203-12.

Sauer, B. (1994). Site-specific recombination: developments and applications. *Curr Opin Biotechnol*, **5**(5): 521-7.

Sauer, B. and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci U S A*, **85**(14): 5166-70.

Saviola, B. and Bishai, W. R. (2004). Method to integrate multiple plasmids into the mycobacterial chromosome. *Nucleic Acids Res*, **32**(1): e11.

Schindler, D. and Echols, H. (1981). Retroregulation of the int gene of bacteriophage lambda: control of translation completion. *Proc Natl Acad Sci U S A*, **78**(7): 4475-9.

Schuch, R., Nelson, D. and Fischetti, V. A. (2002). A bacteriolytic agent that detects and kills Bacillus anthracis. *Nature*, **418**(6900): 884-9.

Schweizer, H. P. (1992). Allelic exchange in Pseudomonas aeruginosa using novel ColE1-type vectors and a family of cassettes containing a portable oriT and the counter-selectable Bacillus subtilis sacB marker. *Mol Microbiol*, **6**(9): 1195-204.

Semsey, S., Blaha, B., Koles, K., Orosz, L. and Papp, P. P. (2002). Site-specific integrative elements of rhizobiophage 16-3 can integrate into proline tRNA (CGG) genes in different bacterial genera. *J Bacteriol*, **184**(1): 177-82.

Semsey, S., Papp, I., Buzas, Z., Patthy, A., Orosz, L. and Papp, P. P. (1999). Identification of site-specific recombination genes int and xis of the Rhizobium temperate phage 16-3. *J Bacteriol*, **181**(14): 4185-92.

Shah, H. N. and Collins, D. M. (1990). Prevotella, a new genus to include Bacteroides melaninogenicus and related species formerly classified in the genus Bacteroides. *Int J Syst Bacteriol*, **40**(2): 205-8.

Shah, H. N. and Gharbia, S. E. (1993). Ecophysiology and taxonomy of Bacteroides and related taxa. *Clin Infect Dis*, **16 Suppl 4**: S160-7.

Shimada, K. and Campbell, A. (1974a). Int-constitutive mutants of bacteriophage lambda. *Proc Natl Acad Sci U S A*, **71**(1): 237-41.

Shimada, K. and Campbell, A. (1974b). Lysogenization and curing by int-constitutive mutants of phage lambda. *Virology*, **60**(1): 157-65.

Shimizu-Kadota, M. (2001). A method to maintain introduced DNA sequences stably and safely on the bacterial chromosome: application of prophage integration and subsequent designed excision. *J Biotechnol*, **89**(1): 73-9.

Shimizu-Kadota, M., Kiwaki, M., Sawaki, S., Shirasawa, Y., Shibahara-Sone, H. and Sako, T. (2000). Insertion of bacteriophage phiFSW into the chromosome of Lactobacillus casei strain Shirota (S-1): characterization of the attachment sites and the integrase gene. *Gene*, **249**(1-2): 127-34.

Shoemaker, N. B., Anderson, K. L., Smithson, S. L., Wang, G. R. and Salyers, A. A. (1991). Conjugal transfer of a shuttle vector from the human colonic anaerobe Bacteroides uniformis to the

ruminal anaerobe Prevotella (Bacteroides) ruminicola B(1)4. *Appl Environ Microbiol*, **57**(8): 2114-2120.

Shoemaker, N. B., Wang, G. R. and Salyers, A. A. (2000). Multiple gene products and sequences required for excision of the mobilizable integrated Bacteroides element NBU1. *J Bacteriol*, **182**(4): 928-36.

Siboo, I. R., Bensing, B. A. and Sullam, P. M. (2003). Genomic organization and molecular characterization of SM1, a temperate bacteriophage of Streptococcus mitis. *J Bacteriol*, **185**(23): 6968-75.

Smith, C. J. and Hespell, R. B. (1983). Prospects for development and use of recombinant deoxyribonucleic acid techniques with ruminal bacteria. *J Dairy Sci*, **66**(7): 1536-46.

Smith, D. E., Tans, S. J., Smith, S. B., Grimes, S., Anderson, D. L. and Bustamante, C. (2001). The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature*, **413**(6857): 748-52.

Smith, M. C., Burns, R. N., Wilson, S. E. and Gregory, M. A. (1999). The complete genome sequence of the Streptomyces temperate phage straight phiC31: evolutionary relationships to other viruses. *Nucleic Acids Res*, **27**(10): 2145-55.

Smith, M. C. and Thorpe, H. M. (2002). Diversity in the serine recombinases. *Mol Microbiol*, **44**(2): 299-307.

Smith, M. C., Till, R., Brady, K., Soultanas, P. and Thorpe, H. (2004). Synapsis and DNA cleavage in phiC31 integrase-mediated site-specific recombination. *Nucleic Acids Res*, **32**(8): 2607-17.

Springer, B., Sander, P., Sedlacek, L., Ellrott, K. and Bottger, E. C. (2001). Instability and site-specific excision of integration-proficient mycobacteriophage L5 plasmids: development of stably maintained integrative vectors. *Int J Med Microbiol*, **290**(8): 669-75.

Stappenbeck, T., Hooper, L. and Gordon, J. I. (2002). Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A*, **99**: 15451-15455.

Stark, W. M., Boocock, M. R. and Sherratt, D. J. (1992). Catalysis by site-specific recombinases. *Trends Genet*, **8**(12): 432-9.

Stoll, S. M., Ginsburg, D. S. and Calos, M. P. (2002). Phage TP901-1 site-specific integrase functions in human cells. *J Bacteriol*, **184**(13): 3657-63.

Streisinger, G. and Owen, J. (1985). Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics*, **109**(4): 633-59.

Subramaniam, S., Tewari, A. K., Nunes-Duby, S. E. and Foster, M. P. (2003). Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. *J Mol Biol*, **329**(3): 423-39.

Sulakvelidze, A., Alavidze, Z. and Morris, J. G., Jr. (2001). Bacteriophage therapy. *Antimicrob Agents Chemother*, **45**(3): 649-59.

Summers, W. C. (2001). Bacteriophage therapy. *Annu Rev Microbiol*, **55**: 437-51.

Sun, J., Inouye, M. and Inouye, S. (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**(13): 4171-81.

Swalla, B. M., Cho, E. H., Gumport, R. I. and Gardner, J. F. (2003). The molecular basis of co-operative DNA binding between lambda integrase and excisionase. *Mol Microbiol*, **50**(1): 89-99.

Tamada, H., Harasawa, R. and Shinjo, T. (1985). Isolation of a bacteriophage in Fusobacterium necrophorum. *Nippon Juigaku Zasshi*, **47**(3): 483-6.

Tan, P., Wang, X. and Wang, J. (1994). [Rumen bacteria degrading toxic mimosine and dihydroxypyridine compounds in China]. *Wei Sheng Wu Xue Bao*, **34**(5): 379-84.

Tanaka, K., Nishimori, K., Makino, S., Nishimori, T., Kanno, T., Ishihara, R., Sameshima, T., Akiba, M., Nakazawa, M., Yokomizo, Y. and Uchida, I. (2004). Molecular characterization of a prophage of Salmonella enterica serotype Typhimurium DT104. *J Clin Microbiol*, **42**(4): 1807-12.

Tarakanov, B. V. (1976). [Biological properties of Streptococcus bovis bacteriophages isolated from lysogenic cultures and sheep rumen]. *Mikrobiologiia*, **45**(4): 695-700.

Tavares, P., Santos, M. A., Lurz, R., Morelli, G., de Lencastre, H. and Trautner, T. A. (1992). Identification of a gene in Bacillus subtilis bacteriophage SPP1 determining the amount of packaged DNA. *J Mol Biol*, **225**(1): 81-92.

Teather, R. M. (1985). Application of gene manipulation to rumen microflora. *Can. J. Anim. Sci.*, **65**: 563-574.

Thomson, A. M. and Flint, H. J. (1989). Electroporation induced transformation of Bacteroides ruminicola and Bacteroides uniformis by plasmid DNA. *FEMS Microbiol Lett*, **61**(2): 101-104.

Thomson, A. M., Flint, H. J., Béchet, M., Martin, J. and Dubourguier, H.-C. (1992). A new Escherichia coli - Bacteroides shuttle vector, pRR1207, based on the Bacteroides ruminicola plasmid replicon pRR12. *Current Microbiology*, **24**: 49-54.

Thorpe, H. M., Wilson, S. E. and Smith, M. C. (2000). Control of directionality in the site-specific recombination system of the Streptomyces phage phiC31. *Mol Microbiol*, **38**(2): 232-41.

Thorson, M. L. (2003) In *Biology*Virginia Tech, Virginia.

Thyagarajan, B., Guimaraes, M. J., Groth, A. C. and Calos, M. P. (2000). Mammalian genomes contain active recombinase recognition sites. *Gene*, **244**(1-2): 47-54.

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. and Calos, M. P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol*, **21**(12): 3926-34.

Tobe, T., Hayashi, T., Han, C. G., Schoolnik, G. K., Ohtsubo, E. and Sasakawa, C. (1999). Complete DNA sequence and structural analysis of the enteropathogenic Escherichia coli adherence factor plasmid. *Infect Immun*, **67**(10): 5455-62.

Tojo, N., Sanmiya, K., Sugawara, H., Inouye, S. and Komano, T. (1996). Integration of bacteriophage Mx8 into the Myxococcus xanthus chromosome causes a structural alteration at the C-terminal region of the IntP protein. *J Bacteriol*, **178**(14): 4004-11.

Twort, F. W. (1915). An investigation on the nature of the ultra-microscopic viruses. *Lancet*, **2**: 1241-1243.

Unniraman, S., Prakash, R. and Nagaraja, V. (2002). Conserved economics of transcription termination in eubacteria. *Nucleic Acids Res*, **30**(3): 675-84.

Vander Byl, C. and Kropinski, A. M. (2000). Sequence of the genome of Salmonella bacteriophage P22. *J Bacteriol*, **182**(22): 6472-81.van de Guchte, M., Daly, C., Fitzgerald, G. F. and Arendt, E. K. (1994). Identification of int and attP on the genome of lactococcal bacteriophage Tuc2009 and their use for site-specific plasmid integration in the chromosome of Tuc2009-resistant Lactococcus lactis MG1363. *Appl Environ Microbiol*, **60**(7): 2324-9.

Van Mellaert, L., Mei, L., Lammertyn, E., Schacht, S. and Anne, J. (1998). Site-specific integration of bacteriophage VWB genome into Streptomyces venezuelae and construction of a VWB-based integrative vector. *Microbiology*, **144 (Pt 12)**: 3351-8.

Ventura, M., Canchaya, C., Pridmore, R. D. and Brussow, H. (2004). The prophages of Lactobacillus johnsonii NCC 533: comparative genomics and transcription analysis. *Virology*, **320**(2): 229-42.

Ventura, M., Foley, S., Bruttin, A., Chennoufi, S. C., Canchaya, C. and Brussow, H. (2002). Transcription mapping as a tool in phage genomics: the case of the temperate Streptococcus thermophilus phage Sfi21. *Virology*, **296**(1): 62-76.

Vepritskiy, A. A., Vitol, I. A. and Nierzwicki-Bauer, S. A. (2002). Novel group I intron in the tRNA(Leu)(UAA) gene of a gamma-proteobacterium isolated from a deep subsurface environment. *J Bacteriol*, **184**(5): 1481-7.

Vercoe, P. E. and White, B. A. (1997) In *Gastrointestinal Microbiology*, Vol. 2 (Eds, Mackie, R. I., White, B. A. and Isaacson, R.) Chapman and Hall, New York, pp. 321-370.

Vieira, J. and Messing, J. (1991). New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene*, **100**: 189-194.

Voet, D. and Voet, J. G. (1990) *Biochemistry,* John Wiley & Sons, Inc, USA.

Wagner, P. L. and Waldor, M. K. (2002). Bacteriophage control of bacterial virulence. *Infect Immun*, **70**(8): 3985-93.

Wallace, R. J. (1994). Ruminal microbiology, biotechnology, and ruminant nutrition: progress and problems. *J Anim Sci*, **72**(11): 2992-3003.

Wang, F. S., Whittam, T. S. and Selander, R. K. (1997). Evolutionary genetics of the isocitrate dehydrogenase gene (icd) in Escherichia coli and Salmonella enterica. *J Bacteriol*, **179**(21): 6551-9.

Wang, P. W., Chu, L. and Guttman, D. S. (2004). Complete sequence and evolutionary genomic analysis of the Pseudomonas aeruginosa transposable bacteriophage D3112. *J Bacteriol*, **186**(2): 400-10.

Wang, X., Wilkinson, B. J. and Jayaswal, R. K. (1991). Sequence analysis of a Staphylococcus aureus gene encoding a peptidoglycan hydrolase activity. *Gene*, **102**(1): 105-9.

Wang, Z., Xiong, G. and Lutz, F. (1995). Site-specific integration of the phage phi CTX genome into the Pseudomonas aeruginosa chromosome: characterization of the functional integrase gene located close to and upstream of attP. *Mol Gen Genet*, **246**(1): 72-9.

Ware, C. E., Bauchop, T., Hudman, J. F. and Gregg, K. (1992). Cryptic plasmid pBf1 from Butyrivibrio fibrisolven AR10: Its use as a replicon for recombinant plasmids. *Curr. Microbiol*, **24**: 193-197.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. (1987) *Molecular Biology of the Gene,* The Benjamin/Cummings Publishing Company, Inc., California.

Weinbauer, M. G. and Rassoulzadegan, F. (2004). Are viruses driving microbial diversification and diversity? *Environ Microbiol*, **6**(1): 1-11.

Weisberg, R. A. and Landy, A. (1983) *Lambda II,* Cold Spring Harbour Laboratory Press, NY.

Wells, J. E. and Russell, J. B. (1996). Why do many ruminal bacteria die and lyse so quickly? *J Dairy Sci.*, **79**(8): 1487-95.

White, B. A. (1998). Molecular biology of the rumen microbial ecosystem: prospects for genetic manipulation. *AgBiotech News and Information*, **10**(6): 171-178.

Whitehead, T. R. (1992). Genetic transformation of the ruminal bacteria *Butyrivibrio fibrisolvens* and *Streptococcus bovis* by electroporation. *Letters in Applied Microbiology*, **15**: 186-89.

Whitehead, T. R., Cotta, M. A. and Hespell, R. B. (1991). Introduction of the Bacteroides ruminicola xylanase gene into the Bacteroides thetaiotaomicron chromosome for production of xylanase activity. *Appl Environ Microbiol*, **57**(1): 277-82.

