MOLECULAR BIOLOGY OF ANTIBIOTIC RESISTANCE ELEMENTS IN HUMAN *HAEMOPHILUS* SPP.

Zaini Mohd Zain

St. Hilda's College

Thesis submitted for the degree of D.Phil at the University of Oxford

Trinity Term 2004

Faculty of Clinical Medicine

University of Oxford



ABSTRACT

MOLECULAR BIOLOGY OF ANTIBIOTIC RESISTANCE ELEMENTS OF HUMAN *HAEMOPHILUS* SPP.

Zaini Mohd Zain

St. Hilda's College, University of Oxford

D. Phil Thesis, Trinity 2004

Antibiotic resistance in *Haemophilus influenzae* emerged in the 1970s and rapidly rose in prevalence worldwide. The principal source of resistance was a large conjugative resistance element often referred to as a plasmid. The origins of this resistance element that integrates and conjugates have not been determined. The main aim of this thesis was to investigate the origin of this integrative and conjugative element (ICE) through comparative analysis of whole sequenced elements. Two Haemophilus resistance ICEs, ICE*Hin1056,* from the UK, and ICE*Hin299,* from Greece, were completely sequenced. Two further ICEs, ICE*Hin2866* from the USA and ICE*Hpa8F* from a UK *Haemophilus parainfluenzae* were compared. These ICEs consisted of highly homologous sequences that were predicted to form functional modules. The properties of these predicted modules accounted for replication, conjugation and integrative and excisive recombination with $tRNA^{leu}$. As the sequence diversity of resistance associated genes common to these ICEs were highly conserved compared to core ICE sequences they probably have been recently acquired by these ICEs suggesting a relatively recent origin. Further comparative analysis of these ICEs revealed a common evolutionary origin with genomic islands (GIs) found among bacteria belonging to β- and γ-*Proteobacteria*. These GIs share coherently organised core genes indicating that they have originated from a distant common ancestor. This finding contradicts the hypothesis that GIs were derived from independently evolving modules. Furthermore, the preservation of a coherent core gene structure between distantly related GIs suggests these core genes acting together confer a particular fitness advantage, which has contributed to the dissemination of adaptive genes such as antibiotic resistance genes. The methods used to compare these genomic islands could in the future be used to classify all genomic islands.

ACKNOWLEDGEMENTS

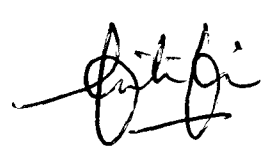
I am enormously grateful to Dr. Derrick Crook who has given me tremendous supervision and guidance, his moral support and enthusiasm, and of course his good-humoured made my life in Oxford, a memorable one indeed. I also wish to thank Dr. Sarah Turner from the Centre for Environmental and Hydrology, Oxford, for her criticisms and ideas on genomic islands. I am also grateful to Dr. Ana Cerdeno-Tarraga from Wellcome Trust Sanger Institute, Hinxton, Cambridge, for teaching me to annotate nucleotide sequences.

Many thanks to Prof. David Ferguson, from Cellular Pathology Division of the NDCLS, Oxford University, for his diligent examination of pilus on *H. influenzae* strains. Also, my gratitude goes to Dr. Sarah White from the Wellcome Trust of Human Genetics, Oxford, for rescuing my DNA from harmful effects of short-wave UV. I would also like to thank for the bioinformatic assistance provided by the officers at the Oxford University Bioinformatics Centre. I am also very grateful to Dr. Lucielle Mansfield for her constant encouragement and friendship, and to Mrs. Betty Coles for taking care of my welfare throughout my stay for the last few years, and helping me in the editing. My appreciations also go to Dr. Rosalind Harding and Dr. Naiel Bisharat for helping me to understand tree topologies, and to Mr. David Griffiths for his ever-willingness to spare consumables in times of emergencies. I do not forget to thank for the technical guidance given by Dr. Ioanna Dimopoulou and Dr. Joanne Russell during the beginning of my research in the laboratory. Also, my appreciation goes to Dr. Nicola Jones for helping me to edit my Chapter one.

I must also acknowledge the Government of Malaysia for the financial support and to the Department of Veterinary Services of Malaysia for recommending me to the scholarship and a study leave. Lastly, but not least, lots of thanks to my husband Rahmat, for his support and willingness to share the sufferings with me doing Ph.D as well, and to my daughter Afiqah, for understanding our situation as students again, and for her patience being home alone during most of our times spent at work.

Note of declaration

I, Zaini Mohd Zain hereby, declare that the work carried out in this thesis was solely carried out by myself. The sequences of *H. influenzae* 2866 were kindly provided by Dr. Arnold Smith from Seattle Biomedical Research Institute, USA. Dr. Thomas Inzana from Virginia Tech, Blacksburg, USA, provided the genome sequences of *H somnus* 2336, and the sequences of ICE*Hpa8F* was given by Dr. Lucielle Mansfield from NDCLS, Oxford University.



"***This thesis is dedicated to Rahmat and Afiqah***"

CONTENTS

ABSTRACT 2

ACKNOWLEDGEMENTS 3

CONTENTS 5

LIST OF ABBREVIATIONS 12

CHAPTER ONE 14

***HAEMOPHILUS INFLUENZAE* AND ANTIBIOTIC RESISTANCE** 14

1.1. Introduction 14

1.2. Historical background 15

1.3. Characteristics of *Haemophilus* spp. 15

1.4. *H. influenzae* genome 16

1.5. Virulence factors 17
1.5.1. Lipopolysaccharide 18
1.5.2. OMPs 19
1.5.3. Pili 19
1.5.5. Capsule 20

1.6. Pathogenesis 21

1.7. Genetic variation between strains 23

1.8. Clinical manifestations of *H. influenzae* infection 24
1.8.1. Serotype b infections 25
1.8.2. Infections caused by non-type b serotypes 26
1.8.3. Infections associated with non-typeable strains 26

1.9. Control and prevention of *H. influenzae* infection 27
1.9.1 Host defences 27
1.9.2. Immunisation 27

1.10. Antibiotic resistance in *Haemophilus* spp. 29
1.10.1. Changing trends in antimicrobial susceptibility 29
1.10.2. Prevalence of antibiotic resistant *H. influenzae* 30
1.10.3. Medical impact of antibiotic resistance 31

1.11. Detection of antibiotic resistance 31

1.12. Mechanisms of antibiotic resistance 31

1.13. Ampicillin resistance in *H. influenzae* 32

1.13.1. Types of β-lactamases in *H. influenzae* 32
1.13.2. TEM β-lactamase 33
1.13.3. ROB β-lactamase 33

1.14. Tetracycline resistance in *H. influenzae* 33

1.15. Chloramphenicol resistance in *H. influenzae* 34

1.16. Resistance to other antimicrobials 35
1.16.1. Kanamycin resistance 35
1.16.2. Trimethoprim resistance 35
1.16.3. Multi-drug resistance 35

1.17. Mobile genetic elements 36

1.18. Transposon mediated resistance 36
1.18.1. Tn*10* 37
1.18.2. Tn*3* 38

1.19. Conjugative and non-conjugative plasmids 39

1.20. Resistance plasmids (elements) in *H. influenzae* 41

1.21. Transfer of plasmids 42
1.21.1. Transfer by transformation 42
1.21.2. Transfer by conjugation 43
1.22.3. Transfer by transduction 44

1.23. Restriction and modification system 44

1.24. Site-specific recombination of *H. influenzae* plasmids 45

1.25. Excisive recombination of plasmid 46

1.26. Origin of resistance genes and conjugative plasmids 46

1.27. Role of plasmids in gene transfer and bacterial evolution 47

1.28. Genomic islands 48

1.29. Integration sites for GIs 48

1.30. Detection of GIs 49

1.31. Historical background on sequencing 50

1.32. Importance of complete genomic sequences 51

1.33. Complete plasmid sequences 52

1.34. Various strategies of genomic sequencing 52
1.34.1. Sequencing by primer walking 53

1.34.2. Clone-by-clone sequencing 53
1.34.3. Whole genome shotgun strategy 54

1.35. Objective of this thesis 54

CHAPTER TWO 55

MATERIALS & METHODS 55

2.1. Bacteria strains and plasmids 55
2.1.1. Bacteria 55
2.1.2. Plasmids 55
2.1.3. Bacterial sequences 57

2.2. Media 58
2.2.1. Solid culture media 58
2.2.2. Liquid culture media 59
2.2.3. Antibiotics selective media 59
2.2.4. X-Gal and IPTG plates 59

2.3. Chemical, restriction enzymes and DNA modifying enzymes 60

2.4. Bacterial growth conditions 60
2.4.1. Growth of *E. coli* and *H. influenzae* strains 60
2.4.2. Culture and storage of bacteria 61

2.5. Colony counts 61

2.6. Antibiotic susceptibility test by Etest 62

2.7. Conjugation 62
2.7.1. Conjugation on solid media 62
2.7.2. Conjugation in liquid media 63
2.7.3. Measurement of point mutation 63
2.7.4. Purification of transconjugants 63

2.8. Electron microscopy examination of pili 64

2.9. Chromosomal DNA extraction 64
2.9.1. Extraction of chromosomal DNA from *H. influenzae* 64
2.9.2. Extraction of crude DNA by boiling method 65

2.10. Plasmid extraction 65
2.10.1. Extraction by mini-prep method 65
2.10.2. Extraction of plasmid DNA by caesium chloride-ethidium bromide gradients 66

2.11. Removal of contaminating chromosomal DNA 68

2.12. Plasmid shearing 68

2.13. Agarose gel electrophoresis **69**
2.13.1. Preparation of agarose gel 69
2.13.2. Gel electrophoresis 69
2.13.3. Observation of DNA in agarose gel 69
2.13.4. Recovery and purification of DNA from agarose gel 70

2.14. Restriction enzymes digestion **70**

2.15. Polymerase chain reaction (PCR) **71**
2.15.1. Selection of oligonucleotide primers 71
2.15.2. Dilution of primers 71
2.15.3. Preparation of PCR mixture 71
2.15.4. Thermal cycling condition 72
2.15.5. Observation of amplified products 72

2.16. Purification of PCR products and DNA **72**
2.16.1. Purification by spin column 72
2.16.2. Polyethylene-Glycol (PEG) purification 73
2.16.3. Ethanol/sodium acetate (NaOAc) precipitation 73

2.17. DNA manipulations **74**
2.17.1. Fragmentation of DNA by sonication 74
2.17.2. Creating blunt-ended DNA 74
2.17.3. Phosphatase treatment of cloning vectors 75

2.18. DNA ligations **75**

2.19. Bacterial transformation **75**
2.19.1. Transformation by heat-shock 75
2.19.2. Transformation of PCR products into TOP10 *E. coli* 76

2.20. Selection of recombinant clones **77**
2.20.1. Confirmation of cloned inserts 77
2.20.2. In-well lysis 77

2.21. Construction of plasmid library **78**

2.22. DNA sequencing **79**
2.22.1. Principles of DNA sequencing 79
2.22.2. Templates for sequencing 79
2.22.3. Cycle sequencing 80
2.22.4. Automated sequencing 80

2.23. Computer software for sequence assembly **80**
2.23.1. Staden 80
2.23.2. Phred and Phrap 81

2.24. Computer programs for sequence annotation **81**
2.24.1. GeneMark and Glimmer 81
2.24.2. Artemis 81

2.25. Programs for sequence comparison **83**

2.25.1. GCG 83
2.25.2. Blixem 83
2.25.3. Repeat sequences 83
2.25.4. Artemis Comparative Tools 83
2.25.5. ClustalX and ClustalW 84
2.25.6. Molecular Evolutionary Genetic Analysis (MEGA) 84
2.25.7. SplitsTree decomposition 84

CHAPTER THREE **86**

PHENOTYPIC CHARACTERISTICS *H. INFLUENZAE* STRAIN 1056 **86**

3.1. Introduction **86**

3.2. Objectives **86**

3.3. Materials and Methods **87**
3.3.1. Bacterial strains 87
3.3.2. Minimum inhibition concentration test 87

3.4. Results and discussion **87**
3.4.1. Minimum inhibition concentration (MIC) 87
3.4.2. To determine the rate of plasmid transfer to *recA*$^+$ and *recA*$^-$ recipients 88
3.4.3. Re-transfer from a transconjugant 89
3.4.4. Evidence for incompatibility 90
3.4.5. Electron microscopy for pilus 91

3.5. Conclusion **94**

CHAPTER FOUR **95**

COMPLETE NUCLEOTIDE SEQUENCE OF THE MOBILE ELEMENT ICE*HIN1056* OF *H. INFLUENZAE* 1056 **95**

4.1. Introduction **95**

4.2. Objectives of whole sequencing of ICE*Hin1056* **95**

4.3. Materials and methods **96**

4.4. Results and discussion **97**
4.4.1. Primer walking 97
4.4.2. To determine the optimum length of time for sonication 97
4.4.3. Construction of plasmid libraries 97
4.4.4. Sequence assembly of ICE*Hin1056* 99
4.4.5. Assignment of potential coding regions of ICE*Hin1056* 101
4.4.6. General properties of ICE*Hin1056* 102
4.4.7. Overview of the genetic organisation of ICE*Hin1056* 103
4.4.8. Putative replication module 111
4.4.9. Antibiotic resistance module 112

4.4.10. Putative transfer module 114
4.4.11. Ampicillin resistance module 116
4.4.12. Identification of putative origin of replication 117
4.4.13. Identification of putative origin of transfer 120
4.4.14. Nucleotide sequence accession number 121

4.5. Conclusion 122

CHAPTER FIVE 123

NUCLEOTIDE SEQUENCE AND ANALYSIS OF ICE*HIN299* 123

5.1. Introduction 123

5.2. Objectives 123

5.3. Materials and methods 123
5.3.1. Extraction of ICE*Hin299* 123
5.3.2. Sequencing of ICE*Hin299* 124

5.4. Results and discussion 124
5.4.1. Overall organisation of ICE*Hin299* 124
5.4.2. Description of ICE*Hin299* sequences based on functional module 131
5.4.3. Putative replication module 131
5.4.4. Putative transfer module 132
5.4.5. Putative transfer/integration module 132
5.4.6. Antibiotic resistance module 133
5.4.7. G+C content of ICE*Hin299* 133
5.4.8. Determination of putative *oriV* and *oriT* 134

5.5. Conclusion 136

CHAPTER SIX 138

COMPARATIVE STUDIES ON MOBILE ELEMENTS OF *HAEMOPHILUS* SPP. 138

6.2. Objectives 138

6.3. Materials and Methods 138
6.3.1. Detection of additional elements on *Haemophilus* spp. 138
6.3.2. *Haemophilus* spp. sequences of ICEs 138

6.4. Results 139
6.4.1. Comparison of mobile elements of *Haemophilus* spp. 139
6.4.2. Phylogenetic relationship between the genes in the *Haemophilus* spp. 145
6.4.3. Comparison of ampicillin resistance genes 149
6.4.4. Comparisons of *oriV*, *ter* and *oriT* 152
6.4.5. Comparison of ICE*Hin1056* to other Haemophilus plasmids and genetic islands 158

6.5. Discussion 161

CHAPTER SEVEN 165

ICE*HIN1056* AND ITS RELATIONSHIP WITH OTHER GENOMIC ISLANDS 165

7.1. Introduction 165

7.2. Objectives 165

7.3. Materials and Methods 165
7.3.1. Identification of GIs 165
7.3.2. Construction of virtual core GI sequences 166
7.3.3. Identification of core genes of other GI candidates 167
7.3.4. G+C% content 167
7.3.5. Phylogenetic relationship between GIs 167

7.4. Results 168
7.4.1. Identification of related coherent GIs 168
7.4.2. Features of core genes 169
7.4.3. Phylogenetic relationship between GIs 173
7.4.4. Accessory genes including antibiotic resistance content 174
7.4.5. Hot spots 176

7.5. Discussion 176
7.5.1. A coherent core GI 176
7.5.2. Antibiotic resistance accessory genes and this family of GIs. 181
7.5.3. Major properties of other accessory genes 183

7.6. Conclusions 185

CHAPTER EIGHT 187

CONCLUDING REMARKS AND FUTURE RESEARCH 187

REFERENCES 190

APPENDIX A 210

APPENDIX B 214

APPENDIX C 219

APPENDIX D 221

APPENDIX E 224

LIST OF ABBREVIATIONS

A	adenine
AA/aa	amino acid
Ala	alanine
amp^r	ampicillin resistant
asn	asparagines
BBB	blood-brain barrier
BHI	brain heart infusion
bp	base pair
C	cytosine
^oC	degrees in Celsius
cat	chloramphenicol acetyltransferase
chlor^r	chloramphenicol resistant
CIAP	calf intestine alkaline phosphatase
CO_2	carbon dioxide
CSF	celebral spinal fluid
CsCl	caesium chloride
DNA	deoxyribonucleic acid
DNTP	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
EDTA	ethylene diamine tetra-acetic acid
EM	electron microscopy
EMBL	European Molecular Biology Laboratory
EtBr	ethidium bromide
G	guanine
g	gram
g	gravitational force
GC	guanine and cystine
GI	genomic island
HEPES	N-2-hydroxyethylpiperazine-N-2-ethanosulfonic acid
HTH	helix-turn-helix
ICE	integrative and conjugative element
IHF	integration host factor
IPTG	isopropylthio-β-D-galactosidase
IS	insertion sequence
kb	kilobases
kDa	kilodalton
L	litre
LB	Luria broth
Leu	leucine
LPS	lipopolysaccharide
Lys	lysine
Mda	megadalton
MgCl	magnesium chloride
MH	Muller Hinton
MIC	minimal inhibitory concentration
µg	microgram
µl	microlitre
ml	millilitre

MLEE	multi-locus enzyme electrophoresis
MLST	multi-locus sequence typing
NAD	nicotinamide adenine dinucleotide
NaOAc	sodium acetate
NCBI	National Centre for Biotechnology Information
OMP	outer membrane protein
ORF	open reading frame
oriT	origin of transfer
oriV	origin of replication
PAI	pathogenicity island
PBS	phosphate buffer saline
PBP	penicillin binding protein
PCR	polymerase chain reaction
PEG	polyethylene-glycol 8000
Phe	phenylalanine
PI	isoelectric point
Pro	proline
PRP	polyribosyl ribitol phosphate
UK	United Kingdom
USA	United States of America
UV	ultraviolet
R plasmid	resistant plasmid
RAPD	random amplified polymorphic DNA fingerprinting
REA	restriction endonuclease analysis
RFLP	restriction fragment length polymorphism
rif^r	rifampicin resistant
RNA	ribonucleic acid
rpm	revolution per minute
s	seconds
Ser	serine
$strep^r$	streptomycin resistant
T	thymine
TE	tris-EDTA
TER	Terminator/stop codon
ter	site for termination of replication
tet^r	tetracycline resistant
TMHMM	transmembrane helix motif
Tn	transposons
TSB	tryptose soy broth
W	watt
X-Gal	5-Bromo-4-chloro-3-indolyl-β-D-galactosidase

CHAPTER ONE

HAEMOPHILUS INFLUENZAE AND ANTIBIOTIC RESISTANCE

1.1. Introduction

The genus *Haemophilus* includes commensal species with a wide range of pathogenic activities. They are small, non-motile Gram-negative bacteria in the family *Pasteurellaceae* (Barrow and Feltham, 1993; Holt *et al.*, 1994). The family also include *Pasteurella* and *Actinobacillus*, two other genera of bacteria that are commensals of animals.

Species in the genus of *Haemophilus* are host-specific. In humans, the *Haemophilus* spp. constitute up to 10% of the normal bacterial flora of the upper respiratory tract and are also occasionally found as normal flora of the conjunctivae and genital tract (Turk, 1982; Turk, 1984). The species of *Haemophilus* that are of clinical importance in humans are *Haemophilus influenzae,* the causative agent for meningitis and septicaemia; *Haemophilus aegypticus* causing conjunctivitis and Brazilian purpuric fever; and *Haemophilus ducreyi* which is the major cause of human genital ulcer disease (chancroid) in many developing countries. Other species of *Haemophilus* rarely infecting humans include *Haemophilus aphrophilus, Haemophilus haemolyticus, Haemophilus parainfluenzae, Haemophilus paraphrophilus*, and *Haemophilus segnis. H. parainfluenzae* is an abundant commensal species that is relatively non-pathogenic, but is rarely described as causing pneumonia or bacterial endocarditis. *H. aphrophilus* is a member of the normal flora of the mouth and occasionally causes bacterial endocarditis.

In animals, species of *Haemophilus* are host-specific; some examples include *Haemophilus somnus, Haemophilus parasuis and Haemophilus paragallinarum. H. somnus* infects the respiratory tract of cattle causing pneumonia, but infection in the

genital tract of a pregnant cow may cause abortion (Miller *et al.*, 1983). *H. parasuis* and *H. paragallinarum* causes respiratory tract infection in swine and poultry, respectively.

1.2. Historical background

H. influenzae is the major human pathogen in the genus of *Haemophilus.* It was first characterised by Robert Koch in 1883 in patients with conjunctivitis and was identified by Pfeiffer nine years later (Killian *et al.*, 1981). The organism was named accordingly because it was mistakenly thought to be the cause of the disease influenza during the influenza pandemic of 1890. The name 'Haemophilus' was given due to the fact that haemin is required for growth.

1.3. Characteristics of *Haemophilus* spp.

Haemophilus spp. are facultative anaerobes, non-motile, and do not form spores. The average size of the bacterium is approximately, 0.3-0.5 μm by 0.5-1.0 μm and the morphology varies from coccobacilli to pleomorphic depending upon cultural environment. As the culture ages, filamentous forms usually develop. Although *H. influenzae* stains poorly with Gram's stain, it is Gram's negative.

Haemophilus (meaning "loves heme") requires a precursor of heme in order to grow. *H. influenzae* prefers a complex medium and requires preformed growth factors that are present in blood. Specifically, it requires X factor, which is supplied by iron containing pigments, and V factor, usually in the form of nicotinamide adenine dinucleotide (NAD) (Barrow and Feltham, 1993). In the laboratory, it is usually grown on chocolate blood agar and grows best at 35-37°C under slight CO_2 tension (5% CO_2). Heme is required for the growth of *H. influenzae* type B but non-typeable *H. influenzae*

(NTHI) strain grows in medium containing free iron and protoporphyrin IX in place of heme (Sanders *et al.*, 1994).

Most strains of *Haemophilus* spp. are usually non-capsulated. Capsule is found in some types of *H. influenzae,* although most strains of *H. influenzae* are devoid of capsules. On colourless medium such as Columbia agar and Levinthal agar, iridescence can be observed in colonies of capsulated *H. influenzae* (Pittman, 1930).

1.4. *H. influenzae* genome

The *H. influenzae* Rd strain KW20 chromosome was the first complete genome sequence of a free-living organism (Fleischmann *et al.*, 1995). This strain was also the source of the first DNA restriction enzyme, *Hind*III. The circular Rd chromosome is 1,830,137 bp long with an approximately 38% G+C nucleotide content. The sequence has 1,743 open reading frames (ORFs) of which, 63 contain frameshifts or stop codon when compared to homologues from other species and therefore, may be pseudogenes (Fig. 1.1).

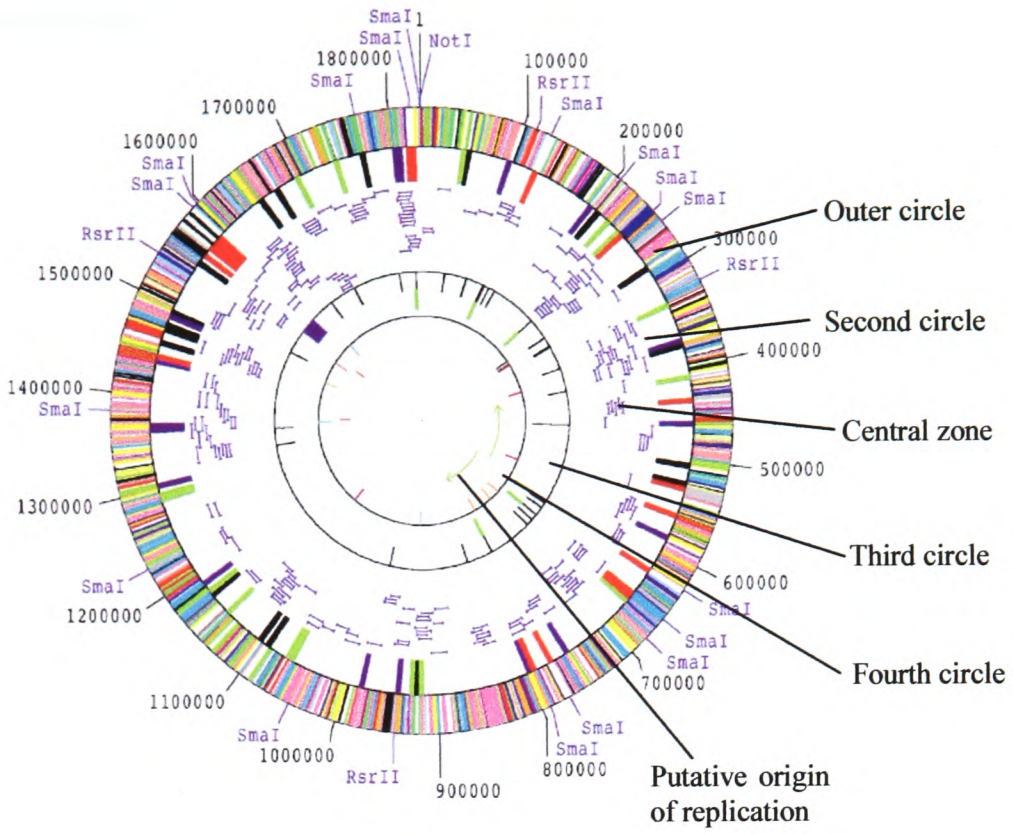


Fig. 1.1. A circular representation of *H. influenzae* Rd chromosome. Outer circle: the location of each predicted coding region containing a database match (colour coded by functional grouping) as well as selected global features of the genome including a unique *Not*I restriction endonuclease site (designated as nucleotide 1). Second circle: GC content bias (red =>42%, blue =>40%, green =>36%, black =>34%). Central zone: the purple I-bars within the circle indicate the large-insert (lambda phage) clones used as a framework to map the sequence contig relative to each other. Third circle: RNA genes (ribosomal RNAs, tRNAs). Fourth circle: simple repeats the green arrows show the putative origin of replication (Fleischmann *et al.*, 1995).

1.5. Virulence factors

H. influenzae does not produce any demonstrable exotoxins (van der Zwan *et al.*, 1978). Apparently, each step in the pathogenesis of *H. influenzae* infection depends on the expression of a combination of several specific virulence determinants. These determinants include lipopolysaccharide (LPS), pili, outer-membrane proteins (OMPs), IgA1 proteases and capsule.

1.5.1. Lipopolysaccharide

Lipopolysaccharide (LPS) is a major suprastructure of glycoprotein on Gram-negative bacteria, which contributes greatly to the structural integrity of the bacteria, and protects them from host immune defences and a major virulence determinant. A feature of *H. influenzae* LPS is that surface exposed epitopes of the oligosaccharide are subject to high-frequency on-off switching of express (phase variation). This heterogeneity may be an advantage to the bacteria allowing them to better confront different host compartments and microenvironments and to survive the host immune response.

H. influenzae LPS consists of two covalently linked regions, lipid A and core oligosaccharides (Richards *et al.*, 2001). The toxicity of LPS was shown to be attributed mainly to the activity of the lipid A portion. Since the lipid A portion is embedded in the outer membrane, its endotoxicity is exerted mostly when the organism is lysed. To activate host cells, the LPS binds a plasma LPS-binding protein (LBP) to form a LPS-LBP complex, which then binds the host cell receptor, CD14 (Lazou Ahren *et al.*, 2001; Wang *et al.*, 2002). CD14 then interacts with Toll-like receptor proteins culminating in the transduction of a cytoplasmic signal (Chow *et al.*, 1999; Medzhitov *et al.*, 1997). Through the activation of a complex cascade of events, the production of cytokines is triggered. The activity of cytokines and complements may lead to septic shock.

The oligosaccharide portion of LPS extends outward from the bacterial surface. Between strains of *H. influenzae*, the oligosaccharide displays significantly different patterns of sugars. The oligosaccharide can influence interactions of the bacteria with host cells and components of the host immune system (Richards *et al.*, 2001).

1.5.2. OMPs

Strains of *H. influenzae* express between 10 and 20 OMPs (Murphy and Apicella, 1987). Among all, the porin protein P2 is the most abundant in the outer membrane of *H. influenzae* type b (Murphy and Bartos, 1988). Other major OMPs are P5 and P6 proteins (Munson and Granoff, 1985). The P5, which has a 50% identity and 65% similarity to the OmpA porin of *Escherichia coli* (Munson *et al.*, 1993), is thought to be involved in the invasion of the mucosal epithelium (Chanyangam *et al.*, 1991), while the P6 has been showed to possess immunogenic properties and constitutes a promising vaccine candidate antigen (Murphy *et al.*, 1992; Nelson *et al.*, 1988).

1.5.3. Pili

Many *H. influenzae* isolates express pili (fimbriae), which mediate adherence to epithelial cells and facilitate colonisation (Guerina *et al.*, 1982; Pichichero *et al.*, 1982). The expression of pili in *H. influenzae* is regulated by phase variable, spontaneously switching from piliated to non-piliated and vice-versa (Connor and Loeb, 1983; Guerina *et al.*, 1982; Pichichero *et al.*, 1982). Generally, strains isolated from invasive infections do not express pili, while those colonising the nasopharyxn more commonly do (Mhlanga-Mutangadura *et al.*, 1998).

1.5.4. Immunoglobulin A1

H. influenzae and a number of other mucosal pathogens, such as *Neisseria meningitis*, *Neisseria gonorrhoeae*, and *Streptococus pneumoniae* constitutively secretes immunoglobulin A1 (IgA1) protease (Reinholdt and Kilian, 1997), an enzyme that cleaves and inactivates human IgA1 to facilitate colonisation (Plaut, 1978).

1.5.5. Capsule

The most important virulence factor in pathogenic *H. influenzae* is the capsule (Roberts *et al.*, 1981). Capsulated *H. influenzae* isolates are classified according to their polysaccharide capsule (Pittman, 1931). There are a total of six structurally and antigenically distinct capsular types (serotypes) designated a through f (typeable strains). Strains that lacks capsule and do not agglutinate with antisera against the capsular serotypes are considered as non-capsulated (non-typeable or NTHI).

Amongst all the serotypes, serotype b is the most virulent. The polysaccharide capsule of serotype b is made up of polyribosyl ribitol phosphate (PRP) known to be the major virulence factor of *H. influenzae* (Moxon and Maskell, 1992). The capsule of *H. influenzae* has anti-phagocytic properties. This is especially prominent with the type b capsule.

The classification method of serotyping of *H. influenzae* was primarily based on serological techniques. One disadvantage of this technique is that cross-reactions between the serotypes may occur. Molecular methods using polymerase chain reaction (PCR) have been developed as a confirmative method to determine capsular type (Jordens *et al.*, 1993). The PCR method for typing the strains of *H. influenzae* are now being adopted for differentiating strains of capsulated (a-f) from non-capsulated and non-capsulated variants of capsulated strains (deletion mutation) of *H. influenzae* (Falla *et al.*, 1994; Jordens *et al.*, 1993).

All encapsulated (typeable) strains possess the *cap* genetic locus, a segment of approximately 18 kb comprised of a region that shows serotype specificity flanked by regions sharing homology among all capsular types (Kroll, 1992). Most types a, c-f strains possess one copy of the *cap* locus. In contrast, most type b clinical isolates possess two copies in tandem. The *cap* locus, either as the tandem duplication of most type be

strains or the single copy of strains of other capsular types, is flanked by IS-like sequences in most strains, suggesting that the region was required by a primordial strain through transposition (Kroll *et al.*, 1991).

One of the conserved regions of the *cap* locus contains the *bex* genes, which are responsible for exporting the capsular polysaccharide (Kroll *et al.*, 1990). The tandem duplication of the *cap* locus in type b strains is characterised by a mutation in the *bexA* gene of one copy. This duplication results in an unstable gene region, with relatively high potential for recombination events. During recombination of the *cap* locus, mutants may be formed that possess a single copy of the *cap* locus containing the *bexA* mutation, which results in non-encapsulated mutants. On the other hand, such a recombination event may generate mutants possessing additional (n) copies of the *cap* locus with n-1 copies of *bexA*. Such mutants are hyper-encapsulated and show increase pathogenicity in the infant rat model (Kroll and Moxon, 1988).

1.6. Pathogenesis

H. influenzae typically colonises human respiratory mucosal surfaces and occasionally the female genital tract (Gilsdorf *et al.*, 1997). From the nasopharyxn, organisms are transmitted from one individual to another by airborne droplets or by direct contact with secretions (Farley and Stephens, 1992). Invasion by *H. influenzae* begins with colonisation of the upper respiratory mucosa that is mediated by pili (fimbriae) (St.-Geme-III, 1994). Inability to express fimbriae results in impaired ability of *H. influenzae* to colonise the nasopharynx (Weber *et al.*, 1991). In addition, presence of fimbriae is associated with a capacity to agglutinate human erythrocytes (van Alphen *et al.*, 1988). The fimbrial gene cluster, is found on an 8-kb fragment of *H. influenzae* type b, but is missing in Rd strain (van Ham *et al.*, 1994).

Fimbriated *H. influenzae* type b isolates have also been found to show efficient adhesion to mammalian extracellular matrices (ECMs) (Virkola *et al.*, 1996), which may promote the invasion of *H. influenzae* type b strains into the circulation. The encapsulated organisms penetrate the epithelium of the nasopharynx and gain access to the blood by passing through the respiratory epithelia and vascular endothelia (Rubin and Moxon, 1983). Specific interactions are believed to occur between *H. influenzae* and the blood-brain barrier (BBB). The BBB is a single layer of unique endothelial cells, which is largely responsible for the maintenance of biochemical homeostasis within the central nervous system (Betz and Goldstein, 1986). These endothelial cells exhibit continuous tight junctions and a marked paucity of pinocytosis (Patrick *et al.*, 1992). It is believed that the *H. influenzae* cells adhere to the BBB and then translocate across or between the cellular tight junctions, to enter the celebral spinal fluid (CSF). Damage to the BBB enhances *H. influenzae* entry into the CSF. Once in the CSF, the population of *H. influenzae* may continue to expand and infect the meninges of the brain, causing meningitis.

The invasive type b *H. influenzae* is able to resist phagocytosis by polymorphonuclear leukocytes because of the presence of polyribosyl ribitol phosphate (PRP) capsule. The capsule is an important factor enabling its survival in the blood, and the resultant bacteraemia (Noel *et al.*, 1992). Bacteraemia seems to precede the entry of the organisms in the CSF and meningitis (Moxon *et al.*, 1977; Moxon and Murphy, 1978).

The less invasive non-typeable strains may induce an inflammatory response at mucosal locations, such as the middle ear, the sinuses, the conjunctiva, or the lower respiratory tract, which is thought to contribute to disease at these sites. Following

colonisation there may be local invasion of the respiratory tract by contiguous spread (St.-Geme-III, 1994).

1.7. Genetic variation between strains

Besides different capsular types, there are other classification systems that could be used to subtype *H. influenzae* strains. Typing of strains provides epidemiological information about the spread of individual clones of the organisms, which helps to identify hyper-virulent strains or distinguish between recurrent and repeated infections (Edmonson *et al.*, 1982; Kostman *et al.*, 1993). In addition, characterising bacterial populations with subtyping techniques may also aid in phylogenetic or evolutionary studies (Aparicio *et al.*, 1996). Methods frequently used for subtyping include: biotyping, whole-cell protein profiling, outer membrane protein (OMP) subtyping, multi-locus enzyme electrophoresis (MLEE) or multi-locus sequencing typing (MLST).

In the case of biotyping, *H. influenzae* strains are classified according to a series of biochemical reactions such as ornithine decarboxylase activity, indole production and urease activity. The biotypes are designated as I through VI (Killian, 1976). This classification has been superseded by superior methods.

Diversity of *H. influenzae* isolates can be studied by looking at their whole cell protein profiles. Using protein profiling, it was found that *H. influenzae* type b belongs to a subset of single clone, whereas non-typeable strains were mostly heterogeneous (Musser *et al.*, 1985; Paterson *et al.*, 1987). This technique has now been replaced by DNA sequence based methods.

Methods utilising the whole genome provide useful information for epidemiological studies (Jordens *et al.*, 1993; Jordens, 1998; Mitsuda *et al.*, 1999). These include random amplified polymorphic DNA fingerprinting, RAPD (Jordens *et al.*, 1993),

restriction fragment length polymorphism, RFLP (van Belkum *et al.*, 1994), restriction endonuclease analysis, REA (Bruce and Jordens, 1991; Omikunle *et al.*, 2002), ribotyping (Leaves and Jordens, 1994) and PCR-ribotyping (Smith-Vaughan *et al.*, 1995).

Multi-locus genotyping method is based on characterising unique combinations of virulence genes or alleles of virulence genes (Musser *et al.*, 1986; Musser, 1996). This method has shown that most *H. influenzae* type b diseases are caused by nine related clones that have no close genetic relationships to strains of other serotype (Musser *et al.*, 1990).

The most recently developed typing scheme is multi-locus sequencing typing (MLST). This scheme provides an unambiguous characterisation of both encapsulated and non-capsulated isolates of *H. influenzae.* This typing was developed based from the principles of multi-locus enzyme electrophoresis (MLEE), but assigns alleles at multiple housekeeping loci directly by DNA sequencing, rather than indirectly via the electrophoretic mobility of their gene products (Maiden *et al.*, 1998; Spratt, 1999). Although both MLEE and MLST are capable of differentiating between capsulated strains from non-capsulated strains, MLST is able to show related lineages between encapsulated strains of *H. influenzae* (Meats *et al.*, 2003). Through this scheme, it was found that recombination appeared to have more impact on the divergence of non-capsulated isolates compared to encapsulated isolates. This could be due to their increased ability to be transformed and/or increased opportunities for non-capsulated isolates to meet other *H. influenzae* within the nasopharyxn (Meats *et al.*, 2003).

1.8. Clinical manifestations of *H. influenzae* infection

Diseases caused by *H. influenzae* infection can be classified as invasive infections and non-invasive infections (Table 1.1). The invasive infections are usually associated with

serotype b, whereas the non-invasive are usually caused by non-typeable (NTHI) *H. influenzae.*

Table 1.1. Important infections caused by *H. influenzae*

Invasive infection[a]	**Non-invasive infection**[b]
Bacteraemia	Conjunctivitis
Epiglottitis	Otitis media
Meningitis	Respiratory disease (chronic bronchitis)
Pneumonia	

[a]90% caused by *H. influenzae* type b, [b]mostly caused by untypeable strains

1.8.1. Serotype b infections

Amongst all the encapsulated strains of *H. influenzae*, serotype b is found to be the most virulent and most invasive. Studies confirmed that *H. influenzae* type b strain was responsible for approximately 85% of all invasive *H. influenzae* disease in children (Booy *et al.*, 1993; Wenger *et al.*, 1992). It was the leading cause of bacterial meningitis and other systemic diseases in infants and young children under the age of 5 years (Turk, 1984; Uduman *et al.*, 2000).

Epidemiological data of 3,931 patients from 21 countries worldwide reported that meningitis was the leading disease manifestation of *H. influenzae* type b (Fig. 1.2) (Peltola, 2000). Other diseases caused by type b *H. influenzae* include epiglottitis, cellulitis, osteomyelitis, and arthritis in older children and also adults (Turk, 1984).

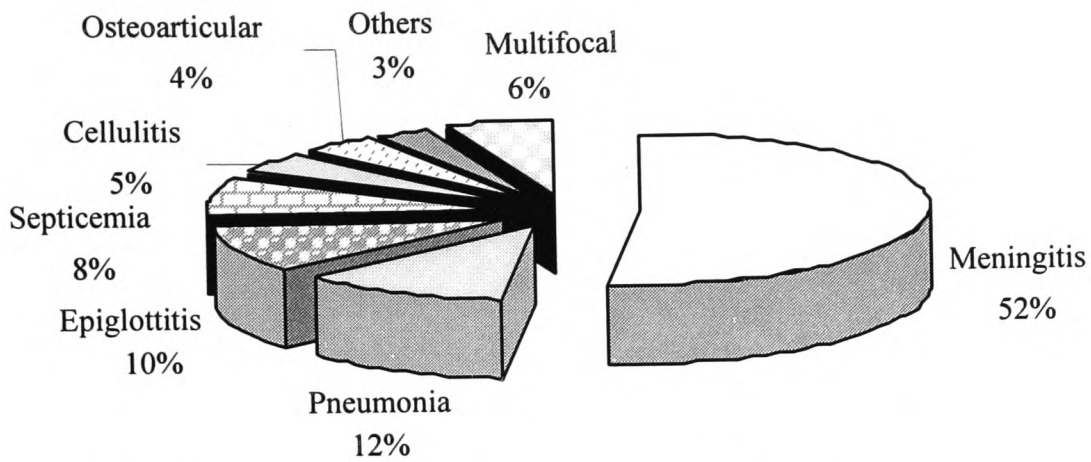


Fig. 1.2. Worldwide spectrum of *H. influenzae* type b disease (non-bacteraemia pneumonia excluded). Data collected in the 1970s to 1990s from 3,931 patients in 21 studies from various parts of the world (Peltola, 2000).

1.8.2. Infections caused by non-type b serotypes

The other *H. influenzae* serotypes, a, c, d, e and f, have been shown to be associated with invasive disease at low frequency (Musser *et al.*, 1990). Non-type b strains rarely cause serious diseases in children (Heath and McVernon, 2002). Amongst all serotypes, the serotype f is regarded to be the most common cause of the non-type b invasive infections (Gonzalez Lopez *et al.*, 2000; Urwin *et al.*, 1996). Over a period of six years from 1989 to 1994, the incidence of invasive *H. influenzae* due to serotype f rose from 1% to 17% with overall mortality of 30% among adults and 21% among children (Urwin *et al.*, 1996).

1.8.3. Infections associated with non-typeable strains

Non-capsulated or non-typeable *H. influenzae* strains are normally commensal organism in the human nasopharyxn and occupy this niche as their natural habitat (Howard *et al.*, 1988; Lerman *et al.*, 1979). Their principle association is with otitis media in children. They are found in the middle ear fluid of over a third of cases with otitis media undergoing tympanocentesis. They may also cause acute or chronic sinusitis in patients of all ages, and occasionally, causes meningitis in adults (Makela, 1988). Another disease with which NTHI are associated is chronic bronchitis. They are commonly isolated from

the bronchial secretions of such cases. The bacteria can persist in lower respiratory tract of these patients for months and appear to escape immunological defence because the major outer membrane protein changes during persistence (Groeneveld *et al.*, 1988).

1.9. Control and prevention of *H. influenzae* infection

1.9.1 Host defences

Exposure to the organism induces the host to produce antibodies against the capsule and antigens. These antibodies are both bactericidal and promote phagocyctosis and are highly effective in preventing infection. During the first six months of life, infants are protected against *H. influenzae* infection due to maternal IgG antibodies. However, children between six-months and five years are most susceptible to *H. influenzae* type b infection. Healthy children above the age of 5 years are rarely infected. The presumed reason for this age distribution is the acquisition of humoral immunity to *H. influenzae* type b with increasing age. This protection is most likely acquired as a result of sub-clinical infection or asymptomatic colonisation. It has been reported that, in addition to acquired humoral immunity involving bactericidal antibody, innate immunity mediated by C-reactive protein together with complement is sufficient to kill *H. influenzae* (Weiser *et al.*, 1998).

1.9.2. Immunisation

Outbreaks of serious infection due to *H. influenzae* type b can be prevented by vaccination or prophylactic therapy. In 1985, the first *H. influenzae* type b vaccine was prepared from purified PRP capsular material of the bacteria (Anderson *et al.*, 1977). This PRP vaccine consists of type b capsular polysaccharide that elicits a strong primary antibody response. This vaccine, however, was poorly immunogenic in young children

and did not provide long-term immunity. It was not efficacious in children under 18 months of age (Peltola *et al.*, 1977).

In 1991, a polysaccharide-diptheria toxoid (PRP-T) conjugate *H. influenzae* type b vaccine was developed. This vaccine consists of PRP conjugated to a tetanus toxoid protein, which induces persistent antibody level in children and proved to be more effective than the earlier vaccine (Eskola *et al.*, 1987).

Routine immunisation was introduced in the UK in 1992. With the implementation of the vaccination programme, the incidence of invasive disease declined dramatically and remained at very low rates until 1998 (Heath and McVernon, 2002). Despite the 93% immunisation uptake rate, the *H. influenzae* type b continues to be a public-health threat in the UK (Garner and Weston, 2003; Pushparajah *et al.*, 2003). In 1999, *H. influenzae* type b disease had started to increase again in children, with the greatest number of cases occurring in vaccinated children (Fig. 1.3) (http://www.hpa.org.uk/infections) (Ramsay *et al.*, 2003). The reduction in vaccine effectiveness partly coincided with the introduction to the UK of a new vaccine (combined Hib and acellular pertussis) (McVernon *et al.*, 2003). Factors such as vaccine uptake rate in immigrant populations, effectiveness of conjugate vaccines, effect on individual immunity and immunological memory, carriage rates, and herd immunity could have contributed to the immunisation failure (Heath and McVernon, 2002).

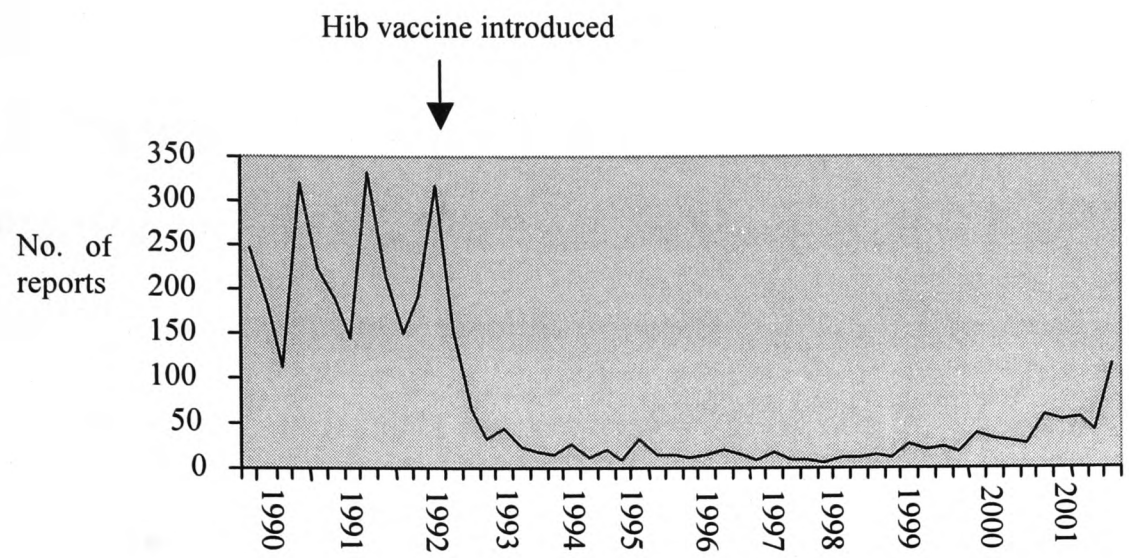


Fig. 1.3. Prevalence of *H. influenzae* type b reported to Communicable Disease Surveillance Centre (CDSC) for England and Wales, 1990-2002.

1.10. Antibiotic resistance in *Haemophilus* spp.

1.10.1. Changing trends in antimicrobial susceptibility

In many countries, over-prescription of antibiotics and increasing use of disinfectants for routine hygiene and heavy use of antibiotics for veterinary medication and/or growth promotion in agriculture, aquaculture and animal husbandry have been shown to contribute to the selection pressure for antibiotic resistance bacteria (Guillemot, 1999; Okeke and Edelman, 2001; Watson *et al.*, 1999).

It is assumed that as a result of increased exposure to antibiotic for treatment of respiratory infections (mainly otitis media), *H. influenzae* is exhibiting increased antibiotic resistance (Fallon, 1982; Gessner, 2002; Philpott-Howard and Williams, 1982; Williams and Moosdeen, 1986). Resistance to chloramphenicol and ampicillin amongst *H. influenzae* isolates has risen since the 1970's to the extent that neither antibiotic can be recommended as a first-line therapeutic agent for meningitis (de Groot *et al.*, 1991; Jorgensen, 1992). Fortunately, *H. influenzae* is almost uniformly susceptible to third

generation cephalosporins, but there is always a possibility that it will acquire broader antibiotic resistance (Cohen, 1992).

1.10.2. Prevalence of antibiotic resistant H. influenzae

The increasing prevalence of antibiotic resistance in *H. influenzae* is recognised worldwide. The first report of ampicillin resistant in *H. influenzae* was in 1972 (Mathies, 1972). Between 1977 and 1981, the prevalence of ampicillin resistance increased from 1.6% to 6.6% (Philpott-Howard and Williams, 1982). However, over the past decade, the level of resistance to ampicillin has not changed significantly (PHLS, 2002). Up to 2000, ampicillin was the most common antibiotic to which *H. influenzae* type b is resistant (Table 1.2).