Whitehead, T. R. and Flint, H. J. (1995). Heterologous expression of an endoglucanase gene (endA) from the ruminal anaerobe Ruminococcus flavefaciens 17 in Streptococcus bovis and Streptococcus sanguis. *FEMS Microbiol Lett*, **126**(2): 165-9.

Whitford, M. F., McPherson, M. A., Forster, R. J. and Teather, R. M. (2001). Identification of bacteriocin-like inhibitors from rumen Streptococcus spp. and isolation and characterization of bovicin 255. *Appl Environ Microbiol*, **67**(2): 569-74.

Williams, B. J., Golomb, M., Phillips, T., Brownlee, J., Olson, M. V. and Smith, A. L. (2002). Bacteriophage HP2 of Haemophilus influenzae. *J Bacteriol*, **184**(24): 6893-905.

Williams, K. P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res*, **30**(4): 866-75.

Wishart, D. S., Stothard, P. and Van Domselaar, G. H. (2000). PepTool and GeneTool: platform-independent tools for biological sequence analysis. *Methods Mol Biol*, **132**: 93-113.

Wisniewski, J., Hrebenda, J. and Bielecki, J. (2000). Insertional knock-out of protein translocation systems common for Yersinia enterocolitica and Listeria monocytogenes. *Acta Microbiol Pol*, **49**(1): 31-42.

Wong, C. M., Klieve, A. V., Hamdorf, B. J., Schafer, D. J., Brau, L., Seet, S. G. M. and Gregg, K. (2003). Family of shuttle vectors for ruminal Bacteroides. *J Mol Microbiol Biotechnol*, **5**(2): 123-32.

Wong, C. M., Seet, S. G. M., Klieve, A. V., Hudman, J. F. and Blackall, L. L. (Unpublished). Characterisation of two *Bacteroides* species from the sheep rumen.

Wuenscher, M. D., Kohler, S., Goebel, W. and Chakraborty, T. (1991). Gene disruption by plasmid integration in Listeria monocytogenes: insertional inactivation of the listeriolysin determinant lisA. *Mol Gen Genet*, **228**(1-2): 177-82.

Wulff, D. and Rosenberg, M. (1983) In *Lambda II*(Eds, Hendrix, R. W., Roberts, J. W., Stahl, F. W. and Weisberg, R. A.) Cold Spring Harbor Laboratory, USA, pp. 53-74.

Wyckoff, H. A. and Whitehead, T. R. (1997). Improved electroporation protocol and vectors for *Streptococcus bovis*. *World Journal of Microbiology and Biotechnology*, **13**: 269-272.

Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. and Gordon, J. I. (2003a). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science*, **299**(5615): 2074-6.

Xu, J., Hendrix, R. W. and Duda, R. L. (2004). Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell*, **16**(1): 11-21.

Xu, J., Hu, J., Wang, J., Han, Y., Hu, Y., Wen, J., Li, Y., Ji, J., Ye, J., Zhang, Z., Wei, W., Li, S., Yu, J. and Yang, H. (2003b). Genome organization of the SARS-CoV. *Genomics Proteomics Bioinformatics*, **1**(3): 226-35.

Xu, M. Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A. and Shub, D. A. (1990). Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science*, **250**(4987): 1566-70.

Xue, G. P., Denman, S. E., Glassop, D., Johnson, J. S., Dierens, L. M., Gobius, K. S. and Aylward, J. H. (1995). Modification of a xylanase cDNA isolated from an anaerobic fungus Neocallimastix patriciarum for high-level expression in Escherichia coli. *J Biotechnol*, **38**(3): 269-77.

Xue, G. P., Johnson, J. S., Bransgrove, K. L., Gregg, K., Beard, C. E., Dalrymple, B. P., Gobius, K. S. and Aylward, J. H. (1997). Improvement of expression and secretion of a fungal xylanase in the rumen bacterium Butyrivibrio fibrisolvens OB156 by manipulation of promoter and signal sequences. *J Biotechnol*, **54**(2): 139-48.

Yagil, E., Dolev, S., Oberto, J., Kislev, N., Ramaiah, N. and Weisberg, R. A. (1989). Determinants of site-specific recombination in the lambdoid coliphage HK022. An evolutionary change in specificity. *J Mol Biol*, **207**(4): 695-717.

Yang, H. Y., Kim, Y. W. and Chang, H. I. (2002). Construction of an integration-proficient vector based on the site-specific recombination mechanism of enterococcal temperate phage phiFC1. *J Bacteriol*, **184**(7): 1859-64.

Yang, Q., Berton, N., Manning, M. C. and Catalano, C. E. (1999). Domain structure of gpNu1, a phage lambda DNA packaging protein. *Biochemistry*, **38**(43): 14238-47.

Ye, Z. H., Buranen, S. L. and Lee, C. Y. (1990). Sequence analysis and comparison of int and xis genes from staphylococcal bacteriophages L54a and phi 11. *J Bacteriol*, **172**(5): 2568-75.

Ye, Z. H. and Lee, C. Y. (1989). Nucleotide sequence and genetic characterization of staphylococcal bacteriophage L54a int and xis genes. *J Bacteriol*, **171**(8): 4146-53.

Yeo, C. C., Yiin, S., Tan, B. H. and Poh, C. L. (2001). Isolation and characterization of group II introns from Pseudomonas alcaligenes and Pseudomonas putida. *Plasmid*, **45**(3): 233-9.

Yin, S., Bushman, W. and Landy, A. (1985). Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. *Proc Natl Acad Sci U S A*, **82**(4): 1040-4.

Yu, A., Bertani, L. E. and Haggard-Ljungquist, E. (1989). Control of prophage integration and excision in bacteriophage P2: nucleotide sequences of the int gene and att sites. *Gene*, **80**(1): 1-11.

Zdobnov, E. M. and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, **17**(9): 847-8.

Zhang, J. and Madden, T. L. (1997). PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res*, **7**(6): 649-56.

Zhang, J. X. and Flint, H. J. (1992). A bifunctional xylanase encoded by the xynA gene of the rumen cellulolytic bacterium Ruminococcus flavefaciens 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. *Mol Microbiol*, **6**(8): 1013-23.

Zhang, N. and Brooker, J. D. (1993). Characterization, sequence, and replication of a small cryptic plasmid from Selenomonas ruminantium subspecies lactilytica. *Plasmid*, **29**(2): 125-34.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol*, **7**(1-2): 203-14.

Zhao, S. and Williams, K. P. (2002). Integrative genetic element that reverses the usual target gene orientation. *J Bacteriol*, **184**(3): 859-60.

Zhu, H., Paradis, F. W., Krell, P. J., Phillips, J. P. and Forsberg, C. W. (1994). Enzymatic specificities and modes of action of the two catalytic domains of the XynC xylanase from Fibrobacter succinogenes S85. *J Bacteriol*, **176**(13): 3885-94.

Zimmer, M., Sattelberger, E., Inman, R. B., Calendar, R. and Loessner, M. J. (2003). Genome and proteome of Listeria monocytogenes phage PSA: an unusual case for programmed + 1 translational frameshifting in structural protein synthesis. *Mol Microbiol*, **50**(1): 303-17.

Zimmerly, S., Hausner, G. and Wu, X. (2001). Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res*, **29**(5): 1238-50.

Appendix 1

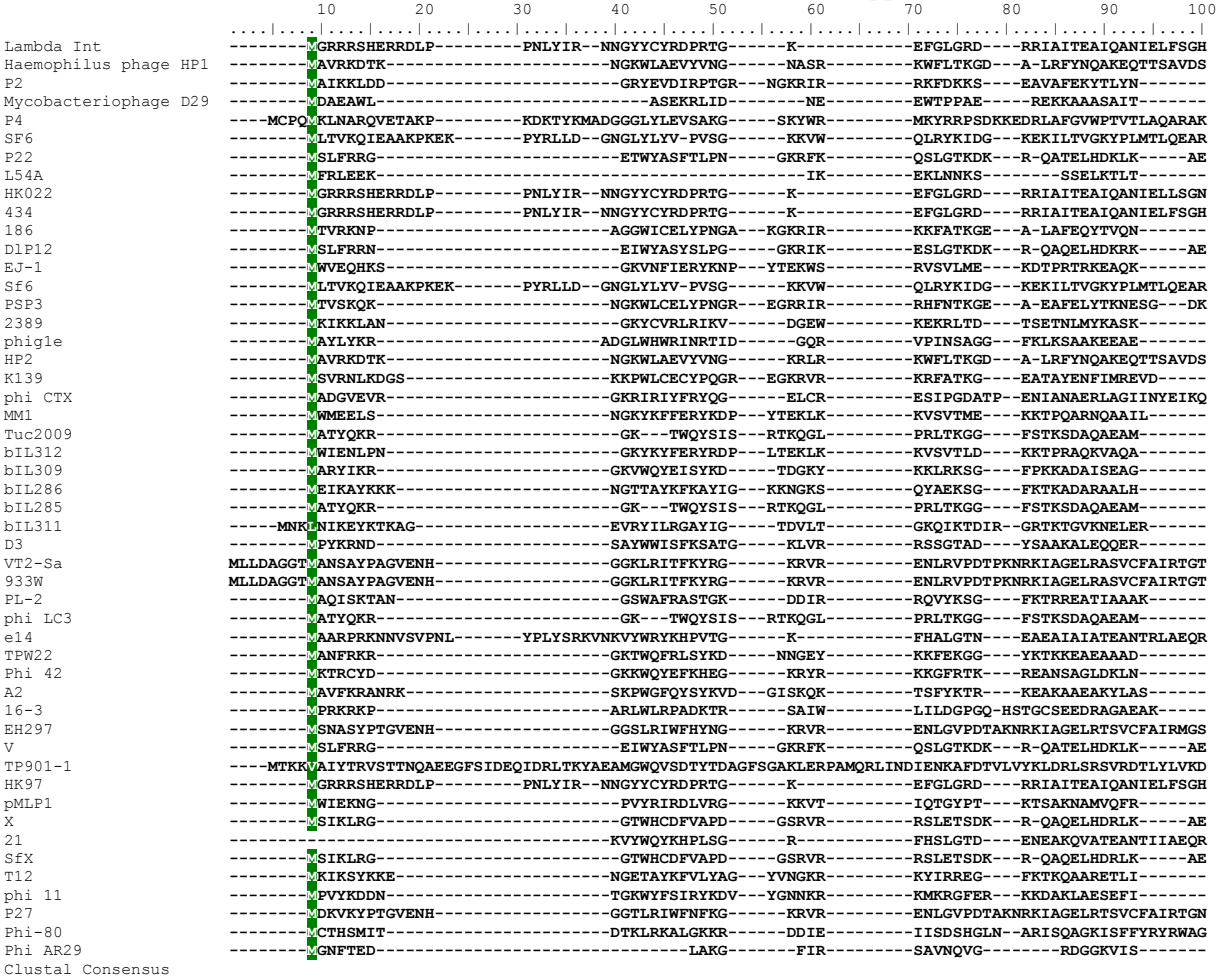


Figure 3.1A: The alignment of ϕAR29 Int protein against 50 known bacteriophage integrase.

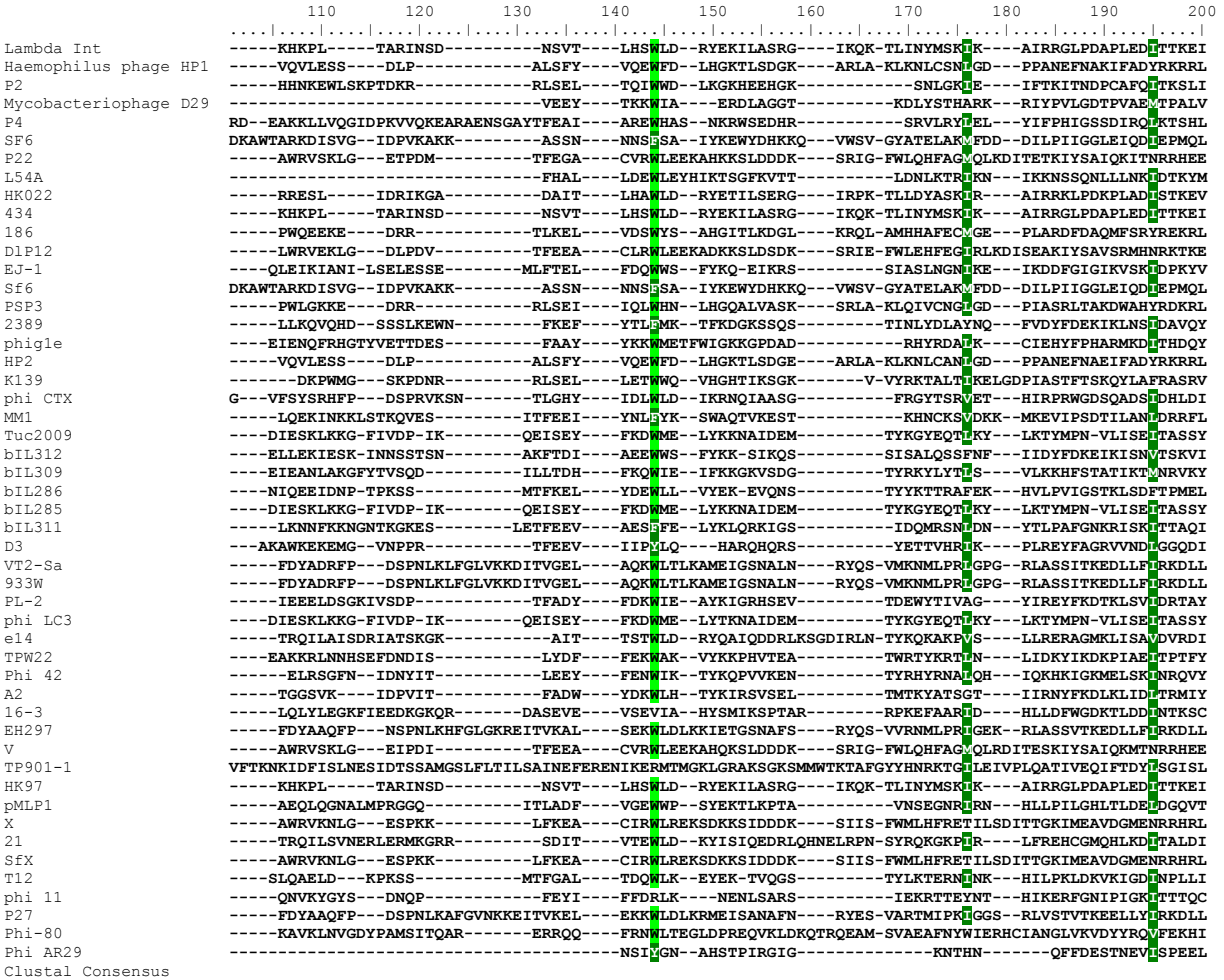


Figure 3.1A: continue

Figure 3.1A: continued

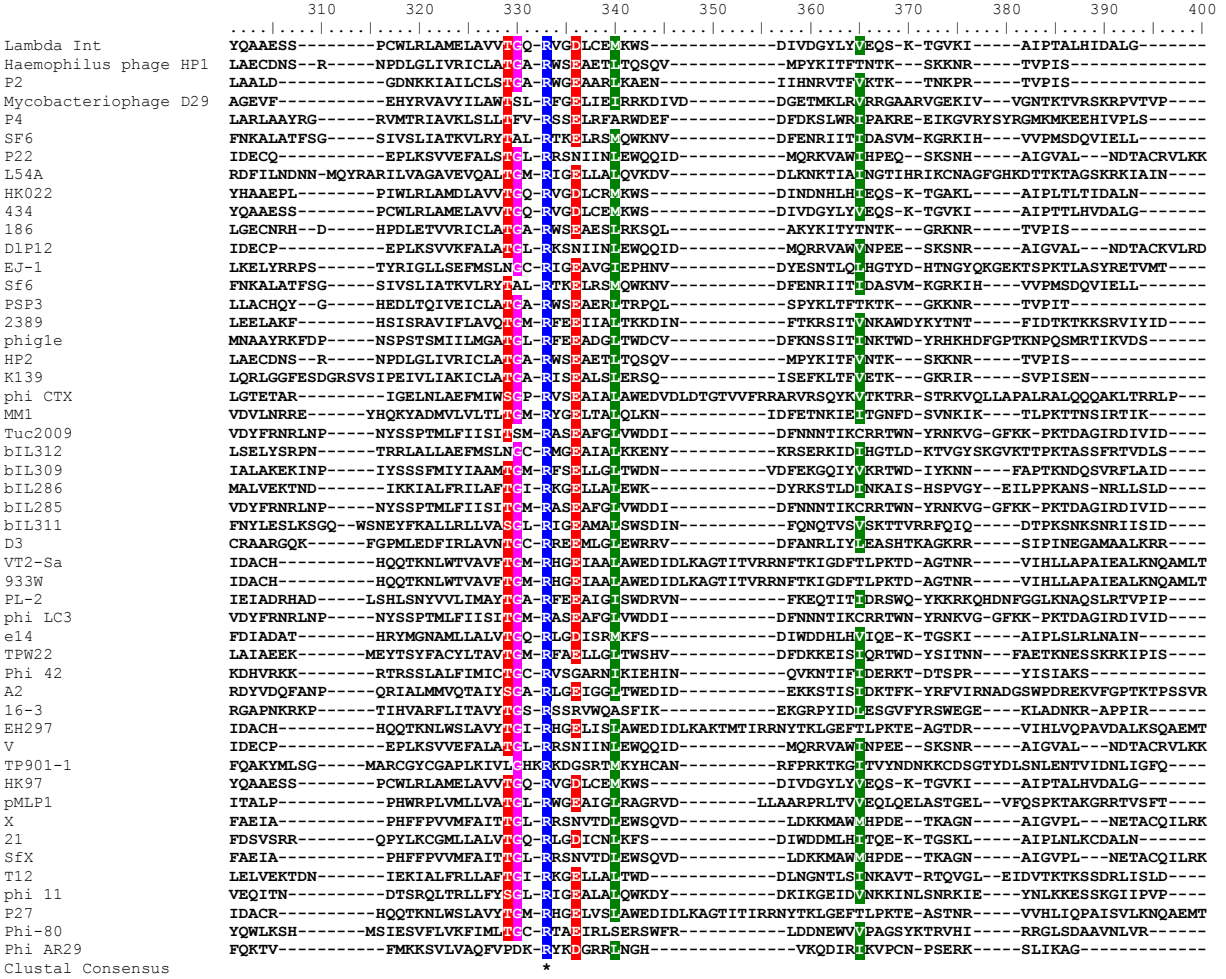


Figure 3.1A: continued

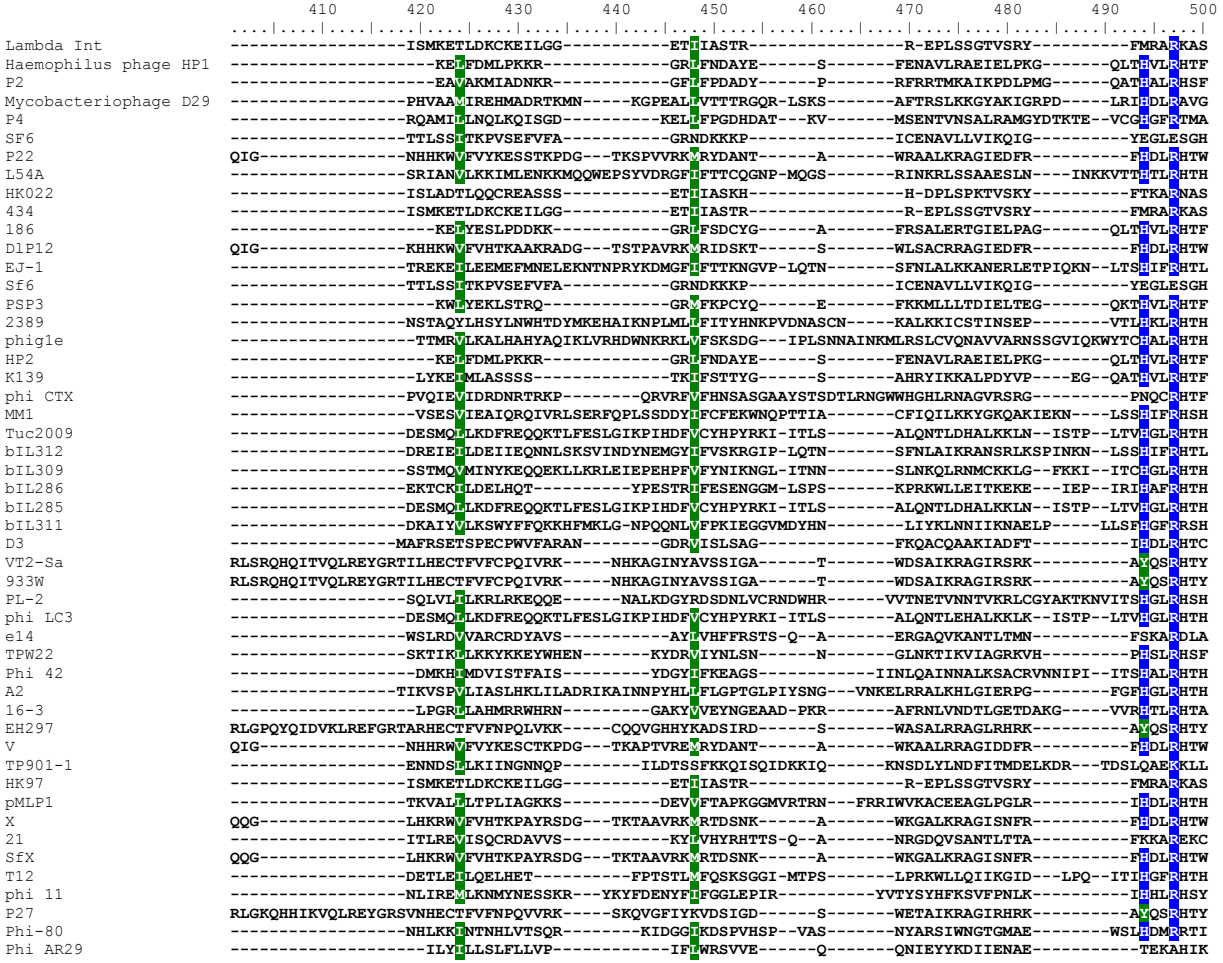


Figure 3.1A: continued

Figure 3.1A: continued

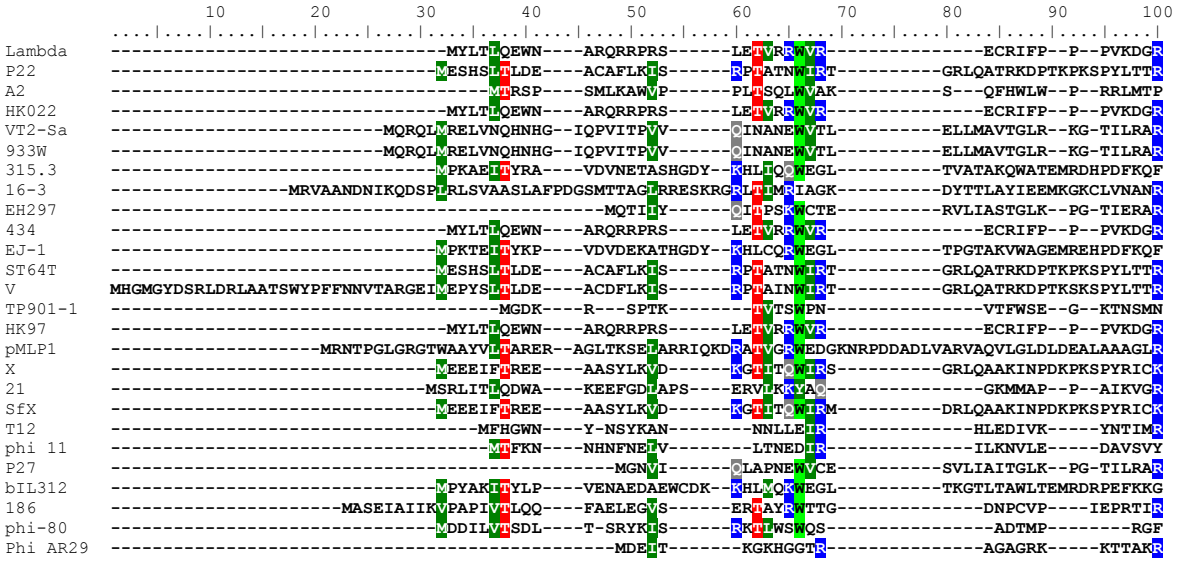


Figure 3.2A: The alignment of ϕAR29 Xis protein against 26 known bacteriophage excisionase.

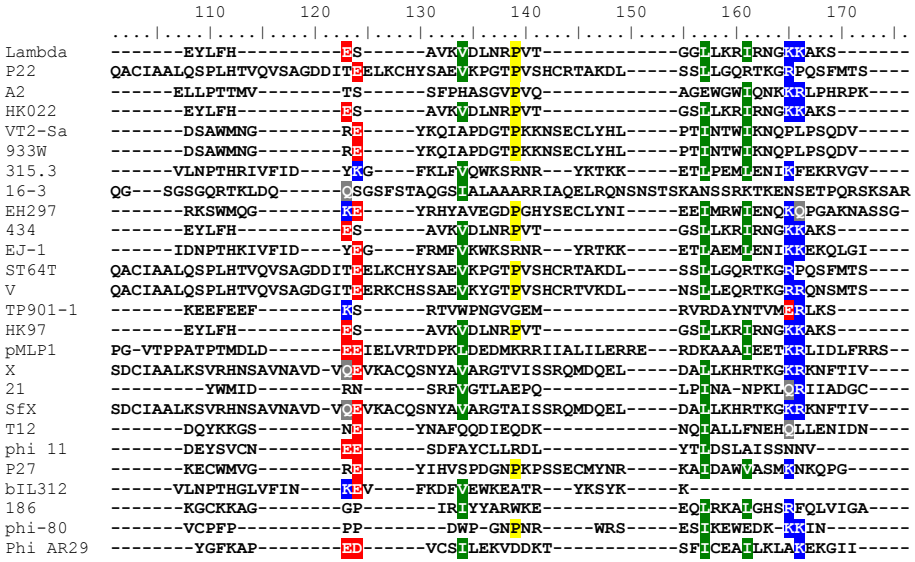


Figure 3.2A: continue

Appendix 2

Table 4.1A: Primers used for the sequencing of phage ϕA29 genome and their characteristics.

Primer prefix (Application)	Primer	Sequence 5'-3'	Tm (°C)
SSP(Primer-walking for ϕAR29 Genome)	SSP1	TGAAAGCCTCCATAACCGACTC	55
	SSP1 rev	TGCGTCTGTGTATCCATCTAAAT	52
	SSP2 for	ATTCCTTTCATTCATTACTTACCAC	51
	SSP2 rev	AAGGTGCGAAGTTTGAGTGAGAA	53
	SSP3 for	TCACCGAGAAAGCAAGAAGA	50
	SSP3 Rev	ACATTGATTAGTTCTTTGACATA	46
	SSP4 for	ATAACTCTAACTCCACCTCCTAATA	53
	SSP4 rev	CTGCCGTTTGGAATAAGGAGAGA	55
	SSP5 for	GAATCGGGAGCAGACAAGTA	52
	SSP5 rev	TCGTGATGGAAACTCTGACTC	52
	SSP5 II for	ACCGTGCTTCCTAATCTCCTTC	55
	SSP6 for	CGTGTTGCCCGAGCCAGTG	57
	SSP6 rev	GCTATTGCTGTCATTGCTCTTG	53
	SSP6B for	CGTGTTGCCCGAGCCAGT	55
	SSP7 rev	TCCTGAAAATGCGACTAAGCC	52
	SSP6C	CAGATGCCAACACTATGCTTATT	52
	SSP8 rev	TGGCTGCTGATGTGTATTGGGAT	55
	SSP8 for	ATGCATTTGTTTATCTCTGTGTCCT	53
	SSP8for rev	GAGGAACGTGGGAAAGCAGAT	54
	SSP9 for	AGCGTCCTGTTCCTCCTGTTC	56
	SSP10 for	CGGCACGGACATATTCGA	50
	SSP10 rev	ACCCACGTTACGACCAATCTC	54
	SSP10b rev	CGGGCTATTTCGGTTTACTGAT	53
	SSP11 for	CCGCCACATGATTGATAGATAGTC	56
	SSP11 rev	GCAGTAAAAGCAATCGCAGGTG	55
	SSP12 rev	GAGCAGGAGTTCAAAGAGAGTGTC	57
	SSP12 for	CGGCTTTCATTGTTACTACTGC	51
	SSP12 for/rev	GCAGTAGTAACAATGAAAGCCG	51
	SSP13 (rev)	TGAGGGAGGACGGGAACAG	55
	SSP13 for	GCACCCTTTCGGATAGACG	52
	SSP14 rev	GGATGAAAAAGAAAGAGCAGTAT	50
	SSP14 for	AGCCCAGCCAAACCGACT	52
	SSP15 rev	GACCTTGACAGATGCGATGA	52
	SSP15 for	TTCCGCTTCCGAGTTGACA	51
	SSP16 rev	TTGAACTCGGATAAAGAACATG	49
	SSP16 for	CCTTTACGAATGGCGAATCTA	50
	SSP17 rev	AAAGCTGTTGAGAATCCTACTAT	50
	SSP17 for	CACCGCTTTGAGAACCACATA	52
	SSP18 rev	AAGTACACGCAGTCGGATAGAT	53
	SSP18 for	AGTCAAGCCCTCCAAGAATGTA	53
	SSP19 rev	CTCCTGCCACCCGAAAGTAT	54
	SSP19 for	CCTGGACACTCGGCTAACAT	54
	SSP20 rev	GGGACGGCATCAAGGTTC	52
SSC(Sequencing for HindIII digestion phage ϕAR29 genomic library)	SSC43 for	CCTGTTCGGTAGCGGTAAGTC	56
	SSC43 rev	ACAGCGAGGGTTGGGAGGAC	58
	SSC46F for	GCATTCCGTTTACCACCATTAG	53
	SSC46F rev	TCTTTGGAGGAGTTTATCGCTGTAG	56