Subsequent reports showed emergence of multi-drug resistant *H. influenzae* isolates. Clinical isolates *H. influenzae* resistant to a variety of antimicrobial agents such as ampicillin, chloramphenicol, tetracycline and cotrimoxazole has been reported in many countries (Baquero, 1996; Jorgensen, 1992).

Table 1.2. Antimicrobial resistance in invasive *H. influenzae* type b infection isolated in England and Wales.

Year	n	Resistant (%)					β-lactamase +ve		
		Amp[a]	(%)	Chlor[b]	(%)	Trim[c]	(%)	n	(%)
1992	610	118	19%	11	2%	20	3%	117	19%
1993	252	51	20%	6	2%	9	4%	50	20%
1994	84	25	30%	0	0%	0	0%	23	27%
1995	77	10	13%	1	1%	4	5%	10	13%
1995	62	14	23%	2	3%	2	3%	14	23%
1997	69	13	19%	2	3%	2	3%	13	19%
1998	50	19	38%	0	0%	2	4%	19	38%
1999	68	14	21%	1	1%	10	15%	13	19%
2000	113	22	19%	1	1%	5	4%	22	19%

[a]ampicillin, [b]chloramphenicol, [c]trimethoprim. Source: Haemophilus Reference Unit

1.10.3. Medical impact of antibiotic resistance

Antibiotic resistance is increasingly regarded as a major problem complicating the management of *H. influenzae* infections. Infections caused by resistant microbes respond poorly to treatment resulting in an increase in morbidity and even mortality of patients. Antibiotic resistance also results in an increase in the cost of medical care (e.g. resulting from prolonged hospitalisations and the use of more expensive antibiotics).

1.11. Detection of antibiotic resistance

Antibiotic resistance in bacteria is most commonly detected by standard laboratory investigations to establish the best choice of treatment. Some of the methods used for detection of antibiotic resistance include: minimal inhibition concentration (MIC) (Barry, 1986); E-Strip test (AB Biodisk, Solna, Sweden); nitrocefin chromogenic cephalosporin test for detection of β-lactamase (Boughton, 1982); iodometric detection (Catlin, 1975), penicillinase paper strip (Jorgensen *et al.*, 1977); chloramphenicol CAT assay; and acidometric method (Scheifele *et al.*, 1976). Besides these methods, PCR detection of β-lactamase provides an important supplement or confirmation to phenotypic methods (Tenover *et al.*, 1994).

1.12. Mechanisms of antibiotic resistance

Generally, antibiotic resistance mechanisms in bacteria can be subdivided as follows:

a. Loss or modification of site of action (e.g. point mutation)

b. Efflux pump (e.g. *tetA*)

c. Enzymatic inactivation of antibiotics (e.g. β-lactamase).

In multi-drug resistance *Haemophilus* spp. strains, more than one of the mechanisms could be present in a single bacterium.

1.13. Ampicillin resistance in *H. influenzae*

Ampicillin resistance in *H. influenzae* is could be due to either one of these mechanisms: plasmid-mediated production of TEM β-lactamase, altered penicillin-binding proteins, and/or diminished permeability (de Groot *et al.*, 1991). The predominant mechanism for *H. influenzae* resistance to ampicillin in clinical isolates is clearly due to β-lactamase production (Gutmann *et al.*, 1988; Heffron *et al.*, 1975).

1.13.1. Types of β-lactamases in H. influenzae

β-lactam antibiotics include penicillin, ampicillin and cephalosporins. The enzyme β-lactamase produced by *H. influenzae* hydrolyses penicillin antibiotics by cleaving one bond of the four-membered ring of the β-lactams. The open-ringed structure of the penicillin is incorporated into the cell wall but sufficiently different from the normal pentapeptide that transpeptidase does not recognise it as a substrate. The cleavage product of penicillin by β-lactamase is different enough that it is not used as a substrate in cell wall synthesis. Hence, the cell is not weaken by the antibiotic thus allowing normal growth (Quintiliani and Courvalin, 1995).

Two types of β-lactamases have been identified in *H. influenzae*; TEM-1 β-lactamase and ROB-1 β-lactamase (Gutmann *et al.*, 1988). ROB-1 β-lactamase produces a plasmid-determined enzyme with broad-spectrum activity like that of TEM-1 but with an isoelectric point (PI) of 8.1 (quite different from the TEM-1 pI of 5.4) (Rubin *et al.*, 1981). These two β-lactamases share limited amino acid sequence homology and have similar substrate profiles and relative rates of ampicillin hydrolysis (Bush, 1989). Ampicillin resistant *H. influenzae* generally produce either one of the two β-lactamases, but rare isolates produce both β-lactamases (Daum *et al.*, 1988; Scriver *et al.*, 1994).

Although both of the genes code for ampicillin-resistance, TEM-1 and ROB-1 enzymes are thought to be unrelated.

1.13.2. TEM β-lactamase

There are 43 naturally occurring TEM β-lactamase derivatives (Bush and Jacoby, 1997). The TEM-1 β-lactamase is the most commonly found in *H. influenzae* (Elwell *et al.*, 1975). TEM-1 β-lactamase is encoded by the *bla* gene. The structural gene for the β-lactamase was considered to reside on a plasmid genome (Heffron *et al.*, 1975; Laufs and Kaulfers, 1977). A plasmid of *H. influenzae* strain G32 (RSF007) was found to share about 48% with the TEM-type β-lactamase transposon, *TnA* (Elwell, 1994).

1.13.3. ROB β-lactamase

Most of the ampicillin-resistant *H. influenzae* isolated in the USA were found to possess ROB-1 enzymes (Daum *et al.*, 1988). ROB-1 enzymes were believed to have originated from *Actinobacillus pleuropneumoniae*, a pathogen that is frequently isolated from pigs (Juteau *et al.*, 1991; Livrelli *et al.*, 1988; Medeiros *et al.*, 1986). The ROB-1 enzyme is most commonly encoded by small non-conjugative R plasmids prevalent in *Enterobacteriaceae* (Brunton *et al.*, 1986; Medeiros and O'Brien, 1975). Plasmids with ROB-1 enzymes in *Pasteurella haemolytica* isolated from cattle and sheep is believed to have originated from *H. influenzae* (Azad *et al.*, 1992).

1.14. Tetracycline resistance in *H. influenzae*

Tetracycline resistance in *Haemophilus* spp. was initially reported in 1975 (Van *et al.*, 1975). Since then, numerous *Haemophilus* spp. with this resistance have been isolated (Kaulfers *et al.*, 1978; Powell *et al.*, 1992). Presently, tetracycline resistant *H. influenzae*

is commonly isolated. In UK, it ranks second (2.7%) as the most frequent antibiotic resistance after ampicillin (7.8%) (Powell *et al.*, 1987). In several of these resistant strains, it was found that the resistance appeared to reside on transposons, Tn*D* (now recognised as Tn*10*) on a resident plasmid (Elwell *et al.*, 1977, Kaulfers *et al.*, 1978).

Resistance to tetracycline often occurs as a result of energy-dependent pumping of the antibiotic from the cell so that the levels of antibiotics are decreased and the ribosome is not inhibited (Speer *et al.*, 1992). Antibiotic efflux is mediated by the Tet membrane proteins, which use an antiport mechanism of transport involving the exchange of a proton for a tetracycline-cation complex (McMurry *et al.*, 1980).

1.15. Chloramphenicol resistance in *H. influenzae*

Resistance to chloramphenicol was first reported in 1972 (Barrett *et al.*, 1972). Since then, increasing numbers of *H. influenzae* type b are resistant to chloramphenicol have been reported (Kinmonth *et al.*, 1978; Simasathien *et al.*, 1980; Uchiyama *et al.*, 1980). In England and Wales, 1% of the *H. influenzae* type b isolated in 2000 was resistant to chloramphenicol (Table 1.2).

Most chloramphenicol resistance in *H. influenzae* is associated with chloramphenicol acetyltransferase production which enzymatically inactive the antibiotics (Powell and Livermore, 1988; Roberts *et al.*, 1980). Although several investigators believed that chloramphenicol acetyltransferase, *cat* gene was plasmid-borne (Roberts *et al.*, 1980; Smith and Kelsey, 1989; van Klingeren *et al.*, 1977), others believed that the resistance genes were integrated into the chromosome (Mendelman *et al.*, 1984, Powell *et al.*, 1988, Stuy, 1980b).

1.16. Resistance to other antimicrobials

1.16.1. Kanamycin resistance

To date, *H. influenzae* isolates resistant to kanamycin alone have not been reported. All the kanamycin resistant *H. influenzae* were found in association with ampicillin-chloramphenicol or ampicillin-chloramphenicol-tetracycline resistance. Resistance to these antibiotics was encoded by a 37 to 44-MDa conjugative plasmid (Levy *et al.*, 1993).

1.16.2. Trimethoprim resistance

H. influenzae resistant to trimethoprim-sulfamethoxazole have also been reported to occur worldwide (Sahm *et al.*, 2000). Resistance to trimethoprim was shown to be due to both permeability changes and to a mutation in the chromosomal dihydrofolate reductase gene (de Groot *et al.*, 1988).

1.16.3. Multi-drug resistance

Bacteria can develop resistance to many antibiotics. Multi-resistance may develop by chromosomally located resistance determinants or mutations in a resident gene; however, it may also develop through the acquisition of resistance genes or an array of resistance genes by horizontal transfer (Rowe-Magnus and Mazel, 1999).

In multi-resistant *H. influenzae* isolates, especially those resistant to both tetracycline and chloramphenicol, resistance epitopes were found to be encoded on a single plasmid and could frequently transfer as a single unit (Mendelman *et al.*, 1984; Roberts *et al.*, 1980; van Klingeren *et al.*, 1977).

1.17. Mobile genetic elements

Bacteria acquire and transfer genetic information as an important mechanism to survive in hostile environments including those containing antibiotics. The acquisition of resistance genes generally involves transfer on mobile elements such as transposons, integrons and plasmids (Gomez-Lus, 1998; Jahn *et al.*, 1979; van Klingeren *et al.*, 1977). These genetic mobile elements are not only able to transfer genetic information to and from within the chromosome, but they are also able to exchange information with other self-replicating genetic elements within the cell (plasmids and phages) (Dionisio *et al.*, 2002). This exchange of information can also occur between strains and species and involves more than just resistance genes (Levy *et al.*, 1993). These elements play important roles in the bacterial evolution, providing mechanisms to generate diversity and for the dissemination of DNA among other bacteria (Gomez-Lus, 1998; Ochman *et al.*, 2000).

1.18. Transposon mediated resistance

Transposons, or 'jumping genes' are known to be responsible for the dissemination of genes encoding resistance to antibiotics and heavy metals and genes that degrade chemicals (Osborn and Boltner, 2002). Transposons are characterised by the presence of short terminal sequences (10-50 bp) at their two ends. These usually form part of flanking repeat sequences known as insertion elements (IS). The size of IS elements usually range between 800 to 1800 bp. Generally, IS elements do not encode any function other than a transposase (*Tnp*) gene (Chandler and Mahillon, 2002). Transposase recognises inverted tandem repeats (ITRs), and uses this for transposition. They can transposase/integrate into numerous non-homologous sequences of DNA, and are capable of inserting at multiple sites in a target DNA. They can also move from one plasmid to another, and in both directions between plasmids and chromosomes in a manner that does not require the

recombination functions of the bacteria. Many transposition events cumulatively result in resistance plasmids (Gomez-Lus, 1998).

Transposons can either be composite mobile elements or non-composite mobile elements. Composite transposons contain other gene(s) for survival and adaptation (e.g. antibiotic resistance genes) that is (are) flanked by IS elements. The IS elements provides transposase actively, although one of the copies is often inactive. An example of composite transposons is Tn*10* (Chalmers *et al.*, 2000). The non-composite transposons, such as Tn*3* do not have insertion sequences at their extremities but only consist of the inverted repeats necessary for recognition by the transposase. All the activities required for transposition, including the transposase and resolvase, are encoded by sequences in the element (West and Tang, 2000).

1.18.1. Tn10

Tn*10* is a Class I, 9,147-kb composite transposon, consisting of a central element of about 6,700 bp of non-repeated sequences that specifies resistance to tetracycline. Both ends of the transposon have inverted repeats of 1,329-bp representing the insertion sequence, IS*10* (Fig. 1.4) (Chalmers *et al.*, 2000; Lawley *et al.*, 2000). The IS*10*-right is a fully functional insertion that can transpose as an individual unit but the IS*10*-left is inactive because it specifies a non-functional transposase protein (Foster *et al.*, 1981; Kleckner, 1989).

To date, there are only two genes involved in expressing resistance. They are *tetA* and *tetR*. The *tetA* is a membrane-bound efflux pump and *tetR* functions as its repressor (Hillen and Berens, 1994). The Tn*10* carries other additional structural genes, *tetC* and *tetD* of unknown functions (Kleckner, 1989). The *tetD* gene product is membrane-associated and inducible by tetracycline, however, no role for *tetC* is known.

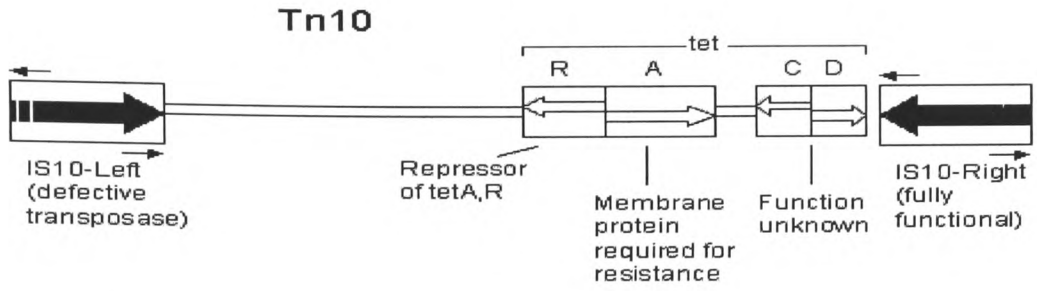


Fig. 1.4. Transposon Tn*10*. IS*10*-left and IS*10*-right flank the *tetR*, *tetA*, *tetC*, and *tetD* genes. The short arrows at the termini of the IS*10* elements indicate nearly perfect inverted repeat sequences. Open arrows within the *tet* genes indicate the direction of transcription; the solid arrow within IS*10*-Right indicates transcription of the transposase gene; the broken arrow within IS*10*-left indicates transcription of the defective transposase genes of this element. This figure was adapted from Mobile DNA (Kleckner, 1989).

1.18.2. Tn3

An example of a transposon commonly found in antibiotic resistance bacteria is the ampicillin resistance transposon, Tn*3*. It is 4,957 kb in size and contains two genes for transposition and one for ampicillin resistance, TEM-1 β-lactamase gene (Roy, 1999). Transposon Tn*3*, is a non-composite transposon that does not have insertion sequences at either extremity, but it has the inverted repeats necessary for recognition by transposase (Fig. 1.5). The inverted repeats are generally 38-bp long, with an outer sequence starting with GGGG and terminating internally with TAAG (Chandler and Mahillon, 2002; Takeya *et al.*, 1979). In this thesis, the features of Tn*3* and Tn*10* present on mobile elements of *H. influenzae* will be considered further in the results chapters.

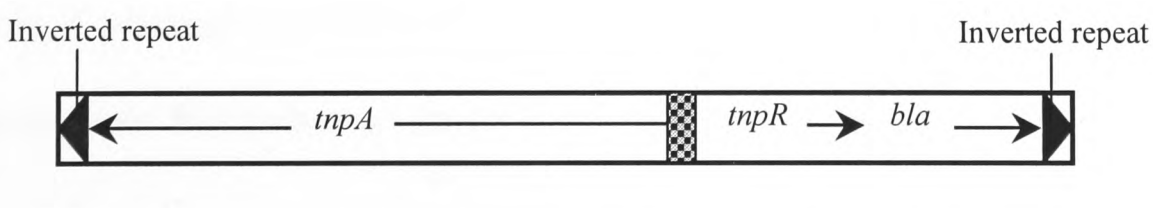


Fig. 1.5. Genetic organisation of Tn*3* family transposons. *tnpA* gene is the transposase and *tnpR* gene is the repressor and resolvase. The crosshatched box is the *res* site.

1.19. Conjugative and non-conjugative plasmids

Plasmids are autonomously replicating extra-chromosomal DNA elements, distinct from the normal bacterial genome and nonessential for cell survival under non-selective conditions (Hardy, 1986). Plasmids play important roles in bacterial survival and diversity. In order to continue their existence and adapt to new environments, a plasmid will have to acquire the ability to survive and propagate, which includes replication, partitioning, post-segregational killing and conjugative transfer as shown in Fig. 1.6 (Thomas, 2000).

In general, bacterial plasmid replicate independently of the host chromosome, although usually they rely on some host-encoded factors for their replication. Plasmid partition systems, one of the maintenance mechanisms, are normally plasmid encoded, although host cellular functions may be required (Austin, 1988; Williams and Thomas, 1992). The mechanism of partition apparently involves the recognition of pairs of plasmid molecules and membrane proteins followed by an active process of separation of the members of the pair to the opposite halves of the dividing bacterial cell (Firshein and Kim, 1997; Thomas, 2000). Unlike in plasmid, the regions of origin of the daughter nucleoids were anchored to the cell membrane at a central position of the cell (Draper and Gober, 2002).

Conjugative plasmid refers to a self-transmissible plasmid that carries genes that can initiate conjugation with another bacterium. Examples of conjugative plasmids include, F plasmid and RSF1010 (Zechner *et al.*, 2000), a 50-kb pRE25, multi-resistance plasmid from *Enterococcus faecalis* RE25 (Schwarz *et al.*, 2001) and a 32-kb pF3028, isolated from *H. aegyptius* (Kroll *et al.*, 2002). These plasmids carry genes for replication, conjugation (*tra* genes) and also genes for mobilization of smaller plasmids. They do not have genes for integration into host chromosome. In contrast, pBtoxis, a toxin-coding

megaplasmid (127-kb) of *Bacillus thuringiensis* subsp. *israelensis* has genes for replication but does not have *tra* genes (Berry *et al.*, 2002). This is one example of a plasmid that is non-conjugative.

Conjugative plasmids should not be confused with conjugative elements because conjugative elements are not able to replicate autonomously even though they can integrate into the host's chromosome (Salyers *et al.*, 1995). One example is CTnDOT, a conjugative transposon from *Bacteriodes sp.*, which self-transfers, integrate and excise using recombinase genes not related to known conjugation plasmids (Bonheyo *et al.*, 2001; Cheng *et al.*, 2000). Another such element is the SXT element found in *Vibrio cholerae*. SXT is different from CTnDOT because of their unrelated conjugative transfer genes. In contrast to conjugative plasmids, the genome of SXT did not contain genes for autonomous replication or segregation, but its genes for the chromosomal recombination resembles those of plasmids (Beaber *et al.*, 2002; Hochhut and Waldor, 1999).

A group of elements that excise and integrate like prophages, and transfer by conjugation like plasmids, has been proposed and is the integrative and conjugative elements (ICEs) (Burrus *et al.*, 2002a). These elements are characterised by their ability to excise by site-specific recombination, transfer by conjugation and recombine into their host's chromosome at a specific site. An example is a 35-kb ICE*St1*, an element of *Streptococcus thermophilus*, which encodes integrative and conjugative functions (Burrus *et al.*, 2002b). Since nomenclature for ICEs has only recently been suggested, some ICEs were initially described as plasmids or transposons and many such elements have not been renamed.

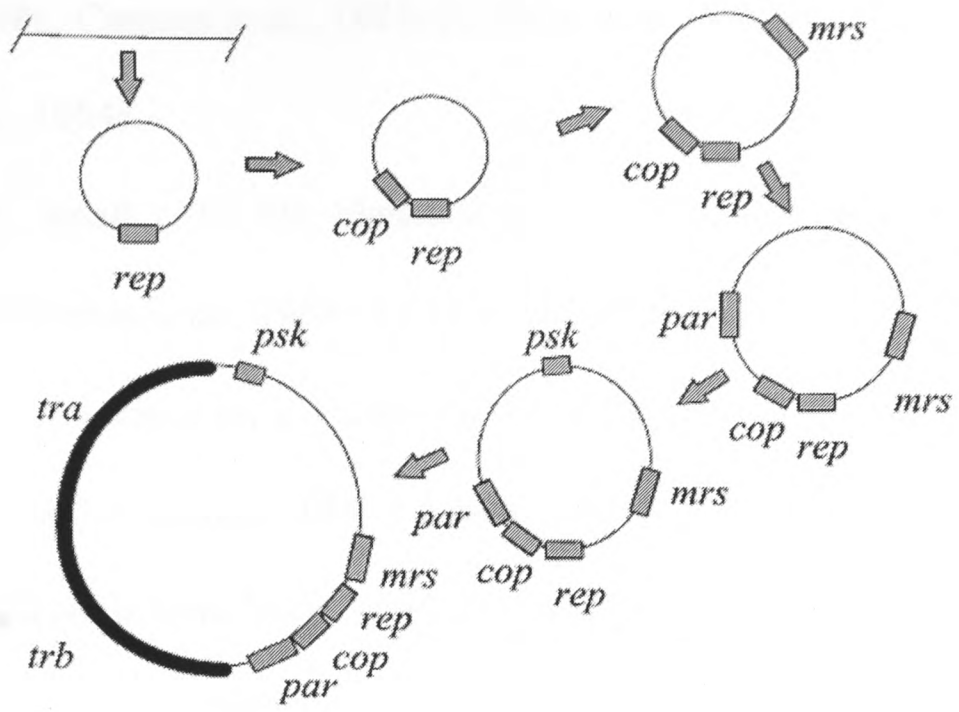


Fig. 1.6. Schematic view of the development of survival functions of a plasmid. The birth of the plasmid is shown as the appearance of a self-replicating circle. Subsequently, gene blocks with specific functions are first recruited by the plasmid through recombination with a range of disparate plasmids and then shuffled by recombination, with fitter combinations being constantly selected from the pool. Gene names are: *rep*, replication; *cop*, copy number control; *mrs*, multimer resolution; *par*, partitioning; *psk*, post-segregational killing; *tra*, DNA processing genes for transfer; *trb*, mating-pair formation genes for transfer (Thomas, 2000).

1.20. Resistance plasmids (elements) in *H. influenzae*

The sudden emergence of ampicillin resistance mediated by β-lactamase in *H. influenzae* was encoded by large conjugative plasmids (Elwell *et al.*, 1975; Elwell *et al.*, 1977). Ampicillin-sensitive Haemophilus strains, by contrast, did not contain plasmid DNA (Elwell *et al.*, 1975; Roberts and Smith, 1980). Similar sized plasmids were also found in tetracycline resistant *H. influenzae* (Jahn *et al.*, 1979; Marshall *et al.*, 1984).

Molecular studies performed on resistant strains of *H. influenzae* showed that antibiotic resistance is usually encoded by genes located on one of two types of plasmids (de Graaff *et al.*, 1976). Large plasmids (>45 kb) were responsible for carrying antibiotic resistance genes in *H. influenzae* (Elwell, 1994). Some large plasmids also encoded genes for multi-drug resistance. Besides β-lactamase that encoded ampicillin resistance, they

carried genes for resistance to tetracycline and chloramphenicol (Brightman *et al.*, 1990; Brunton *et al.*, 1986; Campos *et al.*, 1989; de Groot *et al.*, 1991; Dimopoulou *et al.*, 1992; Mendelman *et al.*, 1984).

Generally, small (<10 kb) plasmids in *H. influenzae* only code for TEM β-lactamase gene (Brunton *et al.*, 1986). Laufs *et al* (1979) found that the structural gene for the β-lactamase was located on a 4.4-kb plasmid (Laufs *et al.*, 1979), while another 4.4-kb plasmid was found to harbour ROB-1 ß-lactamase (Rubin *et al.*, 1981).

Cryptic plasmids have been found in some Haemophilus strains. The function of these plasmids is still unknown. Some researchers believe that resistance plasmids in *H. influenzae* were originally cryptic plasmids that have acquired resistance genes from *H. parainfluenzae* (Laufs *et al.*, 1981).

1.21. Transfer of plasmids

Horizontal transfer of resistance genes is a successful mechanism for the transmission and dissemination of antibiotic resistance among bacteria including pathogens. Three major mechanisms, transformation, transduction and conjugation have been recognised (Davison, 1999).

1.21.1. Transfer by transformation

Transformation involves the direct uptake of exogenous DNA. *H. influenzae* has an efficient natural competence system whereby the cells are able to take up DNA from the environment without the direct interaction of the donor and recipient cells due to the presence of uptake sequences (Danner *et al.*, 1980; Goodgal and Mitchell, 1990; Smith *et al.*, 1999). Transformation was not thought to be not the mechanism for transfer of

antibiotic resistance determinants, because the *H. influenzae* was Rec independent and was resistant to DNAse (Dimopoulou *et al.*, 1992; Saunders and Sykes, 1977).

1.21.2. Transfer by conjugation

Conjugative transfer is the process whereby a DNA molecule is transferred from a donor to a recipient bacterium via a specialised protein complex (Zechner *et al.*, 2000). This process usually requires close physical contact between donor and recipient bacteria, which is usually mediated by plasmids and transposons.

The transfer of antibiotic resistance genes in *Haemophilus* spp. is through conjugation of large plasmids (elements) that are chromosomally integrated (Farrar and O'Dell, 1974; Laufs and Kaulfers, 1977; Stuy, 1980b). Thorne and Farrar (1975) showed that *H. influenzae* transferred an ampicillin resistance determinant to a recipient cell through the process of conjugation (Thorne and Farrar, 1975). Transfer of resistance plasmids through conjugation requires cell-to-cell contact but does not occur in suspension (Stuy, 1980b). Dimopoulou *et al.* (1992) confirmed that transfer of resistance determinants between *H. influenzae* strains occurred by conjugation rather than by transformation (Dimopoulou *et al.*, 1992). Apart from ampicillin, chloramphenicol, tetracycline and kanamycin resistance, genes encoding resistance to other antibiotics (e.g. trimethoprim) could not be transferred by conjugation (Dimopoulou *et al.*, 1992; Elwell *et al.*, 1977; van Klingeren *et al.*, 1977).

Transfer of plasmids between different strains of *H. influenzae* and *H. parainfluenzae* by conjugation have been also reported. Beta-lactamase specifying strains of *H. parainfluenzae* can conjugally transfer their plasmids to *H. influenzae* recipients (Scheifele and Fussell, 1981a; Scheifele *et al.*, 1982). As a result of this observation, Scheifele and Fussell (1986) postulated that *H. parainfluenzae* could play a role as a

vector in the spread of resistance genes to *H. influenzae* (Scheifele and Fussell, 1986). Leaves *et al.* (2000) found that the sequences detectable by PCR for *β-lac*$^+$ amp^r in *H. parainfluenzae* were similar to a family of large plasmids in *H. influenzae* and is consistent with the idea that *H. influenzae* plasmids with *β-lac*$^+$ amp^r have originated from the population of *H. parainfluenzae* (Leaves *et al.*, 2000).

Conjugative plasmid transfer between strains of the same serotype was efficient, but was less efficient when transfer was between different serotypes. According to Stuy (1980), large resistance plasmids transfer less efficiently (~10^{-5}) between *H. influenzae* strains, but transfer between different serotypes are even more inefficiently (Stuy, 1979, 1980b). He reported that conjugation efficiency was unaffected by whether the strains are *recA*$^-$ or *recA*$^+$. Presence of host capsule and the types of restriction enzymes in the host also do not affect conjugation efficiency (Stuy, 1979).

1.22.3. Transfer by transduction

Transduction is the process of DNA transfer mediated by bacteriophage. Fragments of bacterial chromosome are incorporated into phage heads and are transduced into new host cell. The transfer of antibiotic resistance genes by transduction in *Haemophilus* spp. has not been shown. Furthermore, the presence of an efficient conjugative system for transferring resistance genes makes transduction an improbable means of explaining antibiotic resistance gene transfer (Saunders and Sykes, 1977).

1.23. Restriction and modification system

Regardless of how DNA enters a recipient cell, once inside the cell, the foreign DNA may either survives (i.e. replicate and/or recombine with the recipient DNA) or, it may be degraded by restriction endonuclease enzymes. If the DNA survives, it is due to the

modification activity of the donor cell accomplished by sequence specific methylase enzymes. In any cell with a restriction-modification system, both the restriction and modification enzymes have the same sequence specificity (Lewin, 1990).

In *H. influenzae* strain Rd, methylase II and methylase III were found to protect *H. parainfluenzae* transforming DNA from inactivation by *H. influenzae* endonuclease enzymes (Roszczyk and Goodgal, 1975) which suggests that DNA of *H. parainfluenzae* could be transferred into *H. influenzae* without being cleaved by *Hin*d II and *Hin*d III.

1.24. Site-specific recombination of *H. influenzae* plasmids

Plasmids in *H. influenzae* are found integrated in the host's chromosome (Murphey-Corb *et al.*, 1984; Stuy, 1980a). Therefore, usual plasmid isolation procedures are not able to detect extra-chromosomal DNA from clinical isolates. However, following conjugative transfer, integrated plasmids in clinical isolates could be readily detected as closed circular plasmid in recipient (Dimopoulou *et al.*, 1992; Overturf *et al.*, 1987).

Dimopoulou *et al.* have found common sequence (66-bases) structures on the plasmid/element (*attP*) and on host chromosome (*attB*), indicative of site-specific integration (Dimopoulou *et al.*, 2002), see Fig. 1.7 (Leaves *et al.*, 2000). The *attP* of the closed circular element was the point of re-circularisation. Once integrated on the chromosome, the repeat sequences were located at the left (*attL*) and right (*attR*) junctions of the insert. These 66-base sequences were also identical to the 3' of $tRNA^{leu}$, which suggest that the plasmid was integrated on the chromosome at the $tRNA^{leu}$. These *att* sites are shared *attP* of HP1 bacteriophage of *H. influenzae* (Dimopoulou *et al.*, 2002). The conservation of *att* sites between the bacterium, plasmid and bacteriophage suggests that they possibly recombine site-specifically by similar mechanisms, which require an

integrase (*int*). DNA binding protein (IHF) contributes to recombination by promoting the binding of *int* to the *att* sites (Astumian *et al.*, 1989; Craig, 1988).

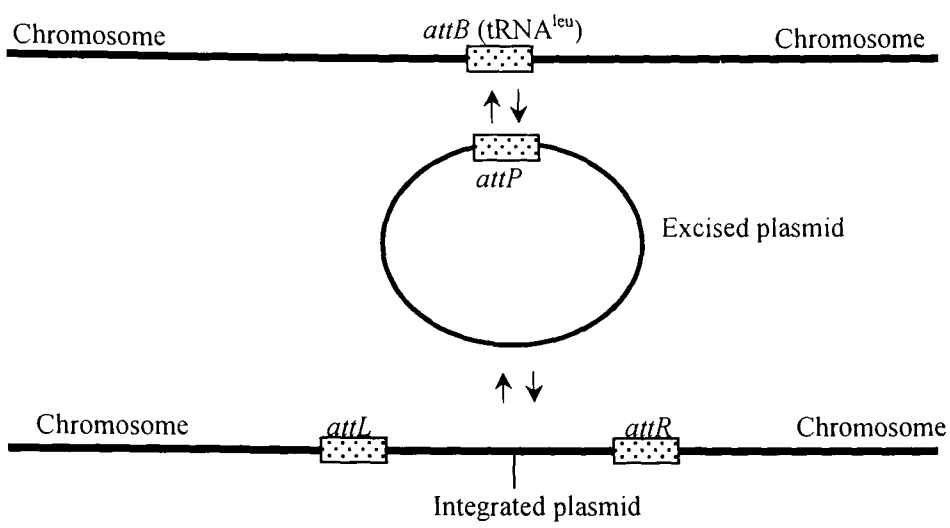


Fig. 1.7. Plasmid and chromosomal PCR primer attachment sites. The orientation at their respective sites on the chromosomes, closed circular (excise) plasmid and plasmid recombined (integrated) with the chromosome; *attB*, chromosomal attachment site; *attP*, plasmid attachment site; *attL*, left plasmid chromosomal junction; *attR*, right plasmid chromosomal junction (taken from Leaves *et al.* (2000) (Leaves *et al.*, 2000).

1.25. Excisive recombination of plasmid

Excisive recombination of an integrated plasmid is likely to involve recombination between *attL* and *attR* to generate intact *attP* and *attB* sites. The excisive site-specific recombination of the temperate bacteriophage, HP1, has been studied (Craig, 1988). Excisive recombination requires *int*, IHF and the Cox protein (Esposito and Scocca, 1994, 1997a). The Cox protein is similar to the Xis protein of *E. coli* phage P2, inhibits the integrative recombination of *attP* and *attB* *in vitro*, and activates the excisive recombination between *attL* and *attR* (Esposito and Scocca, 1994).

1.26. Origin of resistance genes and conjugative plasmids

Several investigators suggested that large conjugative plasmids/elements encoding antibiotic resistance arose as a result of the transposition of resistance genes onto closely related cryptic plasmids/elements that had a common evolutionary origin (Elwell *et al.*,

1975; Laufs *et al.*, 1981; Saunders and Sykes, 1977). Laufs *et al.* (1981) speculated that the multi-resistance plasmids of *H. influenzae* that have emerged in different parts of the world were of multi-clonal evolutionary origin (Laufs *et al.*, 1981).

The genetic relatedness of multi-resistance plasmids (ampicillin-chloramphenicol-tetracycline-kanamycin) of *H. influenzae* isolated in Belgium and Spain was investigated (Levy *et al.*, 1993). Homoduplex analysis showed that they had similar molecular organisation and identical structures, indicating that plasmids in epidemiologically unrelated strains shared a high proportion of shared core sequences. Ampicillin resistant plasmids isolated in Alaska, though different from ampicillin resistant plasmids in the United States, have nevertheless been reported to share significant DNA homology (Mendelman *et al.*, 1985). These observations led to a suggestion that a common genetic element was responsible for the dissemination of this type of resistance in Alaska.

1.27. Role of plasmids in gene transfer and bacterial evolution

Mobile elements, particularly plasmids, have been shown to play important roles in horizontal (or lateral) gene transfer and bacterial evolution. When bacterial genes are transferred between different organisms, chromosomal genes may also be transferred in addition to those forming the mobile elements. Most mobile elements also carry genes that confer potentially useful traits on their bacterial hosts. They rarely carry genes that are absolutely essential for the host bacteria, but frequently carry genes that could improve the 'fitness' of the host under certain circumstances (van Elsas *et al.*, 2000).

Some plasmids are highly promiscuous and have broad host ranges. This characteristic assists the exchange of genes between different genera or even transfer to different kingdoms (Dionisio *et al.*, 2002; Groisman *et al.*, 1992). They mediate gene transfer between organisms dwelling in various environments including plants, (Arnold *et*

al., 2003), soil (Rafii and Crawford, 1988) and marine environments (Dahlberg *et al.*, 1997).

1.28. Genomic islands

Genetic elements referred to as genomic islands (GIs) that are capable of horizontal transfer and integration into host chromosomes are key agents in the prokaryotic genome evolution. Pathogenicity islands (PAIs) are specifically referred to as a class of GIs carrying specific gene clusters contributing to a virulence phenotype, e.g. adherence factors, toxins, iron uptake systems, Type III and Type IV secretion systems (Dobrindt *et al.*, 2004; Hacker and Kaper, 2000; Hentschel and Hacker, 2001). PAIs are usually found on genomes of pathogenic strains or variants, but are absent or rarely present in non-pathogenic strains (Hacker *et al.*, 1997). An example is the DNA cassette encoding Vi capsular polysaccharide found on a 134-kb PAI (SPI-7) of *Salmonella enterica* serovar Typhi (Hacker and Kaper, 2002; Pickard *et al.*, 2003).

Another type of GI observed in prokaryotic genomes is a symbiosis island. This refers to a GI that encodes genes essential for symbiotic interactions with the eukaryotic host. An example is a 502-kb symbiosis island from *Mesorhizobium loti* strain R7A that was introduced in other *M. loti* strains from nodulated Lotus plants (Sullivan *et al.*, 2002).

1.29. Integration sites for GIs

GIs (including PAIs and symbiosis islands) integrate into chromosomes at a range of specific sites of the tRNA loci (Hacker *et al.*, 1997; Hacker and Kaper, 2000). Hacker and Kaper (2002) hypothesized that the conserved nature of tRNAs across different species, and their occurrence in multiple copies within genomes offer reliable targets for integration of mobile elements (Hacker and Kaper, 2002). Most tRNA genes appear to be

targets for GIs integration. For example, the *clc* element integrates at $tRNA^{gly}$ (van der Meer and Sentchilo, 2003), SRL, the PAI of *Shigella flexneri* 2a integrates at $tRNA^{serX}$ (Turner *et al.*, 2001), and the three PAIs of *E. coli* 536 integrate at three different tRNA loci, i.e., $tRNA^{sel}$, $tRNA^{leu}$ and $tRNA^{asn}$ (Dobrindt *et al.*, 2002). The preference of tRNA loci is still unclear, but is believed that specific tRNA may have functions affecting virulence trait and metabolic activity of the GI (Hacker and Kaper, 2002; Reiter *et al.*, 1989).

The recombination of mobile elements such as plasmids and phages with chromosomal DNA is catalysed by site-specific recombinases or integrases, and accessory genes (e.g. excisase, IHF), which encodes their excision and integration activities. Two families of site-specific recombinases have been recognised: the tyrosine and serine recombinase (Stark *et al.*, 1992). The tyrosine recombinase consists of a diverse family of site-specific recombinases found widely distributed in bacteria, plasmids, GIs and bacteriophage. Members of this family usually account for recombination of mobile elements with tRNA genes or their flanking sequences (Esposito and Scocca, 1997b).

1.30. Detection of GIs

Horizontal transfer of genes in prokaryotes can result in the change of microbial genomic composition within a relatively short period (Ochman *et al.*, 2000). The impact could result in an introduction of as much foreign DNA accounting for as much as one-fifth of a bacterial genome (Ochman *et al.*, 2000). The insertion of foreign DNA into the genome can usually be easily recognised by an altered G+C% content that varies from the rest of bacterial genome. They are usually flanked by specific DNA sequences, such as direct repeats. These repeats often resemble the attachment sites for phage integrases and are

often located within tRNA operons (Hacker *et al.*, 1997). GIs also contain genes coding for genetic mobility such as phage genes, insertion sequence elements, integrases, transposases and origins of replication (Hacker *et al.*, 1997; Hacker and Kaper, 2000).

Several methods have been suggested for detecting genes that have recently undergone horizontal transfer. One of the standard methods for detecting GI is to compare two or more genomes of related bacteria; including the pathogenic and non-pathogenic bacteria (Tu and Ding, 2003). The gene clusters present only in pathogenic genomes are candidates for PAIs. Other methods of detection include, guanine-cytosine content (G+C%)(Kerr *et al.*, 1997; Lawrence and Ochman, 1998), codon biases between genomes and GIs (Grocock and Sharp, 2002; Karlin, 2001) and incongruities between phylogenetic trees (Daubin *et al.*, 2002; Daubin *et al.*, 2003).

A prerequisite to understanding the complete biology of mobile elements that allows them to evolve to adapt into new environments is by determination of its entire nucleotide sequence.

1.31. Historical background on sequencing

The 'Genomic Era' began when Fred Sanger and his colleagues first sequenced the whole genome of bacteriophage ϕX174 (5386 bp) in 1977 (Sanger *et al.*, 1977a). Followed soon thereafter by the sequencing of the complete sequence of bacteriophage λ (48,502 bp), cytomegalovirus (229 kb) and 121-kb of chloroplast genomes of *Marchantia polymorpha* (Oda *et al.*, 1992). With the advancement of high throughput equipment, the first genome sequenced using automated-DNA technology was the 186-kb variola (smallpox) virus in the early 1990s (Massung *et al.*, 1993). In 1995, The Institute for Genomic Research (TIGR; Rockville, USA) completed the sequence of *H. influenzae*, the first fully genome-sequenced free-living organism of 1.83 Mbp in length (Fleischmann *et al.*, 1995).

1.32. Importance of complete genomic sequences

Genomic studies are aimed at the acquisition of knowledge of structure and functions of genomes. The scope of genomic research is wide. It involves the determination of complete nucleotide sequence of an organism; the annotation of the sequences obtained; the analysis of the genetic organisation and the interaction of the gene functions; and the study of how and why genomes evolved the way they did (Lee and Lee, 2000). Some of the rationale of construction of genetic maps includes comparisons for taxonomic and evolutionary studies, location and manipulation of newly discovered genes and identification of homologous genes in other organisms.

Since the completion of the sequence of the whole genome of *H. influenzae* in 1995, many prokaryotic genomes have since been sequenced. The availability of the complete nucleotide sequences of a number of bacterial species has stimulated global bioinformatics and experimental approaches to understanding and identifying previously undiscovered functions (Kapfhammer *et al.*, 2002; McClelland *et al.*, 2001; Parkhill *et al.*, 2001).

The 'Genomic Era' has now evolved tremendously and is moving at a rapid pace. To date, genomes of more than 154 bacterial species have been deposited into the nucleotide databases for public assess. Amongst others, genome sequences of bacteria such as *Helicobacter pylori, Mycobacterium tuberculosis, Campylobacter jejuni, Neisseria meningitidis, Yersinia pestis, Bordetella pertusis, S. typhi* and *Streptococcus coelicolor* have been completed. Sequence data of other genomes are being continually added and catalogued in online databases.

Analysis of prokaryotic genomes in databases has provided insights (Dobbins *et al.*, 2004; Fleischmann *et al.*, 1995; Parkhill *et al.*, 2001; Tettelin *et al.*, 2000) that include:

(i) amino acid sequence, which can be determined from DNA sequence and genetic code;

(ii) open reading frame (ORF) which are protein coding regions;

(iii) non-coding DNA types, amount, distribution, information content, and functions;

(iv) gene control/regulatory sequences;

(v) coordination of gene expression, protein synthesis, and post-translational events;

1.33. Complete plasmid sequences

To date, 501 bacterial plasmids have also been sequenced and deposited into the NCBI genome database (http://www.ncbi.nlm.nih.gov/genomes/static/eub_p.html). However, the numbers of sequences deposited are rapidly growing each day. The complete plasmid sequences provide information on the biological properties and genetic arrangement that could have contributed to their virulence (Galli *et al.*, 2001), adaptation to environmental changes (Zhong *et al.*, 2002); and also their usefulness as a vector in recombinant DNA technology (Bartosik *et al.*, 1997; Wright *et al.*, 1997). Sometimes, the complete plasmid sequences also provide information on the mechanisms of replication, as demonstrated by the sequence of *Marinococcus halophilus* (Louis and Galinski, 1997). Comparison of nucleotide sequence data between plasmids has also provided valuable information on the genomic arrangement for evolutionary studies (Katsura *et al.*, 1997; Thomas *et al.*, 1997).

1.34. Various strategies of genomic sequencing

A number of strategies have been used in projects of whole genome sequencing. The strategies used in the projects are broadly primer walking, 'clone-by-clone' or hierarchical (HS) shotgun sequencing and Whole Genome Shotgun (WHS) sequencing (Weber and Myers, 1997).

1.34.1. Sequencing by primer walking

In primer walking, an initial sequencing reaction is performed using a set of primers of a known vector, usually the universal primers. This is usually achieved by cloning the DNA of interest into a cloning vector. The new sequence obtained is then used to design a new set of primers for sequencing along the molecule. This process is repeated until the molecule has been completely sequenced on both strands.

This strategy however, is not only time consuming (approximately, 400 bases at a time, and 2-6 days between sequencing events), but it is also expensive, as so many different new sets of primers are required. Despite these limitations, this is the standard way used by most non-genome labs for sequencing fragments longer than 400 bases.

1.34.2. Clone-by-clone sequencing

In this approach (also known as the hierarchical sequencing (HS) strategy) genomic DNA is cut into large pieces and inserted into BAC vectors to yield a library (Venter *et al.*, 1996). Transformation is usually carried out into *E. coli* for the vector to replicate. The BAC inserts are isolated and mapped by sequence-tagged-sites or by fingerprinting to produce an overlapping tiling path. Each BAC is fragmented randomly into smaller pieces and each piece is cloned into a plasmid and sequenced on both strands. These sequences are aligned so that identical sequences are overlapping and the contiguous sequences are assembled into a finished sequence. This is usually achieved once each strand has been sequenced about four times to produce eight times coverage of high quality data. This strategy has been used for sequencing of bacterial genomes, nematode, *Caenorhabditis elegans* and yeast, *Saccharomyces cerevisiae* (Chen *et al.*, 1993).

1.34.3. Whole genome shotgun strategy

This approach does not involve the step of cloning large ordered fragments in BACs. Fred Sanger first introduced this strategy of sequencing in 1982 (Sanger *et al.*, 1977b). It was used to sequence *H. influenzae* Rd in 1995. This strategy was also used in the sequencing of the whole human genome project (Weber and Myers, 1997). In this strategy, the DNA of the entire genome is fragmented randomly. These fragments are cloned into plasmids and sequenced on both strands. Once the sequences are obtained, they are aligned and assembled together by aligning overlaps. The advantage of the shotgun method is that it requires no prior mapping. However, the main disadvantage is the requirement for substantial information technology (IT) resources. Also, unlike the clone-by clone approach, the relationship between 'contigs' may not be apparent until near the end of the project.

1.35. Objective of this thesis

The aim of this thesis is to investigate the origins of transferable antibiotic resistance in *H. influenzae*.

CHAPTER TWO

MATERIALS & METHODS

2.1. Bacteria strains and plasmids

2.1.1. Bacteria

H. influenzae strain 1056 a capsular type b was isolated from a nasopharyxn of an infant with a family outbreak of *H. influenzae* disease in Guildford in 1986 (Brightman *et al.*, 1990). A strain of β-lactamase positive, ampicillin resistance *H. influenzae* type b, strain 299 was isolated from both the CSF and blood of a patient from the Aglaia Kuriakou Children's Hospital in Athens, Greece in 1991 (Dimopoulou *et al.*, 1997). A streptomycin-resistant, recombination-deficient Rd *H. influenzae* strain nf38 was a recipient used in conjugation experiments, negative for ß-lactamase and chloramphenicol acetyl transferase (Setlow *et al.*, 1972). Other *H. influenzae* strains used in this study are summarised in Table 2.1 and *E. coli* strains used are shown in Table 2.2.

2.1.2. Plasmids

Plasmids used and constructed in this study are shown in Table 2.3. Briefly, p1056 (ICE*Hin1056*) was isolated from transconjugant pX1056 that was obtained by mating *H. influenzae* 1056 to nf38 (Dimopoulou *et al.*, 1992), through a mating experiment using the filtration method as described by Stuy (Stuy, 1979). Plasmid of *H. influenzae* 299, p299 (ICE*Hin299*) was isolated through similar method (Dimopoulou *et al.*, 1997).

Table 2.1. List of *H. influenzae* strains

Strain	Attributes	Source
1056	amp^r, tet^r, chlor^r	C. Brightman
X1056	Transconjugant of 1056 with nf38	D. Crook
299	Greek strain, amp^r	I. Dimopoulou
X299	Transconjugant of 299 with nf38	I. Dimopoulou
nf38	strep^r, *recA*$^-$	J. Setlow
Rd^{sm}*recA*$^+$	Rd strain resistant to rifampicin by selective pressure, capsule deficient serotype d strain, *recA*$^+$	R. Moxon
Rd^{sm}*recA*$^-$	Rd strain resistant to streptomycin by selective pressure, *recA*$^-$	J. Setlow
Rd^{rif}*recA*$^+$	Rd strain resistant to rifampicin by selective pressure, *recA*$^+$	R. Moxon
Rd^{tet}*recA*$^-$	Tetracycline resistant, *recA*$^-$	D. Crook
XT3	Transconjugant of 1056 and Rd^{sm}*recA*$^+$; amp^r, tet^r, chlor^r	D. Crook
XT9	Transconjugant of 1056 and Rd^{sm}*recA*$^-$; amp^r	D. Crook
XT11	Transconjugant of XT3 and Rd^{rif}*recA*$^+$, amp^r, tet^r, chlor^r	D. Crook

Table 2.2. List of *E. coli* strains used

Strains	Attributes	Source
DH5α	F$^-$ Φd*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(r_k^-,m_k^+) *pho*A *sup*E44 λ$^-$ *thi*-1 *gyr*A96 *rel*A1	Invitrogen, Holland
TOP10	One Shot® Chemically Competent *E. coli F' mcr*A Δ*(mrr-hsd*RMS-*mcrBC) Φ80l*acZΔM15 Δ*lacX74 deo*R *recA1 araD139* Δ*(ara-leu)*7697 *gal*U *gal*K *rps*L (Str^r) *end*A1 *nup*G	Invitrogen, Holland
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F' *pro*AB *lac*I^{q}ZΔ15 Tn*10* (Tet^r)]	Stratagene, US
XL10-Gold	Tet^{r}Δ (*mcr*A) 183 Δ(*mcr*CB-*hsd*SMR-*mrr*) 173 *end*A1 *sup*E44thi-1 *rec*A1 *gyr*A96 *rel*A1 *lac Hte* [F' *pro*AB *lac*I^{q}ZΔM15 Tn*10* (Tet^r) Amy Cam^r]^a	Stratagene, US

Table 2.3. List of plasmids/mobile elements and cloning vectors used

Plasmid	Attributes	Size (kb)	Source
ICE*Hin1056*	Isolated from *H. influenzae* 1056, amp^r, tet^r, chlor^r	60	Guildford, Surrey
ICE*Hin299*	Isolated from *H. influenzae 299*, amp^r	54	Athens, Greece
pB7	amp^r	12	I. Dimopoulou
pUC19	amp^r, *lacZ*	2.7	MBI Fermentas, Lithuania
TOPO 2.1	amp^r and kan^r, *lacZ*	3.9	Invitrogen, The Netherlands

2.1.3. Bacterial sequences

The bacterial sequences of GIs retrieved from the following NCBI sites: (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome), DOE Joint Genome Institute (JGI), California, USA (http://www.jgi.doe.gov/) and The Wellcome Trust Sanger Institute (http://www.sanger.ac.uk). The ICE*Hin1056* sequences were obtained from the work carried out in Chapter 4.

The seventeen GIs use in this study are shown (Table 2.4). The bacteria containing these GIs were all members of *Proteobacteria*. *Ralstonia metallidurans* and *Burkholderia fungorum* belonged to β-proteobacteria and all other bacterial species are γ-*Proteobacteria*.

Table 2.4. Bacterial sequences used in this study.

Bacteria	Strain	GI name	Size (bp)	Accession No.
Burkholderia fungorum[a]	LB400	un[b]	70,906	NZ_AAAJ00000000
Haemophilus influenzae	1056	ICE*Hin1056*	59,393	AJ627386
Haemophilus ducreyi	35000HP	un	49,075	NC_002940
Haemophilus somnus[a]	129PT	un	56,442	NZ_AAB002000009
Haemophilus somnus[a]	2336	un	64,831	NZ_AACJ00000000
Photorhabdus luminescens	TT01	un	140,163	NC_005126
Pseudomonas sp.	B13	*clc*	105,027	AJ617740
Pseudomonas aeruginosa	C	pKLC102	103,532	AF257538
Pseudomonas aeruginosa	PA14	PAP1	106,518	AY273869
Pseudomonas aeruginosa	SG17M	PAGI-3	114,632	AF440524
Pseudomonas fluorescens[a]	PfO-1	un	70,421	NZ_AAAT00000000
Pseudomonas fluorescens[a]	SBW25	un	101,233	NC_004507
Ralstonia metallidurans[a]	CH34	un	<76,893	NZ_AAAI00000000
Salmonella enterica serovar Typhi	CT18	SPI-7	134,000	NC_003198
Salmonella enterica serovar Typhi	TY2	un	122,000	NC_004631
Xanthomonas axonopodis	306	un	86,019	NC_003919
Yersinia enterocolitica[a]	8081	un	65,285	NC_003222

[a]sequencing of whole genome is ongoing, [b]un = unnamed

2.2. Media

2.2.1. Solid culture media

Chocolate agar was used to culture *H. influenzae* strains. *H. influenzae* type b (Hib) agar was used in conjugation experiments (Barbour *et al.*, 1993). It was supplemented with nicotinamide adenine dinucleotide (NAD, 10 μg/ml, Sigma) and haemin (15 μg/ml, Sigma). The *E. coli* strains were grown in Luria Bertani's (LB) agar. The constituents and procedures for media preparation are listed in detail in Appendix A.

2.2.2. *Liquid culture media*

E. coli in broth were grown in LB broth. For the growth and dilution of Haemophilus strains, brain heart infusion (BHI) with NAD (10 μg/ml) and haemin (15 μg/ml) was used.

2.2.3. *Antibiotics selective media*

Stock antibiotic solutions were prepared as shown in Appendix A. LB agar in aliquots of 20 ml in universal bottles were melted in boiling water and allowed to cool to 50°C. For antibiotic selections in matings or growth of antibiotic-resistant strains in liquid media, the desired filter-sterilised antibiotics were added to the media at the same time as the NAD and haemin. For growth of strains on solid agar, the antibiotics were added and gently mixed into the molten agar, and immediately poured were immediately poured into 99-mm disposable plastic Petri-dishes. The final concentrations of antibiotics were as follows: ampicillin, 50 μg/ml; chloramphenicol, 20 μg/ml; rifampicin, 20 μg/ml; streptomycin, 50 μg/ml; and tetracycline, 50 μg/ml. The agar was allowed to set at room temperature and the agar plate was dried for at least 20 minutes before use.

2.2.4. *X-Gal and IPTG plates*

The methods for the preparation of stock X-Gal (5-Bromo-4-chloro-3-indolyl-β-D-galactosidase, MBI Fermentas, Lithuania) and IPTG (Isopropylthio-β-D-galactosidase, MBI Fermentas, Lithuania) are shown in the Appendix C. For use in transformation procedure, X-Gal agar was prepared by spreading 40 μl of 2% X-Gal onto the surface of a dry LB agar plate with a sterile glass spreader. When IPTG was required, the X-Gal was diluted with some broth (either SOC or BHI) before the addition of 40 μl of 100 mM IPTG (otherwise, the mixture of X-Gal and IPTG produced a white precipitate). Both

solutions were then spread evenly with a glass spreader. However, IPTG was not required when using TOP10 *E. coli* cells in cloning libraries. The mixture was allowed to absorb into the agar for approximately 10 minutes before drying at 37°C for 15 minutes prior to bacterial culture.

2.3. Chemical, restriction enzymes and DNA modifying enzymes

Chemicals used in the experiments were mostly obtained from Sigma Biochemicals (Dorset, UK); restriction endonuclease and other DNA modifying enzymes were purchased from MBI Fermentas (Lithuania, France), Gibco (UK) and New England Biolabs (Massachusetts, USA). Media and agar were obtained from Oxoid (Basingstoke, UK), unless otherwise stated. The composition of chemicals and solutions are given in the Appendix A.

Tissue culture sterile water (Sigma, Dorset, UK) was used throughout, and where applicable. Most chemicals were of AnalaR grade and supplied by BDH or Sigma (UK). Exceptions to these suppliers are indicated in the text when the specified reagents are mentioned. The constituents of solutions are described in Sambrook (Sambrook *et al.*, 1989) and exceptions to this are described in the text.

2.4. Bacterial growth conditions

2.4.1. Growth of E. coli *and* H. influenzae *strains*

Bacteria, namely *E. coli* were generally sub-cultured on LB agar (without the addition of antibiotics) when recovered from –80°C freezer (Sanyo, Japan). Subsequent culture was made onto agar media containing selective antibiotics, if appropriate. *E. coli* cultured in broth were incubated at 37°C in air with orbital shaking at approximately 100 rpm (New Brunswick Scientific, USA).

H. influenzae grown either on chocolate agar or Muller Hinton agar were incubated at 5% CO_2 in a 37°C incubator (Sanyo, Japan) while *H. influenzae* strains grown in broth were incubated at 37°C in air. *H. influenzae* colonies are well visualised after 48 hours.

2.4.2. Culture and storage of bacteria

Bacterial strains for storage were picked from single colony from fresh agar plates. *H. influenzae* strains were cultured into BHI broth (containing NAD and haemin) and *E. coli* strains were cultured into LB and grown overnight. Equal volume of 30% glycerol was added to the culture and stored at –80°C. Alternatively, fresh pure colonies on agar plates were scraped and generously re-suspended into tryptone soy broth (TSB) with 15% glycerol and stored at –80°C. Commercially prepared competent *E. coli* strains were stored immediately at –80°C upon arrival.

2.5. Colony counts

A preliminary estimation of the concentration of the bacterial cells was made. A ten-fold dilution was made by mixing 1 ml of broth culture to 9 ml of the phosphate buffered saline (PBS) used as diluent. When counting colonies of Haemophilus strains, BHI broth was used.

A series of 10-fold dilution was carried out up to the estimated dilution. One hundred micro-litre of each of the diluted cells was plated using a sterile glass spreader onto agar plates containing the appropriate antibiotics. The diluted cells were allowed to absorb into the agar before incubation, by drying at 37°C for 5-10 min. The agar plates were then incubated at 37°C overnight. For each dilution, a colony count of the bacterial

cells was carried out in duplicate. Only agar plates containing between 30-300 colonies were used for colony counting.

2.6. Antibiotic susceptibility test by Etest

Overnight culture of *H. influenzae* strains on chocolate agar were picked and were suspended in phosphate buffered saline (PBS) to a 0.5 McFarland turbidity. A sterile cotton swab was dipped into the inoculum's suspension and excess fluid was removed by pressing the swab against the inner wall of the glass container. The entire surface of a completely dried Muller Hinton agar (with NAD and haemin) was swabbed twice, by rotating the plate approximately 90 degrees each time to ensure an even distribution of the inoculum. With a pair of sterile forceps, an Etest® strip (AB Biodisk, Sweden) was slowly placed on the surface of the inoculated agar. The agar plate was incubated overnight at 37°C in CO_2 atmosphere.

When the bacterial growth becomes distinctly visible, the minimal inhibitory concentration (MIC) value was read at the point of intersection between the inhibition ellipse edge and the Etest strip.

2.7. Conjugation

2.7.1. Conjugation on solid media

Donor strain (strain 1056) was grown overnight on Muller Hinton agar containing NAD (10μg/ml), haemin (15 μg/ml) and ampicillin (8 μg/ml). Recipient strains were similarly grown on Hib agar containing either streptomycin (20 μg/ml) or rifampicin (20 μg/ml). A dense suspension of cells was made in 1 ml of BHI broth. These cells were harvested by centrifugation and the pellet was re-suspended in BHI broth to a final OD of approximately 0.8 (measured visually with MacFarland's standard). To 1 ml of BHI

broth, 4 µl of donor cells and 40 µl of recipient cells were added. The cells were gently mixed, and 100 µl of the mixture was spread onto duplicate plates of Hib agar without antibiotics. The inoculum was allowed to dry in a short incubation (5-10 minutes) at 37°C, prior to incubation in CO_2 for 8 hours.

The 'mated cells' were harvested with BHI broth to a volume of approximately 1 ml. Serial dilutions were carried out to determine viable counts and plated on selective agars containing antibiotics for enumeration of colony-forming-unit (cfu) of transconjugants, donors and recipients. The rate of transfer is measured by the number of transconjugants per donor cell.

2.7.2. Conjugation in liquid media

Mating of cells in liquid was carried out concurrently, by incubating the original mixed culture in BHI broth at 37°C for the same period of 8 hours. Viable counts were determined as above.

2.7.3. Measurement of point mutation

Donor or recipient cells that were usually susceptible for a particular antibiotic were tested for point mutation by spreading 100 µl of the re-suspended and washed cells onto Hib agar containing the relevant antibiotic. The plates were incubated for viable counts.

2.7.4. Purification of transconjugants

Approximately 60 transconjugants were plated to Hib agar containing tetracycline, ampicillin or chloramphenicol. Viable counts were carried out following overnight incubation.

2.8. Electron microscopy examination of pili

Presence of pilus by examined by transmission electron microscopy (TEM). Prof. David Ferguson, Dept. of Pathology, at John Radcliffe Hospital, Oxford, kindly undertook the EM examination. In brief, the *H. influenzae* colonies grown overnight on Hib agar were re-suspended with approximately 1 ml of phosphate buffered saline (PBS). Copper grid coated with formavar and carbon was flooded with the suspension and was left for two minutes. Excess fluid from the copper grid was discarded, before a drop of water was added. The grid was negatively stained with 1% methyl tungstate, dried and examined using TEM (JOEL 1200EX, Japan).

2.9. Chromosomal DNA extraction

2.9.1. Extraction of chromosomal DNA from H. influenzae

Pure colonies of *H. influenzae* grown on BHI agar were harvested (Sterilin, UK) and were generously re-suspended in 0.2 ml of Tris-EDTA (TE) buffer. A volume of 0.45 ml of GES lysing solution consisting of 5 M guanidium thiocyanate, 100 mM EDTA and 0.5% v/v sarkosyl were added to the suspension and mixed by inversion and allowed to lyse at room temperature for 3 minutes. Cold 7.5 M ammonium acetate (0.25 ml) was added and was mixed by several inversions before it was stored on ice for 10 minutes. Subsequently, 0.25 ml of chloroform: isoamylalcohol (24:1) was added and mixed thoroughly by hand for 3 minutes. The phases were separated by centrifugation at 12,000 g for 20 minutes, and the aqueous upper layer was transferred to a fresh-labelled micro-centrifuge tube. Cold isopropanol of 0.5x volume was added and was gently mixed by inversion for 1 minute or until the DNA threads were visible. Following centrifugation for 5 minutes, the supernatant was discarded and the pellet was re-suspended in 0.2 ml TE buffer.

A 2.5x volume of cold absolute ethanol was added to re-precipitate the DNA. If necessary, the DNA suspension was stored at –20°C for 15-30 minutes. Excess alcohol in the tube was washed off by centrifugation and drained onto a paper towel. The pellet was washed three times with 0.5 ml of 70% ethanol by inversion and centrifugation at 12,000 g for 5 minutes. Finally, the DNA pellet was re-suspended in 0.2-0.3 ml of TE buffer and was allowed to dissolve overnight at 4°C.

2.9.2. Extraction of crude DNA by boiling method

One colony of bacterial cells from a fresh culture of bacteria was suspended into 50 µl of sterile distilled water and boiled for 10 minutes. The boiling lysate was allowed to cool to room temperature for 5 minutes before it was spun at a 16,000 g for 30 seconds. The supernatant was collected and transferred to a fresh tube. This DNA extract (1 µl) was then used as template in PCR amplifications.

2.10. Plasmid extraction

2.10.1. Extraction by mini-prep method

Plasmid extraction was performed according to the instruction provided by a plasmid extraction kit, NucleoSpin Plasmid® kit, (Macherey-Nagel, Germany). Briefly, bacteria from a single colony of an agar plate was picked up and cultured into LB broth with ampicillin (50 µl/ml). The broth was then incubated in a shaking incubator at 80 rpm for 17-18 hours. Subsequently, 1.5 ml of the cultured medium was pipetted into a micro-centrifuge tube, and centrifugation was carried out at 15,000 g (Eppendorf, Hamburg, Germany) for 30 seconds. The supernatant was discarded and the cell pellet was re-suspended by vortexing vigorously in buffer A1 (previously added with Rnase). The suspension was gently mixed with a buffer A2, and was allowed to lyse completely at

room temperature for a maximum of 5 minutes. When binding buffer A3 was added to the lysate, a colloidal precipitate formed was removed by centrifugation for 10 minutes at full speed. Following centrifugation, the supernatant was loaded into a NucleoSpin Plasmid column that had been placed in a 2 ml collecting tube. It was centrifuged again for 1 minute at 16,000 g and the flow-through was discarded.

Pre-warmed (50°C) washing buffer AW, was added to the column and centrifugation was repeated, followed by ethanolic washing buffer A4 for another minute. The residual ethanol in the column was minimised by a further two-minute centrifugation. Finally, buffer AE was added to the column and centrifugation was carried out for 1 minute to elute the plasmid DNA. The extracted plasmids were stored at –20°C until use.

2.10.2. Extraction of plasmid DNA by caesium chloride-ethidium bromide gradients

The method was adapted from the method of Sambrook (Sambrook *et al.*, 1989). This purification of plasmid DNA was carried out according to its buoyant density. A solution of caesium chloride is subjected to a very high centrifugal force such that the ions it contains are forced towards the bottom of a tube. Diffusion counteracts this setting up a gradient of increasing CsCl concentration and thus density down the tube. Any macromolecules present within the CsCl solution will move to, and form bands at that point where the local density is the same as their own buoyant density. Since proteins have a low, RNA a high and DNA an intermediate density, separation is achieved. By incorporating an EtBr into the solution, circular plasmid DNA can be separated from linear chromosomal DNA. This happens because dye binds to and thus displaces fewer of the dense Cs ions from the plasmid DNA. This is thus denser than the chromosomal DNA under these conditions and therefore bands beneath it in the tube.

Following overnight incubation of *H. influenzae* strain X1056 on chocolate agar, the colonies were scraped and emulsified with 1 ml of PBS in a centrifuge tube. The bacterial suspension was then transferred into 500 ml of MH broth supplemented with ampicillin (25 μg/ml), haemin and NAD, and was incubated in a shaking incubator overnight at 800 rpm.

Plasmid DNA was extracted using plasmid extraction kit, Nucleobond AX 500 (Macherey-Nagel, Germany). Briefly, the cultured cells were harvested by centrifugation at 670 g (Sorvall RC, USA) for 10 minutes at 4°C. The cells were re-suspended in the given buffer and then lysed with a buffer containing RNAse. Disrupted cells consisting of chromosomal DNA and cell proteins were precipitated in the form of homogenous suspension. The bacterial lysate was then clarified through a NucleoBond filter and the clear lysate was loaded into a NucleoBond BAC 100 column. The flow-through from the column was allowed to adsorb onto the column twice and washed three times before eluting with 6 ml of elution buffer.

To the eluted plasmid DNA, 6 g of solid CsCl was added and allowed to dissolve. Ethidium bromide of 0.48 ml in volume was added to the solution and was mixed well. The refractive index of the mixture was checked to be approximately 1.388 by using a refractometer (Bellingham and Stanley, England). The mixture was then divided and placed into three Beckman polycarbonate centrifuge tubes (size 13 x 15 mm) and was centrifuged overnight (approximately 16 hours) in an ultracentrifuge (Beckman T100, USA) at 215,000 g at 15°C. Following centrifugation, a fine band consisting of plasmid DNA underneath a bright band of chromosomal DNA seen under UV light was collected and put into a fresh micro-centrifuge tube.

Equal amount of water-saturated 1-butanol was added to the collected plasmid DNA. The two phases was thoroughly mixed, and centrifugation was carried out at 670 g

for 3 minutes at room temperature. The aqueous lower phase was removed and the above extraction was repeated until all the pink colour of EtBr in both the upper and lower phase disappeared.

To approximately 200 µl of the extracted plasmid DNA, 5 µl of glycogen (MBI Fermentas, Lithuania), 20 µl of sodium acetate (NaOAc) and 500 µl of cold 95% ethanol was added and mixed gently. The mixture was put in ice for 45 minutes before centrifugation at a 16,000 g for 15 minutes. The pellet was washed twice with 70% ethanol, dried and reconstituted with 25 µl of TE buffer.

2.11. Removal of contaminating chromosomal DNA

Plasmid DNA was precipitated with 1/10 volume of 3M NaOAc; pH 5.2 followed by 1 µl of glycogen (MBF Fermentas, Lithuania) and was mixed gently and cooled in ice for an hour. The mixture was centrifuged for 10 minutes and washed twice with 70% alcohol. The pellet was finally dried and diluted with deionised water. A 1/5 volume of lambda buffer was added to the plasmid, followed by exonuclease I and lambda exonuclease (New England Biolabs, US). The mixture was allowed to react at 37°C overnight and inactivated at 80°C for 20 minutes. To the mixture, 1/10 volume of NaOAc and glycogen were added and incubated in ice for an hour. The mixture was spun at 15,000 g for 10 minutes before washing with 70% alcohol twice. The pellet was dried and diluted with an appropriate amount of distilled water. It was allowed to dissolve at 37°C for 30 minutes.

2.12. Plasmid shearing

In order to obtain fragments of the element of sizes between 500 bp to 2.0 kb, a comparative study was carried out to determine the optimum length of time required to sonicate the DNA of the plasmid. Purified whole plasmid DNA (20 µl) was made up to

1/100 dilution with sterile distilled water and divided into five micro-centrifuge tubes. Each of the tubes was subjected to sonication at various durations of 10, 15 and 20 s. One tube was not sonicated and was kept as a control. The power of sonication was kept constant at 15%. Following sonication, the sheared DNA was gel-electrophoresed (0.7% agarose gel) and the sizes of the sonicated fragments were checked under UV-light. The DNA that produced a heavy band of smearing between 500 bp to 2.0 kb was noted and the sonication time was selected.

2.13. Agarose gel electrophoresis

2.13.1. Preparation of agarose gel

Most agarose gel electrophoresis used was carried out using 0.7% agarose (Sigma type II; Sigma, UK) prepared in 1x TBE buffer by heat dissolving (see Appendix). Ethidium bromide (final concentration of 0.5 μg/ml) was carefully added to the slightly cooled gel mixture, which was subsequently cast and allowed to set at room temperature.

2.13.2. Gel electrophoresis

Plasmid samples (5 μl) were mixed with 2 μl of 6x loading buffer (Sigma, USA) and were loaded into the wells of the agarose gel. A 1-kb ladder (Gibco, UK) was usually used as a molecular weight marker. Electrophoresis was carried out horizontally at 100 volts with 1x TBE as the electrophoresis buffer.

2.13.3. Observation of DNA in agarose gel

Following electrophoresis, DNA on agarose gels was viewed using a transilluminator of Gel Doc 1000 image analyser (Bio-Rad, USA). For preparation of DNA for cloning, gels

with DNA separated by restriction enzyme were viewed and excised over 365 nm (long) wavelength of TL33-E Transilluminator (UVP, UK).

2.13.4. Recovery and purification of DNA from agarose gel

DNA fragments on agarose that were digested with restriction enzymes, were excised from the agarose gel using a sterile scalpel into separate clean micro-centrifuge tubes. The DNA in the gel block was eluted using a gel extraction kit, MinElute Gel Extract (Qiagen, USA) with some modifications. Following solubilisation with QG buffer, the dissolved mixture was transferred into the columns and spun for 1 minute at 16,000 g. The flow-through was discarded and the column was washed with PE buffer for 1 minute. The column was dried by centrifuging for another minute, before it was air-dried for another minute to get rid of residual ethanol. The DNA was eluted by adding distilled water to the column and spun at 15,000 g for 1 minute. The DNA obtained was kept at –20°C until use.

2.14. Restriction enzymes digestion

Restriction enzymes used in most of the experiments were purchased from MBI Fermentas (Lithuania). All the restriction endonuclease digests were performed in the presence of buffers supplied by the manufacturer. To every μg of DNA, 5-10 units of enzyme were added and the digestions were performed at 37°C for 1 hour, unless otherwise stated. Restriction digest using *Bsg*I required S-adenosylmethionine for optimal activity. For this particular enzyme, the reaction was heated at 65°C for 20 minutes to inactivate the enzyme.

2.15. Polymerase chain reaction (PCR)

2.15.1. Selection of oligonucleotide primers

Pairs of PCR primers were selected from sequence data using certain criteria. Firstly to ensure similar T_m (melting temperature) values for paired primers, the length of oligonucleotides should be between 15-22 bp with a similar guanine and cytosine ratio of between 50 to 60%. The nucleotide sequence was tested for secondary structures and complementary pairing using a website: (http://www.basic.nwu.edu/biotools/OligoCalc.html). The primers used were purchased from either Alta Bioscience (The University of Birmingham, UK) or MWG-Biotech AG (Germany).

2.15.2. Dilution of primers

Oligonucleotide primers were diluted with sterile distilled water to 100 pmoles/µl according to the instruction of the manufacturer as stock solution and stored at –20°C. The primers used for PCR were diluted to 100 µM as working concentration, while primers used for sequencing were further made to 1:15 dilution. The lists of primers used in the whole study are shown in the Appendix B.

2.15.3. Preparation of PCR mixture

For each PCR (20 µl) contained: 2.0 µl of 10x PCR buffer, 3.0 µl of 2 mM dNTP mix (MBI, Fermentas), 1.6 µl of 25 mM $MgCl_2$, 1.0 µl of 100 pmol/µl forward primer, 1.0 µl of 100 pmol/µl reverse primer, 0.5 µl of 2 unit Taq polymerase (MBI Fermentas, Lithuania), 1.0 µl of the template DNA and 9.9 µl of sterile distilled water.

2.15.4. Thermal cycling condition

Amplification of the DNA samples was carried out using an MJ Research Thermal Cycler (USA). The DNA was initially denatured at 94°C for 2 minutes prior to a series of 30 cycles of denaturing at 94°C for 1 minute, annealing at 58°C for 1 minute and extension at 72°C for 2 minutes. At the end of the cycle, the DNA was given a further extension of another 4 minutes at 72°C before it was cooled at 4°C.

2.15.5. Observation of amplified products

Amplified products were separated by electrophoresis in 0.7% agarose gel, and run at 100 mA on a BioRad PowerPac 200 (USA) for approximately 30 minutes. The DNA on the gel were visualised using Gel Doc image analyser (BioRad, US) and the image of the gel was printed to keep as a record.

2.16. Purification of PCR products and DNA

2.16.1. Purification by spin column

PCR product extracted from agarose was purified using Nucleospin Extract (Macherey-Nagel, Duren, Germany). Briefly, one volume of the excised gel was mixed with four volumes of NT2 buffer and transferred into a column fixed into a collection tube. The column was spun in a centrifuge for 1 minute. All centrifugation was carried out at 15,000 g unless otherwise stated. The flow-through was discarded and NT3 buffer was added into the column and spun for another 1 minute. The column was finally rinsed with NT3 and the buffer was completely removed by a two-minute centrifugation.

2.16.2. Polyethylene-Glycol (PEG) purification

2.16.2.1. Samples in 0.2 ml micro-centrifuge tubes

A volume of 60 µl of 20% polyethylene-glycol 8000 (PEG) solution was added to 20 µl of plasmid DNA or PCR products. It was briefly mixed by vortex and incubated at 37°C for 10 minutes. The mixture was centrifuged for 15 minutes at 15,000 g and the pellet was collected. It was then rinsed twice with 70% ethanol and air-dried. A volume of 20 µl of distilled water was used to re-suspend the DNA pellet. Prior to ligation, agarose electrophoresis was performed to 5 µl of the DNA samples to ensure the presence of the DNA.

2.16.2.2. Samples in microAmp tray

Similar procedure was carried out as above, except that samples in microAmp trays were centrifuged using Allegra GR Centrifuge (Beckman Coulten, USA) at 5,000 g at 4°C for 1 hour and the DNA pellets were washed with 150 µl of 70% ethanol by centrifuging at 5,000 g for 20 minutes. The residual ethanol was drained by inverting the tray on absorbent paper towels and spun at 1,000 g for 1 minute. Washing of the pellets with 70% ethanol was done twice before drying at 65°C in a thermal-cycler for at minimum of 10 minutes.

2.16.3. Ethanol/sodium acetate (NaOAc) precipitation

2.16.3.1. Samples in micro-centrifuge tubes

For sequenced samples that were carried out in 0.2 ml micro-centrifuge tubes, 2.0 µl of 3 M NaOAc, pH 4.6 and 50 µl of 95% ethanol was added to the sample in 1.5 ml micro-centrifuge tube. The mixture was incubated at room temperature for 15 minutes before centrifugation at 15,000 g for 20 minutes. The ethanol was removed and the pellet was washed twice with 150 µl of 70% ethanol by centrifugation for 10 minutes. The

supernatant was discarded and the pellet was dried at 65°C for at least 10 minutes prior to sequencing.

2.16.3.2. Samples in microAmp tray

The amount of NaOAc and ethanol used was similar to the samples that were carried out in individual micro-centrifuge tube. However, for samples in microAmp trays, centrifugation was carried out at 5,000 g at 4°C for 1 hour. The pellets were dried by inverting the microAmp tray onto absorbent paper towels and were spun for 1 minute at 1,000 g. The samples were washed with 150 µl of 70% alcohol by centrifuging at 1,000 g for 1 minute and dried at 65°C. The trays were sealed and stored at –20°C until ready for sequencing.

2.17. DNA manipulations

2.17.1. Fragmentation of DNA by sonication

In a 1.5 ml micro-centrifuge tube, 50 µl of *H. influenzae* plasmid DNA was made up to 200 µl with distilled water. The tube with plasmid solution was place into an ice-bath and sonicated for 10 s at 15% power (Soniprep 150, MSE, England). Five micro-litres of the sonicated samples were run through agarose gel electrophoresis to determine the sizes of the fragments. The remaining sheared plasmid solution was kept at –20^{0}C until use.

2.17.2. Creating blunt-ended DNA

Sonicated DNA plasmid was blunt-ended using T4 DNA polymerase (MBF Fermentas, Lithuania). Briefly, 5x reaction buffer (50µl), 2 mM dNTPs (10 µl), and T4 polymerase (5 µl) were added to the sheared plasmid DNA (200 µl) and were incubated for an hour at

37°C. The mixture was PEG precipitated with two washings of 70% ethanol. The pellet was dried and re-suspended in 20 µl of distilled water.

2.17.3. Phosphatase treatment of cloning vectors

Following restriction enzyme digestion, the vector DNA was added with 1 µl of calf intestine alkaline phosphatase, CIAP (MBI Fermentas, Lithuania) and 1 µl of CIAP buffer. The incubation at 37°C was continued for 1 hour and the reaction was inactivated at 85°C for 15 minutes.

2.18. DNA ligations

Ligation of fragmented plasmid DNA into pUC19 cloning vector was carried out using T4 DNA ligase and ligation buffer provided by a ligation kit (MBI Fermentas, Lithuania). A ratio of 10:1 of sheared DNA to cloning vector was used in a 20 µl reaction consisting of 2 µl of 10x ligation buffer, 2 µl of PEG 4000, 2 µl of T4 DNA ligase and distilled water. The ligation reactions were performed at 22°C overnight. For ligation of sticky ends, PEG 4000 was omitted and 1 µl of T4 DNA ligase was added.

2.19. Bacterial transformation

2.19.1. Transformation by heat-shock

The *E. coli* competent cells used in the experiments were mainly obtained from Stratagene (USA) while *E. coli* TOP10 strains were provided in PCR cloning kit, TOPO TA Cloning (Invitrogen, The Netherlands).

Transformations of DNA into competent *E. coli* cells by heat-shock were carried out at 42°C for a very short period of time. Frozen competent *E. coli* cells were thawed on ice and mixed gently by hand. Depending on the strains of the *E. coli*, between 50-100 µl

of the cells were added to a pre-chilled 15-ml Falcon 2059 polypropylene tube. To the cells, 4 µl of β-mercaptoethanol was added and incubated on ice for 10 minutes. However, for transformation of DNA into *E. coli* XL1-blue subcloning-grade, mercaptoethanol was omitted. Between 4-5 µl of the ligation mix was added to the competent cells. Following 30 minutes incubation in ice, heat pulse treatment was given to the mixture in a water-bath set at 42°C for 45 s for most *E. coli* strains. However, for TOP10 cells heat pulse was carried out for 30 s. Transformed cells were then plated onto selective LB agar plates following instructions of the supplier of the competent cells.

2.19.2. Transformation of PCR products into TOP10 E. coli

Prior to transformation of PCR products into TOP10 *E. coli* cells (Invitrogen, Netherlands), TOPO cloning reaction was set up by adding 1.0 µl TOPO vector to a 1.0 µl mixture consisting of PCR product (2.0 µl), salt solution (1.0 µl of 1.2 M NaCl, 0.06 M $MgCl_2$) and distilled water (1.0 µl). The reaction mixture was gently mixed and incubated for 5 minutes at room temperature.

Subsequently, 2.0 µl of the TOPO Cloning mixture was added to a vial of iced-thawed component cells and was gently mixed before incubating in ice for 5 minutes. The cells were heat shocked at 42°C in a water-bath for 30 s without shaking. Immediately after the heat-shock, the tube was chilled in ice and added with 250 µl of SOC medium. The cap of the vial was tightened and incubated for an hour at 37°C in a shaking incubator at a horizontal speed of 15 rpm. Transformed cells were plated onto LB agar containing X-Gal and ampicillin (50 µg/ml) and incubated overnight.

2.20. Selection of recombinant clones

Recombinant colonies were picked based on the colour of the colonies. Ampicillin was used for a selection of bacteria containing the cloning vector. IPTG was used to induce the β-galactosidase gene within TOPO to breakdown the X-Gal into its blue-indole derivative. In the absence of an insert, a functional α-peptide was produced that complemented the gene product of *lacI*^{q}*Z*ΔM15, to produce a function β-galactosidase protein producing a blue-indole derivative. If an insert was present in the TOPO multiple cloning site, the *lac*Z α-peptide was interrupted and complementation does not occur leaving the colony to remain white. Therefore, when screening for cloned fragments only white colonies were investigated further. Ten to 15 white colonies were picked for PCR to assess the presence of the inserts using the primers initially used for the amplification of the PCR fragment.

2.20.1. Confirmation of cloned inserts

Between 10-15 well-separated bacterial transformants were picked up and sub-cultured onto appropriate agar plates and incubated overnight. The inserts were confirmed by either one of these methods:

2.20.2. In-well lysis

This is a rapid method to determine the presence of plasmid following transformation. A loopful of an overnight culture was re-suspended in 100 µl of TE buffer in a well of a micro-titre plate. To the suspension, 20 µl of the mixture of P2 lysis buffer (Qiagen, Germany) and loading dye (Sigma, USA) in the ratio of 3:2 was added. The mixture was agarose electrophoresed and viewed under UV-illumination for presence of plasmid.

2.20.3. Amplification of cloned fragment by PCR

Transformants/recombinants were amplified by PCR using the appropriate primers. Electrophoresis was carried out on the PCR products and observed under UV light. The product size shown on agarose gel determined the fragment size of the cloned inserts.

2.20.4. Digestion of clone fragment by restriction endonuclease enzymes

Plasmids were extracted from the transformants/recombinants using commercial plasmid extraction kit, NucleoSpin Plasmid® kit (Macherey-Nagel, Germany). Appropriate restriction endonuclease enzymes were used to cut the plasmids and fragment size was determined by agarose gel electrophoresis.

2.21. Construction of plasmid library

A random plasmid library was constructed by sonicating the whole plasmid (prepared as described in Chapter 2) at 15% power. A small volume of the fragments was electrophoresed to confirm the sizes of the fragments. The sonicated fragments were end-filled by adding T4 DNA polymerase (MBI Fermentas, Lithuania) for an hour at 37°C prior to polyethylene-glycol (PEG) precipitation. The fragments were ligated into *Sma*I digested pUC19 at 22°C overnight and transformed into XL-10 Gold *E. coli* cells (Stratagene, USA), according to the methods of the manufacturer. Transformed colonies were plated onto ampicillin-LB agar plates (50 μg/ml) that were previously coated with IPTG and X-Gal. White colonies were picked and cultured. Following plasmid extraction, the inserts were sequenced using the M13F and M13R primers.

2.22. DNA sequencing

2.22.1. Principles of DNA sequencing

This procedure utilises base-specific termination during *in vitro* synthesis of DNA and therefore also called the chain termination method. The basis for terminating the elongation of the DNA chain is the use of a nucleoside triphosphate containing a ribose sugar reduced in both the 2' and 3' position. Accordingly, such a nucleoside triphosphate is called dideoxyribonucleoside triphosphate, or ddNTP. Because of the missing hydroxyl group at the 3' position of the sugar, DNA polymerase cannot form a phosphodiester with the next dNTP, and the growing chain terminates after the incorporation of a ddNTP. The ddNTP, however, still has the 5' position of the sugar free to react with the 3' end of the previously incorporated dNTP. This reaction occurs at a low frequency. But when such a ddNTP is mixed with the four normal dNTPs at the appropriate concentrations, chain termination will occur over a range of hundreds of base positions. Consequently, the four normal dNTPs are mixed with the respective ddNTPs in four different reaction tubes. If all four reactions are compared, termination occurs randomly at all positions of the growing DNA chain. In the reactions, one of the normal dNTPs is labelled at its alpha position with ^{32}P, in order to incorporate radioactivity with the newly synthesized DNA. When such products are analysed, a balanced level of radioactive signals is achieved for all bands on the ladder gel, since the longer chains incorporate more label but terminate less frequently than the shorter chains.

2.22.2. Templates for sequencing

Templates for sequencing were prepared from either plasmids with inserts, or from PCR products of cloned fragments. Prior to sequencing, plasmids used were purified by using

Nucleospin® Extract 2 in 1 kit (Macherey-Nagel, Germany), while the PCR products were purified by PEG precipitation.

2.22.3. Cycle sequencing

Samples for sequencing was prepared in a 5.0 µl reaction volume. Between 200-500 ng of DNA template from a cloned plasmid, or 30-90 ng of a PCR product were added to a reaction mixture consisting of 2.0 µl Terminator Ready Reaction Mix (BigDye Terminator V.3, PE Applied Biosystems, USA) and 1 µl of sequencing primers.

Cycle sequencing was carried out in a MJ Research Thermal Cycler. The DNA underwent a denaturing stage at 96°C for 30 seconds, an annealing stage at 50°C for 10 seconds and an extension stage at 60°C for 4 minutes. The cycle was repeated for 30 cycles and held at 4°C until use. Finally the DNA for sequencing was purified from the BigDye by sodium/ethanol precipitation technique.

2.22.4. Automated sequencing

Sequencing analysis was carried out by the chain termination method (Sanger *et al.*, 1977) at the Zoology Department of the University of Oxford using the ABI 377 (PE Applied Biosystems, USA) automated sequence reader.

2.23. Computer software for sequence assembly

2.23.1. Staden

Sequence data generated from automated nucleotide sequencing were assembled and edited using the Staden programme software on a UNIX platform. This package of programmes was designed for sequence handling and analysis was used in this project

(Staden, 1996). Staden was developed at the Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.

2.23.2. Phred and Phrap

Phred and Phrap are programs that have been integrated into the Staden program for database handling (Ewing and Green, 1998; Ewing *et al.*, 1998). Phred reads the raw sequence trace files produced by a DNA sequencer performs base calling and assigns quality values, i.e. error rates for each base. These quality values reflect the estimated probability that a base is erroneously called. Phrap uses the quality information provided by Phred, makes pair-wise comparisons of reads and performs the assembly.

2.24. Computer programs for sequence annotation

2.24.1. GeneMark and Glimmer

Potential coding sequences were predicted using programs that are available through Internet websites, i.e. GeneMark (http://opal.biology.gatech.edu/GeneMark/gmhmm2_prok.cgi) and Glimmer (ftp://ftp.tigr.org/pub/software/Glimmer/).

2.24.2. Artemis

Artemis is a software package for the annotation of DNA sequences. It is a power-tool that allows visualisation of the sequence features and analyses of the sequence within its six-frame translation (Rutherford *et al.*, 2000). It was developed by The Welcome Trust Sanger Institute, Cambridge (http://www.sanger.ac.uk/Software/Artemis). To determine the potential protein-coding regions, the nucleotide sequence was analysed for the position of the initiation codons (ATG, GTG or TTG) and of stop codons in all possible

reading frames. The coding sequences (CDS) were confirmed by referring to the codon usage table of *H. influenzae* Rd (http://www.kazusa.or.jp/codon). The nucleotide sequences of the element were entered into Artemis and each sequence of the putative CDS were translated into amino acids and searched for homology with other sequences in the EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/embl.html) and GenBank (http://www.ncbi.nlm.nih.gov) based on the FASTA (Pearson and Lipman, 1988; Pearson, 1990), BLASTP and BLASTX (Altschul *et al.*, 1990; Altschul *et al.*, 1997) algorithms. The default settings used for FASTA are: Program, fasta3; database, protein; an expectation (E) upper value of 10, and lower value set to the default. The E value represents the number of hits one can 'expect' to find by chance when searching a database of a particular size. For the BLAST algorithm, the default settings used was for bacteria; alignment via pairwise, PSI-blast with inclusion threshold of 0.005 and was formatted semi-automatically. Known genes and putative functions were assigned for individual ORFs by inspection of the search outputs.

Potential proteins were examined against the protein feature prediction tools: Transmembrane helix motif (TMHMM) (http://cbs.dtu.dk/services/TMHMM-2.0/), SignalP Prediction Server (http://www.cbs.dtu.dk/services/SignalP/); Protein Motif databases: Prosite (http://www.expansy.ch/prosite/), and Pfam (http://www.sanger.ac.uk/Softwate/Pfam/index.shtml). The analyses were downloaded into Artemis to further characterise the ORFs. The computer program PSORT (http://psort.ims.u-tokyo.ac.jp/form.html) was used to predict protein localisation sites in prokaryotic cells.

2.25. Programs for sequence comparison

2.25.1. GCG

The GCG (Genetics Computer Group Inc., Madison, Wisconsin) is the programme used for sequence comparison including to the sequences from the GenBank and EMBL databases. This package allows the manipulation and analysis of nucleic acid and protein sequences. This package facilitates comparison between two or more sequences, database searching and retrieval, gene finding and pattern recognition, and translation (http://www.molbiol.ox.ac.uk).

2.25.2. Blixem

Blixem was used in the preliminary stage to predict the position and orientation of the ICE*Hin1056* contigs compared to the position of contigs of another mobile element, ICE*Hin299* which was also being whole sequenced (Sonnhammer and Durbin, 1994).

2.25.3. Repeat sequences

Tandem repeat sequences on ICE*Hin1056* were searched using a Tandem Repeat Finder programme available on the website (http://c3.biomath.mssm.edu) (Benson, 1999). Inverted repeats on the sequence were identified using the Palindrome program (http://bioweb.pasteur.fr/seqanal/interfaces/palindrom.html).

2.25.4. Artemis Comparative Tools

The DNA sequences were compared visually by using Artemis Comparison Tool (ACT), developed by the Sanger Institute (http://www.sanger.ac.uk/Software/ACT/). Two or more DNA sequences were compared using ACT that allowed alignment of sequences based on the Artemis program (Rutherford *et al.*, 2000). The nucleotide sequences of the

genomes/plasmids were input and translated by TBLASTX into amino acid for comparison.

2.25.5. ClustalX and ClustalW

The sequence alignment and tree estimation programme ClustalX, was used to align two or more DNA sequences or proteins. It provided an integrated environment for performing multiple sequence and profile alignments and analysing the results (http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html). This programme is able to match the selected sequences, line them up so that the identities, similarities and differences can be seen. Phylogenetic trees were constructed from a total of 100 bootstrapped replicate re-sampling data sets to provide confidence estimates for tree topologies and viewed using TreeView (http://taxonomy.zoology.gla.ac.uk/rod/rod.html).

ClustalW was used to estimate the percentage of sequence homology between genes using pairwise alignment (http://www.ebi.ac.uk/clustalw/).

2.25.6. Molecular Evolutionary Genetic Analysis (MEGA)

A programme to estimate evolutionary distances and reconstructing phylogenetic trees from sequence analyses was used (Kumar *et al.*, 2001). Aligned sequences from ClustalX were converted into MEGA files for construction of UPGMA with bootstrap values of 1000 repetitions.

2.25.7. SplitsTree decomposition

MEGA files saved as nexus files were downloaded into a programme called SplitsTree (http:bibiserv.techfak.uni-bielefeld.de/splits). This programme uses the split

decomposition method to analyse and visualise distances between gene sequences in the ICEs sequences.

CHAPTER THREE

PHENOTYPIC CHARACTERISTICS *H. INFLUENZAE* STRAIN 1056

3.1. Introduction

Bacterial resistance to antimicrobial drugs is increasing worldwide and is producing an impact on medical practice and antibiotic treatment costs of patients (Cohen, 1992; Cosgrove and Carmeli, 2003). Antibiotic resistance is often associated with mobile elements carrying antibiotic resistance genes.

In clinical isolates of *H. influenzae*, antibiotic resistant genes are most often found integrated into the chromosome, but following transfer by conjugation to a recipient strain, extra-chromosomal plasmid is usually detectable in the recipient (Dimopoulou *et al.*, 1997; Stuy, 1979). This type of element that can integrate and excise, and transfer by conjugation has been termed an integrative and conjugative element (ICE) (Burrus *et al.*, 2002a).

3.2. Objectives

Dimopoulou *et al.* (1992) earlier studied the mobilization of mobile elements present in *H. influenzae* and isolated in the UK (Dimopoulou *et al.*, 1992). In this chapter, the phenotypic characteristics of the mobile element found in *H. influenzae* 1056 were investigated. First, to confirm the earlier studies of Dimopoulou *et al.* (1992) and second to study transfer under a range of additional conditions. It was also important to characterise the conjugative properties prior to fully sequencing a mobile element.

3.3. Materials and Methods

3.3.1. Bacterial strains

Donor and recipient strains used in *H. influenzae* in conjugation studies are shown in Table 3.1.

Table 3.1. *H. influenzae* strains used in conjugation experiment

	***H. influenzae* strain**
Donor	1056
	XT3
	XT11
Recipient	Rd^{sm}*recA*$^{+}$
	Rd^{sm}*recA*$^{-}$
	Rd^{rif} *recA*$^{+}$
	XT9

3.3.2. Minimum inhibition concentration test

The minimum inhibition concentration (MIC) was performed, as described in the materials and methods (Chapter 2) on the *H. influenzae* strains 1056 and Rd ($strep^{r}$) against ampicillin, chloramphenicol, tetracycline and streptomycin.

3.4. Results and discussion

3.4.1. Minimum inhibition concentration (MIC)

The sensitivity/resistance of the strains against antibiotics was carried out prior to mating experiments. The MIC of the isolates is shown in Table 3.2. Strain 1056 is resistant to ampicillin, chloramphenicol and tetracycline.

Table 3.2. Minimum inhibition concentration of strains of *H. influenzae*

Strain	Minimum inhibition concentration (μg/ml)			
	Ampicillin	**Chloramphenicol**	**Tetracycline**	**Streptomycin**
1056	96	16	16	0.25
$Rd^{sm}rec^{+}$	0.25	0.25	0.25	>1024

3.4.2. To determine the rate of plasmid transfer to $recA^{+}$ *and* $recA^{-}$ *recipients*

In this experiment, the conjugative properties of this mobile element, p1056 was conjugated with Rd strain. To control for the possible role of transformation and homologous recombination, the frequency of transfer between strain 1056 to both $recA^{+}$ and $recA^{-}$ strains of Rd was studied. The $recA^{+}$ gene product in *H. influenzae* is necessary for transformation and homologous recombination (Stuy, 1980b). In its absence, acquisition of resistance genes by transformation will not occur.

Conjugation experiments were carried out both on solid and in liquid media. The result showed in Table 3.3 shows that the rate of transfer of the p1056 by conjugation on solid medium is $\sim 10^{-6}$ per donor cell. This result was in agreement with the findings of Stuy (1979) and Dimopoulou *et al.* (1992), who obtained a frequency of transfer of 10^{-7} per donor cell (Dimopoulou *et al.*, 1992; Stuy, 1979). No significant difference in the efficiency of transfer between the $recA^{+}$ and $recA^{-}$ recipients was observed. This essentially eliminates transformation as the main mechanism of transfer. This is consistent with the findings of Dimopoulou *et al.* (1992) where it was shown that DNAse had no effect on transfer efficiency (Dimopoulou *et al.*, 1992).

The efficiency of transfer in liquid medium was 3-log orders lower than on solid media. A similar observation was also obtained during transfer of chloramphenicol and tetracycline resistance into *H. influenzae* recipients by other investigators (van Klingeren *et al.*, 1977). Stuy (1979), however, failed to observe any conjugative transfer in liquid suspension (Stuy, 1979).

Transconjugants selected on tetracycline and streptomycin resistant plates were picked and checked for ampicillin and chloramphenicol resistance. Co-transfer of antibiotic resistance was observed, consistent with the successful transfer of an element as a unit.

Table 3.3. Determination of the rate of transfer of antibiotic resistance of *H. influenzae*. Matings of *H. influenzae* strain 1056 (amp^r, tet^r, chlor^r) with *H. influenzae Rd* (strep^r) on solid and liquid medium.

	Donor[a] *H. influenzae* 1056 (cfu)	**Recipient[b] *H. influenzae* Rd (cfu)**		**Trans-conjugant (cfu)**	**Rate of transfer (No. of transconjugants per donor cell)**
		***recA*$^+$**	***recA*$^-$**		
Selection	Tet	Sm	Sm	Tet&Sm	
Mating experiment on solid medium					
1056 x Rd^{sm}*recA*$^+$	7.3×10^{10}	5.6×10^{11}	-	1.7×10^{5}	2.3×10^{-6}
1056 x Rd^{sm}*recA*$^-$	5.3×10^{10}	-	2.1×10^{11}	6.1×10^{4}	1.2×10^{-6}
Mating experiment in liquid medium					
1056 x Rd^{sm}*recA*$^+$	2.3×10^{10}	2.7×10^{11}	-	1.2×10^{2}	5.2×10^{-9}
1056 x Rd^{sm}*recA*$^-$	3.1×10^{10}	-	2.9×10^{11}	1.1×10^{2}	3.5×10^{-9}

Tet: tetracycline (2 μg/ml), Sm: streptomycin (20 μg/ml). Results represent the average of experiments done in duplicate twice. Within experimental variation was less than half of a log. [a]No. of cells resistant to streptomycin $\leq 10^{-6}$ cfu; [b]No. of cells resistant to tetracycline $\leq 10^{-10}$ cfu.

3.4.3. Re-transfer from a transconjugant

In this experiment, the transconjugant XT3 (amp^r, tet^r, chlor^r in strep^r host) that was obtained from the mating of Rd^{sm}*recA*$^+$, was used as the donor and Rd^{rif}*recA*$^+$ was used as the recipient. The experiment was carried out using either solid or liquid medium. Table 3.4 shows the summary of the results. High frequency transfer of $\sim 10^{-1}$ per donor cell was achieved on solid medium. The rate of transfer in liquid medium was $\sim 10^{-7}$ per donor cell.

The 5-log increase in the frequency of transfer of the transconjugant XT3 over the original isolate 1056, is dramatic. This difference has been previously reported by Thorne and Farrar (1975), Stuy (1979) and Deneer *et al.* (1982) (Deneer *et al.*, 1982; Stuy, 1979; Thorne and Farrar, 1975). The difference in frequency of transfer has been inferred to indicate that the resistance element is chromosomally integrated in the primary isolate while predominantly extra-chromosomal in the transconjugant. This is plausible as excision of the element is likely to be rate limiting in conjugal transfer.