	SSC46R for	CGGCAGACAACATACTTCGCTC	57
	SSC46R rev	TACCGCTACCGAACAGGCACTAT	57
S2C(Sequencing for HindIII digestion phage ϕAR29 genomic library)	S2C1 For	TCGCTGTTTCTGACTTCTCCGT	55
	S2C1 Rev	ACGGAGAAGTCAGAAACAGCGA	55
	S2C2 For	CGCTTCGGAGTTCGTTTCTA	52
	S2C2 Rev	TAGAAACGAACTCCGAAGCG	52
	S2C3 For	GCCGCCCTCTACGCCGA	57
	S2C3 Rev	CGGCGTAGAGGGCGGC	56
	S2C4 For	GGGTCTTCACACGCCACA	52
	S2C4 Rev	TGTGGCGTGTGAAGACCC	52
	S2C6 For	TGTTCCTCCATTTCCCTACG	52
	S2C6 Rev	CGTAGGGAAATGGAGGAACA	52
	S2C8 For	TCTTTGGATTTCTTCGCTGACA	51
	S2C8 Rev	CTGCCGCCTTTCTAGTTTCCATA	55
	S2C10 For	TCCCGGCATTGTATCAGTTA	50
	S2C10 For2	CAGCCACCCATTTAGAGCACAT	56
	S2C10 For3	CGAATACACTACAAAGGCATG	50
	S2C10 For4	TTTGCTTTAATCTCCGAACTACAT	50
	S2C10 Rev	TGGAATGGCAAGGCAATGA	49
	S2C12 For	CCCAAAACTAACACCAGCGT	52
	S2C12 For2	CCTTGTGGTGGCTGGTAGTATT	52
	S2C12 FOR3	CGGAGACAGAAACCAAGACAC	54
	S2C12 Rev	GGCAAACTTCTTAACGCTGGTG	55
	S2C12 Rev2	GCGTCTTGGTCCTCTAGTGC	51
	S2C12 Rev3	TTCTACTTGTGCCATCCCTCTA	53
	S2C14 For	GATAGGGTTGTTAATTTCCGTCA	52
	S2C14 Rev	TAAGTGCTAAAATGCTGTAATCGT	50
	S2C15 For	GAGTTTGCCAGATTGTCAGC	52
	S2C15 Rev	GCTGACAATCTGGCAAACTC	49
	S2C16 For	GCAGTAGACGAAAATAATCACGCT	54
	S2C16 Rev	AGCGTGATTATTTTCGTCTACTGC	54
	S2C16 Rev\Rev2	GAGGCGAAGAGTATTGCTGAGA	55
	S2C16 For\Rev2	ATCGAATCCAATCACAGAAGCTA	52
	S2C16 For\Rev3	TCGTTTGTGGACCTGCTACTA	52
	S2c16 For\Rev4	GGAACGTAAACAGGGACCACTAT	55
2SSP(Confirm ϕAR29 sequence)	2SSP1 for	CCTAGATGTTGTGCGAATGT	48
	2SSP1 rev	CAAGTTTACGGAATGGTGATAC	53
	2SSP2 for	ACCAAGAGGGCAAAATACTGT	50
	2SSP2 rev	AATCAAAAGCAGCACACATAAGA	51
	2SSP3 for	CAGGAGAACGAGGATACAATGATA	52
	2SSP3 rev	CCAGTGTTTAGCCAGTTTAGAAGA	52
	2SSP4 for	CTTCGTTTATTTGTGAGGCTAT	48
	2SSP4 rev	CTTTCGTGGTGGGACTTGC	53
	2SSP5 for	AACACAAGCCAACTAACGATGAT	52
	2SSP5 rev	ACGGTCGGTCATTATCCATTG	54
	2SSP6 for	GCGTAGTGCTGTCCTTTGAG	50
	2SSP6 rev	TCTGCTGTGGGAATGCTCT	50
	2SSP7 for	ACAGCAGACCAAATCATAATA	44
	2SSP7 rev	TCCCTATTCCTTTGGTTGACTC	52
	2SSP8 for	TCACGGAACGGCGATAAACT	55

	2SSP8 rev	TTGAGTGTGGAACGGCTGAG	54
	2SSP9 for	CGCACCCTCTGAAACGA	50
	2SSP9 rev	GCGAAGAAATGGTATGAGGAT	50
	2SSP10 for	TCCGAGCGAAAATCACTAATA	50
	2SSP10 rev	CACAACATTCGCACAACATCTA	51
	2SSP11 for	TCGTGAAATCGCTCCATAC	48
	2SSP11 rev	AGGGACCAGTAAAAACAATGAG	50
	2SSP12 for	AACGCTTTCAACACATTAC	42
	2SSP12 rev	AATGAATGAATAGTAATAGTGTGAT	44
	2SSP13 for	GTCTGGCTTACAATCTTACTCTACTA	48
	2SSP13 rev	AGGTAAGACAATGGAGGAAATAGAGA	50
	2SSP14 for	TCAATACCGTGCTTCCTAATCTC	53
	2SSP14 rev	CTGCCGTTTGGAATAAGGAGAG	54
	2SSP15 for	GGGTTTGTTTTTGCGACTTC	50
	2SSP15 rev	AGCGTGGCGGTTTGTCATA	51
	2SSP16 for	ACCTGTCTTAGCCCATCTCCAC	57
	2SSP16 rev	GATGGAAACTCTGACTCAATGTG	53
	2SSP17 for	AAAACGTTGCTTATTCTTCTTG	47
	2SSP17 rev	TCGGTCAACTTTGTGGAGATG	52
	2SSP18 for	CGTAGATGGTCGTTTCCTTTC	52
	2SSP18 rev	GATGGCTGCTGATGTGTATTG	52
	2SSP19 for	CGGACAAATGGCGAAATA	46
	2SSP19 rev	AAAAGTAGAATGTAAGGATGGTAGTA	52
	2SSP20 for	GCGATAAACTCCTCCAAAGA	50
	2SSP20 rev	GCTCGGTAACGGGCATA	49
	2SSP21 for	GCCTTCCAGTCTTTCATCTTAG	53
	2SSP21 rev	TAGGAGCGAAGTATGTTGTCTG	53
	2SSP22 for	GGTAATAAACTCCCTTCTGACTTC	54
	2SSP22 rev	GATGGTCGGGCGGTCTA	52
	2SSP23 for	AACCCATCAATAAGAATCCCTC	51
	2SSP23 rev	ACCTACAGATTCAAAGCATTACG	52
	2SSP24 for	GGCTTTCTTTATGTTCCCTCTC	53
	2SSP24 rev	AAGAGCAGGAGTTCAAAGAGAGT	53
	2SSP25 for	CGCCCTTTGATTATTGTTATTAC	50
	2SSP25 rev	ACGGGAACAGCAGCGAAT	50
	2SSP26 for	GCATGTTCTTTATCCGAGTTCA	51
	2SSP26 rev	CGACAATAGCAAAAGTAATAACAATA	50
	2SSP27 for	CTGTTCTGTCTCTATCTCCTTA	51
	2SSP27 rev	TAACAAGGCAACTAAAAGAAATAT	47
	2SSP28 for	CATTTTGAACACACCCCATAAG	51
	2SSP28 rev	GCTTTTGTTGCGGTGGA	47
	2SSP29 for	CTCCATCCATTCATTAGGGTCT	53
	2SSP29 rev	TTCGGAGATTAAAGCAAAGTGT	49
	2SSP30 for	CCCGGCATTGTATCAGTTA	49
	2SSP30 rev	TGGAAGAAGAAAGCACTGC	49
	2SSP31 for	GCATCCCAAACGAGAAATC	49
	2SSP31 rev	CCTCGCGTGGTGGTACTC	55
	2SSP32 for	GCCTGCCCAATCTGCTC	52

ϕAR29 Genomic Sequence

```
   1 ttggctttat gtatcaatgt cagttctact tggtagctct gaactcgatc actatatacg
  61 ggaagaagcg ggaagaaata agaagctgcg tctatctcca ttccctcaac tggttctaac
 121 gttgtgtaat aaaagctgtc tagggtagaa cgacactgta gttccatccg cacaataatg
 181 tccgtaggta agtccgttat aattagcttt aacggcacgt tgtcccatat ggtagcacaa
 241 ctggggtagg tatatacttt gctgtccggt atatctatgt tcactccaaa tcttggtatt
 301 cttaatttca ttgcgttaat attgttagaa ttttcgcctt tatgatcttg tttatatcca
 361 gcgtcagccg tttcacccgt tctggattaa tgatgtcgga gactaccccg ccaccgttat
 421 acttgttagg aaccttgatg ccgtcccgct tcatcacata agcgatggcg aaagctgctt
 481 cctccgggat gtttgcacca accgtccggt tcttgtcttg aatccacttt ttaatggcgg
 541 agacaggcgg aaaactcccc gctttcctac cctgctccat ctgtatgaca taatgcggcg
 601 cagtaatagt taccctatca ccgagatcgt ttacttttag atcacgccca aattcaccag
 661 acgctaccaa acctttcgaa acataagatt cgaagatttc ctttttatc tgttctacaa
 721 cctgcaatat ctccttatcc atagttcaat aaatcatctg ttatagaaaa tgttacactc
 781 caaccggact tcattgagtc atagatattc tgtaccttcc tgaaactcaa tccatccaca
 841 tcgaagtgac acacaaaagc ggacattagt ttgtttaaag ccaaatcggt acgcattaat
 901 gtatcaagtt cggcagcgtt atccgtaaga tagtacgatt tatccaagca ctgtaatact
 961 acgttatact ttcgggtggc aggaggcaac ttagacatac caccgtccgg gacatcaaac
1021 gttaagaaca tacctgagat gtcattcact agctcgttaa tagtagacgt atctccgaaa
1081 tagataggca agccgagttt tacggcttcc ccatccataa agtttaatat gtcactgaat
1141 atcatggtaa cttgattatt gcgtcatcgt gaccgttact acacaggcat gcgcccgaat
1201 cgaaatcata ccccatcacc gcatcacctt gtagggttat agaaccttct tggttaaagt
1261 tccgagatag cacggcattg tcgtacatcc ggattttaac tccgtttatg ttaagcgtat
1321 tagaaccata gggaacaatg attctcggct gcggaataca agtaatcacc aaccttgcat
1381 ctttagcatc cgatggatta accgttacgg gtggagcttt ccgaagcgtc atggatacca
1441 tagaatatcc ttctgaggaa gattctaata atcttgtccc tgcatcgaat tttgccattg
1501 catattccca atagttagct acggttaatg ttgcattccc tgcgggtaac ggtttggata
1561 accttaatct atccgactgc gtgtacttaa tgtcgttata aggtcttcca acacgcccag
1621 tatcggtata gaaagcaccc tgttcccaat agtaatcagc taataagata tcctccgggc
1681 ttttctggtc tagctcgtta gtgatatacc tgcgcttctt gaacgctcgg agatacttaa
1741 cacccgctag tggaacgtcg gagaaatcaa taaactttag gtccggcttc gtagacaatg
1801 ctattcccac gtagtgtaca ttagcaggaa cggttagtgt tccgtcattc catccagtat
1861 cggacattag ttttctatcc tcattaaacc aacataaaag gattctatat gctccggata
1921 ctgatatagt atcaccggga tttaccgggt atattgactt cgtagtcata cgtacatctg
1981 agttccttcc gtctttcata acgtcccaag ttagccccgg tgcggtattt aagccgcctc
```