Table 3.4. Determination of the rate of secondary transfer of antibiotic resistance of *H. influenzae*. Matings of *H. influenzae* XT3 (amp^r, tet^r, $chlor^r$, $strep^r$, *recA*$^-$) with *H. influenzae* Rd^{rif}*recA*$^+$ on solid and liquid medium.

	Donor[a] *H. influenzae* XT3 (cfu)	**Recipient[b] *H. influenzae* Rd^{sm}*recA*$^+$ (cfu)**	**Trans-conjugant (cfu)**	**Rate of re-transfer (No. of transconjugants per donor cell)**
Selection	Tet	Rif	Tet&Rif	
Mating experiment on solid medium				
XT3 x Rd^{rif}*recA*$^+$	1.2×10^{8}	3.6×10^{9}	2.0×10^{7}	1.6×10^{-1}
Mating experiment in liquid medium				
XT3 x Rd^{rif}*recA*$^+$	5.0×10^{10}	9.1×10^{11}	2.3×10^{4}	4.6×10^{-7}

Tet: tetracycline (2 μg/ml), Rif: rifampicin (20 μg/ml). Results represent the average of experiments done in duplicate twice. Within experimental variation was less than half of a log. [a]No. of cells resistant to rifampicin $\leq 10^{-9}$ cfu; [b] No. of cells resistant to tetracycline $\leq 10^{-10}$ cfu.

3.4.4. Evidence for incompatibility

In this experiment, the effect of the element in the recipient cell on the proficiency transfer was investigated. Mating experiments involved the donor strain XT11 and recipient strain XT9 and was performed as described above. Strain XT11, the donor was mated with strain XT9. Transconjugants were selected on streptomycin and tetracycline. XT11 was mated into Rd^{sm}*recA*$^+$ as a control.

Table 3.5 shows the rate of transfer. There is 3-log reduction of transfer of the element into a recipient already containing the element over a recipient without the element. Incompatibility is usually associated with both plasmids sharing the same replication of origin (Hardy, 1986). Incompatibility may have different causes including randomisation during replication and partition, and systems that stabilise plasmid maintenance by killing plasmid-free segments from within may give rise to incompatibility effects. Randomisation takes place during replication as well as during partition and leads to distortion of the ratio. However, in this experiment there was no complete exclusion. Such reduced efficiency of transfer has been observed with SXT (M. Waldor, personal communication). The mechanism explaining this reduction is not known.

Table 3.5. Determine of the rate of incompatibility in *H. influenzae*. Matings of *H. influenzae* XT11 (amp^r, tet^r, chlor^r, rif^r, *recA*$^+$) with *H. influenzae* XT9 (amp^r, strep^r, *recA*$^-$) and Rd^{sm}*recA*$^+$ on solid medium.

	Donor[a] *H. influenzae* XT11 (cfu)	**Recipient[b] *H. influenzae* XT9/Rd^{sm}*recA*$^+$ (cfu)**	**Trans-conjugant (cfu)**	**Rate of re-transfer (No. of transconjugants per donor cell)**
Selection	Tet	Strep	Tet&Strep	
XT11 x XT9	8.7×10^{11}	6.1×10^{12}	1.4×10^{6}	1.6×10^{-6}
XT11 x Rd^{sm}*recA*$^+$	3.2×10^{11}	3.9×10^{12}	2.0×10^{9}	6.2×10^{-3}

Tet: tetracycline (2 μg/ml), Strep: streptomycin (20 μg/ml). Results represent the average of experiments done in duplicate twice. Within experimental variation was less than half of a log. [a]No. of cells resistant to streptomycin $\leq 10^{-8}$ cfu; [b]No. of cells resistant to tetracycline $\leq 10^{-10}$ cfu.

3.4.5. Electron microscopy for pilus

The electronmicroscopist was blinded to the samples. The controls consisted of *H. influenzae* Rd and experimental samples consisted of XT3. Each experiment was done in triplicate and repeated once. Long thin pili as seen in Fig. 3.1 were only observed in grids containing XT3, but all the pili were detached from the cells. The pili observed were

morphologically unlike those reported for piliated *H. influenzae* by Guerina *et al.* (1982) and Pichichero *et al.* (1982) (Guerina *et al.*, 1982; Pichichero *et al.*, 1982). Those pili were thought to be involved in cell adhesions.

Conjugative plasmids usually produce pili (Bradley, 1980). In a conjugative plasmid, the presence of pili is believed to be associated with conjugative activity. For DNA transfer by conjugation to take place, physical contact is required between a recipient cell to a donor cell. However, due to the charges associated with their cell outer layers, bacterial cells are naturally repellent to one another. Thus the cells usually require extracellular filaments to form the initial contacts to allow two cells to interact. During this process, the pilus of a donor cell will retract the surface of a recipient cell for intimate association, followed by the formation of mating bridge to serve as a conduit for DNA (Silverman, 1997; Waters, 1999). The presence of pili in *H. influenzae* XT3 is consistent with its functional role in conjugative DNA transfer.

Conjugative pili fall into three basic morphological types which are described as: thin flexible (*IncF*), thick flexible (*IncI*) and thin rigid filaments or rods (*IncN* and *IncP*) (Bradley, 1980). From the electron-micrograph (Fig 3.1), the pili of *H. influenzae* XT3 appear to be thin and rigid. It is apparently similar to the features of the sex pili of the *IncN* and *IncP* plasmids that are thin and rigid (Bradley, 1983). However, visual observation is not a reliable method to classify the type of pilus.

To predict the type of pilus, Bradley *et al.* (1980) have proposed a criterion by calculating the ratio of solid mating frequency to the ratio of liquid mating frequency. From the results obtained above (Tables 3.2 to 3.5), a comparison of transfer frequency between solid and liquid media is tabulated in Table 3.6. A frequency ratio of approximately 1, indicates that the transfer was as efficient in solid as in liquid, which is typical of plasmids with thin flexible pili. If the frequency ratio ranges between 45 and

470, it indicates thick flexible pili, but a plasmid with rigid pili would have a very high ratio of transfer frequency (>470) (Bradley *et al.*, 1980).

The results obtained from the experiments above, shows that there is a very high ratio of transfer frequency in solid medium compared to that in liquid medium. This also suggests that p1056 possess rigid type pili. This renders them fragile and they fragment. The lack of intact pili on EM is also consistent with such fragile rigid pili.

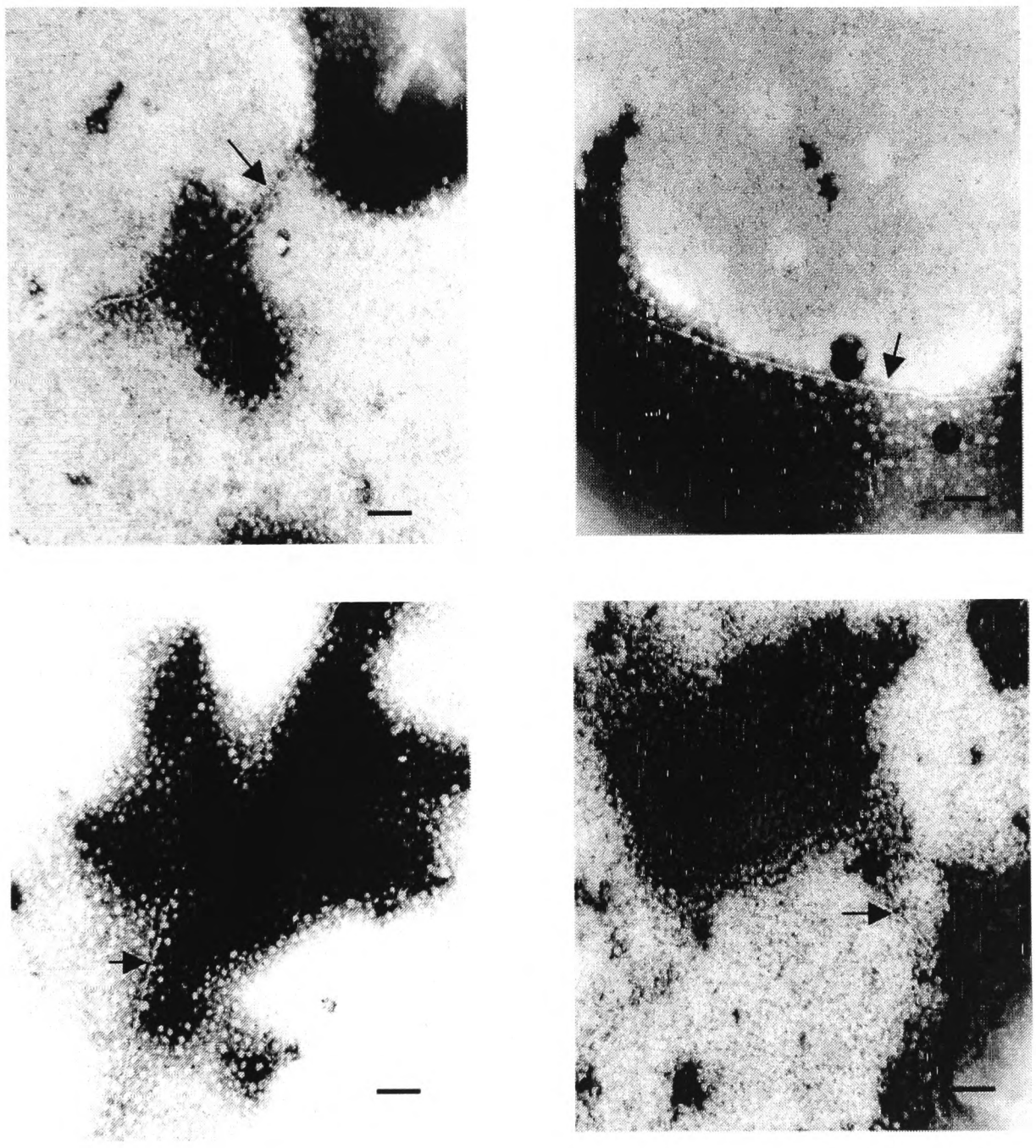


Fig 3.1. Electron micrographs of pili of *H. influenzae* 1056. The arrows on the figures show the long fine pili. The bar represents 100 nm.

Table 3.6. A comparison of the ratio of frequency of transfer in solid and liquid media.

Donor	Recipient	Rate of transfer (No. of transconjugant per donor cell)		Ratio of transfer frequency (No. of transconjugant per donor cell)
		Solid	Liquid	
1056	$Rd^{sm}recA^{+}$	2.3×10^{-6}	5.2×10^{-9}	4.4×10^{4}
1056	$Rd^{sm}recA^{-}$	1.2×10^{-6}	3.5×10^{-9}	3.4×10^{4}
XT3	$Rd^{sm}recA^{-}$	1.6×10^{-1}	4.5×10^{-7}	3.5×10^{7}

3.5. Conclusion

The experiments carried out in this chapter showed that the *H. influenzae 1056* element p1056 was transferred by conjugation to a recipient. The increased efficiency of transfer by a transconjugant as a donor is consistent with excision from the chromosome in the primary isolate and on extra-chromosomal locations in the transconjugant.

The presence of a pilus associated with the presence of p1056 is consistent with it being the sex pilus. From these results, it can be concluded that *H. influenzae* 1056 contains a self-transmissible element. Given the observation of Dimopoulou *et al.* (1992) showing integration at $tRNA^{leu}$ this is an example of an integrative and conjugative element or ICE (Dimopoulou *et al.*, 1992). The element will therefore, be referred to as ICE*Hin1056* (previously p1056).

CHAPTER FOUR

COMPLETE NUCLEOTIDE SEQUENCE OF THE MOBILE ELEMENT ICE*Hin1056* OF *H. INFLUENZAE* 1056

4.1. Introduction

To date, more than 501 complete sequences of plasmid have been deposited into the GenBank, but so far, only four plasmids found in *Haemophilus* spp. have been completely sequenced. Two of these were from *H. influenzae* biotype aegyptius (NC_004058 and NC_004846), one each from *H. ducreyi* (NC_005329) and from *H. paragallinarum* (NC_005245).

Although both plasmids found in *H. influenzae* biogroup *aegyptius* (*H. aegyptius*) 32-kb pF3028 (Kroll *et al.*, 2002), and 32-kb pF3031 (Mukhopadhyay, S. and Actis, L.A., unpublished) were associated with Brazilian purpuric fever none encode virulence traits. A small 6-kb plasmid pNAD1, isolated from *H. ducreyi*, contains the *nadV* gene that confers NAD independence (Martin *et al.*, 2001). Another Haemophilus 6-kb plasmid, p250 was recently deposited in the NCBI and was isolated from *H. paragallinarum* strain HP250, a strain predominantly found in avian hosts. The complete sequence of this plasmid revealed a haemocin (bacteriocin) operon similar to those of *H. influenzae* (Terry *et al.*, 2003). To date, no plasmids of *H. influenzae* type b have been completely sequenced and submitted into the GenBank.

4.2. Objectives of whole sequencing of ICE*Hin1056*

Antibiotic resistance in bacteria has been associated with the presence of plasmids (Khan *et al.*, 1974). Plasmids encoding gene resistance to antibiotics play an important role in the acquisition and transmission of intra-species and inter-species of bacteria. The factors

that influence the rapid spread of antibiotic resistance in bacteria are unclear. An important step in understanding the genetic mechanism involved in the transmission of antibiotic resistance in *H. influenzae*, is the analysis of the complete sequence of an antibiotic resistance element (ICE). The analysis of the nucleotide sequence and the core genes of a resistance plasmid will:

(i) provide an understanding on the genetic relationship between resistant elements;

(ii) provide an understanding on the genetic relationship between elements from different *Haemophilus* spp. from various geographical locations; and

(iii) enable early predictions as to the biological role of the large conjugative *H. influenzae* ICEs.

In this study, the whole nucleotide sequence and annotation of ICE*Hin1056* is produced and investigated.

4.3. Materials and methods

In order to obtain the complete sequence of ICE*Hin1056*, both sequencing strategies: the primer walking and the shotgun methods were used.

Whole ICE*Hin1056* was previously digested with *Pst1* and run in an agarose gel (Dimopoulou *et al.*, 1997). The fragments from the gel were excised and ligated into pBluescript vector and transformed into DH5α *E. coli* cells. A 12-kb fragment clone, designated pB7 was initially sequenced using universal M13 primers. The sequences obtained were used to design subsequent primers to 'walk' into the insert, towards each end of the cloned fragment. As an alternative, shotgun approached was used. The whole element was sonicated, ligated into pUC19 and plasmid libraries were constructed. The inserts were sequenced, assembled and analysed using computer programs.

4.4. Results and discussion

4.4.1. Primer walking

Sequencing by primer walking on pB7 was simple except that it was time consuming because it required many primers to be designed. New primers could only be designed when the previous sequence results were obtained. Therefore, the shotgun approach was then adopted to expedite whole sequencing.

4.4.2. To determine the optimum length of time for sonication

A comparative study was carried out to determine the optimum length of time for sonication. It showed that a 10-s sonication produced fragments of the required sizes. Prolonged sonication resulted in smaller fragments (Fig. 4.1). Fig 4.1 shows that when DNA plasmid was sonicated for 10 s, the fragment sizes were of the desired size. When sonication was carried out for 20 s, the sheared DNA produced a smear, as shown in lane 5. Thus a 10-s sonication was then adopted.

4.4.3. Construction of plasmid libraries

Six random plasmid libraries for ICE*Hin1056* were constructed. Most of the inserts were found to be smaller than 500 bp indicating inefficiency during transformation. Only plasmids with inserts between 500 bp to 2.0 kb length were selected for sequencing as shown in Fig. 4.2.

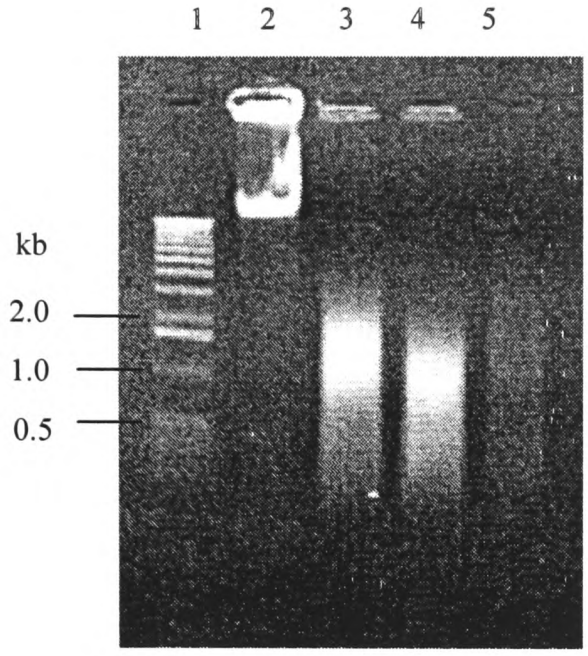


Fig. 4.1. A comparative study to determine the optimum sonication time for plasmid to obtain fragments of sizes between 500 bp to 2.0 kb. Lane 1: 1-kb DNA ladder; lane 2: uncut chromosomal DNA; lane 3-6 shows various sonication durations; lane 3: 10 s; lane 4: 15 s and lane 5: 20 s.

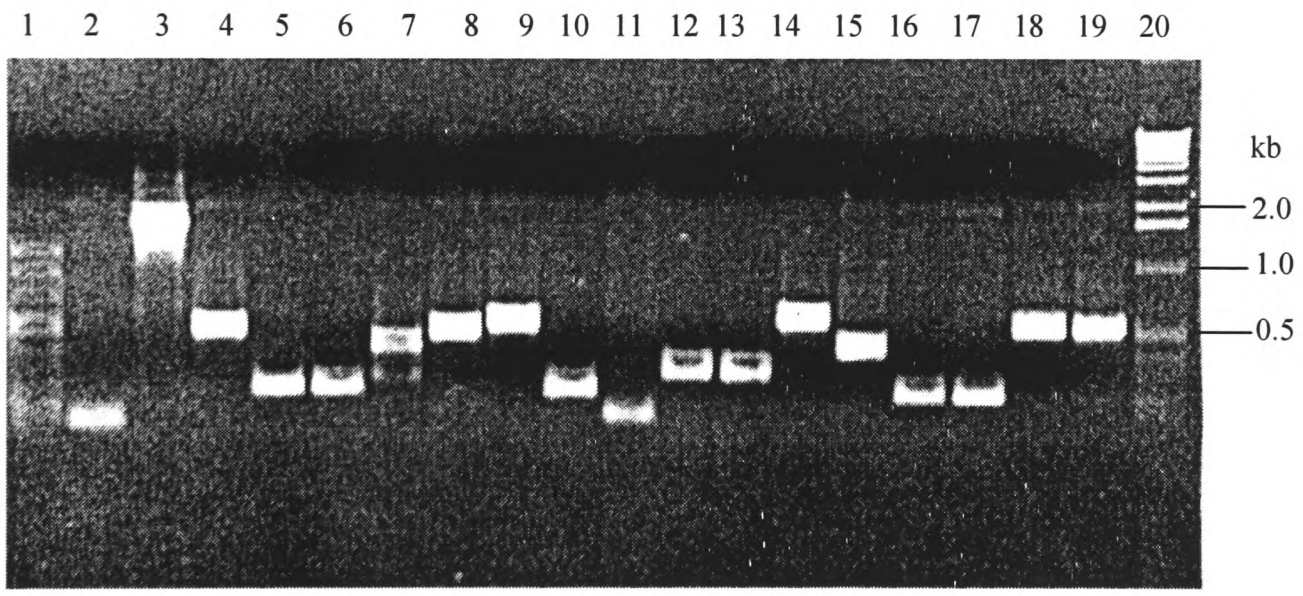


Fig. 4.2. Gel electrophoresis of inserts (by PCR) of plasmid clones derived from the random library. Lanes 1 & 20: 1-kb DNA ladder, lane 2: pUC19 (negative control) and lanes 3-19: pUC19 with random size of inserts.

4.4.4. *Sequence assembly of ICE*Hin1056

All of the sequences obtained from random libraries and by primer walking were assembled. The average usable read length analysed by the ABI 377 sequencer was approximately 500 bp. The sequences were obtained from both forward and reverse strands. In the initial stage of assembly, there were small contigs consisting of one or two sequences and a few contigs with an average length of 8 to 17 kb. When blast searches of these contigs were carried out, most of the small contigs were found to be *H. influenzae* Rd sequences. This suggested that ICE*Hin1056* DNA was contaminated with chromosomal DNA. Improved EtBr-CsCl purified plasmid DNA was used for the later libraries of ICE*Hin1056* (Fig. 4.3). Sequences that belonged to *H. influenzae* Rd chromosomal sequences were removed manually.

A total of 31 assemblies were carried out. Ultimately, 679 sequences were needed to assemble the complete element. Before achieving closure of a circular element, manual editing was needed to re-arrange tracefiles belonging to either of the two large repeats. The presence of the repeats was predicted from examining the template display (Fig. 4.4). The assembly adjacent to this 'pile-up' contained two species of sequences (Fig. 4.5). When these tracefiles were manually edited the assembly resulted in one large contig of 59,393 bp. The average length of the readings was 569 bp, producing a total length of 386,236 bases, i.e. 6.5 fold of the element.

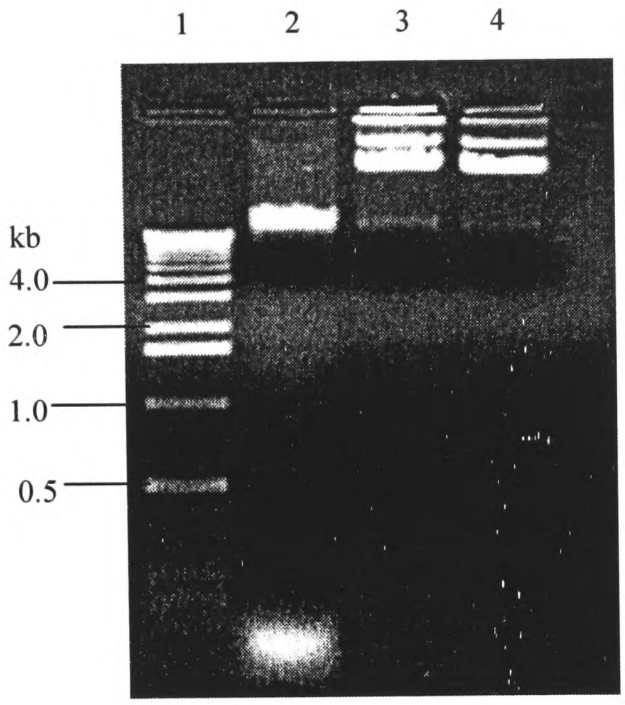


Fig. 4.3. Agarose gel electrophoresis of chromosomal and plasmid DNAs prepared by CsCl-EtBr gradient centrifugation method. Lane 1: 1-kb DNA ladder, lane 2: chromosomal DNA of *H. influenzae* 1056, lane 3: ICE*Hin1056* and lane 4: ICE*Hin299*.

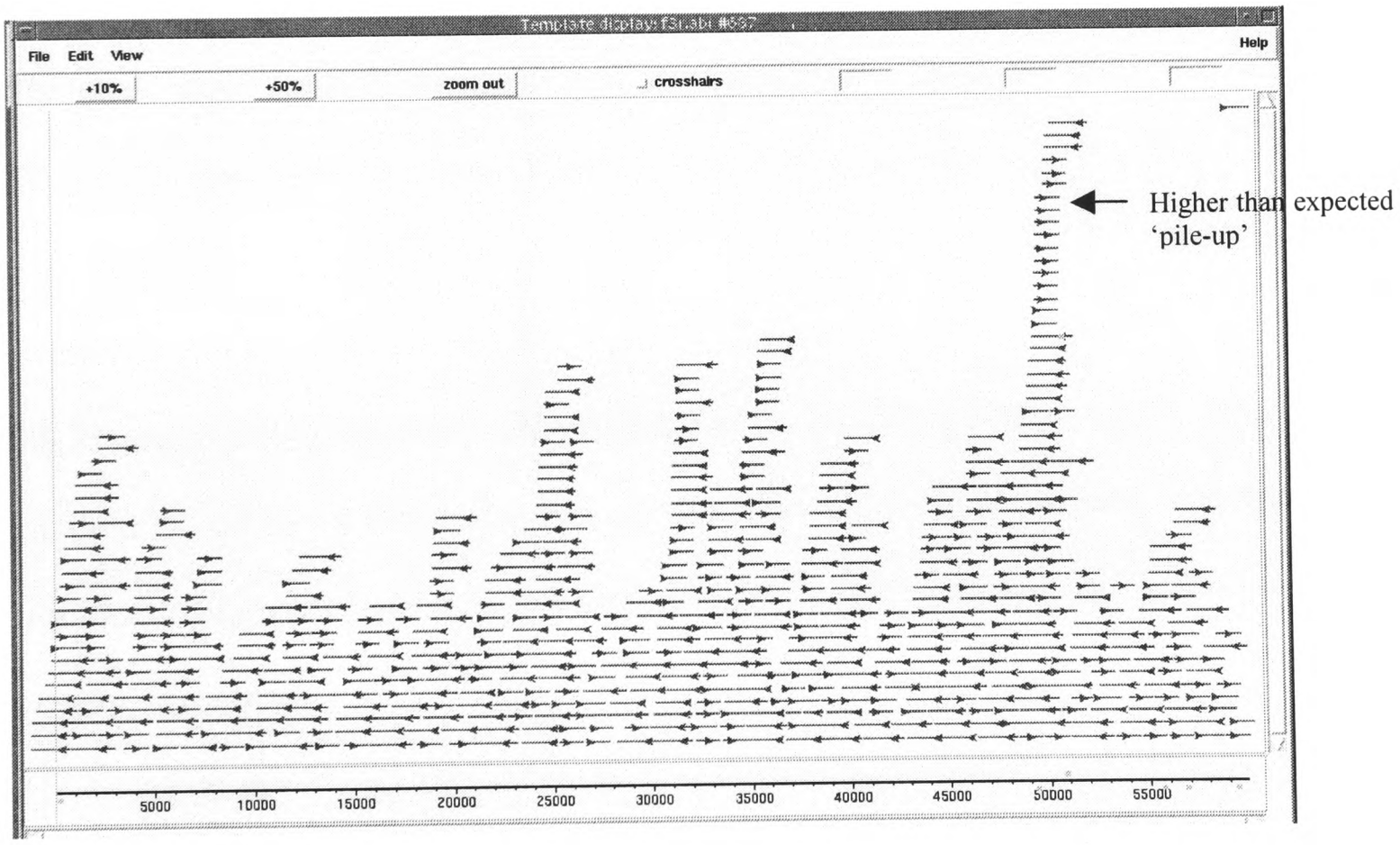


Fig. 4.4. Template display of a single contig of ICE*Hin1056*. Note the pile-up of template sequences at 50-kb position.

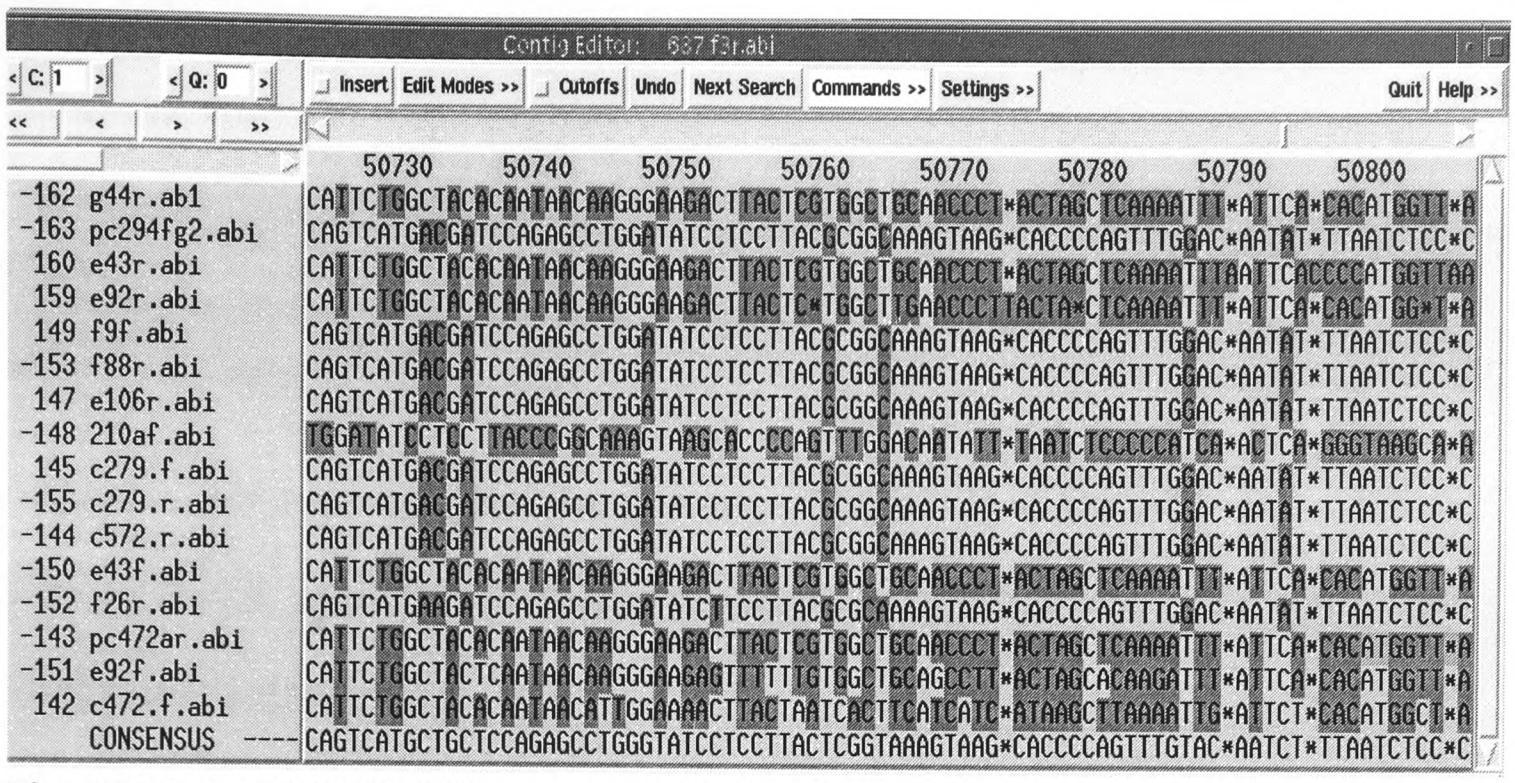


Fig. 4.5. Part of ICE*Hin1056* sequences displayed on contig editor. Note the two patterns of sequences highlighted in dark and clear background.

*4.4.5. Assignment of potential coding regions of ICE*Hin1056

The whole sequence of ICE*Hin1056* was loaded into the Artemis program for annotation.

To predict the coding regions, the entire nucleotide sequence was searched for ORFs >100 amino acids. A total of 64 ORFs were obtained by using ATG, GTG or TTG as the translational start codons and TTA, TAG or TGA as the stop codons. Not all of these ORFs were preceded by Shine-Dalgano sequences. This analysis showed that the element contains 64 ORFs with 48 in the forward direction and 16 in the reverse direction. This directional bias therefore, indicates that the ICE*Hin1056* replication proceeds mostly in forward orientation.

All of the ORFs were translated into amino acid sequences and then subjected to homology search analysis using FASTA, BLASTP and BLASTX, to compare the protein sequences available in NCBI (National Centre Biotechnology Information) and EMBL (European Molecular Biology Laboratory). From the protein homology search, the ORFs of ICE*Hin1056* were shown to encode 62 complete gene proteins and two pseudo-genes.

Amongst the complete genes, 32 showed similarity to genes of *H. ducreyi* 35000HP (acc. no.: AE017153). Of the complete genes, 22 showed homology to genes with known functions, while the remaining genes have not been functionally characterised. Six of the hypothetical proteins did not show any significant matches with any sequences deposited in public databases. No ORFs directly related to bacterial pathogenesis were identified.

*4.4.6. General properties of ICE*Hin1056

The nucleotide position 1 of the mobile element was orientated to correspond to the start of the *attP* sequence. This point was chosen because it is the common location for site-specific recombination and is present in many integrative mobile elements. ICE*Hin1056* has an overall G+C% content of 39.1%, which is slightly higher than the G+C% content of *H. influenzae* Rd (38.2%). The G+C content is unevenly distributed around the circular mobile element (Fig. 4.6). Two regions showed peaks of G+C% content higher than the mean. The first region lies between position 14,581 to 20, 370-bp and has a G+C % content of 45.9%, while the second region lies between 50,667 and 52,733 bp and has a G+C% content of 45.3%. In comparison, the rest of the element has a G+C% content of 36.5%. The anomalies of the G+C content at these regions correlates to the presence of transposons Tn*3* and Tn*10*. Overall percentages of the purines and pyrimidines of the whole element are: A (30.6%), C (19.1%), G (20.0%) and T (30.3%).

Throughout the 60-kb element, nine copies of the 9-bp Haemophilus uptake sequence (5'-AAGTGCGGT) were identified. The genome of *H. influenzae* Rd has 1,465 copies of uptake sequence (Fleischmann *et al.*, 1995). This means that the element has approximately one copy of uptake sequence for every 6.5-kb of the element, in comparison to one copy of the uptake sequence to every 1.2-kb of the genome. Similar numbers of uptake sequence has also been observed in a 24-MDa plasmid isolated from

H. influenzae biotype aegyptius (Kroll *et al.*, 2002). The distribution of the uptake sequence in the whole element is random but is mostly located in the intergenic regions. The general features of ICE*Hin1056* are shown in Table 4.1.

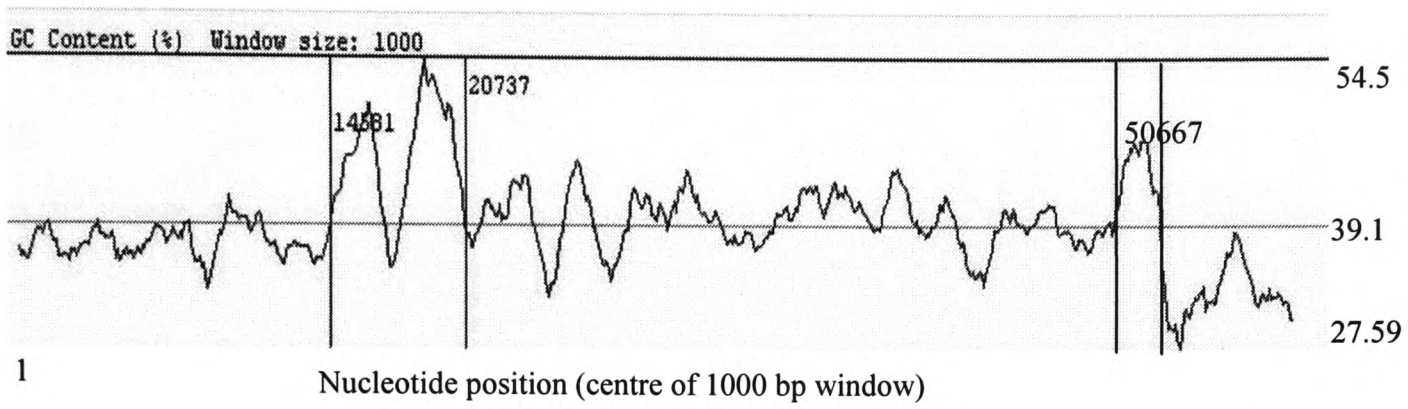


Fig. 4.6. G+C% content of ICE*Hin1056*, based on a 1000-bp window. The coordinates of the regions of high G+C% content are shown. This graph was obtained by using Artemis program.

Table 4.1. Summary of ICE*Hin1056* properties.

Property	**Value**
Total size	59,393 bp
G+C content	39.88%
Predicted coding sequences	62
Pseudo-genes	2
Coding density	84.4%
Average gene length	798
No. of uptake sequence	9

*4.4.7. Overview of the genetic organisation of ICE*Hin1056

Computational analysis of the ORFs of ICE*Hin1056* is illustrated in Fig. 4.7, and the annotations are summarised in Table 4.2. From the annotation of the assembly, it was observed that the genes of the element were arranged in three putative functional modules (Fig. 4.8). The first module that lies from position 1 to 14,719-bp consists of a cluster of putative genes encoding proteins involve in replication. The second module that lies

between 14,720 to 26,581-bp comprises tetracycline and chloramphenicol resistance genes. The third, being the largest module, lies between position 26,582 to 59,393-bp, consists of a number of hypothetical proteins, five homologues of genes associated with DNA transfer and one being an integrase gene. However, within this region, ampicillin resistance genes homologous to Tn*3* is inserted that lies from position 50,985 to 52,585-bp.

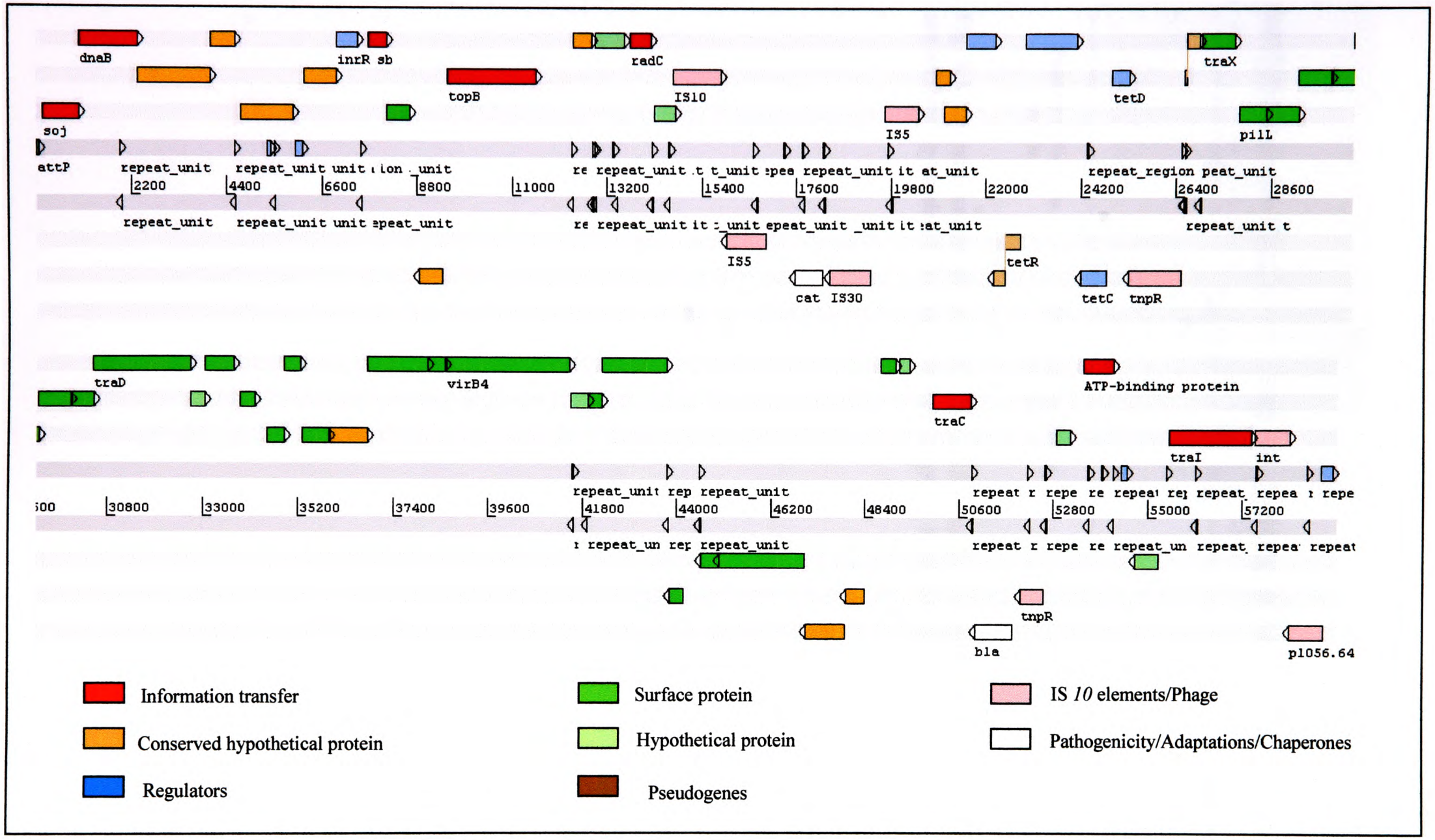


Fig. 4.7. Genetic organisation of a 60-kb mobile element of *H. influenzae*, ICE*Hin1056* generated by Artemis. Genes are represented by arrow boxes. Putative functions of the genes are described in detail in Table 4.2.

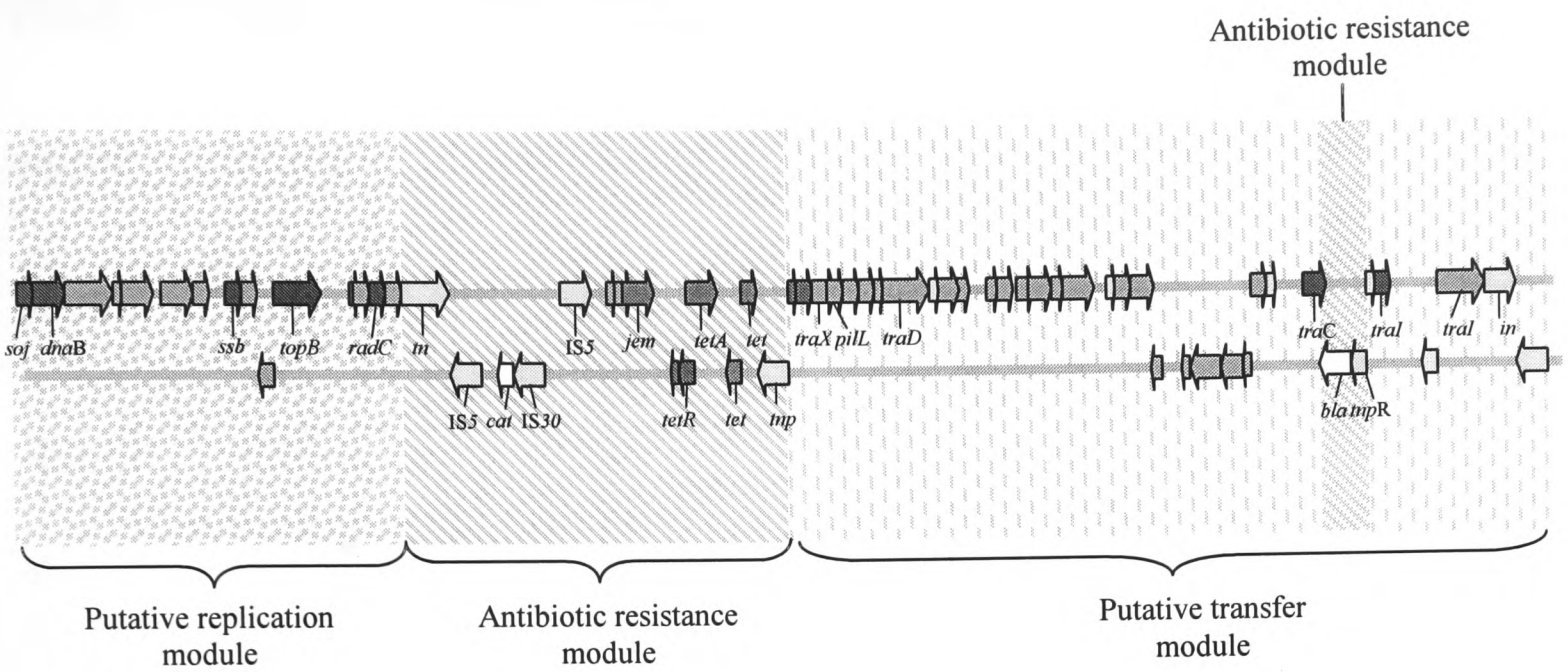


Fig. 4.8. Genetic organisation of ICE*Hin1056* according to their putative functions. The modules are differently shaded to indicate clusters of genes in the same function module.

Table 4.2. List of predicted genes in ICE*Hin1056*. Amino acid similarity of putative proteins encoded by ICE*Hin1056* to sequences in public databases.