2041 tgtcccattc gttaatatca aacacacggt taacatatgg atcaccgtgt agctgaacat

2101 cacctcttac tattacacct gcacctaagt aggcggaacc tgtgatctcc ggaactttcg

2161 tgtaattgtt tacgtttaaa tatgcgtttt tagaaacgaa ctccgaagcg ggtaaattag

2221 cgttcggatc ggttttctga aaactcataa cgtaaaacgg gaattcactt tccataatag

2281 taggattctc aacagctttt cgaagtataa acgattcgtc aaggtaatgt accgctcgta

2341 cagcataccc gccgggaatg tttggaaata ccgaaccact agtagctacg ggtctaatca

2401 ttcttacacg gttcgtagct ggaatcttat tagcttcgta cgcctttccg atatcgcctg

2461 cataggtccc cgattctatc atgtccccgt caagcataat tgtttcgtga ttctccgtta

2521 agcccgtgga atttttggcg ttatccggat aactaacacg tccgagaaga acgtccgaat

2581 tcgcccacat actataaagt ccttgcttat ggctagatac taaggttctg ttaggaactt

2641 ctgtgttccc taaaaactcc gtagcgtaag aaaaatctat atcctcgtat gtgttcccgt

2701 ctatcacgtc atttttgaat attcccttag gaacaacaca tctacgtaac gttaatcctc

2761 cggaattaat taagctaccc gaaacaaaac tatctccgat acggaccaaa ggagaatccg

2821 aagccataat atagtggtcc ctgtttacgt tccctctatg aatgtcagcg aaatactgat

2881 tgtattcaag acgttctacg ttacggaatg taccataagg tttatagtcc tgtctatcag

2941 ttttcatgta catgttggta ttgtacagac cgcccgctaa gaatcgtaac tgcaccgttc

3001 cgagtccgaa gttgtcggta ttcatgatga aactagaatt atccatcacg taatcacccg

3061 atacggattc cggacgaagg tccattaaca aggaaccgct aacatgacct agaattcttg

3121 cacctgctgc ctctagttct gcgggtgtca atgtaggatt cttagcaaaa gcaatcatgc

3181 aaatattata caccgggtga cttaacctgt agatattcgt attgcatact tctgtttccc

3241 cagcatatgc tatcgctcct tttgcaaaat cataatacgc ccaatatatt cttgcgttaa

3301 atccctgtgg cgtgtacaca taggtatctt tgcctattct aagcgtagcg gtatttctca

3361 caaaattctc tgcgtctagc ttcatcgtag aactcgtgaa taaggttcct ttcggcgcat

3421 cctgattata tcgtccttgc tccataggga agttcttcgt attagtagca ggtccacaaa

3481 cgacatccat agttattccg ataaacgaat ccttgatgaa cgtattggct tttgcgtctg

3541 ctgtgaagca tcggaaactc acttttcccg gaatcttagt atttccacta aattccgtta

3601 ttatatctgt atgaacggga ttctcacccg gaaagaattc tcctatatag gcgttatcct

3661 tgattacgat atccggacgg ttcgatatat cacctgcgtt caaccagcat tcgccctcct

3721 gcgataactg tgtttcgtct tttaccgaac cacctacgga ccctttttact ataagtccgc

3781 cgagagagta gatatctctc tcggctacga ttcttcctgc tgaattaata ctgtatttct

3841 ttcccatgat tattttgttt tagatttttg tttctcaact tcatcatgtc tcttgctgat

3901 cgctagaata gcatctgagt agtttatctt cttcgcatct tcgaatgagc agttaaacag

3961 ttcggatgtg atctgtacca ttcctagcac acttttagct tctgtgattg gattcgattc

4021 accggaacca cctgcacctg gtaacagtac acgttccagt tcgtcagcct gcccaatctg

4081 ctccacaatg tacttcgtca gtttcaccat atcagcgaca gtttctggta catatccctc

4141 agtccatccg ctaatacgtt ctaatactgt ttcagctctg cgcgcttcga tcatttgcca

4201 gagagttaca tcttcaactg atggacacgt ataaatgatc tgtccattac gtgttaccca

4261 gcgagaaggt attaagtatt ctgacatata atgcagaagt gcggcttcat cttgcgaaag

4321 actctccaca gcatccggct ttagattcgc tatacgccgt aattgcatca agcgttgata

4381 tctacttgtc agacgaacaa acgtcttcca taaccaacga aatggggtaa ataggcgata

4441 tatctgatat cgcacatatc cccgaaaggt tttaatttcg cttttcttca tcttttcttt

4501 gtttcttgcc gggacgatac ttggcgatca gaaactcagt cgcgtatcgt atagcgtcca

4561 tagcgtgatt attttcgtct actgcctcgt tcgtatcata gagtccggtc atcttgtcaa

4621 atacataaga gtaattatct gcttcgtcct gtatacctcg actaccttgt acaatgtgca

4681 ttttaaactg cttcacttgt gaaatacctg ccatgattga tccttttccc tttatacatg

4741 gaaatatgcg gcatcccaaa cgagaaatct cagcaatact cttcgcctct tgattatccg

4801 caatagtggt aactttatgt agcccattct tgcgcaagac attcgcaata tcccaattca

4861 acagccctgt agagtatgcg atctcttgaa cataaagatc atctttatcg aagccaactt

4921 ttacaatcgc cgtaggatca ccagagaaac cgaagtcaag accaagacac catttacaat

4981 tcgcaggaaa ctccggcaca atatcatatt caggatatac cagtccttcg gttcctcctg

5041 tttcaccaag tccgaagatt cgccaccagt tttcatcagc cttatttctc tcaatctctt

5101 cgatctgctc cggtgtcaaa tatggattat ccttgtaggt gctgacgatt tccaccatgc

5161 cgggaccctt gaaatagtcg tgtgcccaaa acttcttaac cggattaaag tctacataca

5221 gcatcaaacg agtacgaacc gccatttgcc ggaatacttc tttcgggacc ctctgtgctt

5281 cgtttacaaa caggatatca cgtgcaggac caaatacttt cgcagcattc tcacaaccga

5341 aaaactctat ctgtgatcca ttaggaaaag tgtaggtcat ttcggttaaa ttcattgcct

5401 tgccattcca aagaccttca tcttgcaaca tacgtttgaa atcgcgaaac ataccacgtt

5461 tcactcccgg cattgtatca gttacacacg atatgagtaa aggagcttca gactgctcag

5521 caataaggta gagtaactgt aacatactcc atgttttcga agatcgagta ccaccacgcg

5581 aggatactcc acgaatcact ggatttaccg tggcttctaa tagcctgtca aaaacatagg

5641 tcgttttcat tgtccagagc cttctgtttc cccctctcct ttcatcttcc gtttctgcga

5701 tagtgtggag agcttcttaa tattactaac cgattcttct ttcaatactt cgaccttcaa

5761 tgttactcct tccgtcttca catcggttcc agtttgttta ttacgccaac gatcaggaga

5821 tatatttgta agcaagaata tagcagcacc aacattcgcc tctacattct tcactgttac

5881 tattttcttc tttacttttct tcccttcata ttcggtttta gtttcctcaa attcatatcc

5941 gcaagcagct tttgacagag attcaactaa cctctgctct agcttctcct taaactcatt

6001 tttccctttt tttatagcat ccgcaaaatc cgcattttca agccaacgat agtatgtttg

6061 agagtctatg ccgaaatgag cgcaaaagtc tttcaacctc gcaccgccat gctccattag

6121 cccattttca gccacccatt tagagcacat ttcagtcatt tcctttaaat tgtacgccat

6181 gctataatca gttttatatc aatagcaaat ttacccgata tcactcttac agccatgtca

6241 aattgtgtca agtacagatt tacggcgttc ttcatcgtat atcgtattac acagcttgta

6301 ttttaacgaa atgttctcca tccattcatt agggtcttgt gcagtgcttt cttcttccac

6361 gtacttccaa actctcgtgg cacgctccca ccggaacaaa gccttcttta aattctggta

6421 gtgattaatc ctgtagacct gatttgtata tgaggcataa aaatccttct tcaccttgtc

6481 caagccatcc gctttactct cccaatcatt atgactccct atcacaggat caataaaagc

6541 cagctctttg agggggggctt gctcttttct tccattcctc aatcctttca cgaatttctt

6601 taaaccttca aaatgctttg taatatctac tccattgaca tacccgttcc catatctatc

6661 tttaacacac cttaccccaa tctgtaataa atacgaatac actacaaagg catggacatt

6721 aagaattatc cctatttcac gtacatttac ccacgtcttt tcattaaaat ttctatccat

6781 aacacgatta tttttaaaat aaatagtata tttgctatac aatcgtgaat gattggggag

6841 aacaatgctt ttacaccttg ctagttctcc ctatttttt tgatctcttc ttatttctaa

6901 gattttcccg gtcaatcttt ctgccccaca tcatcgagtt atacagggaa acagcatata

6961 aaaaaagttc ctcactactt gcaaggaact ctacttttat agctacttta tgctgaaaag

7021 tttgaatcaa cggaagactt tataaatcta taataaaaga atagaggcta aataaaaagg

7081 aaagagtccg gtgcaatacc gaactctttc gcaaatttgt taataagaca ctgaccaact

7141 tttagagtaa aattcatttt gaacacaccc cataagttat tgatcatcta tttttaattg

7201 ctgattcaat ctactaacac ttaacaaagc gtctatagat tgatttagag tagtacactt

7261 tgctttaatc tccgaactac atctcgtaat ttcattttta acagctttcg cttttctttt

7321 aatttttatc aaaaaagcga tgaagtttgt aaatggaatt aatgcactgg aaaccaaaat

7381 gtatgtccag catggtgata ccgatacttc taaagaacaa catgagagta taatcgataa

7441 tatgaaactt ataatcacag aaaatacaaa gtacaaaatg ctatgtctca ataaagaaaa

7501 atccaaataa ctttgttgct tttcattttc acctaacacc cacccaacaa ctagaaaaag

7561 tgttgtcaag aatgttaata cactccaaaa attaggaatg aactgtgggg ctaaagatgt

7621 gataccaacc aaaaagaggg cagtagtgct atataaaaaa aggaaaagac ttaaagaaga

7681 caaacttttt gattcacaca ttaattttac attttcttct aatcttttct gttctgtctc

7741 tatctcctta actaacagct ctctatctct ttttgctttt tctattttat tattagtact

7801 ttttccttca attactattg gttctaaatg atctaatgta tttttgtcaa ttaataattc

7861 catgcaatct tgaaaagaac taataataaa aagttgaaac ttaaatactt tctctgataa

7921 agtgtgtgta taagctttag catattccac cgcaacaaaa gcaatattta aagtcgcccc

7981 taattgtata agagttgaaa aatcatttag catcatcttc ctctatatta aatagttctg

8041 ttaaattgat tgttacggat tctgtcaaat taaataactc atcaaagaat tctgcagtat

8101 ttcttgctac aactggattg ttaaaattac attcagctgt acgctcttct gtattagtct

8161 caattctata cgaactttca tctcctatag taaaatgaat tggctttttt tccgtatcat

8221 ttgctaaata gggttttact gttgttgttt ttacgtttat aggctttcct tcagatacat

8281 aataggccaa tcttttaaaa agattagatc ttttagctcc atcaatatca aaattattca

8341 acagaatatc aacggaccct ccacgctcaa taaagtcact caatgcttcc acatattcag

8401 atgtattacc gacatgatga cataaagagc ctgcaaaaat acgcaatttt gattttgcat

8461 ttttaaatag ttcaataagt actattaatg catgttcttt atccgagttc aagaaaactc

8521 tattagaatg tgcagttgca agagttgaaa caaattttct ataagtatcc atatttcttt

8581 tagttgcctt gttagggctt aaaaacgttt gcaaatatat aattattaat tgtacaaata

8641 ctaatttaag actgtttttt ttcaaattac attgatatgc tcggcaaatg aaagtatctg

8701 tatctaaaaa ttaaacctat aacattaaga ttatgttgat gacattatga atgattaatt

8761 taatttattg ctttattagt ttcataggca atttccttgc aatacaaata taaccaattg

8821 ttcgtaaaat acaaagtttt gacttttacc tttttttctt tttcatttcc tcaatccaat

8881 tgagatacca ttcacgagct ttttctttag cttcgtcaat gtgaatatca tactgttcag

8941 cactatcatc ctccattaga tcatcacaga tataatctnc tttattaaat ggatcgcgct

9001 ctataaactc ctcagattgg cagaacggac aaggtacatc atccattggc tcatacagat

9061 tgccgttttc atcgcatctg tcaaggtcat atagtttccc gtccacgcaa cacgcatcgg

9121 gatattttgc accccaataa ggaaactggg ggcatggttt cttatttttg ctcatttta

9181 ttcagatttg aacaaaggac gcatccgaat aaagataaag tgtcgaattt taaaattatt

9241 gttaagatgg atgcgccctt tgattattgt tattactttt gctattgtcg aattttaaaa

9301 tttatttaat atggaaatta gtgaaaaaat ggctcaagaa gctttagcga aagttaacgc

9361 taaatgcgct ccattcaagt gtcctatgtg caaacaaaaa ggaaattttc tttttgcgaa

9421 agctgaattt caacatgttg gattcagcag aaaagaaaat cagattaccg ctgaaggcaa

9481 tattgatttc attcctgttg tagcaggcag ctgtccacat tgcggctttg ttgcaacatt

9541 caatttgcag gttttggggt tagtgtaaaa ttacctttta acataactct ttcactaaag

9601 acacccaatt gcttgcattc tccagctaat tgattcttta caagtcgatg ctggtgaatt

9661 ttcaatgttg ttgtttttct tgttttcatt tttatgtatg ttattagtta tactgctctt

9721 tctttttcat cccaaggatt ccaattttcc tctatgggga tttcctcttc atgaaaatgc

9781 ttcatatttt attgttgaat taacttttcg gatatatgta aaggatacca gacttttgct

9841 ttagtatccg tgtcaatacc atgaaaccag accttcccgt tgtctacagt ctcaaaggaa

9901 gttatttcgg ctttctttat gttccctctc gtattgcgat atggaacaat gtctcctact

9961 ttaaatttac tcatttctgt attgttttga aggttacaga ttttgaaatt cggttttcag

10021 tttttccaga tttcctttta tagaatcctt aattatatta atgaggaaat cgcttaatac

10081 cattggaatt cgctgctgtt cccgtcctcc ctcaaatgga tcatcatatt ctataattag

10141 agtaggatta gtgctttcca acacaatagg taaatttggg tattcatctc ttgcgtagta

10201 gttagtatca aactcaaagc aattcaatgc ctgctcatgt tcttctataa gtttgttaag

10261 ccgattggct tcatttaatt tttctttatt catatctgta ttgttttacg ctaattgatt

10321 taacctttttg taaaaagtaa cgaagtcacc aatactcaaa aatacatttg gtaaatcttt

10381 gccaaaatct tcttttgcat ctttcatagc atagctggta aaggtactta ctactatgta

10441 ggttctattt gggaacatat cagcatcttt aactatctct ctgcataaag attccatcaa

10501 agagtttcca atactcatac ctccataatc aaaaaataga taatcgtatt tttcacctct

10561 tgggtacatg tctgtcaaag tcagtcttaa cccatcaata agaatccctc tatcttctaa

10621 aggagattcg acactctctt tgaactcctg ctcttgtgtc tctacaagat aagaaagcat

10681 ggggtctgcg aaatgaataa ttttgccagg taagtaagaa aatgattctt taggcttgtt

10741 tagttcaact aaatatatcc gttctccttc ttctccttcc cattcggaat actgcgttat

10801 cttataaaat ctatcatcta aatcaccatc taaattttta ggcaaagaca cataaatagc

10861 tttatcaaac ttgtttctgt tttttaagat tgtattatgc acaaaatttg aatggcatgg

10921 tcttatgaaa atcggaattg aatttctcgg aaaacagtag tgttcagcat tcttttcgat

10981 tatgaaagac aaaggctcat ctggttggtg aatgtctatt gataagactt taccatcata

11041 catcgaagca aaaagccctc taccagcacc acaatctact actattctat cacctgcgat

11101 tgcttttact gcatcgcaca tttctttgtt tactttgcaa aatatttgat tactcataat

11161 ttgattcctt tctattttgg tattgatcgt cttcccgata tcaggaaaac gttttggtta

11221 ttaaataaaa aaaatagcga tctgatagac cactatgcga gtatcatcgc cgggacgatc

11281 cttttagttc tatgacgtta aacatttctt ttacacgatc agcgatataa tcaccgtatt

11341 tgtttccaaa ctctgtattt ggatcgagat tggtcgtaac gtgggtaata aactcccttc

11401 tgacttcata tcgaagttgt aaaatggttt gtatgacatt tatcccggtc ccgtaatgct

11461 ttgaatctgt aggttcacgt cccaattcat caatagccag attacacata tattcacgat

11521 cagtaaaccg aaatagcccg ttcatacctt tctctgcata catgagagat atttctactg

11581 cactagtgag tctaaagcct atgtgatcgt tattacatcc gtatcgtaaa cggttgatct

11641 tgcctagata acgctgtagc ccttttatca agacagactt gccaactcca ataggtcccc

11701 ataaaagcaa gccttttgag gaatcaagca tcttacttct gcctaataca taatcataca

11761 attcggataa taggaccttg ttgcgttcgt caatgataaa tcccggttct acttccttca

11821 tggagttaat aaactctttt ttccagaagt gttctactcg atcctcattc catgttatct

11881 cctttccttt gatgtgaaat ttaaccgaag gagattgatt tgattccggc tgacttggtt

11941 tcattagagg aatcagctcc ccgactgttc ttattgcttc cattcttttt ttctatttga

12001 aaatcattct tttcccatgt ccttactgct gccttccagt ctttcatctt agaccgcccg

12061 accatccaac cgttagaagt ataatgatcc aaccacttct gtggatcaac attgtttttt

12121 ctttctatgc agtaggcaga aacttcttca atagatggag gaataaactt tctagttttt

12181 acggtttccc ctatattatc tttttgttta gtttctattt tagttttata tatatagtct

12241 ggcgcattgg ttggctgatt ggttcccata ttggttggca gattggctgg ctcatctact

12301 gtcttttggg ctggcttatc taccggaaat tttacggtag ttgaattcac aatcgaattc

12361 tcaaaagctt tttcaaaaga gtacaatcct accacccttt tacttttacc ggatttataa

12421 taaactagtc ctgcattaat caaagaaagc ctggcacgga cgagagtttt ctcatcaata

12481 ttaagagcac aacagagttc aatattcgag caactgaaaa cgtcctccca accctcgctg

12541 ttacaaacgg caactaattc gtagaatagt gcctgttcgg tagcggtaag tcgattacgt

12601 cttcgtgctt ttctcatttt ttctgttaat gtatatccgt ccatagttta atacgcatga

12661 atacagattc ttttactatc agcgacaaaa cgccgtttaa gcttataaca gtaggcaaca

12721 cgtggattcc ctttagctgt gggaacaata gttccattat tgaattttgc gcaagtatct

12781 gggcggataa cttgcttgtc tgatttcttt ttcatatcta ttttccttct aaataatctg

12841 ttactacagc gataaactcc tccaaagacc ggcagacaac atacttcgct cctatactat

12901 caaattcctt ttgataggct ttttggtgat cgctttgtct tcctgtctta gtctttaatt

12961 caattcccat aaaaggataa tacttgttag ggatcaacag aagtaagtca gggaaaccgg

13021 cacgtactcc catctgttta aactttgcag cttcgatagc attccgttta ccaccattag

13081 gagaatgatg caaccttagc ctatattgag gatattgcaa atcgaaccag caaacacaag

13141 ctttttgcaa atcatcttct tcatgttttg gcttcttgcg gatattctta ccgcagtact

13201 gggctttcat ttcttcgaat gtcatatcaa cctttctcct tactcctttg gagtttcttt

13261 ctagttttac gaatcatatc ttcatctctc aaattatatc cccgaatgag gatttctgac

13321 gttttcaagc accggactat cgtctggtat tcttgcttgg tgattgttat tttcatgtgg

13381 gcagtttagg agtagaacct aaataattgc atctgcaata cataaagcac ttcgtacgct

13441 ttctttatgc tctctttacc atttagaata cctccccatg tttgcccgcc aatcttcata

13501 gacaggcagg ctggggtaaa aaggttaaca aagctatctc aatagctcac tcttgcggat

13561 tatagcccta ccagtgacga tagtgctttc catattatga gataatgtac tttgcttaat

13621 tcctatctga tcctcggaca aatggcgaaa tatgcccgtt accgagctga agtaatagtt

13681 ccgcttttcg aagatcaggt agacatggat tactttagtt tttcgcattg tttttattaa

13741 gaacttccaa ataacagcta tttggaatta aatgtaattc ttgtttttt gaatttctat

13801 ctccattaac tgcaaaagtc gatcctcgtc cgaactgggg agatatatcc cagcctcccc

13861 agaactccag tttctaaaac gagtaattgc attgctcatt tcttccgtat ccaattcggc

13921 agaactacga agaatttaa tattacccaa atacttatca gaaacttcct ttatgaatat

13981 gtctttatta cataatatct tgaaatattt ctgctttaca tattctaaag tatttcctgt

14041 ctcagaagca aagtacccta aaattaaatg aaggtatcta ttttgtctgt cagtacgtat

14101 tggtttcttt tctgttaatt ctactatttt cccatttta acaagcaaag cggaacgaga

14161 tttaaactgt tccgcttgca atggattaga aagatcatat aacatttta gaatggaaga

14221 tcatccgctg gtgaaacgct gggagccgaa tcaatttgtt cagcagttgg aacattaggt

14281 tgtggaatat attctttaaa atcaccacaa aagtaattta ccccgtcttt tcgttcttcc

14341 tgtttagggg aacaactcat gacatgagta tgaccatacg tagatggtcg tttcctttct

14401 aatatagcta cattgagata tattttagta ctaccatcct tacattctac ttttttgaat

14461 aattctttcg ggatatccga aagacaaata ctacctgtta gaatcatttc ttttatttt

14521 taaaggttat actatatgat gttgtacttt gtttagcagg aggaaataat tgaaatatat

14581 ctcctgtatc ctcatcaatt tctgtcttag attttgagag tgtcttcaaa tacttttccc

14641 tttccttaca ttgttcatct ataatctttc tttcctccac aagacgttca taaaccggat

14701 cgttacaaac agaaaaatca tatttaacac cagtttcctt tatttggatt actgcaccta

14761 aatagccggg agactcacct ttcccatatt tctcgcattc ttgtattact gcatctttta

14821 tgctttcatc ctttagaaac gtatttattg tttcagaaat acttttcatc tgaaccactg

14881 catcaatagg atttatatta ccatctataa cctttagtat aaaatcatta gcaatctttt

14941 gttgctctgt tttggaagat ggtattctat taaaaaaaat ctctttactc attgcaaatt

15001 cttatttact ttatattgat aatagttctc agatattata ttaatatcct cattggtaca

15061 tttatagttt ttctctacaa gattcataat agaaaaacgt tgcttattct tcttggcgaa

15121 aacttcactt ttataaatcc attccattaa agcttcactt cctaaattag aagcattcaa

15181 actcttgcga ttatcatcta ttggagatga tttggtatat ttcgtagaat cattatccca

15241 atacacatca gcagccatcc caagagcctt acaagaaacc gatatagcat ctgttaaagc

15301 cattttatag cattcatctg aagtgtaagc accattcttt tcattagata caaaagaagc

15361 ccctccaata ccttgtatcc cttcactcca ttcgccttta tatttcacaa aaagattaat

15421 atgcacaaaa cttgcaattt caccatttgc cccttgctca ttccacatct taataatttc

15481 gtaacgccac ccaaatccac aaggtccaaa ttgttctgtc agagttttaa ttctccacat

15541 gggattgata tccgtttttc cttttagacg acctgcagca atcggcttag tcgcattttc

15601 aggaacatct ctaatcctat cataaagttc taaattataa tcgtccatat tcgttataat

15661 taaagtagtt aataaaatag ttcccggata ccgaaccaac ggacaccggg ataattcaaa

15721 acttaaatag cggactggat accgcacgga gtcccttact ccatgattag aattaaacaa

15781 taaattattt gcgtaatata atatctattt cagattgttt atatagtctt tttccaccga

15841 cttctgcggg aattaaataa cctgtcttag cccatctcca caaagttgac cgatcaacct

15901 gtaactgttt acaagcatct ttgggcttta cgaaagtctc ttttttctta gccaaaatag

15961 attcttcaac ttcttctttt gtttttttgaa tgagataatc tgcgaatgct tttaaatctt

16021 caaaattcac attagccgat atcttactcc ccccaagact catgatctct tgaatactca

16081 tatctatccc cctattcttt ttgatgcacc tcttttgaac ttctctctaa aagcatgaac

16141 acagttaaca atagcattat aatacatgat attgtttcgt ttctagtcat ttcgatttgc

16201 aatgctaggt gggtcaccat agcaagagca atgacagcaa tagcattttg aattttatga

16261 atagttttca tagaacatta tttttaaatt aatactaagc ggcaagcatt aaatcaccgt

16321 ccattttagt tctcattata cgtagagctg tttttgcagc aggacgaata ttacaacgtc

16381 tcatgtccat ttgataattc ggcgtaatag caagtataaa gaaccaaaca gagaagaata

16441 actcaatacc gtgcttccta atctccttca aatcaaagtt tcttttagtc ctatcacata

16501 gcaggaataa agtaagctct acgttattgt taatgcctaa cttcttatga atatccctaa

16561 tctgtgcctt tatggtccaa actgactttt tgagcaaatc ggcaatttca tctggggttt

16621 gtccttttgc gacttcatta gctacttgat actcacattg agtcagagtt tccatcacga

16681 aatacgttta gctctaaaaa ctccctttttt atagtccaac tctccttctc tcttgattat

16741 aattccaaat ctgcgtctaa cacgatatcg aattgtactc attattccat cataagcaga

16801 tatcggaaat tctactactt catttagctt catttcactg attgattttg tccaatcacc

16861 agttatttttt ttcacttctt ttgccataag attaattatt tgattattat tagtggataa

16921 gcccggattc gaaccgggaa tgtaaattca agagcctcac tgaatggtta cagaatgtct

16981 ggcttacaat cttactctac taagcgttac caattccgcc acttatccaa ttaaaaaggt

17041 gcgctattct cacgaacggc acaccctaca acacaaacac aaaataaaac acgacaaaac

17101 aactcaatac atctcgacag atgtattagt attgtttagt aactacaaca acttcttcag

17161 tcttttaatt gcctctcctc tctccttatt ccaaacggca gggttagtat ccccatcctt

17221 atttattact tcagctgcca aataccaata atcagcttcg ttatatccat tctcctttga

17281 gaatttacga gcttccttca tcgttgcaca ttcctcaata atattatggg tccccctatg

17341 acaaaccgcc acgctaacaa taatctcact attcttataa tgcatgattt acctccttat

17401 tatactttta aagttagtta gtgcccgcga taccttttac ggattctccc acgtatcgag

17461 acgtgacggg ctgtatgttg aatcacttag atagcgttat agctcgccta acctgctata

17521 tgcttactga taaagacttt tcggacttcc aagtgatata tgtaactaat tcgaaccttc

17581 aaccgatcac ggcattcctg ctacggttga atttcttttc gtatgatcca atatgtcaaa

17641 gaactaatca atgtaccctg aaagcgtttt gctcgcttct ttcgtaggtt ctaacctaac

17701 agagcctcgt aatcttttat tattcgaaga aggttactga taatttcttc tttcgtttct

17761 ttgcttccag ccagcatctg aactgtatat tcatctcgtt ctttcagatc gtccgtgtat

17821 ttccgaagga aaactaagtt ttcgtttatt ttttcttcac tcattatctc caagaactat

17881 cgtagttagc atacttatcg gcgaaaaacg ctttcaacac attaccctgt ttagactcaa

17941 tggctttcgg cttcaatgat tctacatatt catccatctt taagcgagcg tccacccaag

18001 aaatacgcaa ggcagattta agagaataac catactggcg tacatatacc caagctctct

18061 gcatgatggc tttcatatta tatttgccgt cttttactag ttcataatct ctatttctca

18121 ttgtcttacc tatttttagt tatgtaaaat atttggtttt ctcactcaaa cttcgcacct

18181 ttgcagtgtt ggatgttgtt tgatgttgca aagatacata atattttatg tagaacaaga

18241 aatctacata ataatttatg tatttaactt ttattttcaa ttaaacgccg gataaatcac

18301 ataataattt atgtatatat aatgagtgta ataaaagaga ggcttacaga atttctaacc

18361 tataaaggta tagggcaaat aaaatttgcc gaagcagctg gcttgtcaag aggcttcgtg

18421 aatgtattag gtgatggcat tagttcaaaa tcacttaata aaatcagcca agcttttcct

18481 gagttgaata ccttatggtt aaccacagga gaaggcgaaa tgctcaaaac taccaataat

18541 acaccccaat ataatgaagc tacgcccatc caacaagacg tggtttatat cccgttagtt

18601 aatcaattcg cttatgcggg ttatttagat ggatacacag acgcatctta catggagcaa

18661 ttacctaaaa taccatttat agtagataaa gaaggacatg gaaattatat agcctttgag

18721 gtcaaaggag atagtatgaa taatggaacc gaagaaagct atctagaagg cgatagactt

18781 tactgtcgtg aaatcgctcc atacctttgg gcaacttcca aattacatct tcgcaaatgg