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (% aa identity)	E value
p1056.01	141-977	*soj*	Putative chromosome partitioning protein	Similar to *H. ducreyi* 35000HP possible chromosome partitioning related protein SWALL:AAP95855 (EMBL:AE017153) (83.33 in 276 aa)	1.1e-93
p1056.02	979-2334	*dnaB*	Replicative DNA helicase	Similar to *H. ducreyi* 35000HP replicative DNA helicase DnaB*1* or HD0973 SWALL:AAP95854 (EMBL:AE017153) (70.94 in 444 aa)	5.2e-119
p1056.3	2327-4015		Conserved hypothetical protein	Similar to central region of *H. ducreyi* 35000HP hypothetical protein HD0971 SWALL:AAP95852 (EMBL:AE017153) (70.18 in 218 aa)	2.5e-53
p1056.04	4015-4566		Conserved hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0969 SWALL:AAP95850 (EMBL:AE107153) (79.0 in 181 aa)	2.6e-56
p1056.05	4716-5939		Conserved hypothetical protein	Similar to N-terminal region of *H. ducreyi* 35000HP hypothetical protein HD0968 SWALL:AAP95849 (EMBL:AE017153)(68.35 in 237 aa)	1.5e-58
p1056.06	6164-6919		Conserved hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0966 (SWALL:AAP95847) (EMBL:AE017153) (72.44 in 254 aa)	1e-71
p1056.07	6919-7407	*inrR*	Putative integrase regulator R	Low similarity to *H. ducreyi* hypothetical protein SWALL:AAP958445 (EMBL:AE017153) (74.41 in 43 aa), and to *Pseudomonas sp.* B13 integrase regulator R SWALL:Q8341 (EMBL:AJ536665) (28.5 in 134 aa)	5.3e-06; 2.6e-05
p1056.08	7654-8076	*ssb*	Single-strand binding protein	Similar to *H. influenzae* single-strand binding protein ssb or HI0250 (SWALL:P44409) (EMBL:HI04997) (66.66 in 168 aa)	5.1e-40
p1056.09	8100-8621		Putative lipoprotein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0961 SWALL:AAP95842 (EMBL:AE017153) (46.51 in 172 aa)	6.9e-23
p1056.10c*	8837-9394c	*osa*	Hypothetical protein	Similar to the C-terminal region of *S. flexneri* protein Osa SWALL:P29772 (EMBL:PP30471) (36.97 in 119 aa)	8.6e-08
p1056.11	9482-11530	*topB*	DNA topo-isomerase III	Similar to *H. ducreyi* 35000HP DNA topoisomerase III topB1 or HD0954 SWALL:AAP95835 (EMBL:AE017153) (79.11 in 450 aa)	1.3e-139
p1056.12	12415-12843		Conserved hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0951 SWALL:AAP95832 (EMBL:AE017153) (57.63 in 144 aa)	5.4e-33
p1056.13	12919-13599		Hypothetical protein	No significant database matches	
p1056.14	13738-14208	*radC*	Putative DNA-repair-like protein	Similar to *H. ducreyi* 35000HP conserved possible DNA repair protein HD0945 SWALL:AAP95827 (EMBL:AE017153) (41.02 in 156 aa)	3e-20
p1056.15	14301-14780		Hypothetical protein	Very low similarity to *H. ducreyi* 35000HP hypothetical protein HD0944 SWALL:AAP95826 (EMBL:AE017153)	0.87

*c: reverse direction

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (% aa identity)	E value
p1056.16	14720-15838	IS*10*	Transposase of IS*10*	Identical to *E. coli* and *Salmonella* Typhi transposase (IS*10* transposase) R0085 SWALL:Q53371 (EMBL:S67119) (100 in 365 aa)	1e-154
p1056.17	15964-16896	IS*5*	Transposes for IS*5*	Similar *E. coli* transposase *insH* for insertion sequence element IS5 SWALL:P03837 (EMBL:U00039) (55.66 in 300 aa)	1.8e-68
p1056.18	17570-18211	*cat*	Chloram-phenicol acetyl-transferase	Identical to *Pasteurella piscicida* chloramphenicol acetyltransferase CAT SWALL:Q8RQP2 (EMBL:AB082569) (100 in 213 aa)	3e-91
p1056.19	18374-19321	IS*30*	Transposase IS*30*	Similar to *N. meningitidis* transposase, IS*30* family SWALL:Q9JS28 (EMBL:AE002563) (41.5 in 318 aa)	2.3e-44
p1056.20	19653-20438	IS*5*	Transposase of IS*5*	Similar to *Vibrio cholerae* IS5 transposase VCA0472 and VCA0282, SWALL:Q9K218 (EMBL:AE004379) (58.30 in 259 aa)	2.4e-60
p1056.21	20822-21142	*ybeB*	Conserved hypothetical protein	Similar *to Salmonella* Typhimurium, *S. flexneri, E. coli*, and *S. Typhi* YbeB protein or JemB or YdjA or R0078 or HCM1.246c SWALL:Q9S457 (EMBL:AP005147) (100 in 106 aa)	7.7e-42
p1056.22	21024-21521	*ybeA*	Conserved hypothetical	Similar to *S.* Typhi, and *S. flexneri* 2A hypothetical protein R0079 SWALL:Q9K2K3 (EMBL:AF250878) (100 in 165 aa)	6e-63
p1056.23	21529-22215	*ybdA*	Putative transcriptional protein	Similar to *S.* Typhimurium, *S. flexneri, E. coli,* and *S.* Typhi YbdA protein or JemC or YeaA or R0080 SWALL:Q9S456 (EMBL:AP005147) (97.8 id in 228 aa)	1.1e-90
p1056.24c	22193-22815c		Tetracycline repressor protein (pseudogene)	Pseudogene. Identical to *E. coli* tetracycline repressor protein class B from transposon Tn*10 TetR* SWALL:P04483 (100 in 207 aa)	3.1e-85
p1056.25	22897-24102	*tetA*	Tetracycline resistance protein	Identical to *S.* Typhimurium, *S.* Typhi and *S. flexneri* 2a TetA protein SWALL:Q9K2Y4 (EMBL:AP005147) (100 in 401 aa)	1.8e-148
p1056.26c	24215-24808c	*tetC*	Putative transcriptional regulator	Identical *to S. Typhimurium,* and *S. flexneri* 2a putative transcriptional regulator TetC SWALL:Q93F25 (EMBL:AP005147) (100 in 197 aa)	9.5e-73
p1056.27	24896-25312	*tetD*	Putative transcriptional regulator of Tn*10*	Highly similar to *E. coli* transposon Tn*10* TetD protein (SWALL:P28816) (98.55 in 138 aa)	1.6e-53
p1056.28c	25322-26530c	IS*10*	Transposase of Tn*10*	Identical to *Citrobacter freundii* transposase Tnp SWALL:Q8GC95 (EMBL:AJ508060) (100 in 402 aa)	1.9e-170
p1056.29	26582-26931		Conserved hypothetical protein (pseudogene)	Pseudogene. Low similarity to central region *of H. ducreyi* 35000HP hypothetical protein HD0944 SWALL:AAP95826 (EMBL:AE017153) (40.74 in 81 aa)	2.4e-08
p1056.30c	26989-27753c		Putative conserved membrane protein	Low similarity to *P. syringae* membrane protein, putative PSPTO0411 SWALL:Q88AH9 (EMBL:AE016857) (26.19 in 252 aa)	1.3e-08

*c: reverse direction

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (% aa identity)	E value
p1056.31	27852-28481	*pilL*	Putative lipoprotein	Similar to the N-terminal of *P. aeruginosa* pilL2 or Rl086 SWALL:AAP84211 (EMBL:AY273869) (32.98 in 191 aa)	9.9e-15
p1056.32	28485-29225		Putative membrane protein	Similar to *S. Typhi* putative exported protein STY4558 SWALL:Q8Z1K6 (EMBL:AL627282) (35.77 in 246 aa)	1.8e-24
p1056.33	29204-29962		Putative exported protein	No significant database matches	
p1056.34	29978-30484		Putative membrane protein	Similar to *P. aeruginosa* hypothetical protein SWALL:Q8GPT3 (EMBL:AF440524) (42.51 in 167 aa)	4.1e-19
p1056.35	30481-32730	*traD*	DNA transfer protein	Similar to *P. putida* hypothetical 80.3 kDa protein SWALL:Q8VMJ3 (EMBL:AJ344068) (59.04 in 713 aa), and similar to *Providencia rettgeri* TraD SWALL:Q8RL4 (EMBL:AY090559) (31.64 in 610 aa)	1.4e-158; 3.9e-20
p1056.36	32741-33049		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0942 SWALL:AAP95824 (EMBL:AE017153) (46.72 in 107 aa)	7.2e-17
p1056.37	33049-33735		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0941 SWALL:AAP95823 (EMBL:AE017153) (74.56 in 228 aa)	1.2e-73
p1056.38	33866-34216		Putative export protein	Similar to *H. ducreyi* 35000HP hypothetical protein HDO940 SWALL:AAP95822 (EMBL:AE017153) (68.26 in 104 aa)	1.4e-20
p1056.39	34906-35277		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0941 SWALL AAP95820 (EMBL:AE017153) (60.16 in 123 aa)	1.9e-29
p1056.40	35937-36806		Conserved hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0936 SWALL:AAP95818 (EMBL:AE017153) (66.66 in 285 aa)	3.7e-73
p1056.41	36817-38217		Putative exported protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0935 SWALL:AAP95817 (EMBL:AE017153) (61.23 in 485 aa)	5.7e-68
p1056.42	38227-38634		Putative lipoprotein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0934 SWALL:AAP95816 (EMBL:AE017153) (64.34 in 129 aa)	2.1e-30
p1056.43	38650-41520		Conserved hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0933 SWALL:AAP95815 (EMBL:AE017153) (75.40 in 797 aa)	0
p1056.44	38227-38634		Putative lipoprotein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0934 SWALL:AAP95816 (EMBL:AE017153) (64.34 in 129 aa)	2.1e-30
p1056.45	38650-41520	*virB4*	Putative Type IV secretory pathway, *VirB4* components	Similar to *H. ducreyi* 35000HP hypothetical protein HD0933 SWALL:AAP95815 (EMBL:AE017153) (75.40 in 797 aa); and to *H. somnus* 2336 Type IV secretory pathway, VirB4 components (acc. no. ZP_00132732)	0
p1056.46	41528-41935		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0931 SWALL:AAP95813 (EMBL:AE017153) (47.40 in 135 aa)	1.8e-26
p1056.47c	41945-42268c		Hypothetical membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0930 SWALL:AAP95812 (EMBL:AE017153) (39.80 in 103 aa)	5.7e-12
p1056.48	42277-43770		Putative membrane protein	Similar to *S. Typhi* putative membrane protein STY4579 SWALL:Q8Z1I6 (EMBL:AE016848) (42.8 in 500 aa)	6.4e-89

*c: reverse direction

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (% aa identity)	E value
p1056.49c	43826-44155c		Hypothetical membrane protein	No significant database matches	
p1056.50c	44575-44982c		Hypothetical exported protein	No significant database matches	
p1056.51c	44998-46998c		Putative exported protein	Similar to N-terminal of *H. ducreyi* 35000HP hypothetical protein HD0922 SWALL:AAP95805 (EMBL:AE017153) (61.49 in 561 aa)	4.9e-125
p1056.52c	47013-47939c		Putative exported protein	Similar to C-terminal of *H. ducreyi* 35000HP hypothetical protein HD0921 SWALL:AAP95804 (EMBL:AE017153) (74.77 in 218 aa)	1.4e-70
p1056.53c	47951-48388c		Conserved hypothetical protein	Low similarity to C-terminal of *H. ducreyi* 35000HP hypothetical protein HD0920 SWALL:AAP95803 (EMBL:AE017153) (62.50 in 48 aa)	3.9e-08
p1056.54c	48763-49125c		Hypothetical membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0919 SWALL:AAP95802 (EMBL:AE017153) (52.58 in 116 aa)	2.7e-16
p1056.55c	49210-49443c		Hypothetical protein	Low similarity to *H.* ducreyi 35000HP hypothetical protein HD0918 SWALL:AAP95801 (EMBL:AE017153) (37.68 in 69 aa)	0.00054
p1056.56c	49985-50890c	*traC*	Putative primase protein	Similar to the *H. ducreyi* 35000HP probable DNA primase TraC or HD0917 SWALL:AAP95800 (EMBL:AE017153) (71.52 in 288 aa)	9.8e-90
p1056.57c	50985-51845c	*bla*	beta-lactamase TEM precursor	Identical to *Acinetobacter baumannii* beta-lactamase *bla*TEM-1 SWALL:AAP20891 (EMBL:AY263331) (100 in 286 aa)	1.7e-111
p1056.58c	52028-52585c	*tnpR*	Transposon Tn*3* resolvase	Identical to *Salmonella enteritidis* resolvase SWALL:BAC77164 (EMBL:AB103092) (100 in 185 aa)	7.6e-60
p1056.59	52884-53201		Hypothetical protein	No significant database matches	
p1056.60	53491-54189		Putative DNA-binding protein	Similar to *P. putida* hypothetical protein SWALL:AAP57242 (EMBL:AY299015) (33.04 in 227 aa)	8.8e-13
p1056.61c	54706-55260c		Hypothetical protein	No significant database matches	
p1056.62	55497-57404	*traI*	Conjugative relaxase	Similar to N-terminal of *H. ducreyi* 35000HP hypothetical protein HD0900 SWALL:AAP95784 (EMBL:AE017153) (75.68 in 218 aa), and to *P. rettgeri* TraI (SWALL:Q8RL15) (EMBL: AY090559) (26.63 in 473 aa)	1.2e-62; 2.4e-12
p1056.63	57495-58307	*int*	Putative site-specific recombinase	Similar to *H. ducreyi* 35000HP integrase/recombinase HD0897 SWALL:AAP95781 (EMBL:AE017153) (64.81 in 270 aa)	2.7e-70
p1056.64c	58302-59078		Conserved phage-related hypothetical protein	Similar to, although shorter than, Bacteriophage phi CTX ORF28 SWALL:Q9ZXJ7 (EMBL:AB008550) (23.55 in 276 aa)	6.9e-08

*c: reverse direction

4.4.8. Putative replication module

The nucleotide sequence from 1 to 14,719 bp contains 15 genes, only five of which have homology to genes of known functions. Thirteen of fifteen amino acid sequence in this region shares homology to sequences found in *H. ducreyi*. Their ORFs are in syntheny.

The first gene, p1056.01 is highly similar to the putative chromosome partitioning related protein (*soj*) of *H. ducreyi* 35000HP (acc. no. AE017153). *Soj* is also known as chromosome partitioning related protein and belongs to the *parA* family of genes that encode a protein that actively divides and distributes plasmid copies upon cell division (Autret *et al.*, 2001). The second gene, ORF p1056.02 is predicted to encode for a gene homologous to the helicase, *dnaB*. The enzyme encoded by this gene is responsible for unwinding helical DNA strands. The helicase binds DNA to a molecule of adenosine triphosphate (ATP) to begin the unwinding process (Subramanya *et al.*, 1996). The ATP/GTP binding (P-loop) motif is featured in this protein. The protein motifs, pfam indicate the presence of a DnaB helicase-like domain. Little is known about the five conserved hypothetical proteins that lie downstream of the helicase protein; however, it is thought that they are most likely involved in replication. One of these genes, p1056.07 curiously has low similarity (E value 2.6e-05) to the integrase regulator R, *inrR* of *Pseudomonas* sp. B13 (acc. no. CAD60671). This gene is responsible for the activation of transcription of integrase expression (Sentchilo *et al.*, 2003).

The eighth gene in this module, p1056.08, has homology to *ssb* gene that encodes for single-stranded binding protein. The function of this protein is to bind the lagging single-stranded DNA created by DnaB helicase to prevent them from re-annealing so that DNA replication can continue (Lewin, 1990). Two ORFs downstream of the single-stranded protein are hypothetical proteins. One of the genes, p1056.09 is believed to encode membrane lipoprotein based on a motif typical of a lipoprotein-binding site. The

other gene, p1056.10 encodes hypothetical protein, and has low similarity (E value 8.6e-08) to the homologues of C-terminal region of *S. flexneri* Osa protein (acc. no. 1560781). This protein suppresses the oncogenicity activity in *Agrobacterium tumefaciens* but its function in *H. influenzae* is rather doubtful (Chen and Kado, 1994).

The gene, p1056.11 is highly similar to gene that encodes DNA topoisomerase III, *topB* protein of *H. ducreyi* 3500HP (acc. no. AE017153). Topoisomerase is the enzyme that changes a DNA's superhelical. During replication of DNA, the enzyme facilitates the unwinding of the double-helical structure of DNA (Zhang *et al.*, 1995). It recognizes and cleaves regions of single-stranded DNA and holds and protects both free ends of the broken strand from random ligation reactions that could be detrimental to the organism. Homologues of the four genes mentioned, *parA*, *dnaB*, *ssb* and *topB* genes are recognised to play a role together in plasmid replication (Espinosa *et al.*, 2000).

Genes p1056.12 and p1056.13 encode hypothetical proteins with unknown functions, while ORF p1056.14 possibly encodes a DNA repair protein, similar to *H. ducreyi* 3500HP RadC protein (acc. no.: AE017153). This protein functions specifically in the recombination repair that is associated with a replication fork (Saveson and Lovett, 1999). p1056.15 has low similarity to a gene that encode hypothetical protein of *H. ducreyi* follows; however, the amino acid of the protein is predicted by PSORT to be possibly a lipoprotein.

4.4.9. Antibiotic resistance module

The second module is approximately 12 kb and is a transposon Tn*10* variant. It contains tetracycline and chloramphenicol resistance genes. In this region, three types of insertion sequences, namely IS*10*, IS*5* and IS*30* were identified. The IS*10* are encoded by genes p1056.16 and p1056.27, which flanks both ends of this composite transposon. Within the

IS*10*, lies insertion sequences IS*5* (p1056.17 and p1056.20), which flanks chloramphenicol acetyltransferase gene, *cat* (p1056.18) and IS*30* (p1056.19) (Fig. 4.9). This represents a variation on the typical Tn*10* structure.

A cluster of genes (p1056.21 to p1056.27) that lies within the Tn*10* like structure comprises genes involved in the regulation of tetracycline resistance. Gene p1056.21 that encodes ybeB *or* jemB protein, at present has no known function, but is believed to be a member of a family of functional proteins (Chalmers *et al.*, 2000). Next to this gene is p1056.22, which is predicted to be a homologue of Ybe protein. A comparison between the Tn*10* of *S. flexneri* and ICE*Hin1056* (Fig. 4.10) shows that ICE*Hin1056* does not have the *jemA*. *JemA* is a gene that is homologous to the sodium-dependent glutamate permease of *Salmonella* sp. and *E. coli*. Analysis has shown that the functional residue involved in sodium-dependent glutamate interactions is glycine and this is important for protein flexibility (Lawley *et al.*, 2000). Absence of *jemA* suggests that ICE*Hin1056* depends on *tetA* as the only transport system. Downstream is p1056.23, which encodes putative transcriptional protein, YbdA. The N-terminal of this protein contains a metal/DNA-binding domain; but the C-terminal domain is of unknown function (Chalmers *et al.*, 2000).

Unlike the Tn*10* of *S. flexneri* and *S. Typhi,* the Tn*10* of ICE*Hin1056* probably does not have a functional tetracycline repressor (Chalmers *et al.*, 2000; Lawley *et al.*, 2000) as p1054.24 is a pseudogene. Since the repressor gene is mutant, this suggests that tetracycline resistance in ICE*Hin1056* would be constitutively expressed. A membrane-bound efflux pump produces tetracycline resistance and is encoded by *tetA* gene p1056.25 in ICE*Hin1056*. The Tn*10* contains two additional structural genes, *tetC* and *tetD*, however, their functions are still unknown.

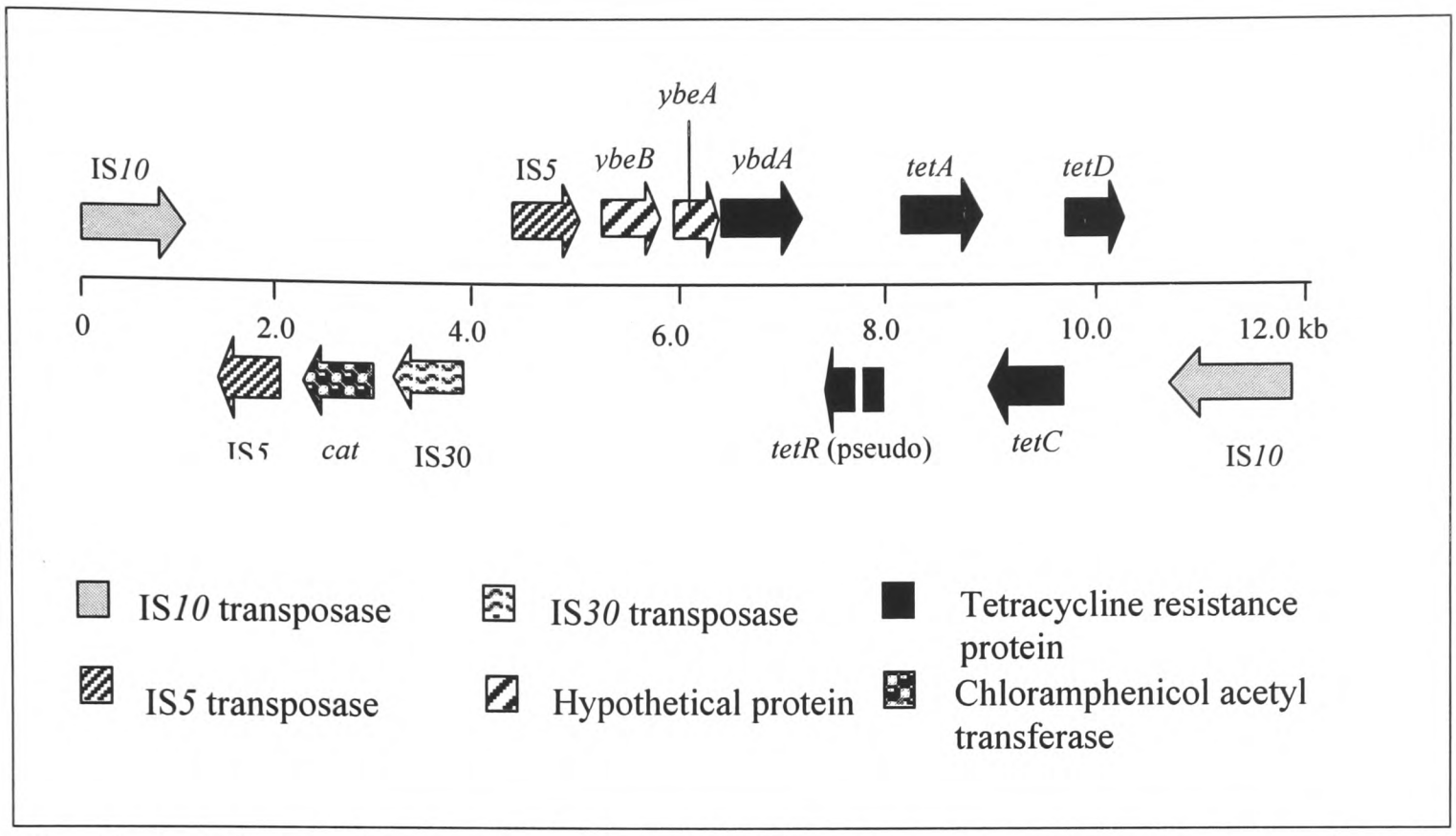


Fig. 4.9. Genetic organisation of Tn*10* insert in ICE*Hin1056*. The Tn*10* transposon consists of *cat* gene and *tet* genes flanked by two inverted IS*10*. Within the Tn*10*, the *cat* gene is flanked by two IS*5*. Note that the *tetR* is a pseudogene.

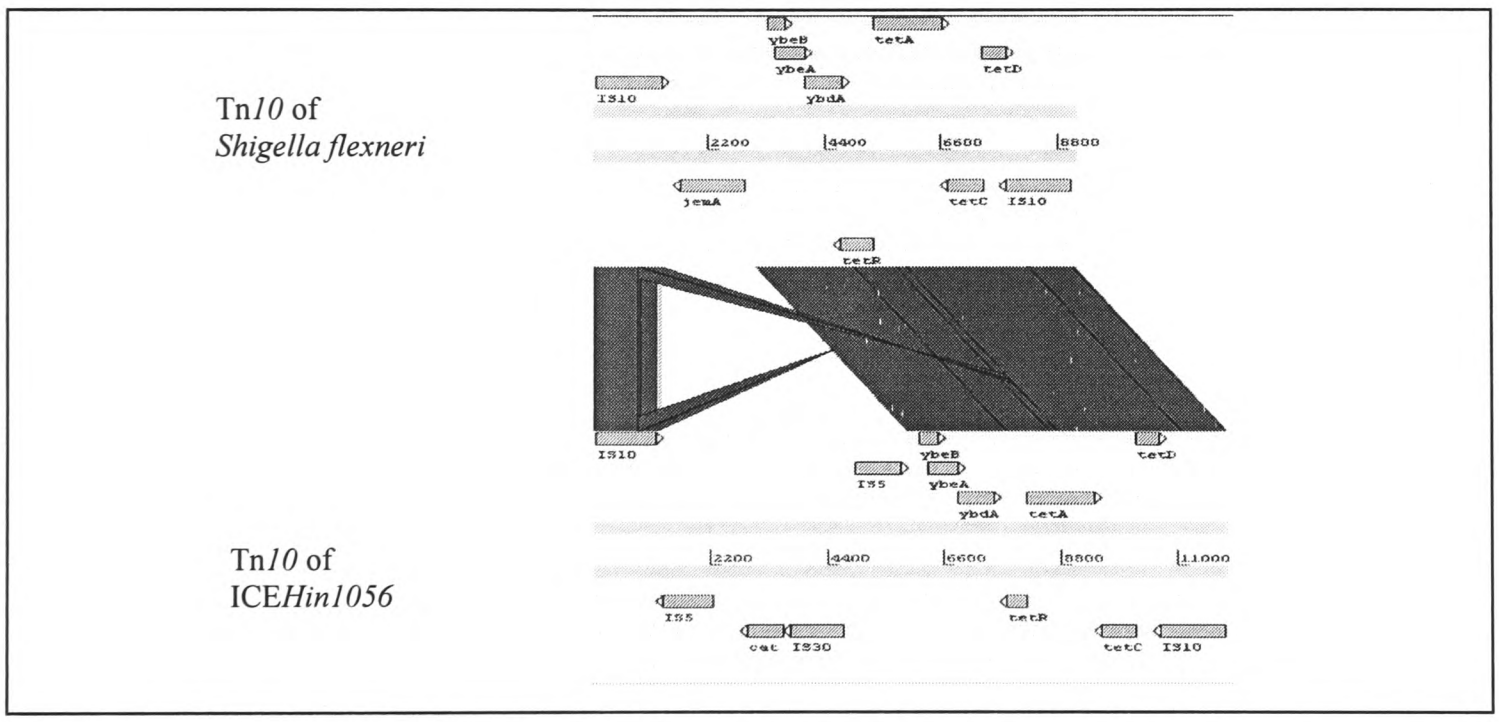


Fig. 4.10. ACT diagram showing comparison of Tn*10* of *S. flexneri* (acc. no. AF162223) and Tn*10* of ICE*Hin1056*. Note that ICE*Hin1056* has *cat* gene flanked by IS*5*, but is devoid of the *jemA* gene.

4.4.10. Putative transfer module

The majority of ORFs that lie between 26,582 and 59,393-bp encodes hypothetical proteins. The gene, p1056.29 has low similarity to a gene that encodes hypothetical

protein of *H. ducreyi*, however, it is a partial gene. Next to this gene ORF p1056.30, is a gene that has low similarity (E value 1.3e–08) to homologues of *Pseudomonas syringae* membrane protein (acc. no.: 12928499) and also to the *S. Typhi* TraX (acc. no.: 2877970). The *traX* is required in pilin acetylation in conjugative transfer (Manchak *et al.*, 2002; Maneewannakul *et al.*, 1995). This feature is confirmed by the prediction of THMM2.0 that showed an array of eight transmembrane helices and Pfam motif prediction of the TraX protein. The TraX protein is required to process the translation of *traA* (propilin polypeptide) for the formation of F-pilus (Manchak *et al.*, 2002).

Downstream of the TraX are series of ORFs that are membrane-associated. They consist of genes encoding proteins involved in export functions (ORFs p1056.31, p1056.32, p1056.33, p1056.38, p1056.41, p1056.50, p1056.51 and p1056.52), and genes that encode membrane protein (p1056.34, p1056.37, p1056.39, p1056.47, p1056.48, p1056.49, p1056.54, and p1056.55). Gene p1056.35 shows similarity to *traD* of *Providencia rettgeri* conjugative genomic island R391 (acc. no. AY090559). Other genes in this region include genes encoding lipoproteins (ORFs p1056.42, p1056.44) and hypothetical proteins (p1056.36, p1056.40, p1056.43, p1056.46, and p1056.53). The motifs of membrane-associated proteins in this region have features similar to the Type IV pilus biosynthesis locus (Srimanote *et al.*, 2002). ORF p1056.31 is a lipoprotein that has high similarity to the N-terminal region of PilL protein of the pathogenicity island PAP1 of *P. aeruginosa* PA14 (acc. no.: AAP84211). The pilin gene, *pilL* is a Type IV pilus and has a short leader sequence. The *pilL* is a gene involved in production of sex pilus and is responsible for the transfer of DNA by conjugation (Sakai and Komano, 2000). The Type IV pilus gene has also been associated with pathogenicity of *Yersinia pseudotuberculosis* (Collyn *et. al*, 2002), however, the Type IV pilus in ICE*Hin1056* is for conjugative transfer rather than for pathogenicity since no virulence genes are found.

p1056.59 did not show matches to any genes in the database. The adjacent p1056.60 (53,491–54,189-bp position) showed similarity to a homologue of *P. putida* hypothetical protein (acc. no. AAP57242). This gene has helix-turn-helix motif (HTH) that suggests it to encode a putative DNA binding protein. The p1056.61 encodes hypothetical protein while ORF p1056.62 encodes protein similar to the hypothetical protein of *H. ducreyi* 35000HP (acc. no. AAP95784). This ORF is also similar to *traI* that encodes putative relaxase of *P. rettgeri* (acc. no. AY090559). p1056.63 appears to be a homologue of integrase/recombinase protein of *H. ducreyi* 35000HP. This integrase gene belongs to the XerC integrase lineage and is most similar to the *xerC* of *P. aeruginosa* (acc. no.: AF285416).

4.4.11. Ampicillin resistance module

This module (coordinates 50,985 and 52,585 bp) consists of genes that encode β-lactamase resistance. ORF p1056.57 is identical to *Acinetobacter baumannii* TEM-1 β-lactamase *bla* precursor (acc. no. AAP20891). Adjacent to this gene, ORF p1056.58 is identical to *Salmonella enteritidis* resolvase *tnpR* (acc. no. BAC77164). This gene also shows a HTH DNA binding motif. However, unlike the Tn*3* of *E. coli*, the Tn*3* of ICE*Hin1056* is truncated because *tnpA* is missing. Also, only one copy of the 38-bp inverted repeat (5'-GGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAG-3') is present. Fig. 4.11 depicts the truncated Tn*3* of ICE*Hin1056* in comparison to the complete Tn*3* of *E. coli* (acc. no. V00613). Although *tnpA* is absent in ICE*Hin1056*, the β-lactamase activity is detectable and 1056 is ampicillin resistant with an MIC of 96 μg/ml.

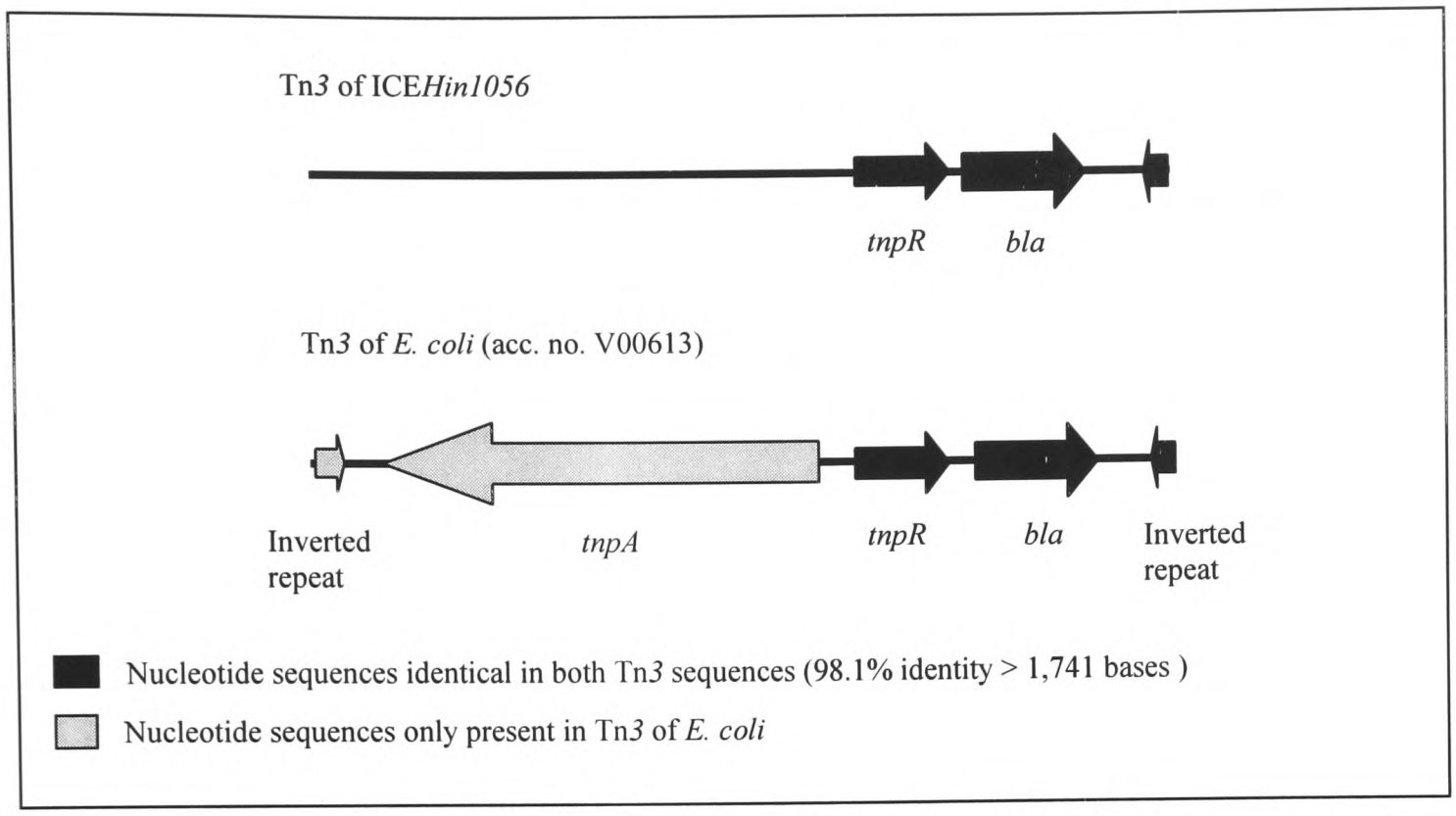


Fig. 4.11. Genetic organisation of Tn*3* transposon. The Tn*3* of ICE*Hin1056* is compared to the Tn*3* of *E. coli* (acc. no. V00613), where the black arrows refer to proteins found on both transposons, and the grey arrows refer to proteins missing in Tn*3* of ICE*Hin1056*.

4.4.12. Identification of putative origin of replication

One of the insights from analysing the bacterial genome or plasmids is that, the base composition of each chromosomal or plasmid strand changes at the origin and terminus of replication (Freeman, 1998; Lobry, 1996a, b). This phenomenon is known as GC skew, which can be produced by calculating (G–C)/(G+C) values for a given length of the entire sequence. The site where there is a switch of the polarity of GC skew often corresponds with *oriV* where initiation of DNA replication takes place. This pattern was well observed in genome sequences such as *Borellia burgdorferi* (Picardeau *et al.*, 1999). The GC skew of the whole sequence of a plasmid of *H. influenzae* biotype *aegyptius* was also obvious, resulting in the recognition of the *oriV* (Kroll *et al.*, 2002). However, the whole genome of *H. influenzae* Rd did not show an obvious GC-skew, as shown in Fig. 4.12 (Uno *et al.*, 2000).

The sequences of ICE*Hin1056* were also analysed for GC skew. The GC skew plot of this mobile element did not reveal clear bioinformatics evidence of *oriV* (Fig. 4.13). The inability to obtain a GC skew in ICE*Hin1056* is most probably due, in part, to the presence of insertions that could have interfered with the GC bias. It is reported that, in the absence of transposition and insertion, a bias favours the accumulation of purines in the leading strand (Freeman, 1998).

Although the GC skew in ICE*Hin1056* is unclear, a possible *oriV* site of the element is predicted to be located at approx. 6000-bp. This prediction is based on the presence of multiple inverted sequences and direct tandem (two 33-bp and four 36-bp) repeats that lie within AT-rich region (Fig. 4.14).

A program Oriloc has been developed to predict the putative origin and terminus of replication in prokaryotic genomes (Frank and Lobry, 2000). It successfully predicted the origin of replication in the complete genomes of bacteria lacking genes involved in replication (*dnaA, priA and recA*), such as, *Bifidobacterium longum* (Schell *et al.*, 2002) and *Blochmannia floridanus* (Gil *et al.*, 2003). The sequences of ICE*Hin1056* were downloaded into this program, but were unable to show a clear signal for the origin of replication. The tentative position of a skew reversal is at position 0, and at approximately at 42-kb position (Fig. 4.15). This is opposite the putative *oriV* deduced from sequence structure data discussed above.

The author of the Oriloc program suggests that the *oriV* for this element is at position 0 (Lobry, personal communication). This programme fails to locate the *oriV* of ICE*Hin1056* probably because the element has low G+C (<50%) content and is further complicated by the presence of transposons. However, it was able to detect the termination of replication (*ter*) at 42-kb position. Existence of *ter* site in ICE*Hin1056* sequence suggests that the element undergoes theta type replication, consistent to the

report of Bussiere and Bastia, 1999 (Bussiere and Bastia, 1999). Initiation of DNA replication occurs at *oriV* and, during theta replication, the replication forks, moving unidirectionally or bidirectionally, are arrested at sequence-specific replication termini (Bussiere and Bastia, 1999; del Solar *et al.*, 1998). No sequences resembling IHF ((A/T)ATCAANNNNTT(A/G) where N is A/C/T/G) was detected in the whole sequence.

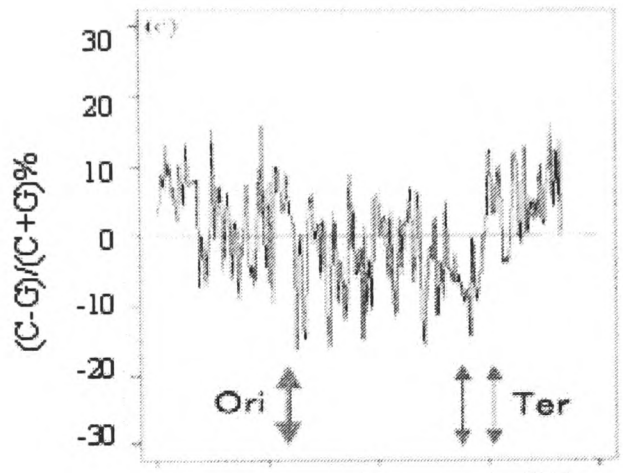


Fig. 4.12. GC skew of the whole genome of *H. influenzae* Rd taken from Uno *et al.* (Uno *et al.*, 2000). Non-overlapping window of 10 kb was used for this analysis. Arrows indicate the replication origin and terminus. No clear shift point is observed in *H. influenzae* Rd.

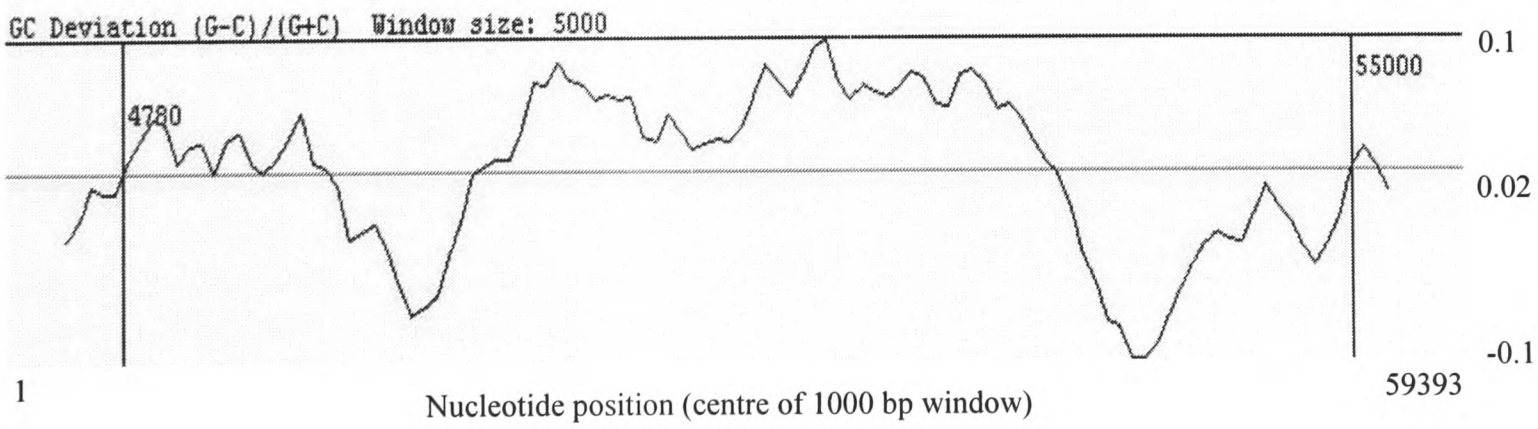


Fig. 4.13. GC deviation [(G-C)/(G+C)] calculated for ICE*Hin1056* based on a 5000-bp window. The GC skew is obscured by the insertion elements. This graph was obtained by using Artemis program.

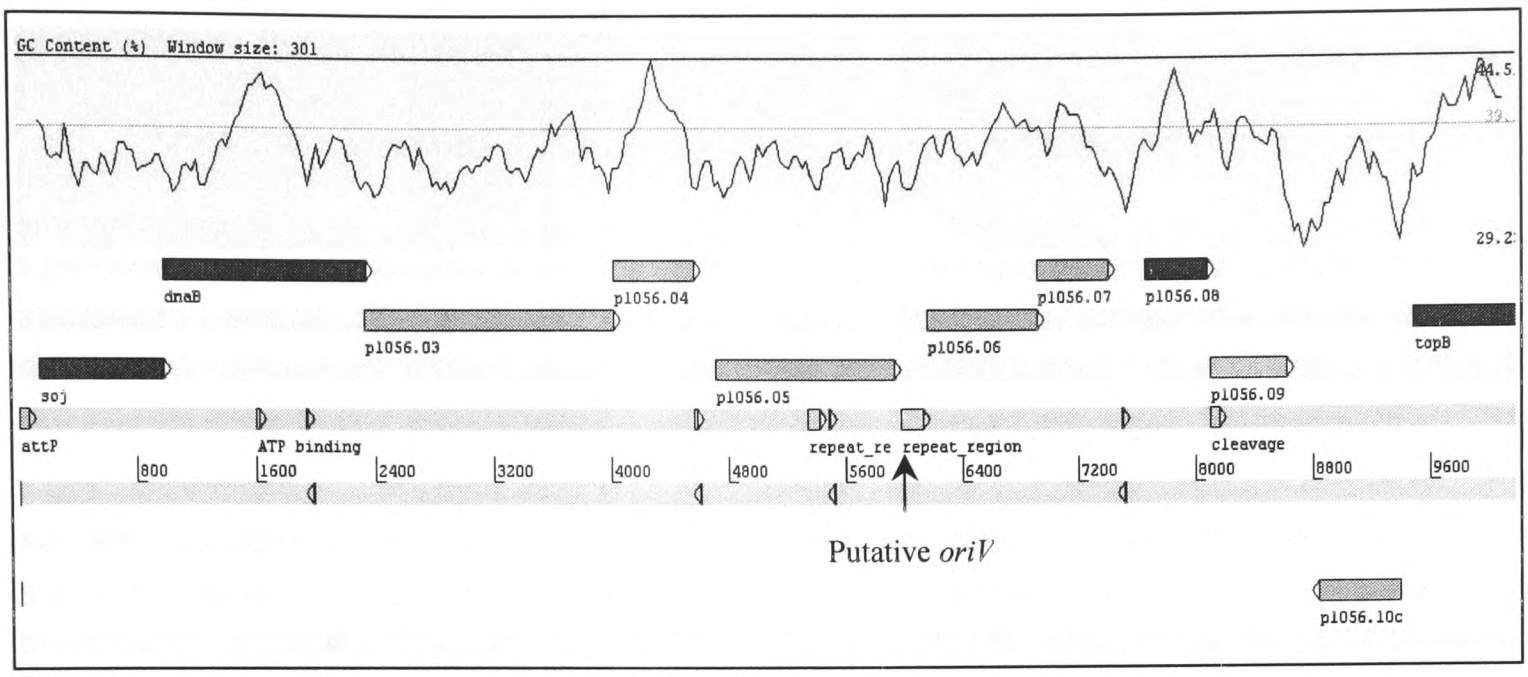


Fig. 4.14. Putative *oriV* of ICE*Hin1056*. Presence of interons in an AT-rich region indicative of possible *oriV* site. This diagram was obtained by using Artemis program.

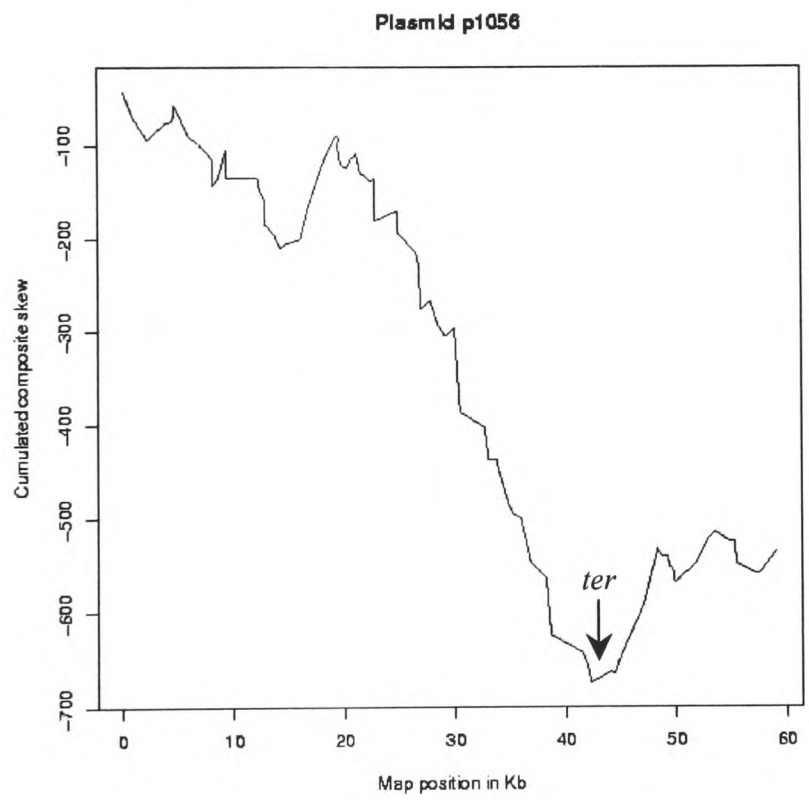


Fig. 4.15. Cumulative composite skew of ICE*Hin1056*. The *oriV* of this element is unclear, however, the site for termination of replication terminus (*ter*), is approximately at 42-kb position shown by arrow. This graph was obtained by using Oriloc program.

4.4.13. Identification of putative origin of transfer

During conjugation, initiation of DNA transfer occurs at a site known as origin of transfer or *oriT*, whereafter cleavage by a nickase, a single-stranded DNA molecule is released

which will ultimately be transferred to a recipient cell. Usually the region close to the *traI* gene contains the possible *oriT*.

From the sequences of the element, it is suggested that the putative *oriT* of the ICE*Hin1056* is between gene p1056.60 and p1056.61 (54190 to 54714-bp position). This assumption is based on reports of others that the *oriT* of conjugative plasmid is usually located in (i) the vicinity of their DNA processing genes, (ii) a non-coding region, and (iii) the presence of repeats (Galli *et al.*, 2001). In ICE*Hin1056* this region contains a 114-bp DNA domain-containing 10.3 sets of 13-bp perfect direct repeats. The *traI* gene, p1056.62, (similar to *P. rettgeri traI*, acc. no. AY090559) and a putative DNA-binding protein gene, p1056.60, (similar to *P. putida*, acc. no. AY299015) are located immediately adjacent to this region (Fig. 4.16). Besides, this region also has a lower G+C content of 34.3% (i.e. higher A+T content) typical of an *oriT*.

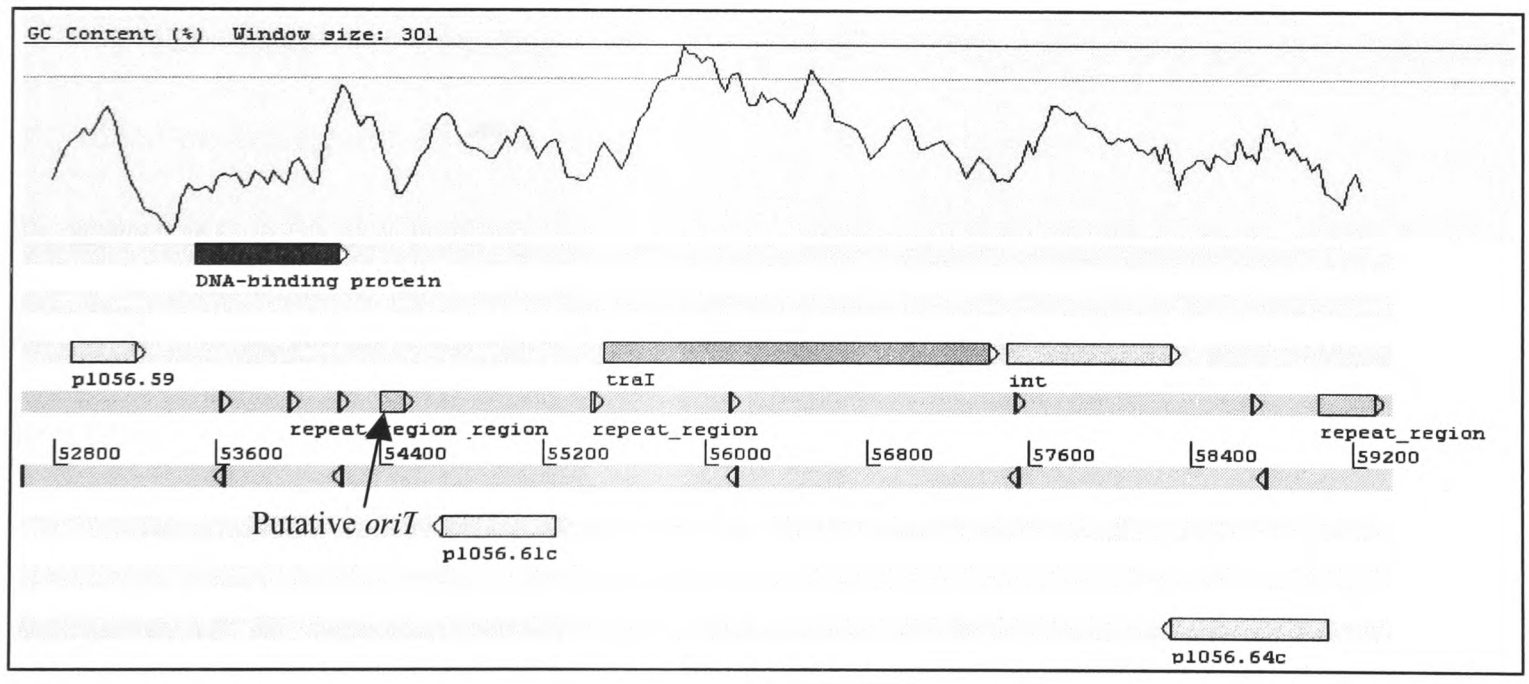


Fig. 4.16. Putative *oriT* of ICE*Hin1056*. Presence of multiple interons, DNA binding site and close to *traI*, indicative of possible *oriT* at 54,396 bp position shown by an arrow. The small arrowheads represent repeats. This diagram was obtained by using Artemis program.

4.4.14. Nucleotide sequence accession number

The complete sequence data of ICE*Hin1056* was deposited in the EMBL/GenBank DNA databases under the accession number AJ627386.

4.5. Conclusion

The complete nucleotide sequence of ICE*Hin1056* was determined in this study. This is the first report of a complete sequence of a conjugative and integrative element encoding multi-drug resistance in *H. influenzae.* The complete sequence of ICE*Hin1056* has provided new understanding of its structure and organisation from which insight into its successful transmission of antibiotic resistance in *H. influenzae* can be predicted. This circular 59,393-bp element was preferably integrated into the chromosome of *H. influenzae* strain 1056 at the 3' end of the $tRNA^{leu}$ site (Dimopoulou *et al.*, 2002). The annotation of the sequenced circular element reveals that the element consists of gene clusters of related functions likely to be involved in integration, conjugative transfer and probable replication and maintenance of a plasmid form. Besides these genes, ICE*Hin1056* also contains genes encoding resistance to antibiotics. No evidence of genes involved with virulence was found on analysing the nucleotide sequence of this element.

The antibiotic resistance genes on ICE*Hin1056* are contained in a Tn*10*-like structure encoding tetracycline-resistance and chloramphenicol resistance. This finding is in agreement with the findings of Jahn *et al.* and Levy *et al.* (Jahn *et al.*, 1979; Levy *et al.*, 1984). Ampicillin resistance is encoded by the TEM β-lactamase gene and is present in a truncated Tn*3* missing *TnpA*.

During the annotation of the genes of ICE*Hin1056*, many of the genes matched the genes of *H. ducreyi* 35000HP and are organised in the same order. A comparison between these two sequences was carried out in chapter seven.