18841 gatttcgtta tagtgcacac cgatggaatc atagttaagc gcattataga tcatgatgtg

18901 gaaaatcaca ctattactat tcattcatta aatgatatgt accctgatcg agttattgat

18961 ttgtgcgatg taaaacagat tttcaatgtt atagaatcag ttagacctag aagaagataa

19021 aaataagaat aacaatcgaa tattaaatta attaaaacac aagattatga aaaaggcact

19081 gctattaatt tcaatctttt tattaccaac atttttgcaa gcatgtagtg atgaagacga

19141 caaccaaaga tgtcaggcaa taactaaaga tggagaccaa tgtaaacgta atgcagaaaa

19201 aggaagcatc tactgttggc aacataaaaa atagcaataa atttaatggg aaattttact

19261 gaagatttag caaaagggtt tatacggtct gctgtgaatc aagtgggacg agatggaggg

19321 aaagtgataa gtaactctat ttatgggaat gcacatagta ccccaataag aggtatcggt

19381 aaaaatacac ataaccaatt tttcgatgaa tcaaccaatg aggttatctc ccctgaagaa

19441 ttaagattaa gagcagaagc agaagggttt aaagtatctt tatttagata taacgctggt

19501 attaaaatag cactctatat tgtttcttta ttttttgcta ttttagtagt accttctatt

19561 attatattaa tatttggtat catgaagttt tttcaaaaaa cagtattcat gaagaaatct

19621 gttttagttg cacaatttgt accagataaa agatataaag atggtcgcag gctgaacgga

19681 catgtaaaac aggatataag aataaaagta ccttgtaatc cttccgagcg aaaatcacta

19741 ataaaagcag gcatattata cattttactc tcattgtttt tactggtccc tatattctta

19801 tggcgctctg tcgttgaaca acagaacata gagtattata aagatattat agaaaatgca

19861 gaaacagaga aagcacatat taaagaagac tttgaattat ttaaagacac agtgagatat

19921 aataaaaaga tgaatgaatt taatgaaaag taccaaaagg cagtagagta tttaaattca

19981 cacaatcaaa caaaatcgga taattaaaaa actagcttat gaaaaagatc attttattag

20041 tatgtgcaat cactgcactt tgttcatgtg ggggatcagg taatcaaaac gagaaaaaag

20101 taagagaagt agtagaagcc aaactgaaaa cagaaatgaa tgattggtct agctatgagt

20161 ttgtttctgc ggaagccatt gatactataa agtatattga taatatcaac tatcgaaaag

20221 aatacttcca aaaaagcatt gaaaacaata aaggggcatc caattatgga ttagattatt

20281 cttcttcaat aactaaagat agtataatac ttattggaat agattctatt caaaatgcga

20341 tgggcgataa agtcaatgaa gatgtagcct atttatacaa gtataaattt agaggtaaga

20401 ataaattagg agctgtaatc ttggacgagt accttatata tatttcgcca aactgggaaa

20461 taattcagat gacgaatgat ccaaagaaac tttataataa tcccggagac ttccccggat

20521 atgttgatct tgttaaaaag aacatgtaat accaataata gcccgtctaa aaaacgggct

20581 tttatttttat taataaatct ctccacatac ctagatgttg tgcgaatgtt gtgcaacggc

20641 ataaaacaga aaatcgcaac tatctaataa taagacaatt acgattttgc aatgtgaccc

20701 cggtgcggtc taatatttaa aaatgcacac tatacaatct catacaaata cattaattat

20761 cagatatata caaaacagca agttgtatta tgttgcattt tagttgtgca attaagcatt

20821 ttcgttgtgc aaaaacagta tatttgcaca accgtaaaat gatagctata tggcaacagt

20881 aaaagcattt attcgatcaa gcaaaaagga taactttgtt aatatacgat ttcgtttatc

20941 cgatggtagg aatatacaat tatttcacac ttccgagatt ttagtacaac cttcaatttg

21001 ggatgagaaa aaggaacaat acaaatccaa agtacttata ccaatccatt gcaaaactag

21061 agaagagttt tataaagaca taacggaacg caaaaaccta attctacgtt tgtacaccga

21121 tcacaagata gaaaccagtg aacaactaaa tcaatatata gatgaatata ttaatccgga

21181 gaagtatatt aacaaggaat caagtgaaag cctccataac cgactcaact tatacataga

21241 gcaatgttat aaggatggaa tatttggcga aggcagaaag aaacattatg atgtattatt

21301 acgagaatta aaccgttttc tcattatcaa caacttaagt gatatcactc caaccgattt

21361 caataatgag aaacttatac tatttcgtga ttttctgttt aatgaatata tattggtaga

21421 taaatatagg gggctttatg ttgacatgaa taatcgcaat attccaacct caccaagagg

21481 gcaaaatact gttgcaacaa aattaaaaaa gctacaggca tttttcaacg aactggaaag

21541 taacgatgaa ataccagtat caccattccg taaacttgga aaacagcgta aaacagtaat

21601 gatgaaggaa caatacgatg aacctatatt tttaactaaa gccgagttca ttaaactaca

21661 aaatacagat gtccccagct ccctacaaga gactaaagac gctttcttac ttcaatgtag

21721 tttaggttgt cgaattggag acttccaaac gctatccttt gataatatct caatagaaga

21781 agatattcct ttcattcatt acttaccaca caaaactcta aaagaaaatg atacaagaat

21841 tgaaatcaaa actcctttaa tgcgttttgc tttaaaaatt ataaagaaat acaatttcaa

21901 tttccctata ttacgctatg tatcaggaga acgaggatac aatgataaga taaaagtttt

21961 gttagaatac tgcaaaatag atagacttgt tgctgtattt gatgaagcga aaagcaaaaa

22021 tgaatataag cccctatatg aagtagcaag tagcaaacta gctcgtaaaa cgcatgtaga

22081 cttaatgaac aaagtacaag ttgataagta tgcagccgga ttacattcaa agaatagtga

22141 cgcagtaaac cgatatacta aactaggagt aaaagatcgc tttattctta tgtgtgctgc

22201 ttttgattgt aatgaatata aagtagataa taagttacat attatctaaa tgttaaaaat

22261 gatattttct taaacttttt cttgatttcg tttgtacgtt atcaaaataa tgcttatctt

22321 tgtaatgtaa tcaaagagat gattacagcg gttaatcttc ttcaccgaga aagcaagaag

22381 agaaacctaa tcggtattca ttatgagact cgcaatttca tttagaatct ggaaagtaag

22441 agtaaacttc gaaattctac tctaaaatcc aaggggaggg aaacctcccc aaccgattag

22501 gattttgcaa atataacaat tatggatgaa ataacaaagg gaaaacatgg gggcacaaga

22561 gctggagcag gacgaaagaa aacaactgcg aaacgctacg gctttaaggc tcctgaggat

22621 gtctgttcaa ttcttgaaaa ggtagacgat aagacttcgt ttatttgtga ggctatatta

22681 aaactagcca aagagaaagg aataatctaa ttattaacaa aagaggtagc ttattcagct

22741 acctcttact ttcttctaaa ctggctaaac actggaaaga aagtaaactt cctagttttt

22801 cttataagac agtttatcta tattatgaag taacatattc caatcagttt caaaccaatt

22861 agtattcaca ttttctctat ctatgataac tctaactcca cctcctaata taagatatat

22921 ttcactaccg ataagagcac ttagattagt acctattctc cttctattaa agaacatact

22981 tactttattt attttatcaa tattttcatt agttactcta actctacaat aactaagttg

23041 actatcatca ataaaacgat tacggcattg agttatcatt acattgtctt tcatatcttc

23101 cattacttat aaatttaata ttatgagtat tcaatatgaa cactcttcaa caaatattaa

23161 aaaagaaatc ggataatcca aaagtgggaa aggatatttt tgattttaag tatcccctag

23221 ccttttcaat ctaaagtgtt ttccttctta ttcgaaacac aagccaacta acgatgatta

23281 aagctataat acttatatag attttatctt tatgcaagtc ccaccacgaa agctctacaa

23341 cagtttcttt ctgattcaac atagcatcga ctttataact caatgaatcc aatctattag

23401 agaattgctg caaagtaata gataatgttt cataaacttc agtcctttcc tgatcctgct

23461 tggatgcagt agtagtactt tctttgaccg gatattgctt tcctgacgaa tcgggagcag

23521 acaagtaaac tgttttattc tcaatcttca gatcactcaa cttgtcagta gtaactttcg

23581 tttgcttatt cacatccagc cgtagcgatt caattaagtt tctcaaatac aaaaaatccc

23641 ctgaatagtc tacctgcttt tgagtttcca tattccggga agtcttgcag gaagcaaacc

23701 atattcccga cattaggaat atggttatat agattagcgc tttcatggtc ggatcactgt

23761 attacgaaga aaattagaaa attcactcct gacatcgaag caggggcacg ccttaatata

23821 ttctgccggc tctatctctc cgctgccgtc caggtccggc gaagtatcac ggtgtccgag

23881 tacttcaatt ataggatatt ccttacagag cttcgcgatc aattcgcgta gtgctgtcct

23941 ttgagctgga gtacgtgtat ctgcaggttt tccagatacg tctagacctc cgatataaca

24001 gatgccaaca ctatgcttat tatacgaaga ctctgaaaat cctttggtat tacaatgcgc

24061 tccgtcaatg gataatgacc gaccgttttc tacggtacca tctaaatcaa ttacaaagtt

24121 atagccaatt tgattaaaac cacgtgcccg atgcatccgg tcaatatcct tagctcgcaa

24181 gtcttgtccg gcacgtgttg ccgagcagtg aatgatgatt gagtctatat cttctctctt

24241 catattcttt cctcctatat aattaaaatc aatattaaaa cttgaattag ttgcccgata

24301 gctccaccga ttaatgttgc cacaatatca agccaatccc attttccacc ccaatccctg

24361 tccttaaact ccatccctgc cgccaatcct gccacaaata aaatagtaaa tagtgctcct

24421 gctggaatag catacagcaa atgtttaggg cgattacttt ccttgatcca gttcattatt

24481 tactctcctt ttcgcttccg ttttttcttt ttcgggtata ttaccaatca tgctactaat

24541 tttagttctg acatagacag ggatgccgaa aacagctccc gaccatatca agcattgcgc

24601 aaagaaccat aatacagtat catgaatgat acctataggc tcaactaaaa aaccagcaac

24661 tgacaccccc actcctgcga agagcattcc cacagcagac caaatcataa tatcgtctct

24721 cgtttccctt ttcatttcct cactctttta acttataaaa catactcgat aaggcttaac

24781 agagtcacaa caatgtttat cctttggatt atctgattcc ttaaataact cgtatacaat

24841 atggaaatcc tctataaaag catctgcttt cccacgttcc tcttcccatc gcttatttt

24901 gttatagtct ggcaatatca acgacccgtt gtatacctgc gttttcagtc ctgtagatgt

24961 gctttttgg tccgctatct tcatatatcg cacaaacgca taataacata ggatcgtata

25021 aagaggaact atattgtatt tttccccggc tattacgaca tctaacgaat aattagaatc

25081 agaatcggtt ccgacaggga tatcactctc tccttttcca cctcctagct cactcgacac

25141 cgaaaagaat gtatcaccgc aaagagctac ctttatatcg agcttatctg cctcttgaat

25201 gcatttgttt atctctgtgt ccttaacatc tgctgcgata tcaaagattt cacgaaactt

25261 cttgattact tcggaaaaac tattcatttg ggtctgtttt aacgtttact gtcgagttag

25321 attcttctaa ttcgttataa atttcacata cttctgttgg aaggtctaaa gcacgcgcaa

25381 tttcccgact caacttacta cgaagttttg tcacggaacg gcgataaact ttctgcatct

25441 ccttcacgac ctcgccggac gcattggaga atgagatcaa agacgagtca accaaaggaa

25501 tagggatgtt ataggcttgt gaagcgatat ccttctttag cggttcattg taagccttat

25561 acagattagc gtcaatcggt acgcccaact ggtctacctt aataaatggc ttatcagtca

25621 gagcattttc atcacggacc aataccgctg atcctgcccc ttgcgctccc attatatctt

25681 ttatcccctt aacaaaagcg tcctgttcct cctgttcggt aaattctccg tgagaaatga

25741 tgctgcacat gtggaaacca cgtgtaagcg tccgttctac ataagtagag ttcatcgctt

25801 ccgcctgcat ctcagactgg actgcatgaa acggagaaag cggatagggc ttcgtagtga

25861 agaagtttat gtacaaaagc tgtccggggt gattctcgat tccgccgaaa aactctactt

25921 catccgcaaa gttatctgga ttgaacgccg ggtaggtaac tgccgtttta tccaattggg

25981 tagacttgat attttgacga tcccaattat taaacactac gtacttatga ataaccggat

26041 tcgttaagta gtctttattc agcccggcac ggacatattc gaaaggaacg ggataaatca

26101 ttttagggcg ataatcgcct ccatactgga caattagagc gcaccctctg aaacgagcta

26161 cgtcatatgc cagcatattc agtatctcgt ccatgttatc tccgtgggcg ttcgtcattt

26221 cgccaaaatc acggttttta aagccttcac attctatcgc ttcgctcagc cgttccacac

26281 tcaaagaggc ggttttgcta gcatatataa gctccgataa aatttggggg tataggtttc

26341 cgtccccata ccccacaatc ttttcggaaa ccttagcgtt aactttgagc gctctatcta

26401 ctattacgtt tactttcttg tgagctatca tattaacgtt tccttttagt ttattccagt

26461 tccttcataa tctcgtctac tgaaacctcc ggcgttgttt ctactgaagc ttccggttca

26521 ggatcagcgg gggtctccgg ttctacgggc ttctcttctt ctggagcttc tgtctccggt

26581 gtttgttccg gttcaggatc agcgggcaaa acggtcggta cgtcctcctg tggctctatt

26641 ggacctaagt cctcaaagta ggatttataa accggatttt ctttcatgat tcgttcggcg

26701 atggcgtccg tacagttaaa tgcacggtaa acaactccgt ccgccacatg attgatagat

26761 agtcccggtt tcattacgta gcgaacgtgt atgcccgtca agtagtgatc ctcataccat

26821 ttcttcgcgt acgcacgatc catatggcac atagggtcca gttttagatg cgtaatactt

26881 ttacagagat tcaaaatctc gaactcgtcc gttaagcgaa ttaattcgcg cacgggcttg

26941 atatcttttg ttacagtttt ctttgctctt gccatgatta tacagttttt aaagagttat

27001 actgtgccgc cgtaatacta taatggaaat ctccgcaaga cccgtccggc gtttttaaag

27061 tagcggttga aactccgtct gttgaactat cagttgataa atctgaaacc tctaacggtg

27121 agttacatcc caaaatgaaa tattggttgt tttttgttct aatagcaacc aaaaacgacc