CHAPTER FIVE

NUCLEOTIDE SEQUENCE AND ANALYSIS OF ICE*Hin299*

5.1. Introduction

Large conjugative plasmids of *H. influenzae* type b carrying the β-lactamase gene, isolated from Greece and the UK, have been analysed for their genetic relatedness (Dimopoulou *et al.*, 1992). Southern hybridisation techniques, using UK ICEs as probes show that elements cut with *Pst*1 and *EcoR*1 shared some homology. However, the extent and degree of relatedness is not known.

5.2. Objectives

In this chapter, molecular characteristics of a mobile element of a *H. influenzae* strain 299 isolated in Greece will be investigated by sequencing the whole element ICE*Hin299*. The information on the complete sequence of this element will be compared to ICE*Hin1056*, and if possible to other Haemophilus ICEs, to better understand the properties of these elements and how they may have evolved.

5.3. Materials and methods

*5.3.1. Extraction of ICE*Hin299

The ICE*Hin299* was grown on chocolate agar for 48 hours before being cultured into MH broth for plasmid extraction. Extraction of the element was carried out using CsCl gradient centrifugation as described in Chapter 2.

5.3.2. *Sequencing of ICE*Hin299

As in ICE*Hin1056*, shotgun-sequencing method was performed on ICE*Hin*299. Nucleotide sequences obtained were assembled, analysed and the contigs were ordered in the same order as ICE*Hin1056*. Gaps between the contigs were filled by 'primer walking' procedure. In regions where nucleotide sequences were shared between ICE*Hin299* and ICE*Hin1056*, primers of ICE*Hin1056* were used to close the gaps, and to improve the quality of the sequences.

5.4. Results and discussion

5.4.1. *Overall organisation of ICE*Hin299

Two random plasmid DNA libraries and six sequence assemblies were carried out. The total number of readings used in the final assembly was 787, with an average length for each sequence being 503 bases. The complete sequence of ICE*Hin299* produced a length of 53,902 bases. It had 38.59% G+C content consisting of 31.8% adenine, 18.4% cytosine, 20.9% guanine and 28.9% thymine. The total number of bases sequenced was 396,396 bases producing approximately 7.4 fold coverage. Of the 67 predicted genes, 53 showed no homology to genes of known functions. As in ICE*Hin1056*, ICE*Hin299* also has nine copies of Haemophilus uptake sequence (Table 5.1). This number of uptake sequence is also similar to the uptake sequence of a plasmid of *H. aegyptius* (Kroll *et al.*, 2002). Whether all other mobile elements of *Haemophilus* spp. constantly have nine copies of the Haemophilus uptake sequence remains to be shown.

As in ICE*Hin1056*, the start of the sequence was positioned at the site of recombination, *attP*. The whole element contains 67 ORFs with 52 ORFs in the forward and 15 in the reverse direction (Fig. 5.1). Unlike ICE*Hin1056*, ORFs of ICE*Hin299* has only one pseudo-gene. Thirty-five of the predicted proteins showed homology to proteins

of *H. ducreyi*. Fourteen genes on this element did not have matches with any sequences in databases. Properties of ICE*Hin299* are shown in Table 5.1.

Table 5.1. Summary of ICE*Hin299* properties[a].

Property	**Value**
Total size	53,902 bp
GC content	38.59%
Coding sequences	67
Pseudo-genes	1
Coding density	88.1%
Average gene length	709 bp
No. of uptake sequence	9

[a]Comparison of properties between ICE*Hin299* to ICE*Hin1056* is shown in chapter 6.

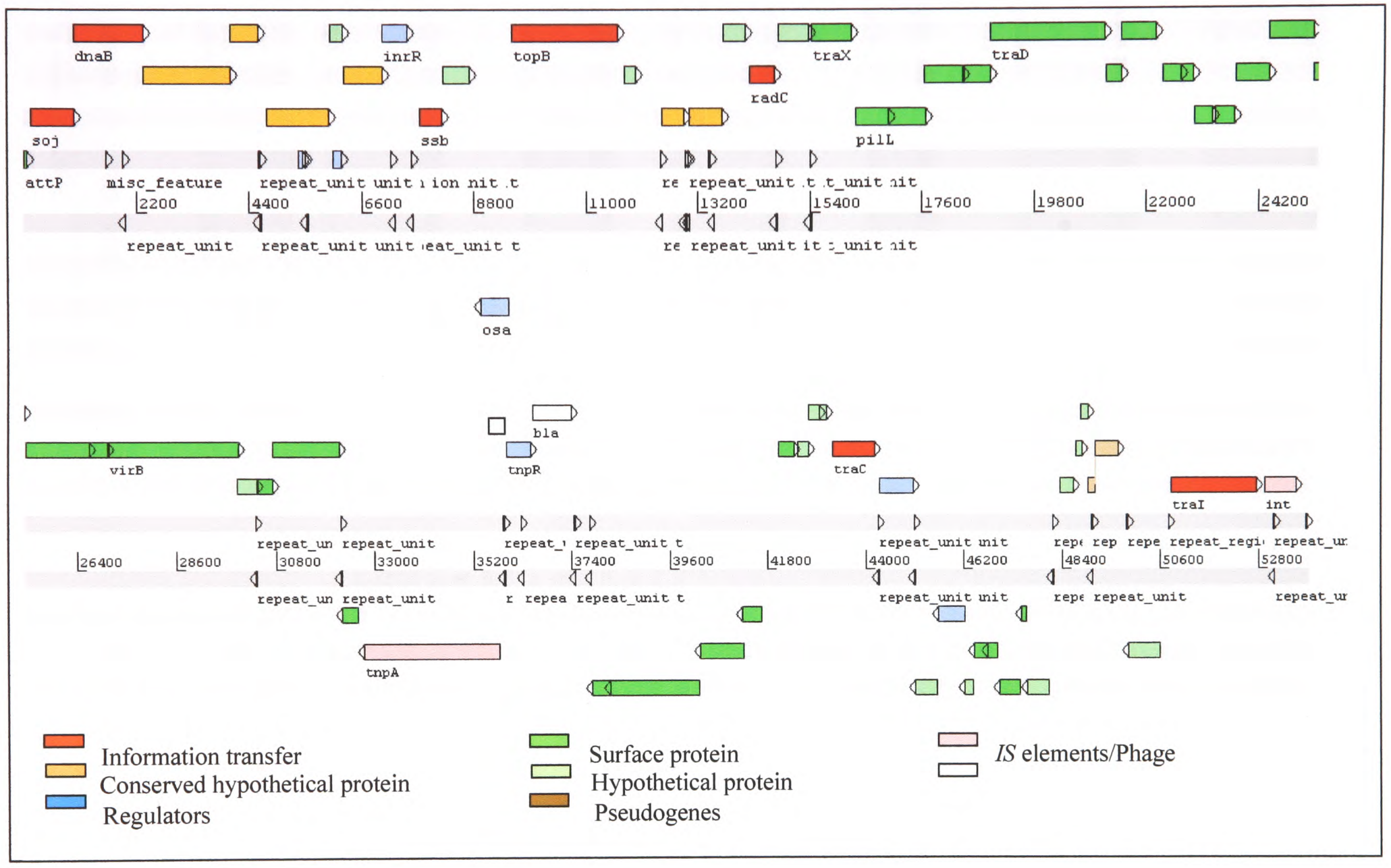


Fig. 5.1. Genetic organisation of ICE*Hin299*. Genes are represented by arrow boxes. Putative functions of the genes are described in Table 5.2.

Table 5.2. Putative ORFs of ICE*Hin299*. Amino acid similarity of putative proteins encoded by ICE*Hin299* to sequences in public databases.

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (aa identity)	E value
p299.01	141-977	*soj*	Chromosome partitioning related protein	Similar to *H. ducreyi* 35000HP possible chromosome partitioning related protein HD0974 SWALL:AAP95855 (EMBL:AE017153) (83.33 in 276 aa)	8.3e-94
p299.02	979-2334	*dnaB*	DNA helicase	Similar to *H. ducreyi* 35000HP DNA helicase DnaB or HD0973 SWALL:AAP95854 (EMBL:AE017153) (71.62 in 444 aa)	1.8e-120
p299.03	2327-4027		Conserved hypothetical protein	Similar to N-terminal of *H. ducreyi* 35000HP hypothetical protein HD0971 SWALL:AAP95852 (EMBL:AE017153) (72.30 in 195 aa)	6.3e-49
p299.04	4027-4578		Conserved hypothetical protein	Similar to *H. ducreyi* 35000HP conserved hypothetical protein HD0969 SWALL:AAP95850 (EMBL:AE017153) (79.0 in 181 aa)	2.7e-56
p299.05	4728-5951		Conserved hypothetical protein	Similar to, although longer than *H. ducreyi* 35000HP hypothetical protein HD0968 SWALL:AAP95849 (EMBL:AE017153) (68.77 in 237 aa)	1.1e-58
p299.06	5968-6231		Hypothetical protein	No significant database matches	
p299.07	6248-7003		Conserved hypothetical protein	Similar to the central region of *H. ducreyi* 35000HP hypothetical protein HD0966 SWALL:AAP95847 (EMBL:AE017153) (72.44 in 254 aa)	1e-71
p299.08	7003-7491	*inrR*	Putative integrase regulator	Similar to, although longer *H. ducreyi* 35000HP hypothetical protein HD094 SWALL:AAP95845 (EMBL:AE017153) (74.41 in 43 aa)	5.3e-06
p299.09	7737-8159	*ssb*	Single-stand binding protein	Similar to *A. actinomycetemcomitans* single strand binding protein Ssb SWALL:Q9F278 (EMBL:AF302424) (72.41 in 145 aa)	5.6e-41
p299.10	8186-8704		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0961 SWALL:AAP95842 (EMBL:017153)(47.09 id in 172 aa)	2.9e-23
p299.11c*	8920-9477c	*osa*	Putative osa protein	Low similarity to C-terminal region of *S. flexneri* protein Osa SWISSPROT:OSA_SHIFL (38.05 in 119 aa)	9.6e-08
p299.12	9553-11613	*topB*	DNA topoisomerase III	Similar to *H. ducreyi* DNA 35000HP DNA topoisomerase III TopB1 or HD0954 SWALL:AAP95835 (EMBL:AE017153) (8.66 in 450 aa)	3.9e-139
p299.13	11750-11974		Hypothetical protein	No significant database matches	
p299.14	12498-12926		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0951 SWALL:AAP95832 (EMBL:AE017153) (57.63 in 144 aa)	5.5e-33
p299.15	13041-13682		Hypothetical protein	Low similarity to C-terminal region *Rhizobium meliloti* hypothetical 50.7 kDa protein SWALL:Q91UT8 (EMBL:PPS304453) (28.98 in 138 aa)	0.0052
p299.16	13705-14124		Putative lipoprotein	No significant database matches	

*c: reverse direction

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (aa identity)	E value
p299.17	14225-14695	*radC*	Putative DNA repair protein	Similar to *H. ducreyi* 35000HP conserved possible DNA repair protein HD0945 SWALL:AAP95827 (EMBL:AE017153) (41.02 in 156 aa)	5.2e-20
p299.18	14788-15384		Putative lipoprotein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0944 SWALL:AAP95826 (EMBL:AE017153) (36.06 in 183 aa)	7.1e-15
p299.19	15442-16203	*traX*	Conserved hypothetical glycine-rich protein	Similar to *P. syringae* putative membrane protein PSPTO0411 SWALL:Q88AH9 (EMBL:AE016857) (25.49 in 251 aa)	3.7e-08
p299.20	16302-16931	*pilL*	Putative exported protein	Similar to N-terminal region of *P. aeruginosa* PilL2 or RL086 SWALL:AAP84211 (EMBL: AY273869) (32.98 in 191 aa)	1.1e-14
p299.21	16935-17675		Putative exported protein	Similar to *S. Typhi* putative exported protein STY4558 or T4256 SWALL:Q8Z1K6 (EMBL:AL627282) (35.77 in 246 aa)	1.6e-24
p299.22	17654-18412		Hypothetical protein	No significant database matches	
p299.23	18428-18934		Putative membrane protein	Similar to *P. aeruginosa* hypothetical protein SWALL:Q8GPT3 (EMBL:AF440524) (42.51 in 167 aa)	4.1e-19
p299.24	18931-21180	*traD*	Putative transfer protein	Similar to *P. putida* hypothetical 80.3 kDa protein SWALL:Q8VMJ3 (EMBL:AJ344068) (59.18 in 713 aa), and to *P. rettgeri* conjugative genomic island R391 TraD SWALL:Q8RL14 (EMBL:AY090559) (31.0 in 610 aa)	7e-159; 2e-76
p299.25	21215-21499		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0942 SWALL:AAP95824 (EMBL:AE017153) (44.0 in 100 aa)	2.5e-15
p299.26	21499-22185		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD094 SWALL:AAP95823 (EMBL:AE017153) (74.12 in 228 aa)	7.9e-73
p299.27	22316-22666		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0940 SWALL:AAP95822 (EMBL:AE017153) (40.44 in 89 aa)	3.2e-05
p299.28	22670-22915		Putative exported protein	No significant database matches	
p299.29	22935-23288		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0939 SWALL:AAP95821 (EMBL:AE017153) (50.0 in 120 aa)	2.1e-18
p299.30	23355-23726		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0938 SWALL:AAP95820 (EMBL:AE017153) (60.97 in 123 aa)	6.9e-30
p299.31	23738-24382		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein SWALL:AAP95819 (EMBL:AE017153) (67.74 in 186 aa)	4.2e-52
p299.32	24382-25251		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0936 SWALL:AAP95818 (EMBL:AE017153) (67.01 in 285 aa)	2.3e-74

*c: reverse direction

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (aa identity)	E value
p299.33	25262-26662	*virB4*	Putative type IV secretion	Similar to *H. ducreyi* 35000HP hypothetical protein HD0935 SWALL:AAP95817 (EMBL:AE017153) (61.44 in 485 aa), and to *H. influenzae* R2866 Type IV secretory pathway, COG3451: VirB4 components (ZP_00157627) (97.0 id in 956 aa)	1.1e-68; 0.0
p299.34	26672-27079		Putative exported protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0934 SWALL:AAP95816 (EMBL:AE017153) (63.56 in 129 aa)	8.7e-33
p299.35	27095-29965		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0933 SWALL:AAP95815 (EMBL:AE017153) (75.65 id in 797 aa)	0
p299.36	29937-30380		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0931 SWALL:AAP95813 (EMBL:AE017153) (47.61 in 147 aa)	1.4e-27
p299.37	30390-30713		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0930 SWALL:AAP95812 (EMBL:AE017153) (39.8 in 103 aa)	5.5e-12
p299.38	30722-32215		Putative membrane protein	Similar to *S. Typhi* membrane protein STY4579 SWALL:Q8Z1I6 (EMBL:AL627282) (42.6 in 500 aa)	5e-89
p299.39c	32271-32600c		Putative membrane protein	No significant database matches	
p299.40c	32758-35763c	*tnpA*	Transposase	Identical to *S. enteritidis* transposase SWALL:BAC77163 (EMBL:AB103092) (100 in 1001 aa)	0
p299.41	35927-36484	*tnpR*	Tn*3* resolvase	Identical to *S. enteritidis* resolvase SWALL:BAC77164 (EMBL:AB103092) (100 in 175 aa)	4.1e-61
p299.42	36532-37392	*bla*	Beta-lactamase TEM1	Identical to *S. flexneri* beta-lactamase SWALL:AAC97980 (EMBL:U48775) (100 in 286 aa)	1.7e-111
p299.43c	37840-38248c		Putative exported protein	Low similarity to *Neisseria gonorrhoeae* Gly1ORF1 precursor Gly1ORF1 precursor SWALL:O30392 (EMBL:AF003941) (22.38 in 134 aa)	0.13
p299.44c	38264-40264c		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0922 SWALL:AAP95805 (EMBL:AE017153) (61.49 in 561 aa)	2.5e-120
p299.45c	40279-41253c		Putative membrane protein	Similar to, although longer than *H. ducreyi* 35000HP hypothetical protein HD0921 SWALL:AAP95804 (EMBL:AE017153) (74.77 in 218 aa)	6.6e-75
p299.46c	41217-41660c		Putative exported protein	Similar to, although longer than *H. ducreyi* 35000HP hypothetical protein HD0920 SWALL:AAP95803 (EMBL:AE017153) (62.50 in 48 aa)	9.4e-09
p299.47	42029-42391		Putative membrane protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD0919 SWALL:AAP95802 (EMBL:AE017153) (52.58 in 116 aa)	1.6e-17
p299.48	42476-42709		Hypothetical protein	Similar to *H. ducreyi* 35000HP hypothetical protein HD091 SWALL:AAP95801 (EMBL:AE017153) (37.68 in 69 aa)	0.00054
p299.49	42712-42960		Hypothetical protein	Low similarity to the central region of *H. ducreyi* 35000HP HD0947 hypothetical protein SWALL:AAP95829 (EMBL:AE017153) (34.0 in 50 aa)	0.35
p299.50	42964-43121		Hypothetical protein	No significant database matches	

*c: reverse direction

Gene ID	Coordinates	Gene name	Predicted protein	Database similarity (aa identity)	E value
p299.51	43253-44206	*traC*	DNA primase	Similar to *H. ducreyi* 35000HP probable DNA primase or HD017 SWALL:AAP95800 (EMBL:AE017153) (71.92 in 317 aa)	3.3e-98
p299.52	44310-45071		Possible type I restriction enzyme M subunit	Similar to *H. ducreyi* 35000HP possible type I restriction enzyme M subunit HD0914 SWALL:AAP95798 (EMBL:AE017153) (66.27 in 252 aa)	1.2e-68
p299.53c	45106-45601c		Hypothetical protein	No significant database matches	
p299.54c	45627-46211c		Resolvase/ integrase	Similar to *H. aegyptius* resolvase/integrase-like protein BPL01 SWALL:Q8VRE8 (EMBL:AF447808) (55.27 in 199 aa)	5.6e-28
p299.55c	46226-46417c		Hypothetical protein	No significant database matches	
p299.56c	46432-46737c		Putative membrane protein	Low similarity *to Fusobacterium nucleatum* subsp. *vincentii* ATCC49256 hypothetical protein FNV1196 SWALL:EAA24216 (EMBL:AABF01000046) (31.46 in 89 aa)	0.79
p299.57c	46741-46935c		Hypothetical protein	No significant database matches	
p299.58c	46997-47455c		Putative lipoprotein	Very low similarity to *Synechococcus elongatus* Tlr2451 protein SWALL :Q8DG72 (EMBL:AP005377) (34.83 in 89 aa)	3.3
p299.59c	47481-47603c		Putative membrane protein	No significant database matches	
p299.60c	47627-48109c		Hypothetical protein	Low similarity to the N-terminal region of *Yersinia pestis* hypothetical protein YP01174 SWALL:Q8ZGV3 (EMBL: AJ414146) (26.38 in 144 aa)	0.29
p299.61	48363-48650		Hypothetical protein	No significant database matches	
p299.62	48710-48844		Hypothetical protein	No significant database matches	
p299.63	48817-48978		Hypothetical protein	No significant database matches	
p299.64	48975-49636		Conserved hypothetical protein (pseudogene)	Pseudogene. Similar to *P. putida* hypothetical protein SWALL:AAP57242 (EMBL:AY299015) (32.57 in 221 aa)	2e-12
p299.65c	49900-50604c		Hypothetical protein	No significant database matches	
p299.66	50841-52745	*traI*	Putative relaxase	Similar to C-terminal region of *Xylella fastidiosa* hypothetical Xf1753 SWALL:Q9PCM5 (EMBL:AE003998) (36.79 in 606 aa), and to *P. rettgeri* (acc. no. AAM08003) (31.0 in 231 aa)	4.9e-77; 4e-18
p299.67	52932-53648	*int*	Integrase	Similar to *A. actinomycetecomitans* integrase SWALL:Q9ZAB5 (EMBL:AF006830) (69.74 in 265 aa)	8.6e-70

*c: reverse direction

5.4.2. *Description of ICEHin299 sequences based on functional module*

The gene organisation reveals four main modules (Fig. 5.2 and Table 5.2). The first module (coordinates 1-15384) shows a cluster of putative genes involves in replication, and the second domain (15385-32600) believed to be the region responsible for gene transfer. The third domain (326001-37392), inserted between the third and fourth domain is responsible for ampicillin resistance. The fourth domain (37840-38248) contains a region putatively involved in transfer and integration (i.e. *int* is present).

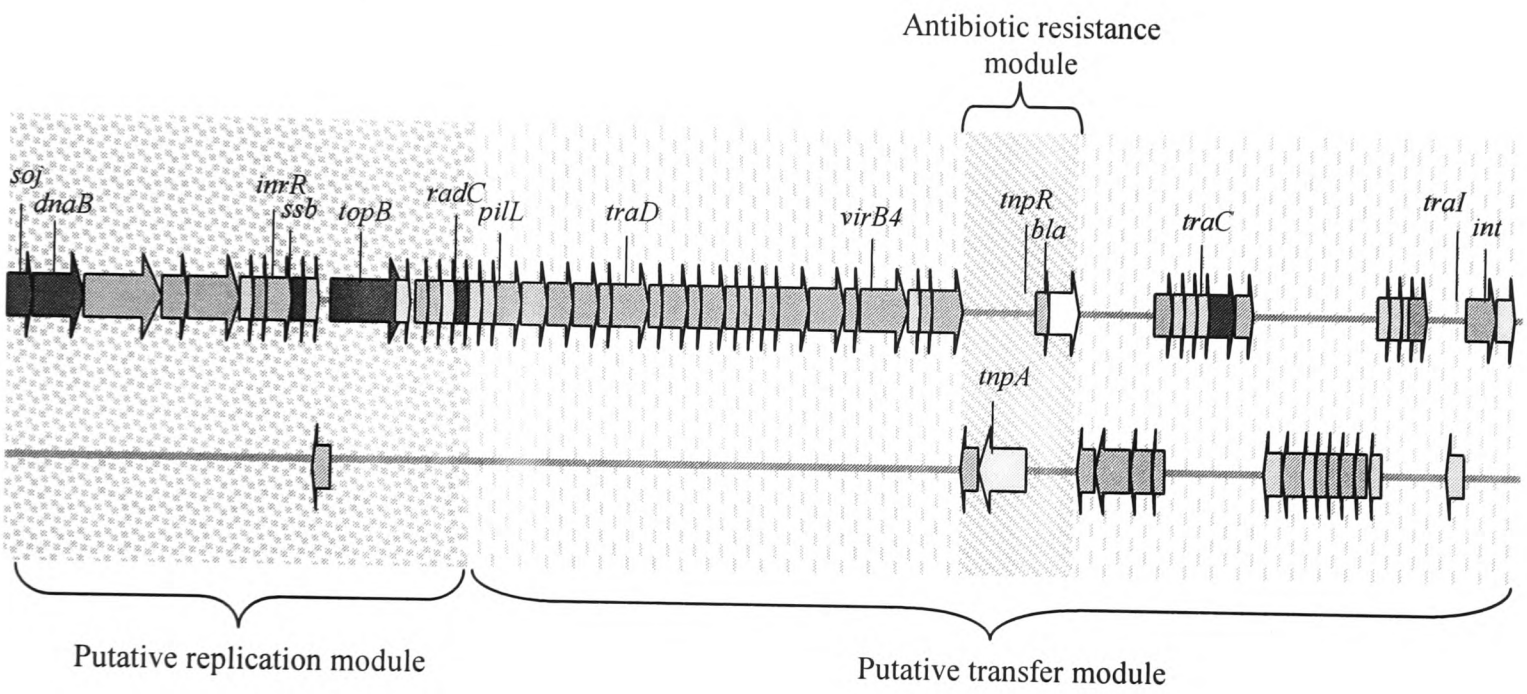


Fig. 5.2. Genetic organisation of ICE*Hin299*, according to putative functions.

5.4.3. *Putative replication module*

The complete sequence analysis shows that approximately 140 bases downstream of *attP* is a gene similar to the partitioning protein *soj* of *H. ducreyi* (p299.01), followed by the helicase gene *dnaB* (p299.02). Following these two genes are five ORFs (p299.03 to p299.07), which encodes for hypothetical proteins, of which four are conserved. These four genes not only shares homology to genes of *H. ducreyi* 35000HP (NC_002940), but they are also in syntheny with ICE*Hin1056*. The p299.08 is similar to the putative integrase regulator *inrR* gene of *H. ducreyi*, associated with the excisive recombination of

the element. Adjacent to this, is a gene similar to the *ssb* of *A. actinomycetemcomitans* encoded by p299.09, followed by a gene with homologue to conserved hypothetical protein of *H. ducreyi* (p299.10). Gene p299.11 encodes for protein that has low similarity (E value 9.6e-08) to Osa protein of *S. flexneri* (SWISSPORT:OSA_SHIFL). How such gene product with oncogenic suppression has a role in this element is unclear. Next to the Osa protein, is a gene similar to *H. ducreyi* topoisomerase III *topB* (p299.12). Genes p299.13 to p299.16 are predicted to produce proteins of unknown functions. Gene p299.17 has homology to RadC DNA repair protein that functions specifically in recombinational repair associated with replication. Presence of *soj, dnaB, ssb* and *topB* in this region is consistent with this module being involved with replication, presumably of the plasmid form of ICE*Hin299*. Surprisingly, no *parB* and *dnaA* genes were detected during the annotation of this element.

5.4.4. Putative transfer module

This is the biggest module (p299.18 to p299.39) consisting of a cluster of hypothetical proteins putative for DNA transfer. The presence of the lipoprotein gene (p299.18), glycine-rich protein (p299.19), PilL (p299.20), TraD (p299.24), VirB4 of Type IV secretory pathway (p299.33) and a series of membrane and exported proteins (p299.21, p299.22, p299.23, and p299.25 to p299.39), are consistent with this module being involved in DNA conjugative transfer.

5.4.5. Putative transfer/integration module

Genes p299.43 to p299.67 consist of 25 ORFs of which many encode hypothetical proteins with unknown functions. Only three ORFs in this module have homologues with known functions. Gene p299.51 encodes a homologue to DNA primase TraC of *H.*

ducreyi 35000HP (acc. no. AAP95800). The gene product of p299.66 is homologous to a conserved hypothetical gene of *Xylella fastidiosa* (acc. no. AE003998) and is also predicted to be a putative relaxase (similar to the TraI of *P. rettgeri* (acc. no. AAM08003). Gene p299.67 is similar to the *int* gene of *A. actinomycetecomitans* (EMBL AF006830). Presence of these genes likely explains the capability of ICE*Hin299* to carry out integrative recombination with its host's chromosome.

5.4.6. Antibiotic resistance module

This module consists of a cluster of genes belonging to the Tn*3* transposon, which has been inserted into the transfer module. Unlike ICE*Hin1059*, ICE*Hin299* has a complete copy of Tn*3*. The organisation of the genes is similar to the archetypal Tn*3* transposon (Grindley, 2002). It is flanked by two 38-bp inverted repeats. However, these two inverted repeat sequences are not identical. Other investigators have also reported similar observation (Takeya *et al.*, 1979). The β-lactamase TEM-1 *bla* gene is identical to the *bla* gene of *S. flexneri* (acc. no. U48775). The *tnpA* gene of Tn*3* on this element is found immediately adjacent to the one end of the transposon and is transcribed towards it. The TEM-1 β-lactamase gene of ICE*Hin299* is similar to the β-lactamase gene of ampicillin resistant *H. influenzae,* which were commonly isolated (Hasegawa *et al.*, 2003; Ling *et al.*, 1989; Scheifele *et al.*, 1982).

*5.4.7. G+C content of ICE*Hin299

Fig. 5.3 depicts the distribution of G+C content in the whole element of ICE*Hin299*. Between positions 32,401 to 37,666 bp, the G+C% was higher than the average percentage. This increased G+C% correlates with Tn*3*. This difference in G+C% is

typical of an insertion of a "foreign DNA" into another DNA sequence (Hacker *et al.*, 1997).

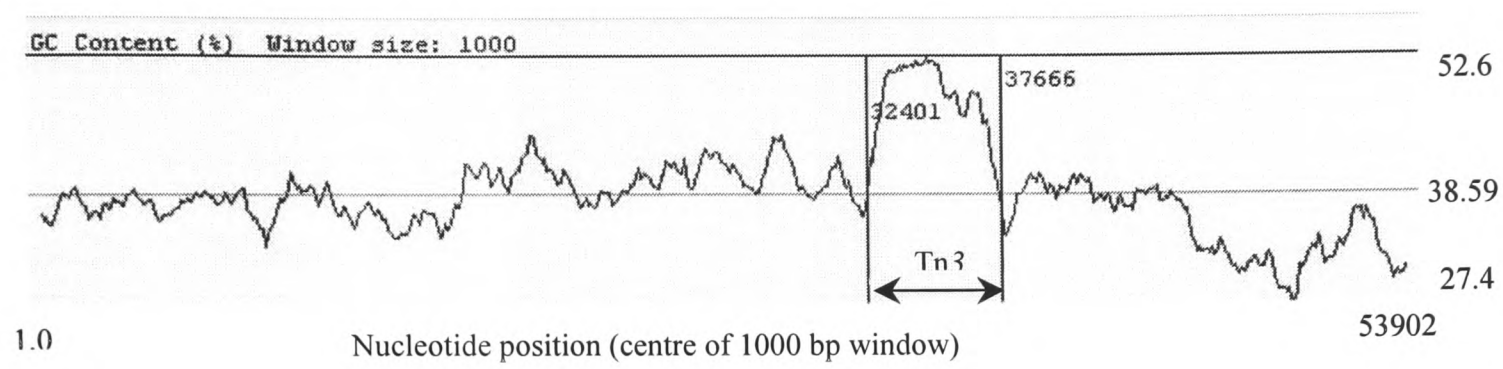


Fig. 5.3. G+C content of ICE*Hin299*, based on a 1000-window. The region between 32,401 and 37,666 bp has high G+C% indicating a recent DNA insertion. This graph was obtained by using Artemis program.

5.4.8. Determination of putative oriV *and* oriT

The *oriV* of most extrachromosomal DNA are usually characterised by structural elements such as, DnaA box sequences, multiple repeat sequences (interons), and an AT-rich region (Galli *et al.*, 2001). The DnaA box sequences act as binding sites for the host DnaA initiation protein. Binding of DnaA protein to the DnaA box consensus sequence, localised within a promoter region, can stimulate or inhibit transcriptional activity, hence the regulation of gene expression and/or stimulation of the initiation of DNA replication (Messer and Weigel, 1997). However, in ICE*Hin299*, no DnaA box (5'-TT(A/T)TNCACA) sequence was detected. If an *oriV* is present it would function independent of a DnaA box. Examples where DNA replication occurs without a DnaA box are plasmid RSF1010 and bacteriophage λ (Konieczny, 2003).

It is estimated that the *oriV* of ICE*Hin299* lies between positions 4,583 to 6195-bp. This prediction is based on the presence of multiple indirect repeats, two 33-bp and four 36-bp direct repeats in an AT-rich region (Fig. 5.4). This region also correlates to the G+C% skew of the element. The process of replication is predicted by bioinformatics to

terminate at approximately 32,000-bp position, as predicted by the Oriloc program (Fig. 5.5) (Frank and Lobry, 2000). This is opposite to the predicted *oriV*. The existence of specific replication termini suggests that ICE*Hin299* undergoes theta-type replication (Bussiere and Bastia, 1999).

Also from the graph depicted in Fig. 5.4, it is predicted that the ICE*Hin299*-specific *oriT* is located upstream of p299.66, at approximately 48,439-bp. This area has multiple repeats, two 13-mer direct repeats and two sets of inverted repeats. This region is adjacent to a gene similar to a relaxase gene of *P. rettgeri* (acc. no. AAM080030). This location is consistent to the observation by Murata *et al.* (2002), that *oriT* is usually located at or close to the end of the *tra* gene cluster (Murata *et al.*, 2002) and adjacent to a conjugative relaxase (Mills *et al.*, 1998). These are consistent to some of the common features found in transfer origins (Lanka and Wilkins, 1995).

The determination of *oriV* and *oriT* based on the characteristics of the nucleotide sequences is a prediction, and therefore, not confirmative. Experimental data is necessary to demonstrate the presence of *oriV* and *oriT* of an element (Mills *et al.*, 1998).

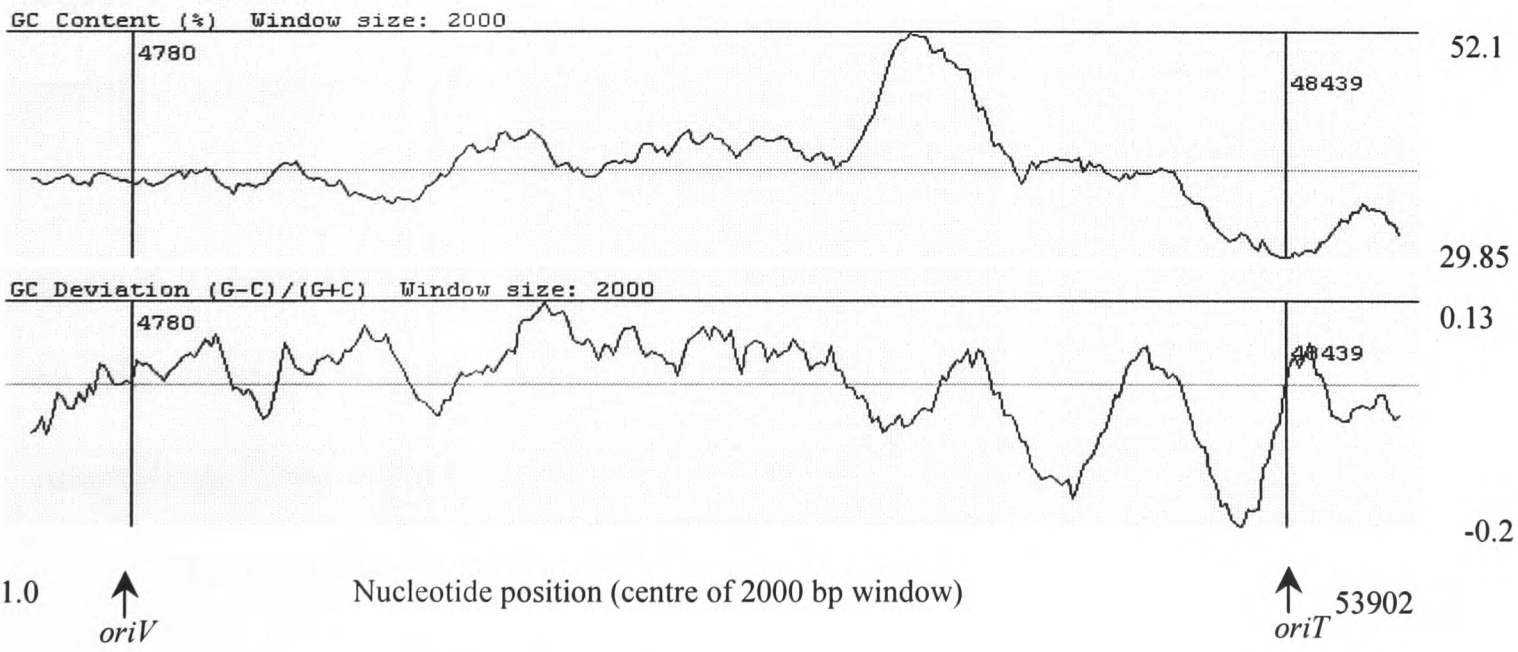


Fig. 5.4. G+C% content (above) and GC deviation [(G – C)/(G + C)] (below) of ICE*Hin299*, based on a 2000-bp window. Position 4,780-bp and 48,439-bp correlates to putative *oriV* and *oriT* respectively. This graph was obtained by using Artemis program.

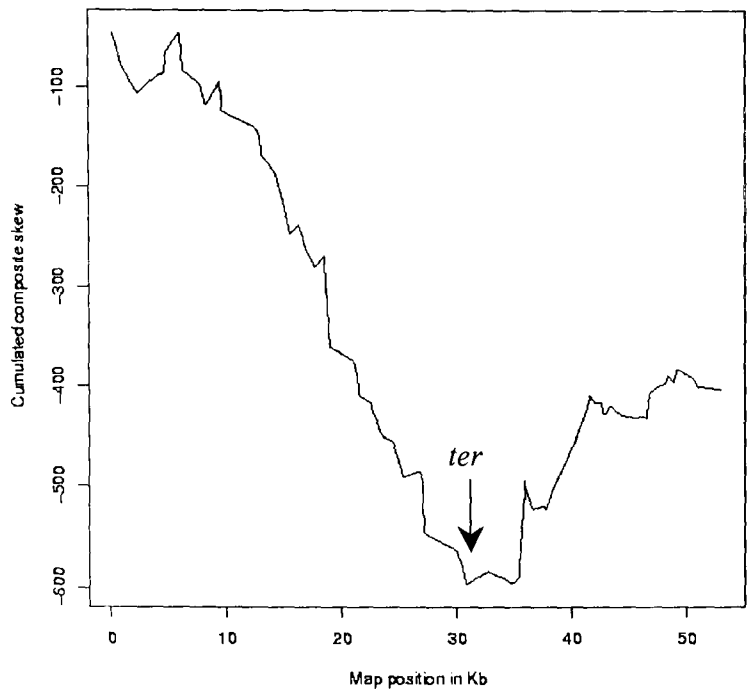


Fig. 5.5.An output of the Oriloc program showing the terminus of ICE*Hin299*. Replication of ICE*Hin299* is estimated to terminate at position 32,000-bp position.

5.5. Conclusion

The primary objective in determining the complete sequence of ICE*Hin299* was to provide a basis to further understand the characteristics of ICEs as the vehicles for dissemination of antibiotic resistance in *H. influenzae* strains. Although ICE*Hin1056* was sequenced as described in Chapter four, the sequencing of another ICE of *H. influenzae* is useful in comparative analysis of this family of element as described in Chapter six.

The complete nucleotide sequence data of ICE*Hin299* yielded a 53,902-bp element. It possesses only one antibiotic resistance gene *bla*, while the rest of the genes on this element are predicted to be mainly concerned with replication, transfer and integration. No genes were predicted to be associated with virulence.

The complete sequence of ICE*Hin299* further emphasizes that the genes of this mobile element are likely arranged in clusters according to functional modules. The detection of clusters of gene domains essentially equivalent to those of ICE*Hin1056*

strengthens the findings that these genes are required for transfer, propagation and integration of this family of ICEs.

CHAPTER SIX

COMPARATIVE STUDIES ON MOBILE ELEMENTS OF *HAEMOPHILUS* SPP.

6.2. Objectives

In the previous chapters, the molecular characteristics of two mobile elements of *H. influenzae* were described. Preliminary analysis of ICE*Hin1056* and ICE*Hin299* shows that these epidemiologically unrelated mobile resistance elements of *H. influenzae* share extensive sequence homology. The main aim of this chapter is to investigate the relationship between mobile elements found in *Haemophilus* spp. to better understanding the extent to which they share common origins.

6.3. Materials and Methods

6.3.1. Detection of additional elements on Haemophilus *spp.*

All genomes and plasmids of *Haemophilus* spp. in the NCBI that have been completely sequenced were compared to the sequences of ICE*Hin1056* using Artemis Comparison Tool (ACT). Those sequences that share sequence homology with ICE*Hin1056* were extracted and compared.

6.3.2. Haemophilus *spp. sequences of ICEs*

The complete sequence of ICE*Hin1056* and ICE*Hin299* obtained from Chapters 5 and 6 were used in this study. A nucleotide sequence of non-typeable *H. influenzae* R2866 genome, which is still incomplete, was kindly made available by Dr. Arnold Smith from Seattle Biomedical Research Institute, USA. Dr. Thomas Inzana, Virginia Tech, Blacksburg, USA provided *H. somnus* 2336 genome sequences, which is also incomplete.

A complete sequence of an element of *H. parainfluenzae* ICE*Hpa8F* was provided by Dr. Lucielle Mansfield, from the Nuffield Dept. of Clinical Laboratory Sciences, University of Oxford, UK.

Nucleotide sequences of complete genomes of *H. ducreyi* 35000HP (NC_002940) and *H. somnus* strain 129PT (NZ_AAB002000009); plasmids of *H. ducreyi,* pNAD1 (NC_005239), *H. aegyptius* pF3028 (NC_004058) and pF3031 (NC_004046), and *H. paragallinarum* p250 (NC_005245) were extracted from the NCBI database. A sequence of a 16-kb genomic island of *H. influenzae*, HiGI1 was obtained from accession number AF198256 (Chang *et al.*, 2000).

6.4. Results

6.4.1. Comparison of mobile elements of Haemophilus *spp.*

During the search for sequence homology of ICE*Hin1056* to the sequences of Haemophilus genomes and plasmids, *H. ducreyi* 35000HP and *H. somnus* 129PT were found to show sequence homology to ICE*Hin1056*. However, only human *Haemophilus* spp. that inhabit the nasopharynx were used for comparative studies. Relationships of ICE*Hin1056* to the elements of *H. ducreyi* and *H. somnus* will be described in Chapter 7.

The nucleotide sequences of four mobile elements of *Haemophilus* spp., ICE*Hin1056*, ICE*Hin299*, ICE*Hpa8F* and ICE*Hin2866* were viewed with Artemis. In Table 6.1, the general features of the four *Haemophilus* spp. are shown. Amongst the four elements, ICE*Hin1056* is the largest in size and has the highest G+C% content. This is attributed to the presence of Tn*10* in ICE*Hin1056* that is absent in the other elements. ICE*Hin299* has the most ORFs (67 predicted genes), but most of these additional ORFs are conserved hypothetical and hypothetical proteins with unknown functions. Although

the element of *H. influenzae* 2866 has the least ORFs, the average length of its predicted genes was the longest (838 bp).

The nucleotide sequences of these four elements were compared using ACT. Two different combinations of sequences were performed to ensure fair comparison between the sequences. Figs. 6.1 and 6.2 show that all the elements share extensive sequence homology with each other, and the genes are arranged mostly in the same order (syntheny). High sequence homologies (95-99%) were observed in the replication and conjugative modules, but in the conjugation/integration module, the percentage of homologies ranged widely. The lowest homology (32%) of sequences in this module was between ICE*Hin1056* and ICE*Hpa8F*, while ICE*Hin1056* shared the highest (98%) homology with ICE*Hin2866*.

Individual genes of each element were compared by pairwise alignment, and the presence of genes unique to an ICE was recorded; antibiotic resistance genes (ampicillin, chloramphenicol and tetracycline in ICE*Hin1056* and ampicillin gene in ICE*Hin2866*, ICE*Hin299* and ICE*Hpa8F*) were excluded in this analysis (Fig. 6.3). The figure shows that 46 predicted genes are conserved in all of the elements. These genes are possibly the essential genes (core genes) in ICEs of human oropharyngeal *Haemophilus* spp. In ICE*Hin2866*, ICE*Hin299* and ICE*Hpa8F*, there are three to six genes that are non-conserved. All of these genes are located 3' of *traC* and 5' to *traI* genes. None of these genes have known functions. These non-conserved genes are most likely accessory genes and have no predicted function. They may also mark a hot spot for insertion of accessory genes. In all the elements except ICE*Hin1056*, the insertion of Tn*3* occurred at the same location.

Table 6.1. General comparative features of mobile elements of four *Haemophilus* spp.

General features	**ICE*Hin1056***	**ICE*Hin299***	**ICE*Hin2866***	**ICE*Hpa8F***
Size	59,393 bp	53,902 bp	53,115 bp	52,725 bp
GC content	39.1%	38.6%	38.8%	38.7%
No. of uptake sequence	9	9	10	9
Open reading frame	64	67	54	60
Per cent of genome coding	84.4%	89.0%	85.8%	89.4%
No. of pseudo-genes	2	1	0	1
Functionally classified*	67% (43/64)	66% (44/67)	53.7% (29/54)	61.6% (37/60)
Conserved with no function	18.8% (12/64)	20.8% (14/67)	15% (9/54)	21.6% (13/60)
Hypothetical protein	10.9% (7/64)	23.8% (16/67)	29.6% (16/54)	15% (9/60)
Average length base	798 bp	725 bp	838 bp	795 bp
Insertion elements:				
IS*10*	2	0	0	0
IS*5*	2	0	0	0
IS*30*	1	0	0	0
Tn*3*	1 (truncated)	1	1	1
Repeat sequences:				
Direct	8	5	4	4
Inverted	56	44	42	45

* as determined by Psort, Pfam, Prosite, HTH motif, HMTMM, FASTA and BLASTP, e.g. transmembrane or DNA binding site etc.

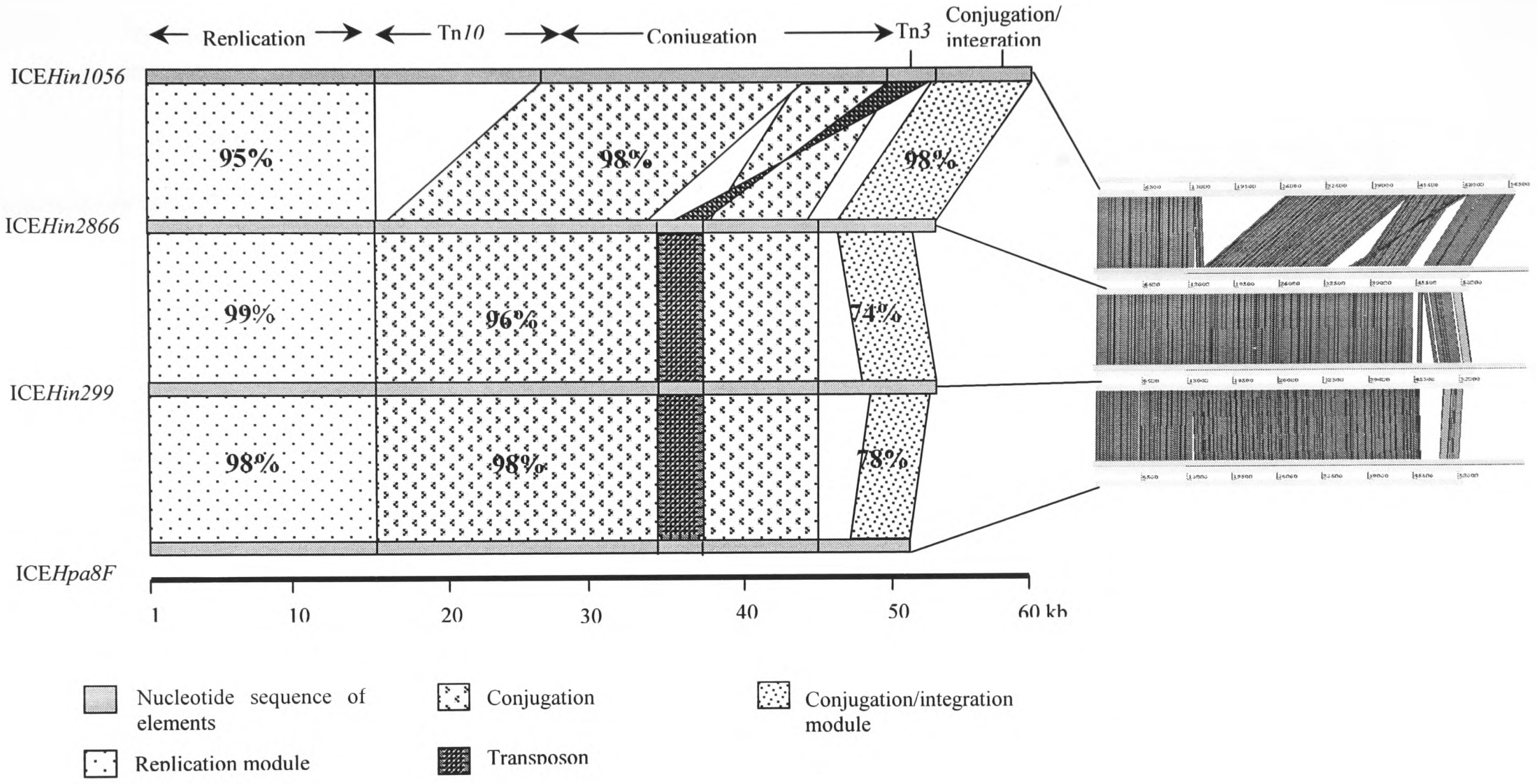


Fig. 6.1. Schematic diagram of ACT of four mobile elements of *Haemophilus* spp. Percentage sequence homology of each functional module between ICE*Hin1056*, ICE*Hin2866*, ICE*Hin299* and ICE*Hpa8F* was compared. The small insert of actual diagram of ACT shows that the genes are arranged in syntheny.

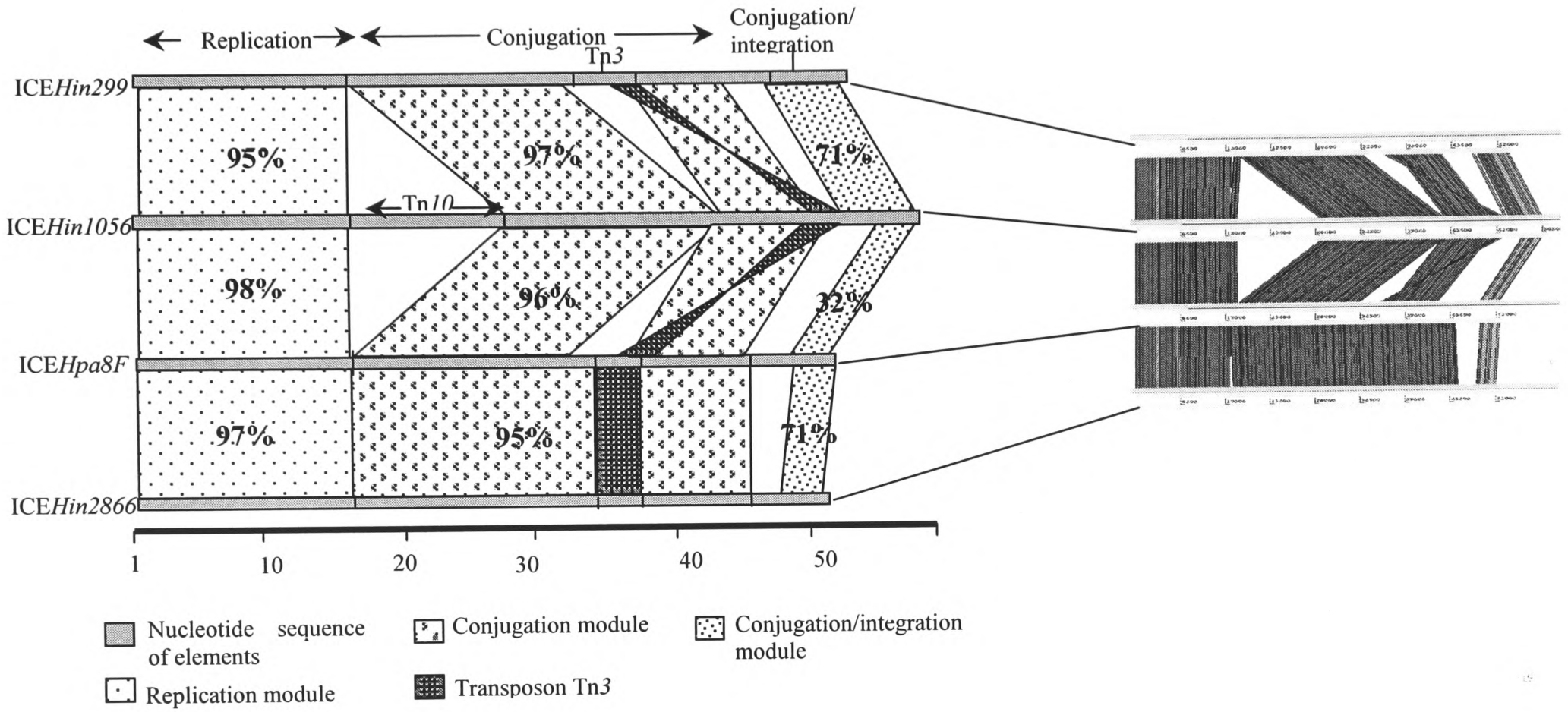


Fig. 6.2. Schematic diagram of ACT of four mobile elements of *Haemophilus* spp. (arranged in a different combination). Percentage sequence homology of each functional module between ICE*Hin299*, ICE*Hin1056*, ICE*Hpa8F*, and ICE*Hin2866* were compared. The small insert of actual diagram of ACT shows that the genes are arranged in syntheny.

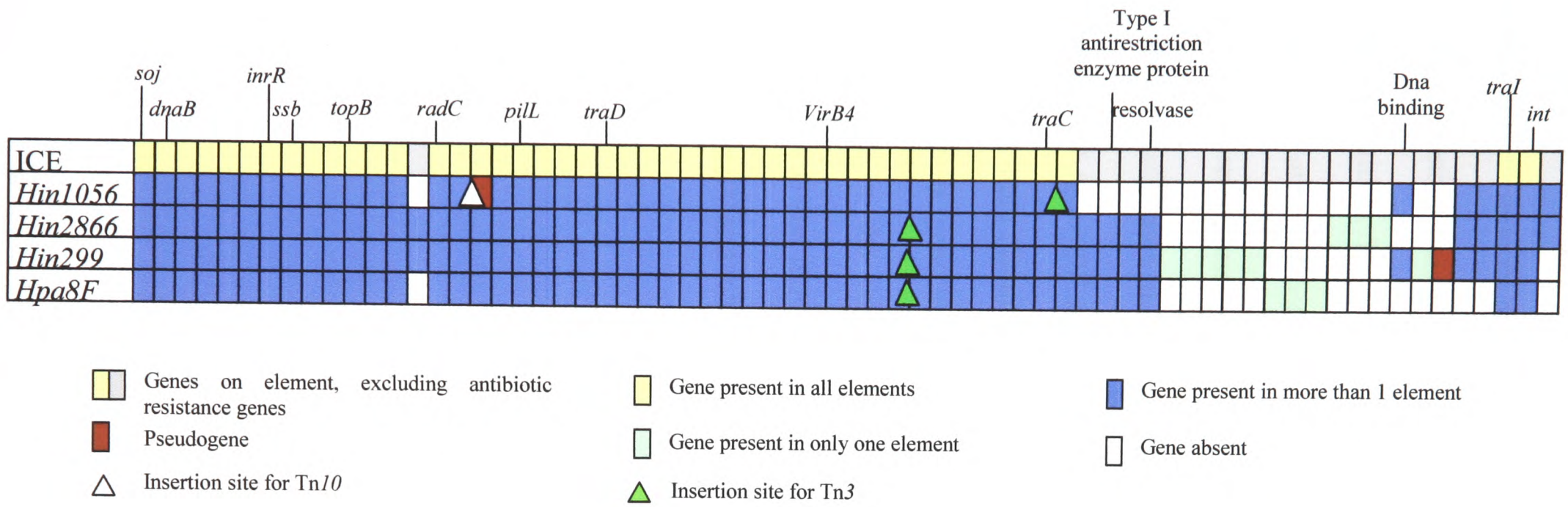


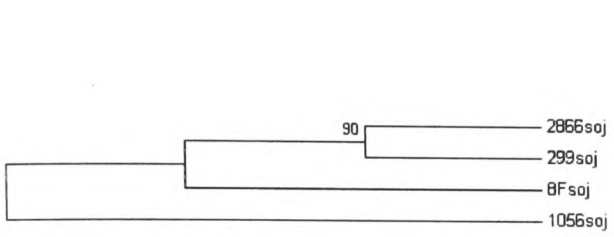
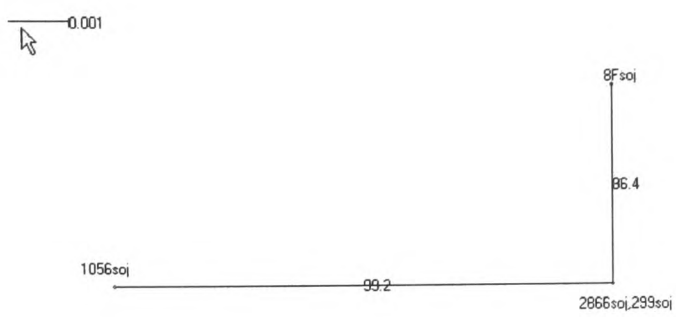
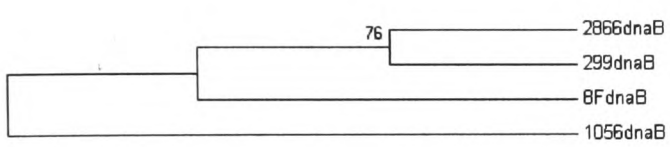
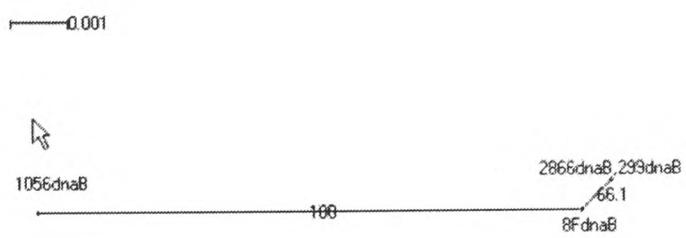
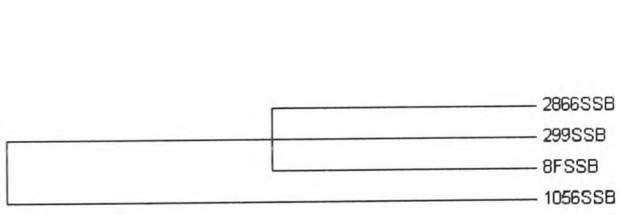
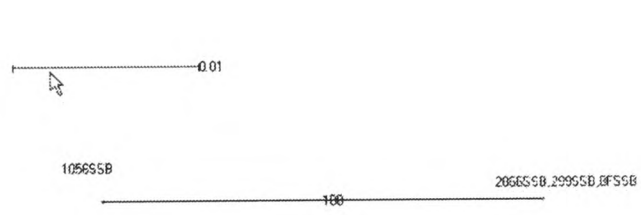
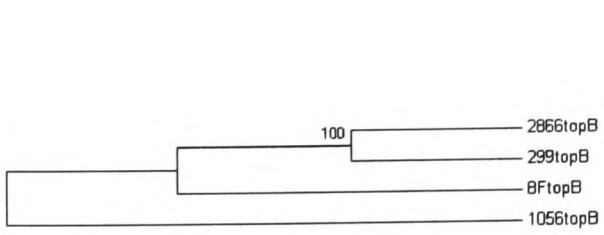
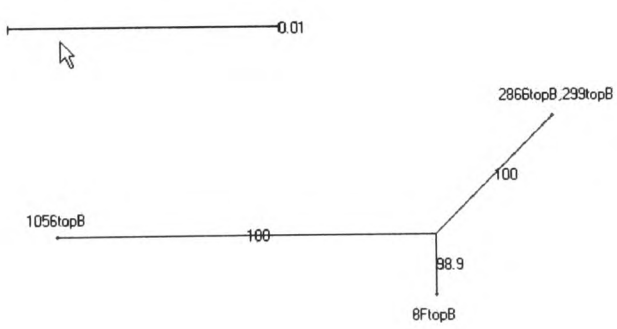
Fig. 6.3. An array of predicted genes present in mobile elements of four *Haemophilus* spp. ICEs. Absence or presence of genes was recorded to reveal conserved genes and differences between ICEs.

6.4.2. *Phylogenetic relationship between the genes in the* Haemophilus *spp.*

Relationships between 12 genes present in all of the elements were determined. Nucleotide sequences of four genes in the replication module, six genes in the conjugation module and two genes in the integration/integration module were chosen. Phylogenetic relationship between the genes was determined using MEGA and SplitsTree decomposition. The results are shown in Table 6.2 and pairwise distances between the genes, created by MEGA, are shown in Appendix C.

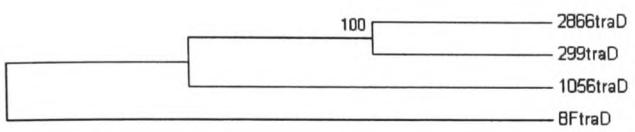
Overall, results showed that genes in ICE*Hin1056* are constantly found to be more distant than genes of other elements. Although the genes in the replication module *soj*, *dnaB*, *ssb* and *topB* are congruent to each other, the genes in ICE*Hin1056* seem to be relatively more distant. Similarly in the conjugation module, the *traD* gene of ICE*Hpa8F* was the only gene showing a slight change in the congruence of the tree topologies. However, in the conjugation/integration module, the relaxase and the primase of ICE*Hin1056* and ICE*Hin2866* were, by comparison, very distantly related to those found in ICE*Hin299* and ICE*Hpa8F*.

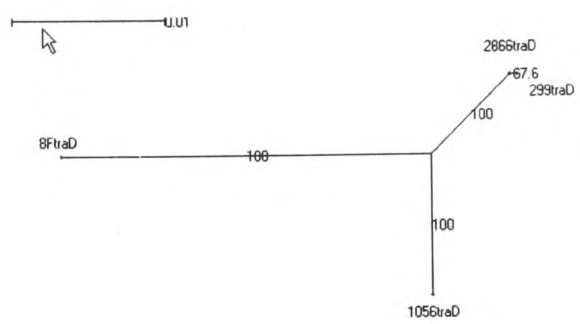
Table 6.2. Phylogenetic relationship of genes in mobile elements of four *Haemophilus* spp.

UPGMA output	SplitsTree decomposition output
soj	
	
dnaB	
	
ssb	
	
topB	
	

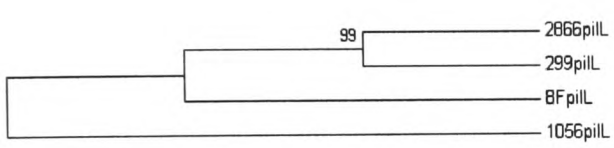
UPGMA output	SplitsTree decomposition output

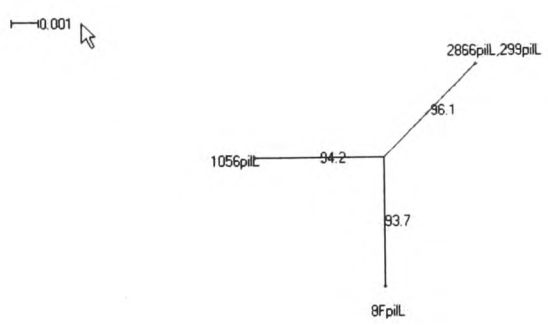
traD



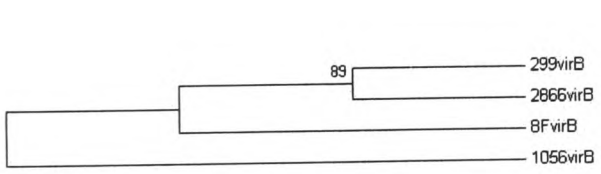


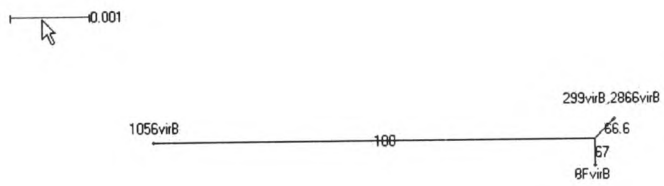
pilL



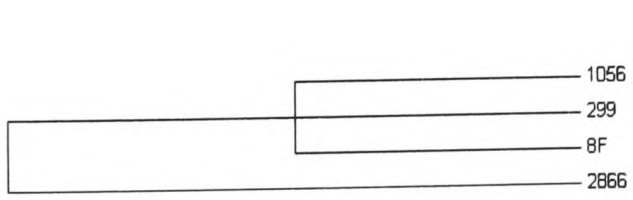


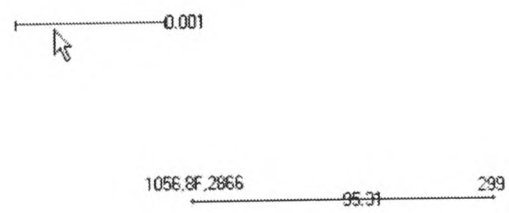
virB4





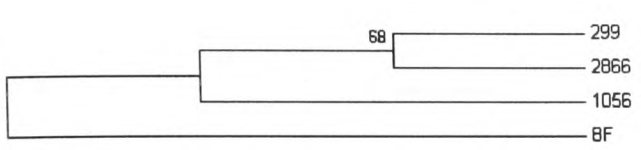
Gene 48

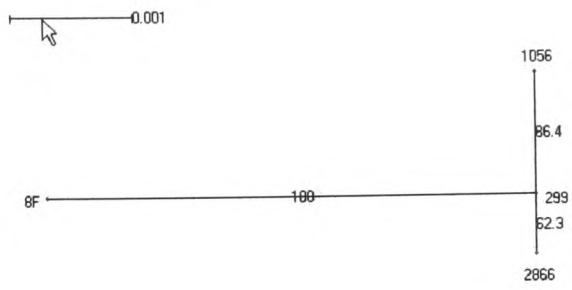




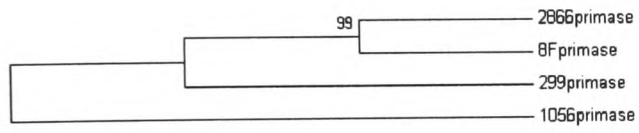
UPGMA output	SplitsTree decomposition output

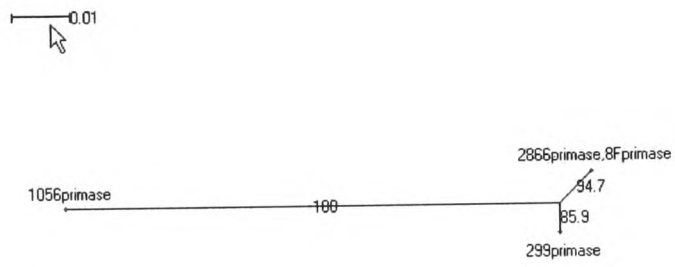
Gene 52



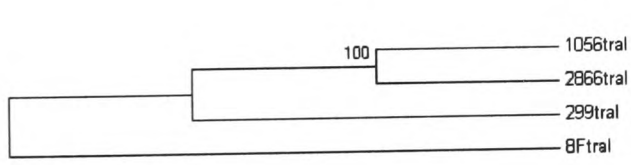


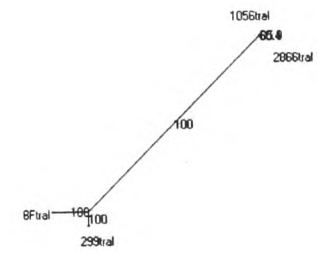
traC



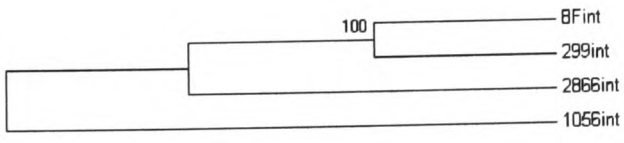


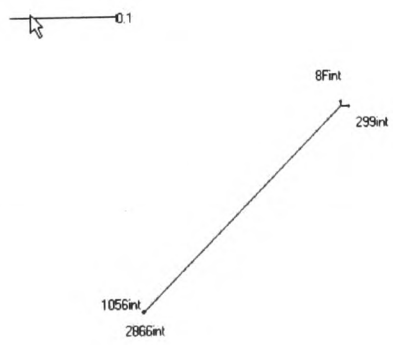
traI





int





6.4.3. *Comparison of ampicillin resistance genes*

Sequence homology of the genes on the Tn*3* transposon of the elements was investigated. The nucleotide sequences of these genes were compared using ACT and pairwise alignment for the percentage of sequence homology were carried out using ClustalW (Fig. 6.4).

One major difference between the Tn*3* ICE*Hin1056* to the other elements is that it does not have the gene that encodes transposase. With the exception of ICE*Hin1056*, the *tnpA*, *tnpR* and *bla* genes of ICE*Hin299*, ICE*Hpa8F* and ICE*Hin2866* are identical. ICE*Hin1056* shares an identical *bla* gene with these ICEs, but has 3% base difference in its *tnpR* gene. However, the ampicillin resistance gene *bla* of the *Haemophilus* spp. is between 1-6% different in nucleotide sequence to the Tn*3* associated *bla* found in *E. coli*.

The sequences (100-bases) flanking the Tn*3* were investigated to determine the preferred site for insertion of the transposon. The truncated Tn*3* of ICE*Hin1056* was located at a different location to those found in the other three ICEs. Tn*3* inserted at the identical site in ICE*Hin299*, ICE*Hin2866* and ICE*Hpa8F*, as deduced from analysing the sequences at the point of insertion (the ampicillin genes of ICE*Hin1056* were excluded in the comparison because ICE*Hin1056* does not have complete sequences). See the nucleotide sequences at the left junction between Tn*3* and the three ICEs, Fig. 6.5. A similar observation was noticed at the right junction of Tn*3*.

A phylogenetic relationship of ampicillin genes between *Haemophilus* spp. and *E. coli* (acc. no. V00613 and plasmid pIGAL1, acc. no. AY424310) are shown in Table 6.3. The ampicillin resistance genes within *Haemophilus* spp. were congruent but are not relatively different from those of *E. coli*.

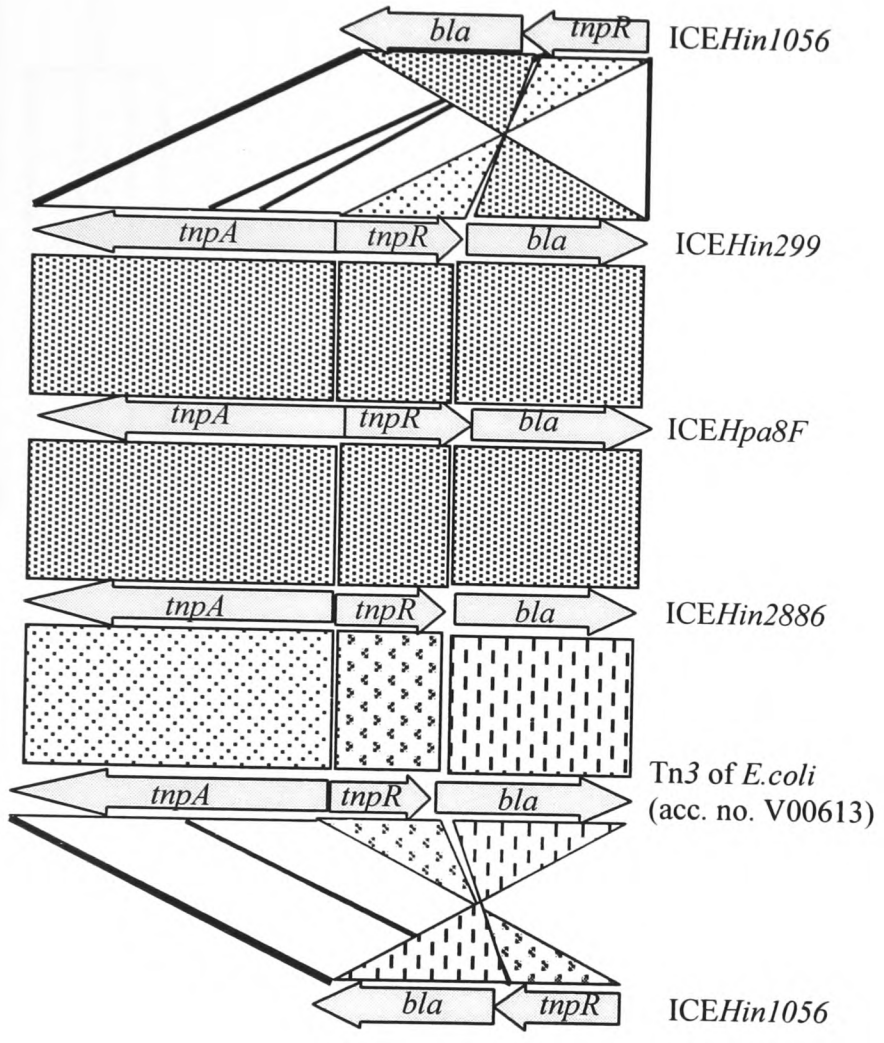


Fig. 6.4. Schematic diagram of Tn*3* transposons of four *Haemophilus* spp. in comparison to *E. coli* (acc. no. V00613).

```
CLUSTAL W (1.82) MULTIPLE SEQUENCE ALIGNMENT
```

Core ICE sequences

Tn*3* terminal repeat

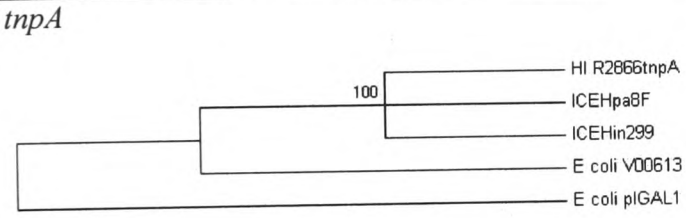
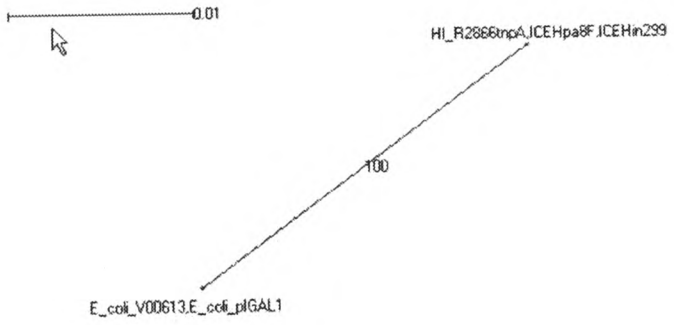
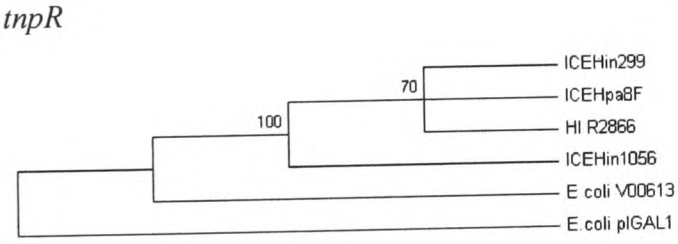
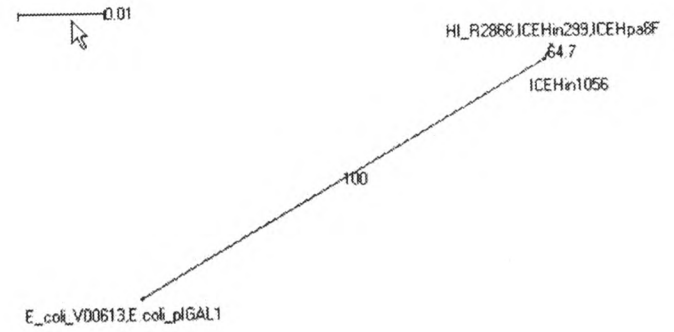
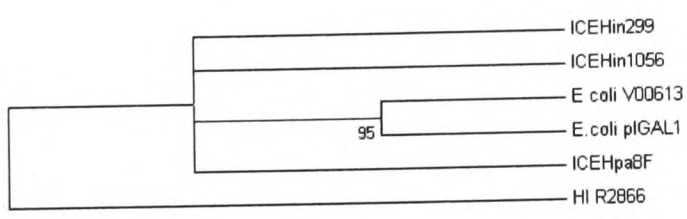
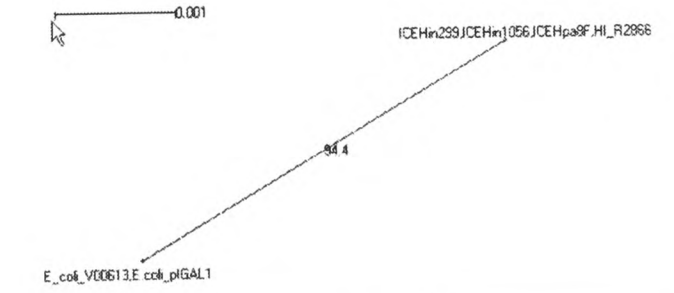
```
2866        ACGTCGCAAATTTTACTAATTGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGC
299         ACGTCGCAAATTTTACTAATTGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGC
8F          ACGTCGCAAATTTTACTAATTGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGC
            ***********************************************************

2866        AACGTTTTCTGCCTCTGACGCCTCTTTTAATGGTCTCAGATGTCCTTTGGTCACCAGTTC
299         AACGTTTTCTGCCTCTGACGCCTCTTTTAATGGTCTCAGATGTCCTTTGGTCACCAGTTC
8F          AACGTTTTCTGCCTCTGACGCCTCTTTTAATGGTCTCAGATGTCCTTTGGTCACCAGTTC
            ************************************************************
```

3'-Part of *tnpA*

Fig. 6.5. Part of the nucleotide sequences of left junction between Tn*3* of three ICEs. Conserved core ICE sequences observed at the junction. Sequences were aligned using ClustalW.

Table 6.3. Phylogenetic relationship of ampicillin resistance genes of *Haemophilus* spp. and *E. coli*

UPGMA output	SplitsTree decomposition
tnpA	
	
tnpR	
	
bla	
	

6.4.4. *Comparisons of* oriV, ter *and* oriT

The putative regions for *oriV* of the Haemophilus ICEs were predicted and shown as depicted in Fig. 6.6. This prediction was based on the presence of iterons type sequences, multiple indirect repeats in an AT-rich region and located in an intergenic region. The sequences 200-bp upstream of the iterons are fairly conserved with three bases differences, however, the nucleotide sequences of the presumed iterons are variable (Fig. 6.7)

Replication terminates at a predicted *ter* site. In all of the elements, the predicted *ter* sites are indicated by the G+C% skew as shown in Fig. 6.8. At this site, all of the elements have multiple indirect repeats. Presence of TMHMM motif, signal peptide with cleavage site and ATP/GTP binding site are predictive of *ter*-associated genes. The nucleotide sequences immediately before and after the predicted *ter* site are highly conserved.

The presence of predicted genes with the following: HTH motifs, TMHMM motifs, the presence of homologues to primase and relaxase and inverted repeats in an AT-rich region predicts a putative *oriT* (Fig. 6.9). This position also maps asymmetrically to the putative conjugative module, thereby, conforming to another of the features of an *oriT* (Lanka and Wilkins, 1995). The indirect repeats predicted to represent *oriT* have a core 8-bp (ATGGGGAG) conserved in all ICEs. However, there are sequences variations between the ICEs (Fig. 6.10). The nucleotide sequence flanking the putative *oriT* of ICE*Hin1056*, ICE*Hin299* and ICE*Hin2866* are highly conserved, but in ICE*Hpa8F*, only the sequences containing the indirect repeat of the putative *oriT* and those immediately adjacent to it are conserved (Fig. 6.10).

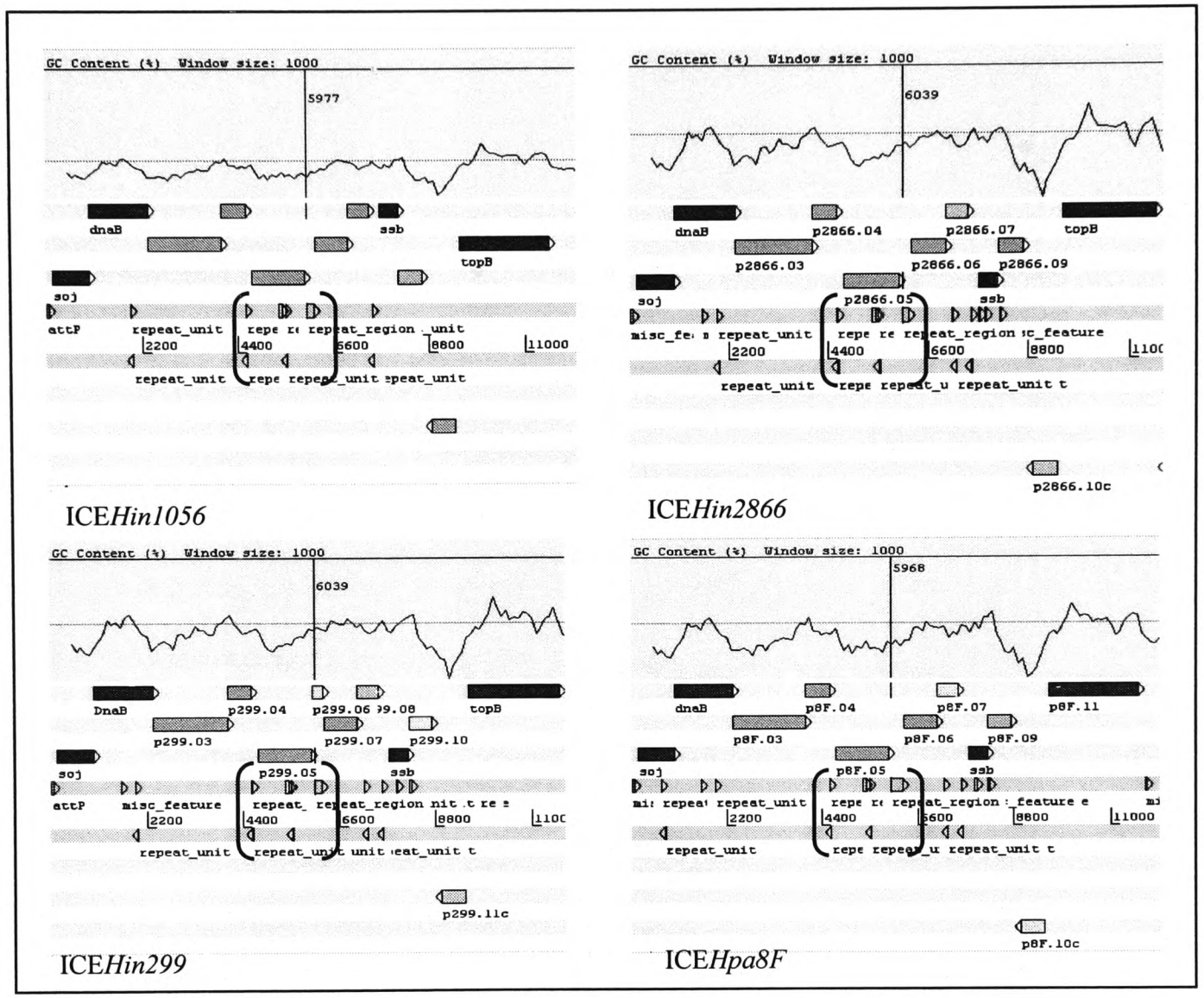


Fig. 6.6. Putative origin of replication (*oriV*) of four mobile elements of *Haemophilus* spp. Presence of direct repeats and multiple indirect repeats in low G+C% region, indicative of the region of an *oriV* (circled). The small arrowheads represent indirect repeats and the short arrow within the circle represents interons. This diagram was obtained by using Artemis program.

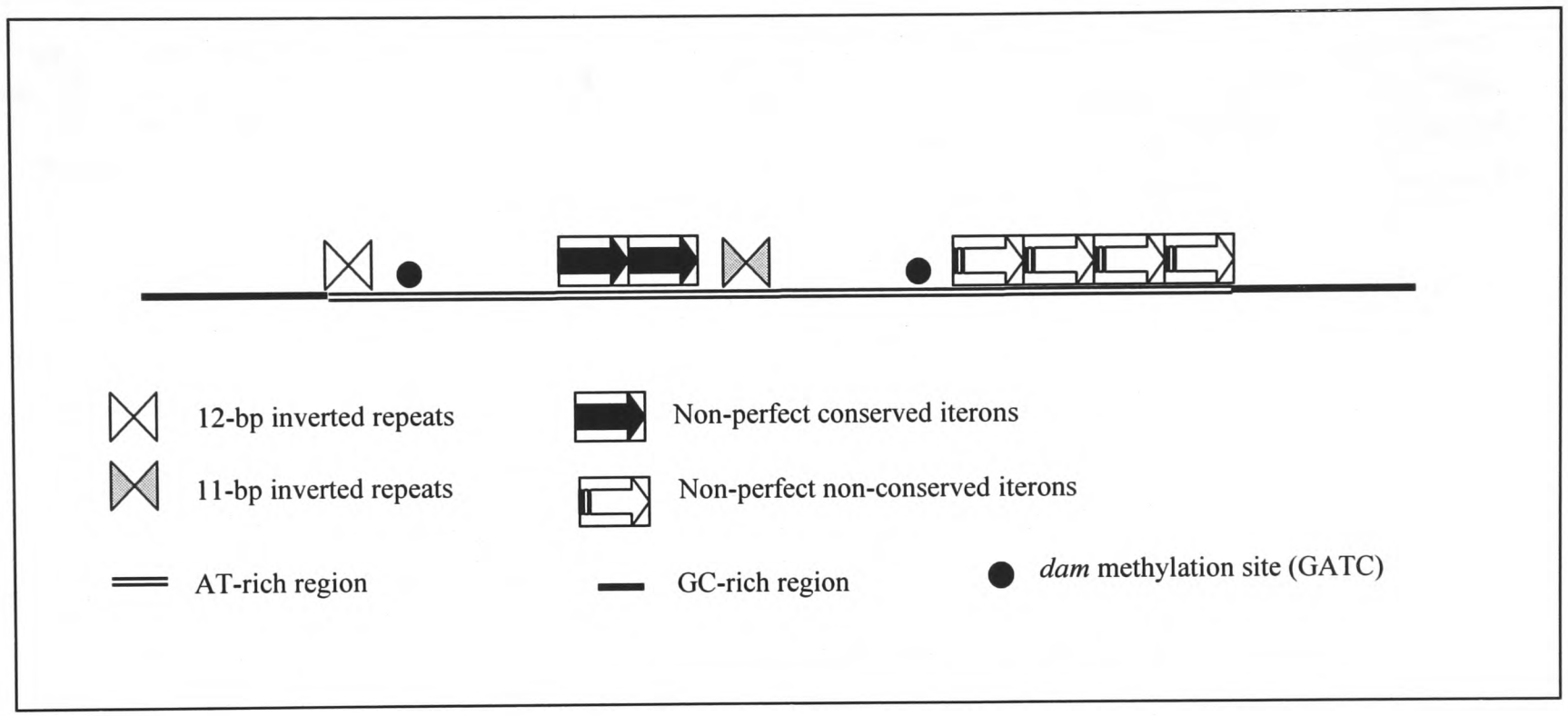


Fig. 6.7. Schematic diagram of the *oriV* region of ICEs.

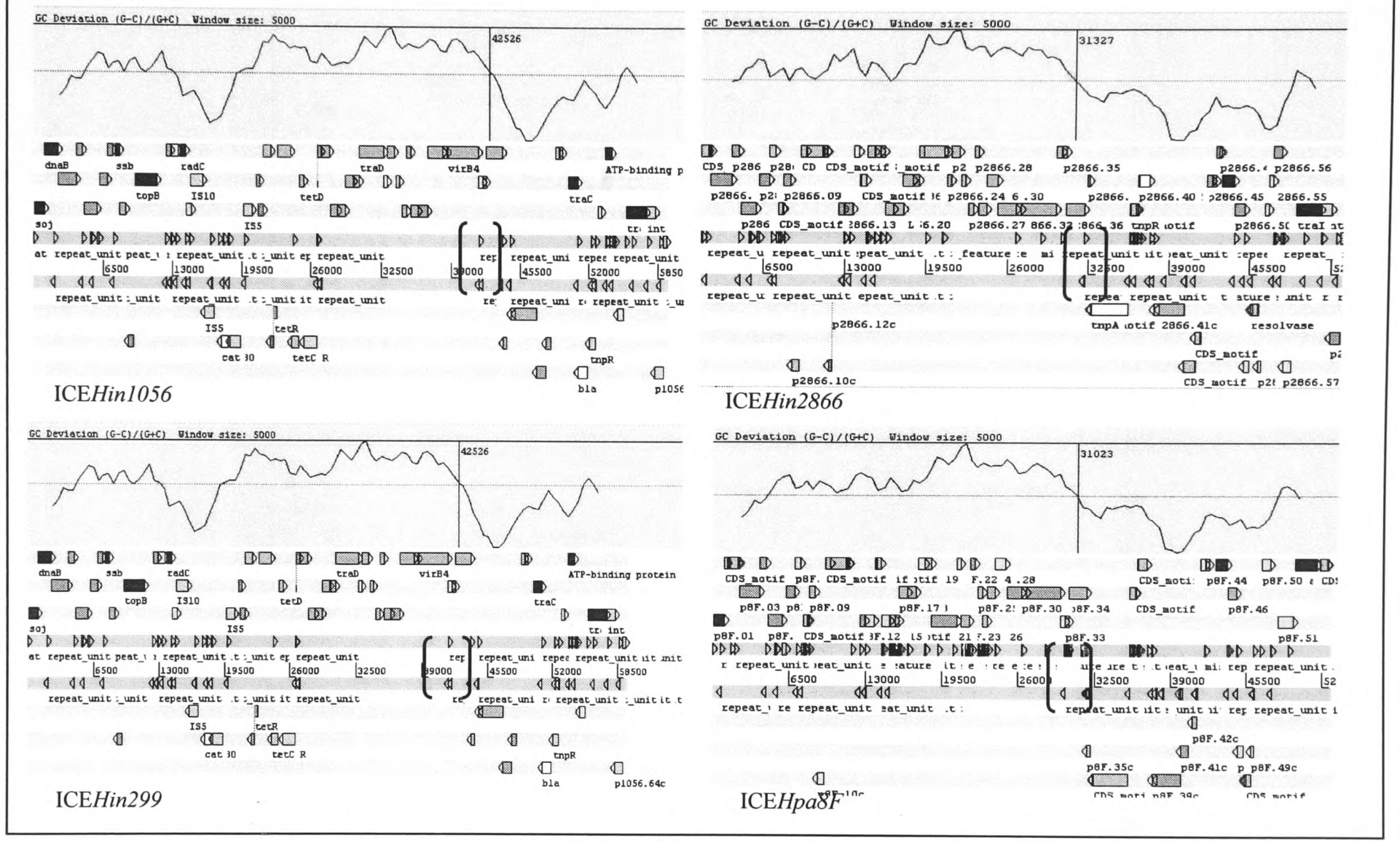


Fig. 6.8. Putative replication termination (*ter*) site of the four mobile elements of *Haemophilus* spp. The *ter* sites are bracketed. The small arrowheads represent indirect repeats and the short arrow within the circle represents interons. This diagram was obtained by using Artemis program.

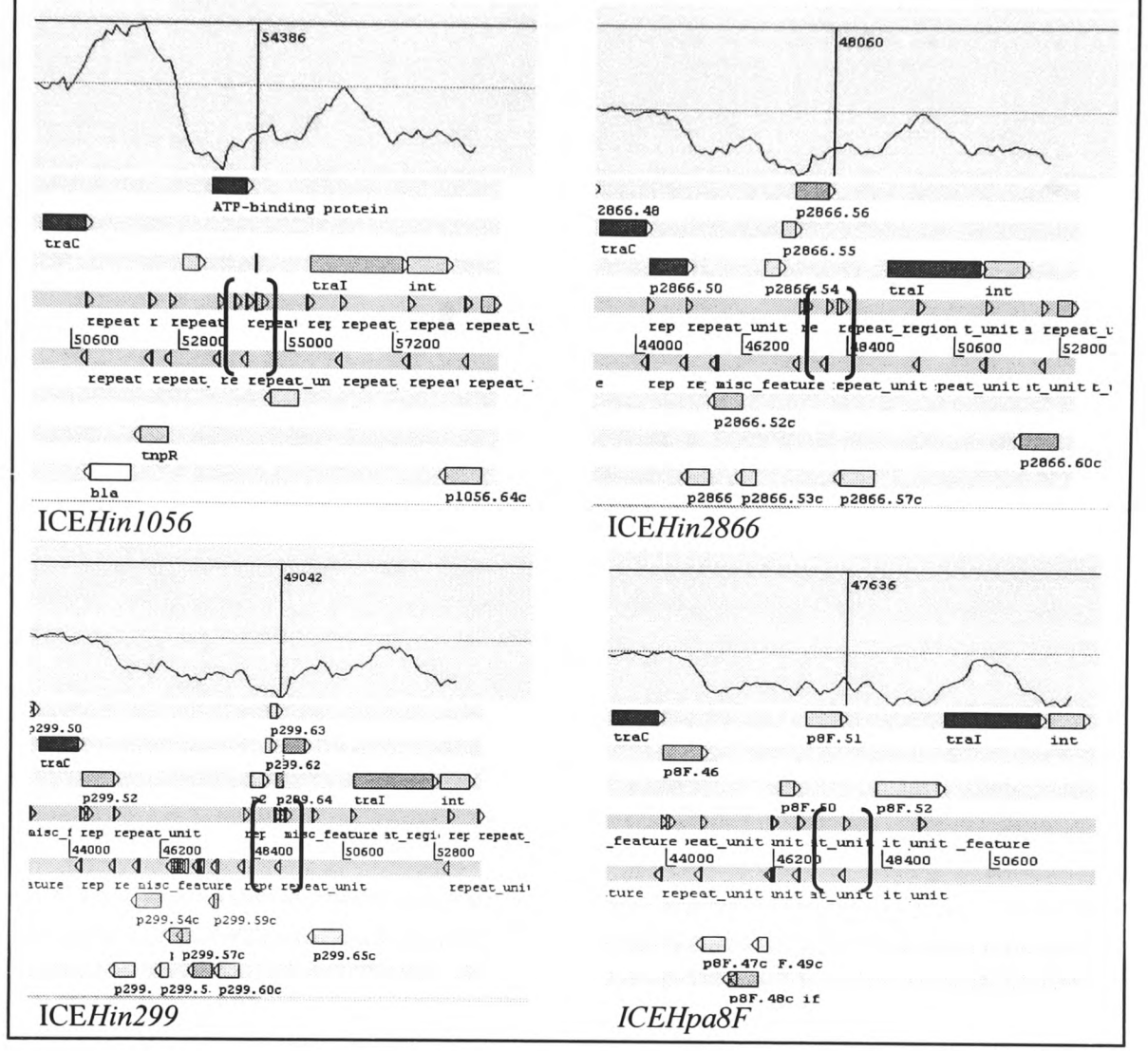


Fig. 6.9. Putative origin of transfer (*oriT*) of four mobile elements of *Haemophilus* spp. Presence of direct repeats and multiple indirect repeats in low G+C% region, indicative of the region of an *oriT* (bracketed). The small arrowheads represent indirect repeats and the short arrow within the circle represents interons. This diagram was obtained by using Artemis program.

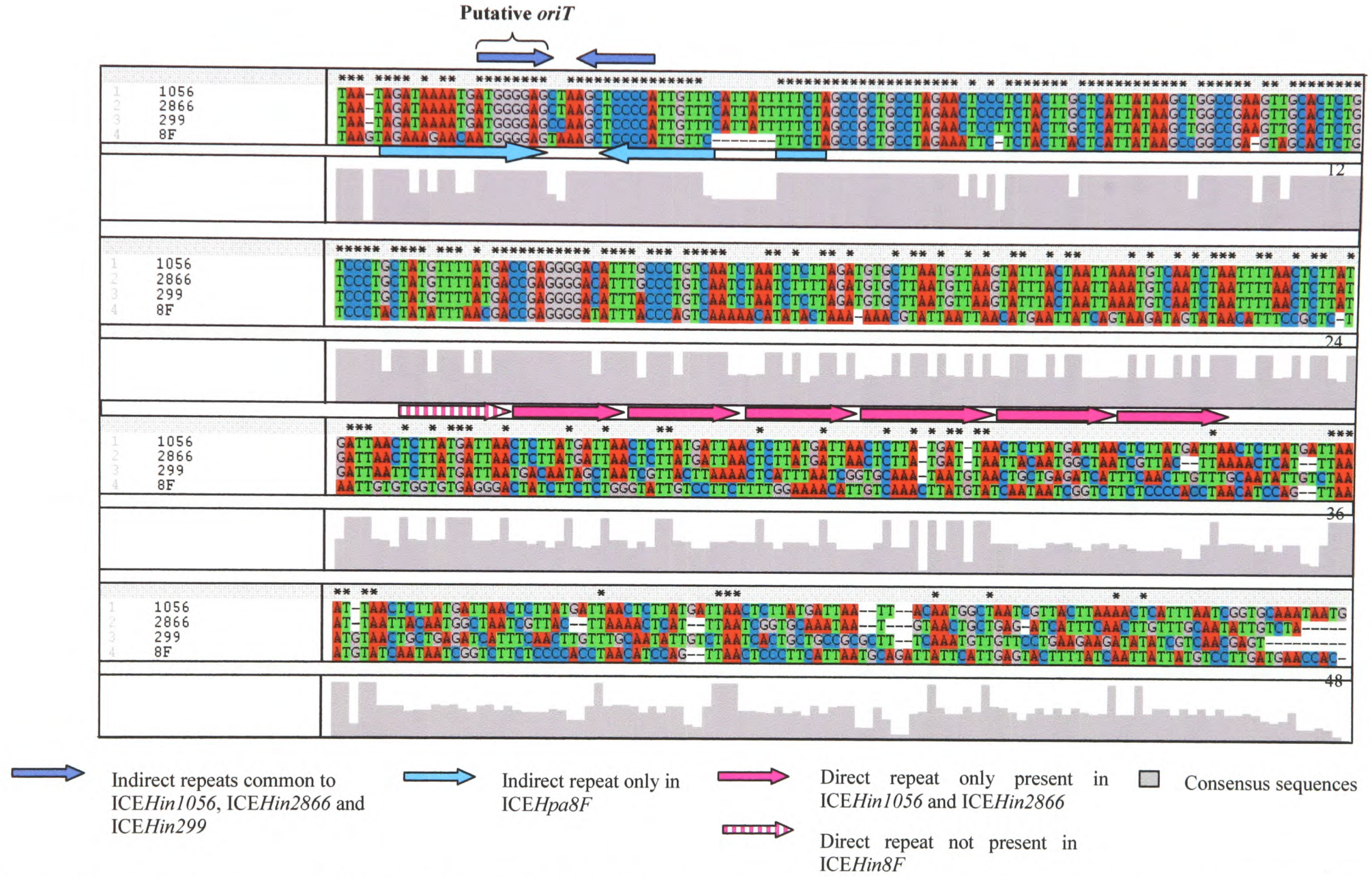


Fig. 6.10. Alignment of sequences of the four elements closed to the *oriT*. Note that the sequences are highly conserved in ICE*Hin1056*, ICE*Hin299* and ICE*Hin2866* but in ICE*Hpa8F*, only the site of *oriT* is conserved. Alignment performed by ClustalW.

6.4.5. *Comparison of ICE*Hin1056 *to other Haemophilus plasmids and genetic islands*

6.4.5.1. Genetic island of HiG1

H. influenzae strain Eagan has been reported to possess a 16-kb genetic island, HiG1 is inserted into the genome at the 3' end of $tRNA^{leu}$ and possesses an integrase gene located downstream of the $tRNA^{leu}$ (AF198256) (Chang *et al.*, 2000).