27181 cagaaaccaa agcgataatg tagctaacta caggtaacga tgaaagaagt ttcatgacta

27241 ctgatatttc caacatagta ggagcattgt cgtttgcccg cgatgcctcc gtcacttgaa

27301 tagagttttt tacagattga acggtatatc cccttgttcc ggctttcatt gttactactg

27361 cctgtcctgt cgtagaagat accgagatac tagaaacatc ttcgtagttc aatattacgg

27421 cttcttctac tccggcaatt ccagagatta aaccgggatt agcgcaatca aacgctaaat

27481 cttgtgctat cttctttaaa caagccataa ttatgatgat ttagacacga gagtattcca

27541 ggtagcctca gtgattgacg cacgggcttc tcctagaaca ttctcaggtg tggttaacgt

27601 aatggcggta tagccgccgt tatcatttgc tgattcttcc agtccggata cctccaaacc

27661 gtaattacag ccatagatac ggtaaacacc tgtttcaacc atttttgcaa ccgccacaag

27721 gcgagaattg agaatggtat tgataaacac gttttccgca ctcgttttct tatacacggt

27781 aaaattaacc gattgttcca gcgcattcgg ggcgttttcg ttaattcttt gggcttctgt

27841 agcatttgca ccctttcgga tagacgctac ccggataacc ttgcccgaag tagtcaacgt

27901 aatagtagcg ataccgttgg cgatggagat agattggata tctgagtagt taataagcaa

27961 caaatcagct attccaactg ctccacctaa acaatcgtag gttattgcac ccgtgatatt

28021 actaatacat cccatgttgt taagttaatt tgttagcttc taaatatgtc caaacagcag

28081 ataaagcaac tatattgcga tcacccctag agttatccgg tgttttcagc gttacggtat

28141 caaaaccgcc agccgccgaa gtatctccgt ccatgcttgc cacttccagc ccggtattaa

28201 gtccggcaac tttaatgtta ccgccgtctt tgagttttgc gaaggctacg taattcccgg

28261 atagcagtga ttccttaatc gctgcaccgt cagaagtttt atcgtagacc gtgatagtaa

28321 ccgtctgttc cattccggct gcaccgtcca gtgtacgcaa agcgtccact acttttgcgc

28381 cgttcttgta acagtcaacc ggaattgcct ttgcaccgga tacaaggacg atagagttta

28441 gtgtcacgcc atcgccaccc attacaaaag aagccaattc cgatttgtta acaaggtaca

28501 gcccagccaa accgactgag ccgcccgcac acccaaaaac gatagcctta tttaatttca

28561 tacatgccat agttttcggt attagttgtt atgctttcgc ttttgttgca agtttcaaaa

28621 tagacggaat agctaccatt acgtccgcag caaacacagt tgtagagtaa tacttgcggt

28681 ctttcgcatc ttgaatgaac ggtttaatgt tcacgcttga atcttccaaa gcgatttgga

28741 tattacgttt cggagtaaac gcgataaacg catcttcgtc tgttgcatct gcgatcatag

28801 acgctgaaac atgaggaagt tcattgatct tgtatccttc taaagtgtac acagcttttc

28861 cgttctcgaa atgttcctga gcagttgtgt tgtccttact ctgaactaag ttcttaaaca

28921 agcgcattac attagaagtc acgaagaact cgctaagttc cttttgatcg gggcgttgtg

28981 aatcaataag tttcttcatg gtatcctcta cgctagccgt agtcaactgc aaaggaagga

29041 tagtttctgc gctgtccttc atttgcttaa tgaaaccacc atttttgaaa atattgtaag

29101 caacatcacc tgtcttggtt ccgtccagcc atgcgagacg tagcaaatca gcctccaaaa

29161 ctttcagtac ttccgaagcc atgaaacctg ccagttgagc ttcatcgaaa tcgtccgaca

29221 aatgaatacc cttagctacc atctttcccc acaaatcttg caaacaaaca acgataggta

29281 actccaaagg ctggaagtca tagtacttta catgatcgga catatctgtg tactcataag

29341 taccttcaca tccagcagac ttacgaagtg ccttatcttt cgctacaaaa gtaacaatag

29401 gcgttttgtt gtcaagtcca gagagaacgg ttgcaccgcg ttccatttcg ccaaccagcc

29461 cgacagtcaa agaaatgacg tcagccagtg agttaatgtt cagattattt aaatctgtaa

29521 atgtcattgc cataatttat agtctcctat gattttagtg tgattacttt tcttttgcga

29581 acttaaccat cgcctcccgc gctttcttgc gtgcttcctc gttggaaagc tgcgtttttt

29641 gtgcatctgt tttcggtttg cctcccaccg tacgagttgc aacaggtggc gttttcgttt

29701 gcttggaaag cattgtttta atctcactca aagatgattc aagagcagtt agacgctgag

29761 aaaattcgtc cggtgtcttg gtttctgtct ccggtgtttc ctccggcttt tcgtccatgt

29821 actctttgaa ctcggctatc ttgccatctt tgatcacgag aacgattttt ccctcttccg

29881 ggatatcaac ggtgatctca ccatcttcca cggtggttcc gtcctctttt actacttcgt

29941 cacctacggc tgcctcttct cctgctgcct taatagtgat cttttcacca ttaaccgttt

30001 ctacgatttc ctctttaagc tccgttttct ttgaaaaggt gctgataatg cttgataaaa

30061 gacccatttt gttctttgat ttttggttat taaaaagcga acttgtcgcg gctggtaagc

30121 ctacaaggtc acacgtgaat aattcttcaa aactcgtcac gtcccatgtt tggttctctt

30181 cgttccacac cttagtgtct aggtctacga ctgaaacgcc tagcatttcc ggttcctttt

30241 caatcatgtc cttcataaat cctgcctcct gcggatagtt ctttaagaga gcttccgaaa

30301 aggtcagatc ggcgtagact actccgtttt cctcaacgaa attagagaaa ctaccgatgt

30361 actggtccag cagatcgtta ccgttgtgtg tccggcgaga atggataggg cgaatactac

30421 cagccaccac aagggacgct agcgattcgg gcgtaataac gattttcccg gtttttaatt

30481 caccgttaac ttcatccgtc caatcgtttg cggtcggtcc cgcctctata attcgtattt

30541 tcttgaaatc cataaattag agatttacta tataaaccat tctatattat aaactatatt

30601 tcaacacaaa aatagctgtt acaacctact tactatccgc aaattgtagc aaatacgaat

30661 tagcccaatg cggaatcaat aacaattgta cgattttgtt gtacatcggt aatatcttgg

30721 actgaaacta tcggatttgg agcattctct accccttcta caaacgcaag tgcgatagct

30781 gccactgtct tgtcggacaa atcaaccgct tgcttagaca tagaccgatt aagattcgta

30841 taagattgag tactgataac atcgaatcct ccccccctgtg catatcggta agcattagaa

30901 ttgccaaaag accgtccccc gtattgctga ttaagagcag acaacgcatt tatagcatga

30961 gatgcacgtt tgttgaggat gtacatattc tcgcctcctt ccgcctcgaa ctgctgtccg

31021 tttgatccgg tgaacgttac accgccctgt gaatgcggag caccaaacac ggtaccacct

31081 ttggcaaact tcttaacgct ggtgttagtt ttgggaactt cttcttttac cttcataatt

31141 gaggctactt gtttcaatcc ggcagcgatc acgatggctg cttgcgctac tccccaaata

31201 ccaccctgcg caatagcctt agatgcacct aagtatgtgt taatcgtagc ctgtgccaga

31261 gcgaacgctt ttccggcttc cgattcctgt cccatgatat tagagatttg cccggcgata

31321 tcggctgtca tttgtaactt ggcgtttacc agttcctttt ctctcttttc ccgtatttgg

31381 gcgtatttgg attcaatcaa agaaatgtcc gccccagtct tctcagcgtt cgcaacttcc

31441 atttggtatt gctgctctag acgaagggat tcacgttcga agtcactagt tatgctggct

31501 tcctcgatgg cacgtttgtt ctctaggtct atggcttccg tttctctctt cttagcggct

31561 tcttcggctt ctagggtggc tacctgttct tggaatgcta tccgctgttc cagcttgatg

31621 ttatcgaatt cctgttgcgt tataagcccc tgctgcaacc tgtatcgctc cttttctaag

31681 atagcttggt ttagcgcgtc ttggtcctct agtgccttct gcttgtctac tataccgata

31741 ttagattctc ttatctttag ctgcaattcg gttattccgt tttcgtagct ttttagaact

31801 tcctgctgaa ctttcttagc ggtttcggcg gctttccttt ctgcgtcttc tttgtctttt

31861 ttgactttct ccgctgccgc tttttgggat ttagcataat cctcggcacg ctttttatcg

31921 gctactgcgg cggctgccat atcagctttt tccagcccgg atacctgact agttagttcc

31981 ttccgttgtg ataaatactg tgctctcttt tcctccaatg ctgctaaagc ctcttgttct

32041 ttcctcctgt cttcatcaga agtgtaagac aattcgtttt gcgtctttat ctgctgatat

32101 ttagcctcta gtatctttaa ttcggaggct tccatctctc tagatatccg tagggcttca

32161 ttagctgcat cttgcctttc cttagaactc ttagtttgat cggcaagaat agccttttgt

32221 tcctccattt ccctacgttg ccgggctaaa actacaataa gatcggtttc agcgttatat

32281 atgtcccgtt ctacttgtgc catccctcta gcagtttcga tagatttaac ggtttcgtcc

32341 gatatcagac ctaaccagtt gtaaacctta atataagcct ctgccagcca ctcaaaaacc

32401 ttgactatct ccacaaacag agcggctacg gcgtccaata ccttagttat gatcaactcg

32461 ataggtgcta atatagtctt aacggaaaca gccagttcgt tgttgcggtc catcaacttt

32521 ccgatagctg agatcaccgc aagaatagcc gatgcaatgg caacaaaggg gttcgccatc

32581 aatgcagcgt tgaatgcctt tatagaagca ataccgccgg acatacctt gaccatttgc

32641 ccggttgcgc cagtcatacc gccgaggctt ccggttgcct tctcgatatc ctctgcatag

32701 ttacccacgt ttctacgggt atctcccacc cctttctcaa ggtctttcag tttgtcggag

32761 atttccttgg tttgggctac catctgtttc ccagcctccg agtttgtccg ctgttcaacg

32821 gacattttgt ttagtgcctt cgtgttaagg gataacttcg ctcgcagggt ttccacgctt

32881 tccgcttccg agttgacaat ggtagtgtgt gctttaatag cggctgtatt ctcggaagtc

32941 tcattcttat tattgtttag ttgtttggtt aacgatatga tcgctgtttc tgacttctcc

33001 gttgccttttt cgaacgccgt ttgatctata agattgtctt tatagttctg acgaagcccg

33061 gcaagtgccg tcttctcggc gtttatctgc ttggttagct gtttcttttc ttctgcaagg

33121 tctaaagact tcttgatcag agcgtccaaa ccttctactg cctcgtccgt cttgaacgaa

33181 aggtctagta atgtaacatt atctgccatt tatttaacga ttaagattaa ttttagttag

33241 ttttacctta cattcctgcg tggaaatgtt gtagtccgtg atcgaccgga catagaaaaa

33301 cgcattgagt tgtttaaacc acaccacacc gttctcttta tagttgtttt cgatgtaata

33361 ataagggatt ttagccttaa ttgtcacgtc caaagcgtcc gaaaacagtc cgtagtattt

33421 ctgaagcgac tgcgtatact cgattgattt gaagtactct acccatgtag acccagcacc

33481 cacctctcct ttacgaatgg cgaatctagg atacgtccca tcttggggat acggcacacc

33541 ggattccacg atgtctccct tagccgtgaa cgatgccttc gaaacctgta gactcttgaa

33601 gaaatccccg atcttgaaca ctgcggtgtt cgggaagccc tcggaatcct cgatcttgtc

33661 ggttgacagg taaaagtcag accaatcctg ccggaattca ttgaaggtga ccggaccatt

33721 ccgatctacg ttgcgggcat cgtcagcgat cagtttgtac acgttgacta tgactcccgt

33781 gtttccgcta tcgtccacgt tcagcgtaaa cttccatccc cattggaaga ccttgcagat

33841 atcggtgaga tacgttatgg catccgatac tccgcagttt ccccattcct gtgtaagtcc

33901 tgcggtattc tgaagcgatg ttatttttcc cgacaggttt acggcttcat ccggtgacat

33961 gaaggatggc ggaagcgcat ccaacttatt gtagtcattg atgtccctgc tgatgtagaa

34021 tccatagaag tgacggatag gtatattcat agtcgtagaa acgtacttgt agcggacctg

34081 cgtgatgtcg ttctgatcat cgtccttttg gaagatcgct accgtctgat tattgtacac

34141 cgctttgaga accacatagt tcagcatcgt gcctatctta gacatatcca gcgtaatgta

34201 tgcgttgttg tccatagtaa caacagccac ggttgattct gtcacgccac ccggtgctgt

34261 catggctgtc cgggtatcgt actccttgat atctagagcg ttccctttaa agtatttcgt

34321 accgtcatca gtctccgatg catacctcca aaacacggta acagacctgc ctacaagtag

34381 atcggaaatc cctttctgac ccaagttctc gatgatcaga cccggtgcat aacctccgta

34441 gtcgatagcg gggaatgcca cgggagtagg agaacctagt gctttcgtga tcagatcgct

34501 agctcggaag aaccgggtat tgcccacatt cgagtcgatg agagtagcct ccacgacttc

34561 ttttggcaac tgcgacatct ttaggtccga ctgcgatagg gcaatgctgt agctctcctc

34621 atcgcaggtc accttagcct tgaaacgctt gttgatagcg atgccaccta tgtaaatcct

34681 agcctcgtac ttcatgtcac gtctcacgta accgaagttt cgcatctgat agaagatacc

34741 atcgttcacc cggtttctcg gtgctttgat gttagccgag taggtacggg tagactcgcc

34801 aaacgagtag ggagatgacg cattgatgga gagtttgacg tctgtcttag tcaagccctc

34861 caagaatgta ccgtttatct gtatctttat atccatgtta gtattggaat tttaatgttg

34921 ctgttttcgc caaaccggat gctgtatact tgactcccgt agtcgaggag cagcgcatct

34981 tggtgtcaaa aggcacgcca tccagccctt tcaccgtcac gtccggagac gttgacagga

35041 catcaagagc gaacttgtta gcctcggtca gttcgaatac gcaggtcagt tccctcttgg

35101 ttaccgtacc accgtccaag ccctgtgtta tcgtaggctg cgtgctccaa ttgtagcagg

35161 agatcgcatc gtaggaccca cacgagttta gccatttcag tgtaatagct ccgcatgcgc

35221 gcacctcttc cggataaacc ttaacagcag ctacggaccc atccggattc tttatcgtga

35281 tcttcttgta ttggcatgca tccgatattc cgtcaccctg catgtaatta aatttgtcag

35341 ccgttccgtc ttgatagata acgtctacat catagtccct atcatgataa cggctgtcta

35401 taaagaaatt atcatccagt gtgtgcgcta atggtgcccg aggtccgtag ttgtcccgga

35461 aatctgtatc agccctggac actcggctaa cattatttct actcgccaag ttcatcacgg

35521 gaatacgaac tatctgagta acactattag ccgttaga