An ACT was performed for HiG1 against ICE*Hin1056* and the result is shown in Fig. 6.11. No significant matches between these two sequences were observed. In addition, pairwise comparison between the sequences did not give any significant similarity. Although this locus has several features of a genetic island, it does not show an organisation or sequence homologies that suggests it shares features in common with the ICEs reported here other than inserting in $tRNA^{leu}$ and possessing an integrase.

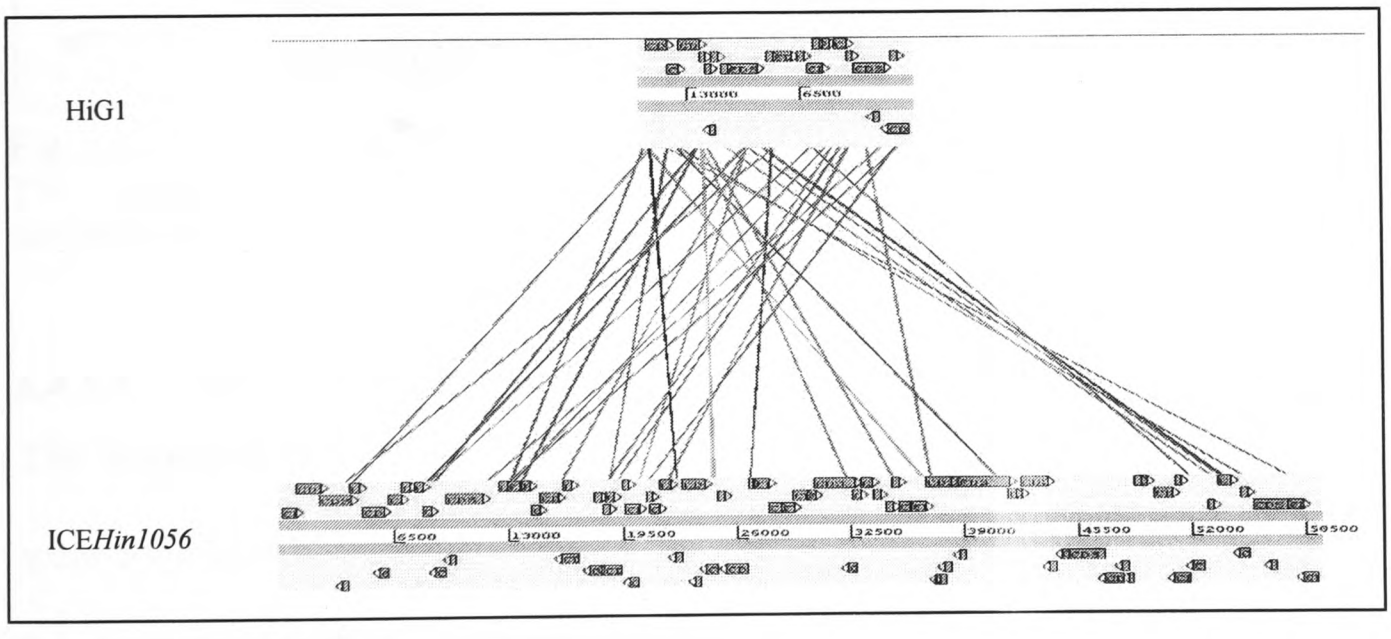


Fig. 6.11. An ACT diagram of sequence comparison between genetic island HiG1 and ICE*Hin1056*. The vertical lines denotes amino acid homology of >25%. No gene similarity was observed.

6.4.5.2. *H.* aegyptius *plasmids*

Two plasmids (pF3028 and pF3031, acc. no. NC_004058 and NC_004846 respectively) have been isolated from *H. influenzae* biotype aegyptius causing Brazilian Purpuric Fever in children. An ACT comparison was carried out between the sequences of ICE*Hin1056* and the plasmids. Fig. 6.12 shows that only *topB* of the *H. aegyptius* plasmids was homologous to ICE*Hin1056*. In addition, no significant similarities at nucleotide level were observed by pairwise alignment.

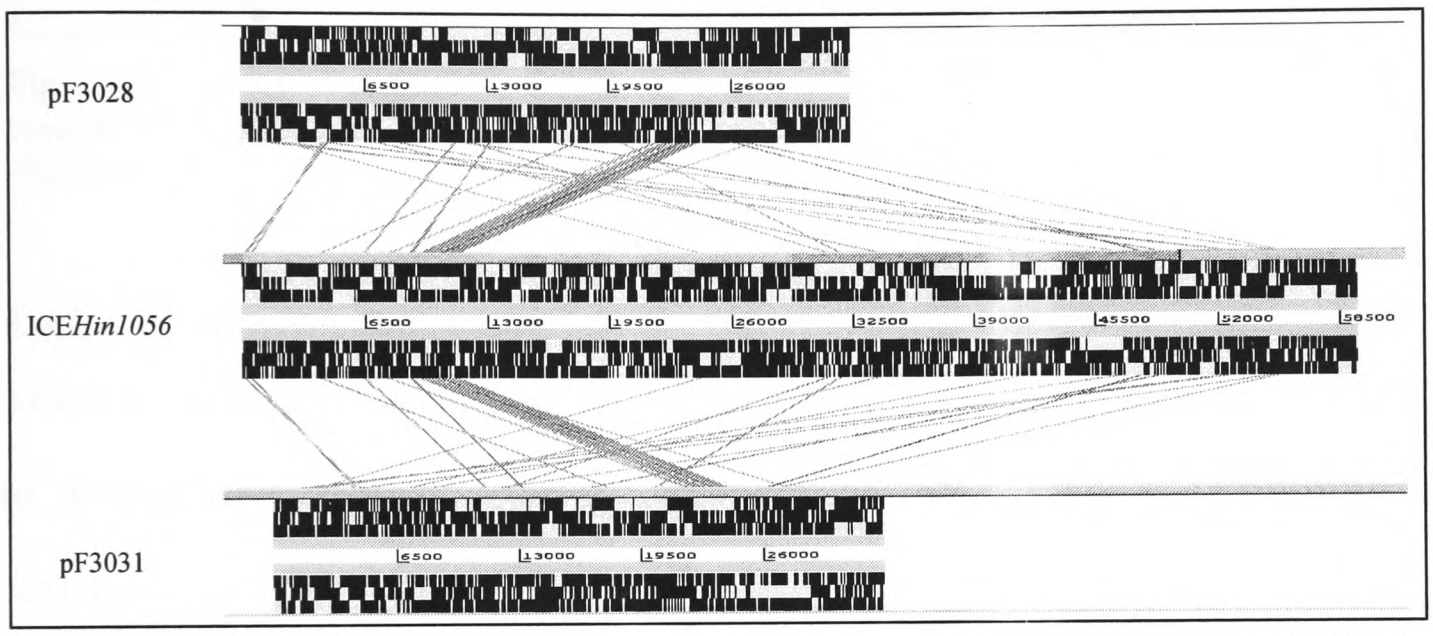


Fig. 6.12. An ACT diagram of sequence comparisons between pF3028, pF3031 and ICE*Hin1056*. The vertical lines denotes amino acid homology of >25%. The thick line indicates the gene similarity between the sequences.

6.4.5.3. H. ducreyi *pNAD1*

The sequences of a 5.6-kb plasmid pNAD1, isolated from *H. ducreyi* strain ATCC27722 were compared to ICE*Hin1056*. The ACT diagram in Fig. 6.13 shows that pNAD1 do not have any gene similarities to ICE*Hin1056*.

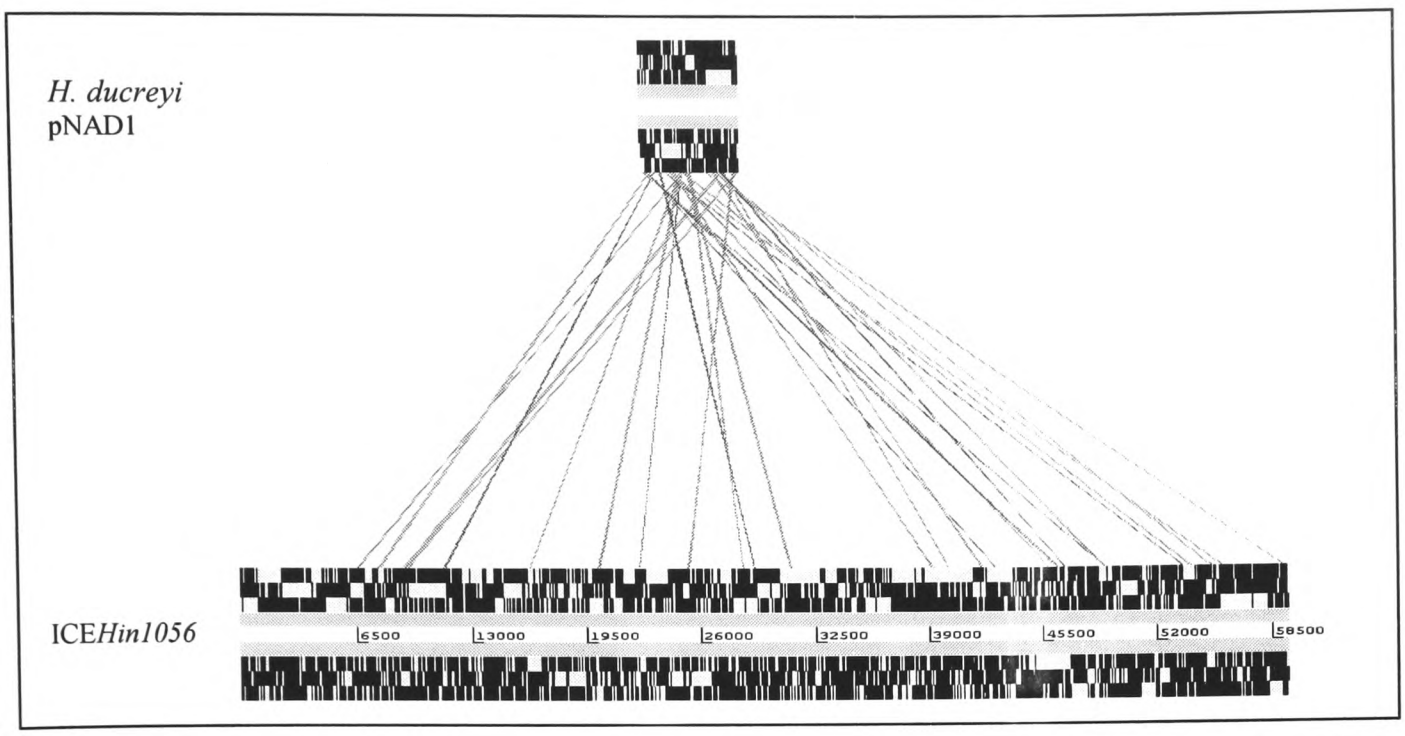


Fig. 6.13. An ACT diagram of sequence comparison between pNAD1 and ICE*Hin1056*. The vertical lines denotes amino acid homology of >25%. No gene similarity between the sequences is observed.

6.4.5.4. H. paragallinarum p250

The nucleotide sequence of a plasmid isolated from an avian Haemophilus was compared to ICE*Hin1056*. Plasmid p250 encodes haemocin in *H. paragallinarum* HP250. It contains a putative integrase, a putative replication protein (*repB*) and five haemocin genes (Terry *et al.*, 2003). Haemocin in *H. paragallinarum* is believed to aid in colonization by inhibiting other Gram-negative bacteria in the respiratory tracts of chickens.

The result of ACT between p250 and ICE*Hin1056* is shown in Fig. 6.14. No genes of p250 were similar to ICE*Hin1056*. This observation indicates that p250 does not have shared sequences with ICE*Hin1056*.

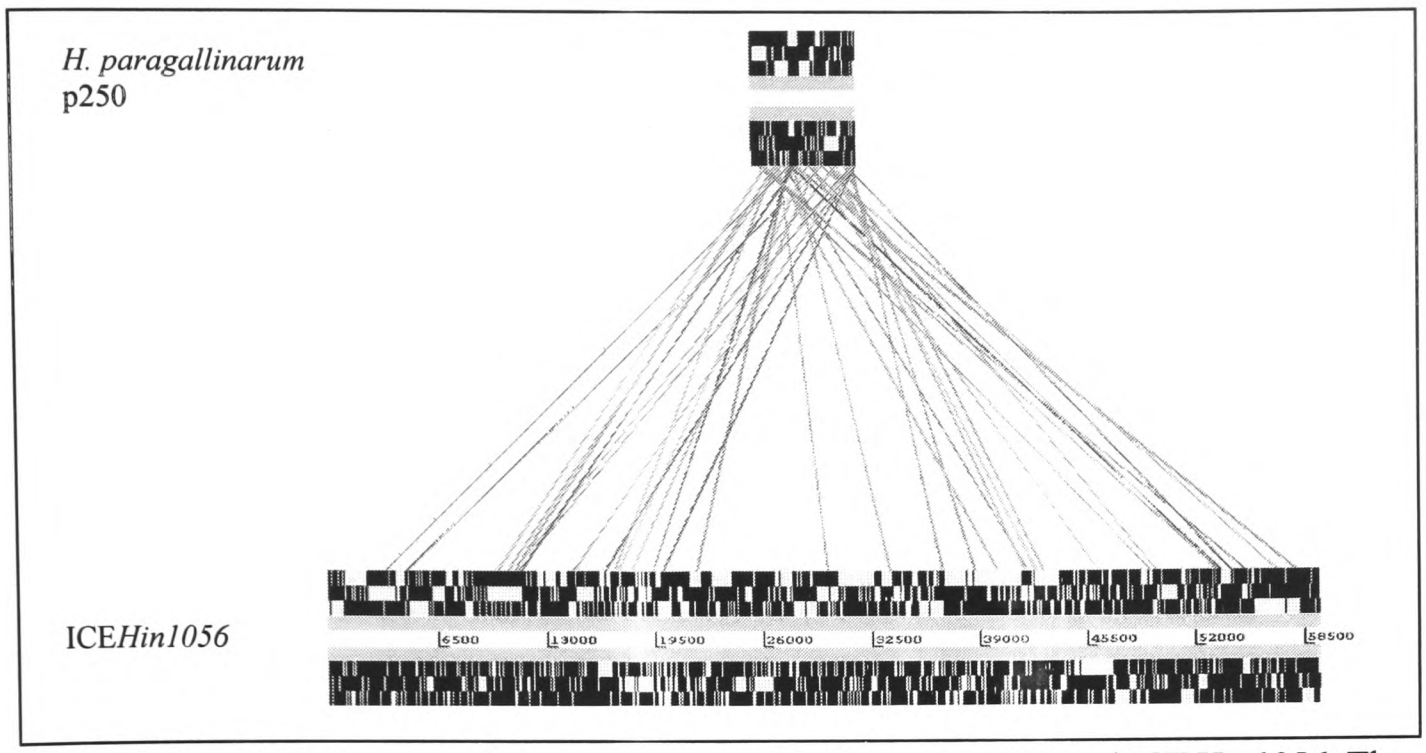


Fig. 6.14. An ACT diagram of sequence comparison between p250 and ICE*Hin1056*. The vertical lines denotes amino acid homology of >25%.

6.5. Discussion

The mobile elements encoding antibiotic resistance ICE*Hin1056*, ICE*Hin299*, ICE*Hpa8F* and ICE*Hin2866* have been shown to possess nucleotide sequences that are highly homologous and are arranged in the same order. The predicted genes with homology to genes with known functions are highly conserved and are organised into putative modules for replication, conjugative transfer and integration.

The phylogenetic relationships of the genes in the replication and conjugation modules of the four elements were essentially congruent as demonstrated by the conserved tree topologies. Some of the ICE*Hin1056* and ICE*Hpa8F* genes have measurably diverged, but it is not possible to determine whether this has been a result of point mutation or recombination. The genes in the conjugation/integration module are divergent and, showed different tree topologies to those in the other modules. This strongly suggests that the genes in this module have undergone recombination and re-assortment. These genes are located between the putative *oriT* and the integration/recombination site, *attP*. The data available in this comparative study suggests

that the conjugation/integration module consists of one of two divergent lineages that have recombined with the rest of the highly conserved ICE core sequences. How this has happened is unclear, but may relate to the ICEs preferred host species. The ICE*Hin1056* variant module is over-represented in *H. influenzae* compared to *H. parainfluenzae* (Dr D. Crook, personal communication).

In all the ICEs, the *bla* gene are carried on Tn*3*. Tn*3* is inserted in the same specific site in all the ICEs, except ICE*Hin1056*. This site specificity for insertion of Tn*3* is not surprising, as transposons have preferred sites of insertion. The event that has led to the truncated Tn*3* in ICE*Hin1056* is unclear. Though, such defective transposons have been frequently observed in other organisms. The highly conserved nucleotide sequence of the Haemophilus derived Tn*3* transposons when compared with those found in *E. coli* is consistent with a recent acquisition by Haemophilus. It also suggests that *E. coli* has not recently donated Tn*3* to oropharyngeal haemophili.

The sequences at the *oriV* region were highly conserved. It is located in the replication module, between *dnaB* and *ssb*. Though this is in a region predicted to contain a gene, there are no clear functional homologues for the predicted gene. Therefore, this may represent an intergenic region without any stop codons. In this AT-rich region, multiple inverted repeats and two regions of putative iterons were observed. This is consistent with the features of an *oriV*, described in various reports (del Solar *et al.*, 1998; Espinosa *et al.*, 2000). Iterons play important role in the binding site of Rep proteins. Surprisingly, no Rep protein and DnaA protein-binding sequence have been detected in this region. It is possible that ICEs of *Haemophilus* spp. does not have DnaA box, and are similar to that reported for *P. aeruginosa* plasmid R18, which carries out replication without a DnaA box (Shah *et al.*, 1995). Despite the lack of a box, the deletion

of *dnaA* in *P. aeruginosa* reduced the plasmid copy number. Therefore, the lack of a DnaA box does not necessarily mean that DnaA has no role in replication.

The replication process in all the elements putatively terminates at a common *ter* site, indicated by the descending G+C% in the GC–skew graph. At this site, the presence of predicted genes with a TMHMM motif, a predicted cleavage site and multiple inverted repeats are features typical of where replication terminates (Lanka and Wilkins, 1995). Presence of a predicted *ter* site on all of the elements suggests that the ICEs carry out theta replication (Bussiere and Bastia, 1999; del Solar *et al.*, 1998; Espinosa *et al.*, 2000).

The sequences of the *oriT* region were conserved in ICE*Hin1056*, ICE*Hin299*, and ICE*Hin2866,* but in ICE*Hpa8F* only sequences around the inverted repeat of the predicted *oriT* were conserved. There can be variation in the sequences of *oriT* between related plasmids from different species and this may explain why ICE*Hpa8F* differs from the other three ICEs all found in *H. influenzae*. Also, it is possible that the minimal *oriT* consists of only the core 8-bp shared by all the inverted repeats. Besides the inverted repeats, the high A+T content of this region, proximity to the predicted *oriT* to *traC*, and *traI* at the extreme end of the putative conjugative module suggest that DNA transfer process begins at this region (Lanka and Wilkins, 1995).

The predictions for the location of the *oriV*, *ter* and *oriT* of the elements are mainly based on the features and motifs provided through bioinformatics, and are not really necessarily reliable predictions. The ultimate test of whether these are the predicted structures will depend on experimental proof.

Other species of Haemophilus, *H. ducreyi* 35000HP and *H. somnus* 129PT have sequences on their genome that share sequences homologous to ICEs, indicating that the strains of *H. ducreyi* and *H. somnus* apparently carry mobile elements that are still

integrated on their chromosome. Further analysis of these elements is described in Chapter 7.

In contrast, the genetic island of *H. influenzae*, HiG1 and plasmids of *H. aegyptius*, *H. ducreyi* and *H. paragallinarum* do not share homology to ICEs. Although HiG1 has inserted into $tRNA^{leu}$, it clearly has no detectable homologies using the TBLASTX algorithm.

CHAPTER SEVEN

ICE*Hin1056* AND ITS RELATIONSHIP WITH OTHER GENOMIC ISLANDS

7.1. Introduction

GIs which are part of the horizontal or flexible gene pool often constitute a large part (over 10%) of bacterial genomes (Hacker *et al.*, 1997; Hacker and Carniel, 2001; Spencer *et al.*, 2003). Genes important in habitat specific adaptation cluster on these island (Dobrindt *et al.*, 2004; Hacker and Carniel, 2001). Well-recognised examples include island with genes involved in pathogenesis (Hacker and Kaper, 2000), avirulence (Alfano *et al.*, 2000), chemical degradation (van der Meer and Sentchilo, 2003), antibiotic resistance (Klockgether *et al.*, 2004) and metabolic functions (Larbig *et al.*, 2002). Hitherto, the evolutionary origin of these islands has not been elucidated (Dobrindt *et al.*, 2002).

7.2. Objectives

The objective of this study was to investigate the relationship between ICE*Hin1056* and other GIs among all bacteria. The presence of a coherent conserved core structure was investigated among homologous GIs.

7.3. Materials and Methods

7.3.1. Identification of GIs

The identification of sequences homologous to ICE*Hin1056* was achieved by searching the NCBI database using the TBLASTX algorithm. The input ICE*Hin1056* sequence consisted of the entire element without the antibiotic-resistance associated (Tn*10* and

Tn*3*) sequences. Sequences producing significant alignments (i.e. E value $<10^{-5}$) were individually interrogated by TBLASTX for contiguous sequences homologous to ICE*Hin1056*. Twenty candidates GIs were identified.

Preliminary analysis of the ORFs of the candidate GIs were compared for amino acid sequence homology using ACT. However, four potential GIs that were found to contain short sequence of shared homology, or were from incompletely sequenced genomes were excluded for further analysis. The remaining 17 GIs (including ICE*Hin1056*) with extensive homology were used for further investigation. A preliminary investigation of phylogenetic relationships between these 17 GIs was undertaken using ClustalX for multiple alignment and tree estimation. The amino acid sequences of two genes present in all 17 GIs were aligned. This yielded four clusters. One well-known representative GI (ICE*Hin1056*, SP1-7, PAP1 and PAGI-3) of each cluster was chosen to determine all genes shared by these GIs and to identify potential core GI genes.

7.3.2. Construction of virtual core GI sequences

Core GI sequences were defined as those present (BLASTP E value $<10^{-5}$) in at least three of the four GIs SPI-7, PAP1, ICE*Hin1056* and PAGI-3. The presence or absence of GI genes was visually scored by pair-wise comparison between each of the GIs using ACT and individually aligning potential homologues using the BLASTP algorithm. Thirty-three core genes fitting the definition were identified (numbered 1 to 33, Fig. 7.1). The sequence of each gene was used to construct four virtual core GIs each representative of ICE*Hin1056*, SPI-7, PAP1, and PAGI-3. Virtual core GIs consisting of either amino acid sequences or nucleotide sequences was constructed for each of the four GIs respectively. For any of the genes missing from a GI, the homologue from PAP1 was used.

7.3.3. *Identification of core genes of other GI candidates*

The concatenated nucleotide sequence of core genes (containing 33 core genes) from each of these four representative GIs were used to re-interrogate the databases using TBLASTX to ensure that no additional GIs were missed during the preliminary screening using ICE*Hin1056* sequences. No other GIs were found.

A virtual coherent GIs were used to compare to each candidate GI using ACT. The presence and absence of core genes was visually scored by pair-wise comparison between the candidate GIs and the virtual coherent element. A coherent evolutionary related GI was defined as possessing ≥25 (~75%) of the 33 core genes. Of the 33 core genes, only 15 genes were shared by all the GIs.

7.3.4. *G+C% content*

The percentages of G+C% content of each GIs and their core sequences was determined through Artemis, and the G+C% content of the host's whole genomes was accessible through GenBank website (http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html)

7.3.5. *Phylogenetic relationship between GIs*

Phylogenetic relationships between the amino acid sequences of each of the 15 core genes common in all the GIs were analysed using ClustalX. The concatenated amino acid sequences of all 15-shared genes in each GI were aligned. The analysis used the following settings: exclude position with gaps, correct for multiple substitutions, phylip format and bootstrap neighbour-joining tree of 100. The output of the analysis was viewed using TreeView (Page, 1996). The congruence of tree topologies was compared to determine whether there was evidence of divergent evolution.

7.4. Results

7.4.1. Identification of related coherent GIs

Twenty potential GIs with homology to the non-resistance associated sequences of ICE*Hin1056* were identified from interrogation of the NCBI database using the TBLASTX algorithm. Four of these with extensive homology (ICE*Hin1056*, SPI-7, PAP1 and PAGI-3) and representing the diversity of these GIs were selected to identify the predicted core element (i.e. genes present in at least three of four) of these GIs. Thirty-three genes fulfilled this definition. Virtual core elements of 33 genes each representing these four GIs were constructed and used in a pair-wise comparison with the 20 candidate and related GIs. Sixteen GIs of the 20 met the definition of a coherent element (24 of the 33 (~75%) core genes being present). The species and strains harbouring the coherent GIs are listed Table 2.4. As the GI in *B. fungorum*, (acc. no. NZ_AAAJ00000000) shares nucleotide sequence identity to the *clc* element of *Pseudomonas* sp. B13 (van der Meer and Sentchilo, 2003), it was not analysed as a separate GI.

The following complete genomes contained contiguous sequences homologous to < 24 of the 33 genes present in a virtual coherent GI: *X. fastidiosa* 9a5c, (acc. no. NC_002488, 11 genes); and *P. syringae* pv. *tomato* DC3000 (acc. no. NC_004578, 17 genes). These, therefore, did not meet the definition of a coherent GI. An incomplete genome sequence of *A. vinelandii* (NZ_AAAU00000000) contained sequences suggesting the presence of a homologous GI, but the unfinished sequence was unsuitable for determining whether it contained a coherent GI. Lastly, a *P. fluorescens* strain currently undergoing whole genome sequencing contained a homologous GI, but was not analysed further due to publication restrictions. The three GIs of *E. coli* strain 536 (PAI I_{536}, PAI II_{536}, and PAI III_{536}; acc. nos.: AJ488511, AJ494981, X16664, respectively),

which were investigated as negative controls, exhibited no homology with a virtual core element when viewed with ACT.

7.4.2. Features of core genes

The elements were all predicted to, or are known to, integrate into tRNA genes as listed in Fig. 7.1. There were five different tRNA genes that were sites of insertion for these GIs, with six of 15 inserting in $tRNA^{gly}$. The core genes of each of the GIs were largely in the same order (i.e. syntheny) (Fig. 7.2). The amino acid sequence homology between GIs at the extremes of homology was as low as 25% to 30% for many of the shared core genes indicating wide evolutionary divergence. Of the 33 core genes, ten had homology to genes with known function. These included four genes, *parA*, *dnaB*, *ssb* and *topB* among genes 1 to 10 (Figs. 7.1 and 7.2). Homologues to these genes are recognised to play a role together in plasmid replication (Espinosa *et al.*, 2000; Gerdes *et al.*, 2000). One of the genes in this region *inrR* has recently been shown to regulate integrase function in the *clc* element and controls excisive and integrative recombination of that element with $tRNA^{gly}$. Genes 11 to 32 are mostly of unknown function, but four exhibit homology to *pilL* (core gene 11), *virB4* (core gene 27), *traD* (core gene 17) and a relaxase or *traI* (core gene 32); homologues of these four are recognized to play a role in conjugative DNA transfer (Zechner *et al.*, 2000).

Integrases were found in all GIs other than that in *R. metallidurans*. As the DNA sequence of this organism was not complete, it is premature to conclude that an integrase associated with this GI is absent. The remaining GIs possess an integrase of either the P4 or XerC/D lineages. SPI-7, the pathogenicity island found in *S. Typhi*, possesses copies of both types of integrase (Fig. 7.2). The XerC/D integrase that was immediately adjacent to the innermost $tRNA^{phe}$ sequence present at the right end of the GI (the putative *attR)* was

selected as the integrase shown in Fig. 7.2, because only this integrase was consistently present in SPI-7 like GIs identified in a range of *Salmonella enterica* serovars Typhi (Pickard *et al.*, 2003).

Fifteen of the core GI genes were conserved across all the GIs (Fig. 7.1). The G+C% content of these were determined and ranged widely from 40.2% for *H. somnus* 2336 to 69.6% for *X. axonopodis*. Compared to the average G+C% of their host's genome, nine GIs had higher percentage content. These consisted of all those found in the *Pasteurellaceae* and *Enterobacteriaceae* (except SPI-7 in *S. Typhi,* which had a lower content than its host genome). Similarly, the GIs found in *Pseudomonas* sp. B13, B. *fungorum,* (genomic G+C% = 61.8%), *R. metallidurans* and *X. axonopodis* had higher G+C% content than their host's genome. However, the GIs found in both *P. fluorescens* and the GIs PAGI-3, PAP1 and pKLC102 found in *P. aeruginosa* had lower G+C% content than their hosts.

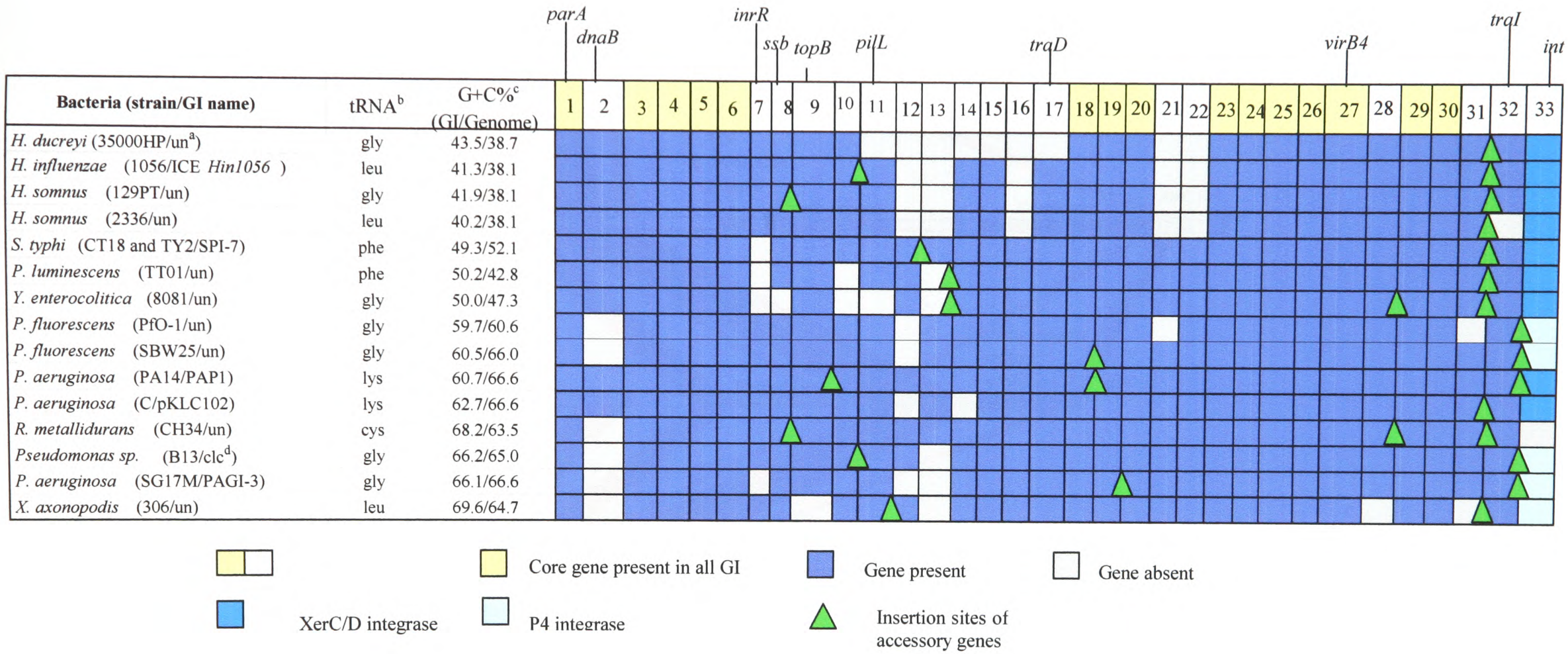


Fig. 7.1. An array of predicted genes present in related coherent GIs found in Proteobacteria. [a]unnamed, [b]tRNA site of GI integration, [c]G+C% content of GI and its host genome, [d]*clc* element is present in the complete sequence of *B. fungorum*.

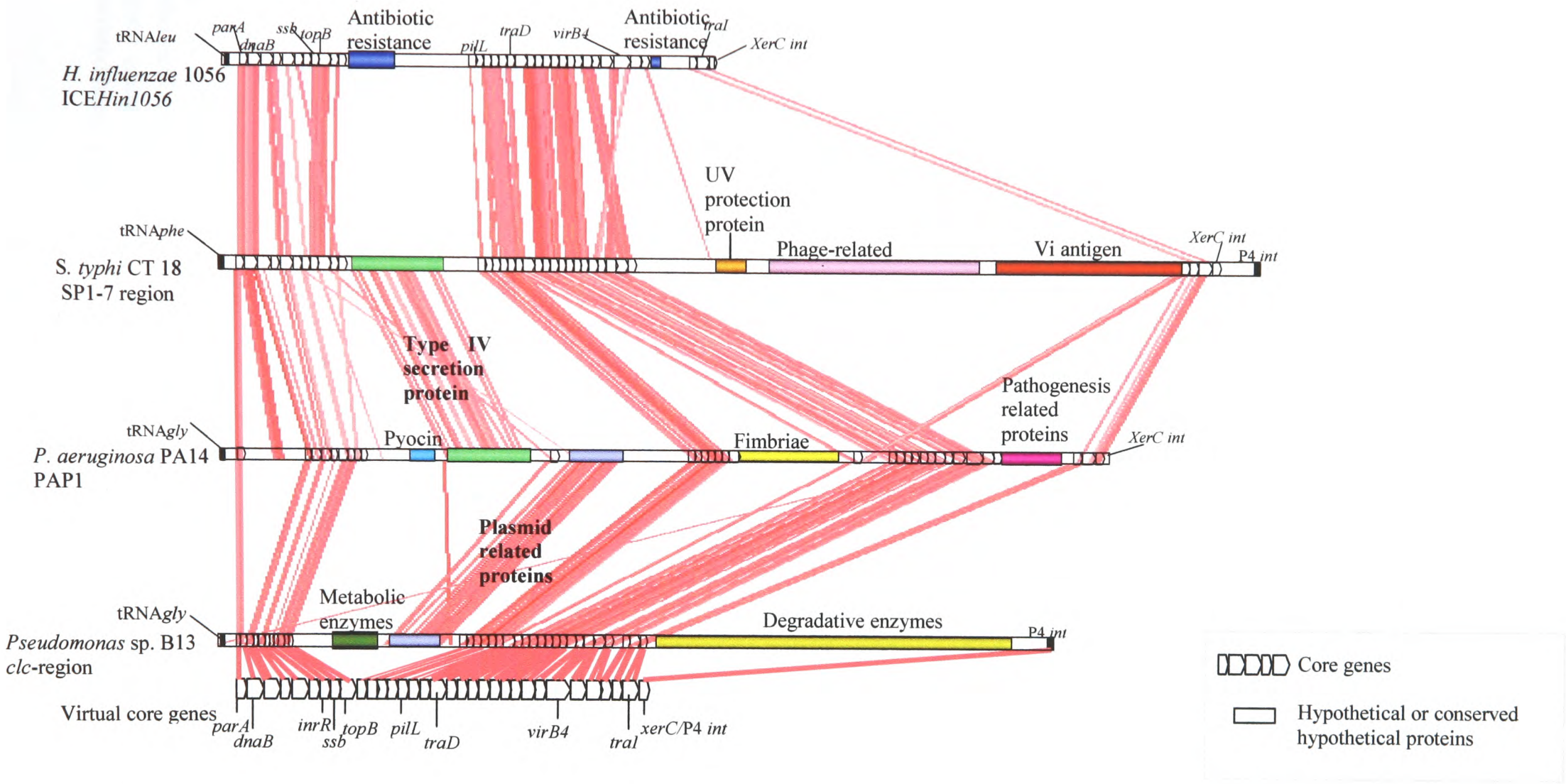


Fig. 7.2. A modified ACT diagram of the homologous regions identified between four GIs ICE*Hin1056*, SP-7, PAP1, *clc* and the core genes of the virtual GI. Homologous sequences (>25% amino acid identity) are indicated by red lines joining regions on the five schematic representations of the GIs.

7.4.3. *Phylogenetic relationship between GIs*

The phylogenetic relationships between each of the 15 genes common to all the GIs exhibited congruent structure with a few minor exceptions. The trees for each of four representative core genes (genes 1, 6, 18, 27) are available in Appendix E. The phylogenetic relationships of the GIs as determined by alignment of the concatenated amino acid sequences of all 15 conserved genes is depicted in Fig. 7.3 and is congruent with that for each of the individual genes.

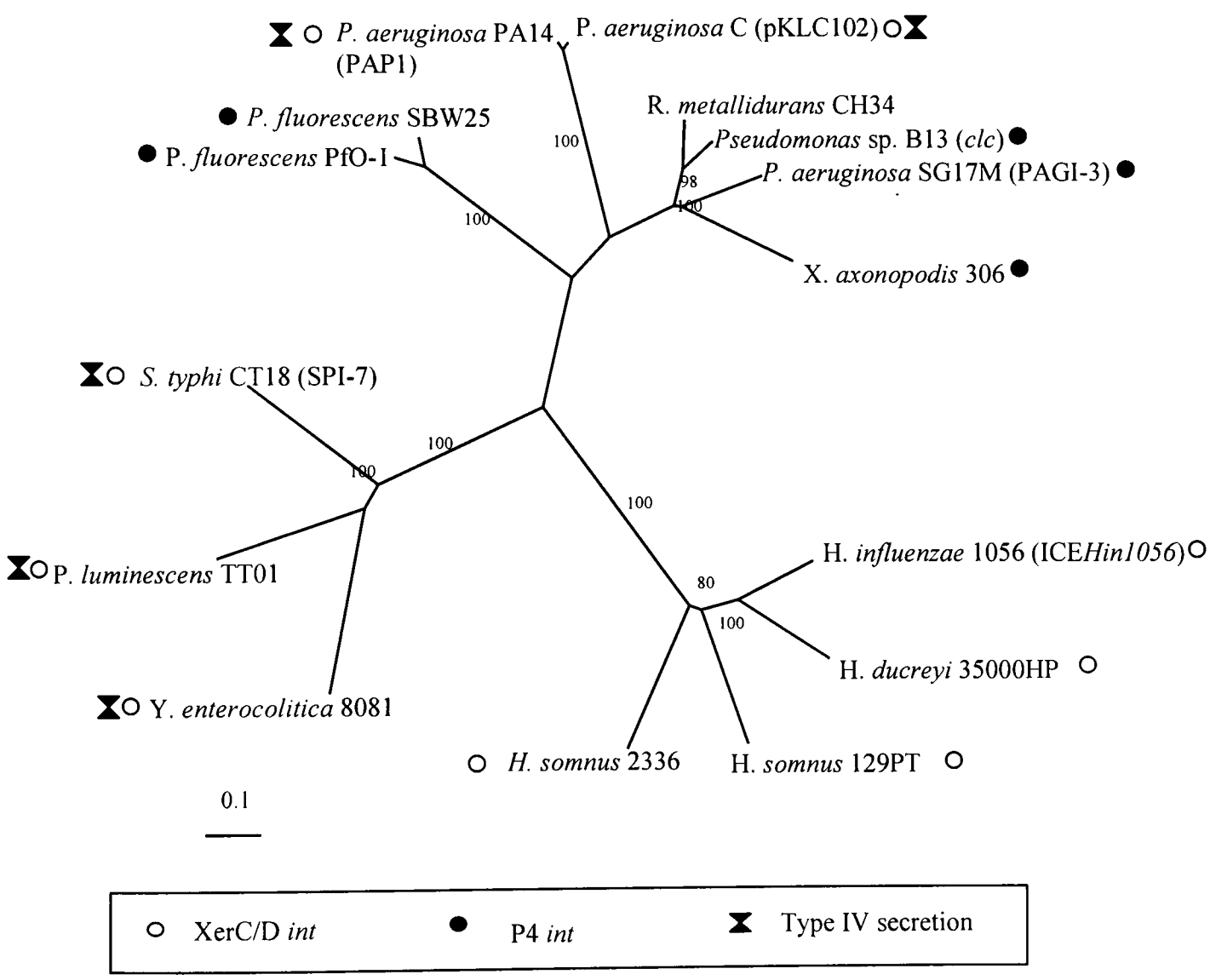


Fig. 7.3. A SplitTree generated by ClustalX analysis of concatenated core genes of 15 GIs. The amino acid sequences of the 15 predicted genes common to all 15 GIs were concatenated and aligned using ClustalX. The alignment of each of the genes alone was congruent to the one illustrated here. The names of the GIs, where available, are given in the parentheses.

7.4.4. *Accessory genes including antibiotic resistance content*

The size of the coherent virtual element was approximately 18 kb, while the whole GIs ranged from 49 kb for *H. ducreyi* to 140 kb for *P. luminescens* (Table 7.1). The variation in the sizes of these related GIs was attributably largely to differences in accessory gene content.

Nestled between conserved core genes of each GI was a wide range of different accessory genes (Figs. 7.1 and 7.2). The sites of insertion were largely clustered in two regions, either between core genes 9 to 14 or between core genes 31 to 33. The major attributes of the accessory genes are listed in Table 7.1 and include: antibiotic, metal and antiseptic resistance, degradation of chemicals, Type IV secretion systems, two component signalling systems, biofilm regulation, Vi antigen capsule synthesis, cytolethal toxin production, anti-restriction systems and a wide range of metabolic functions. Many of the accessory genes do not share homology to genes of known function (Fig. 7.2). The resistance genes found in ICE*Hin1056* were inserted at two different sites. A Tn*10*-like transposon that contains the *tetA* and *cat* genes lies between core genes 10 and 11. Tn*3* that contains the *bla* gene is inserted between core genes 31 and 32 (Figs. 7.1 and 7.2).

Table 7.1. Molecular and phenotypic characteristics of GIs of bacterial pathogens causing diseases in human, animals and plants.

Bacteria (strain/GI name)	Size (bp)	Accession No.	Major attributes of accessory genes
H. ducreyi (35000HP/un[a])	49,075	NC_002940	Type I antirestriction, cytolethal distending toxin
H. influenzae (1056/ICE*Hin1056*)	59,393	AJ627386	Antibiotic resistance
H. somnus (129PT/un)	56,442	NZ_AAB002000009	Type I antirestriction, virulence protein
H. somnus (2336/un)	64,831	NZ_AACJ00000000	Antibiotic, metal and antiseptic resistance
Y. enterocolitica (8081/un)	65,285	NC_003222	Type IV secretion, arsenic resistance, hemagglutinin-related protein
S. typhi (CT18 and TY2/SPI-7)	134,000	NC_003198	Type IV secretion, uv protection, Vi antigen
P. luminescens (TT01/un)	140,163	NC_005126	Metabolic enzymes, Type IV secretion
P. fluorescens (PfO-1/un)	70,421	NZ_AAAT00000000	Biofilm regulators, fimbrial proteins, metabolic enzymes
P. fluorescens (SBW25/un)	101,233	NC_004507	Metabolic enzymes
P. aeruginosa (PA14/PAP1)	106,518	AY273869	Pyocin, type IV secretion, fimbrial protein, two component signal system
P. aeruginosa (C/pKLC102)	103,532	AY257538	Type IV secretion, virulence protein
R. metallidurans (CH34/un)	>76,893	NZ_AAAI00000000	Degradative enzymes, metabolic enzymes, pathogenesis-related protein
Pseudomonas sp. (B13/c/c[b])	105,027	AJ617740	Metabolic enzymes, degradative enzymes
P. aeruginosa (SG17M/PAGI-3)	114,632	AF440524	Plasmid transfer proteins, type I restriction-modification, metabolic enzymes
X. axonopodis (306/un)	86,019	NC_003919	Phage-related proteins, efflux pump, hemolysin

[a]un = unnamed, [b]*clc* element is present in the complete sequence of *B. fungorum*.

7.4.5. Hot spots

Some 'hot spot' areas were apparent where insertion of accessory genes was mainly concentrated (Fig. 7.1). Most of the insertion of accessory genes occurred in two regions: between the *ssb* and *pil*L genes (core genes 9 and 14); and the region close to the *int* gene (core genes 31 and 33). It is noted that all of the islands have insertions in the region close to the *int* gene. The reason for their preference is unknown.

The regions where no or few insertions occurred are apparently made up of clusters of genes that are cohesive. It is possible that the functions of these clusters of genes would be interrupted by insertion of accessory genes. The region to the left of the first 'hot spot' contains gene cluster putatively involved in replication. It has been previously shown in enterobacteria that the order of genes close to replication origins is more conserved (Sharp *et al.*, 1989). The following region between core genes 14 to 31 contains a gene cluster likely encoding DNA transfer. Insertion of accessory genes in this region may disrupt conjugative transfer.

7.5. Discussion

7.5.1. A coherent core GI

A family of related and coherent GIs with a common evolutionary origin among diverse Proteobacterial hosts has been recognized for the first time. These GIs have diverged widely as demonstrated by as much as 30% divergence (40% to 70%) in the G+C% content between GIs and amino acid homologies as low as 25% to 30% between homologous GI genes. These GIs have acquired diverse accessory genes with habitat specific adaptive functions typical of GIs or PAIs (Dobrindt *et al.*, 2004; Hacker and Carniel, 2001). Other features typical of Proteobacterial GIs are insertion into tRNA genes and a G+C% content different to their host's genome (Dobrindt *et al.*, 2004).

It has been hypothesized that GIs are mosaics of independently evolving genes or clusters of genes (Burrus *et al.*, 2002a; Toussaint and Merlin, 2002). This is referred to as modular evolution. In the reports by Burrus *et al.*, it was demonstrated that heterogeneous modules were assorted between different GIs (Burrus *et al.*, 2002a, b). This lack of coherent structure is probable following repeated recombination arising from frequent horizontal spread of GIs. This would be expected to erase coherent structures between GIs over time. The finding reported here, at least, identifies one family of related GIs that challenges this hypothesis. A coherent element with a common evolutionary origin has been found. The cut-off value of 75% (core GI sequences in at least three of the four GIs) may have limited the identification of only one family of GI. Perhaps, by reducing the cut-off value, more than one family of GI could be identified, is yet to be determined.

The accessory genes of these related GIs are inserted in a manner apparently consistent with modular evolution. However, variable modules consisting of accessory genes nestle in a conserved module that is a coherent core GI. This forms the apparent mosaic structure of this family of GIs. The presence of a coherent core GI suggests a fitness property of the conserved core genes acting collectively that has ensured their survival as a coherent whole. It is unclear what this may be; yet it is reasonable to surmise that it is the ability of these GIs to transmit and propagate better as a coherent whole. The role of accessory genes in the survival of this coherent element appears to be interwoven with allowing organism specific adaptation to habitats. This will necessarily involve complex interactions between the core GI, accessory genes and the survival of the host bacterium in new hostile environments.

Bio-informatics analysis of the core genes and review of the functional properties of some of the well-characterised examples of this family of GI provides insight on the possible function of the core GI genes. An integrase is present in all but one (*R.*

metallidurans) of these GIs. These integrases are likely to encode integration with tRNA genes and all belong to the family of tyrosine integrases. The only GI where this has been conclusively shown is the *clc* element found in *Pseudomonas* sp. B13, which contains a homologue of the P4 integrase (Ravatn *et al.*, 1998). A curious aspect of the tyrosine integrases found in these GIs is that they belong to two different lineages, a P4 lineage or a XerC/D lineage. Both types of integrase are known to encode recombination with tRNA genes in other mobile elements (Mantri and Williams, 2004). How members of these GIs have acquired either of these two divergent lineages of integrases is intriguing and the available data provides no obvious explanation. The observation that SPI-7 found in *S.* Typhi contains both integrase lineages, while SPI-7 like GIs found in other *S. enterica* serovars do not (Pickard *et al.*, 2003) may suggest that there are ready opportunities for exchanging integrase lineages.

How GIs transfer horizontally has not been clearly explained (Dobrindt *et al.*, 2004). Though, it has been recognised that the *clc* element is a GI and is capable of conjugative transfer (van der Meer and Sentchilo, 2003). The observation made here that another conjugative element, ICE*Hin1056*, found in *H. influenzae*, is also a related GI strengthens this observation. Both these GIs transfer by conjugation at frequencies of 10^{-6} to 10^{-7} (transconjugants/donors) (Dimopoulou *et al.*, 1992; Ravatn *et al.*, 1998). The genes encoding this property have not been identified in either GI. However, the genes encoding this function are likely to be present among the core genes as none of the accessory genes present in either GI is shared. Also, none of their accessory genes is known to encode conjugative functions. Furthermore, four of the core GI genes are homologous to *pilL, traD, virB4* and *traI* and are interspersed along a contiguous 22-gene segment of the GI. These genes are known to be involved in DNA conjugative transfer systems (Zechner *et al.*, 2000). Therefore, it can be inferred that core genes are likely to

encode transfer functions that would explain how this family of GI transfers between hosts. Before this can be firmly concluded, however, this needs to be formally investigated and demonstrated experimentally.

Autonomous replication would be an unexpected property of GIs (Dobrindt *et al.*, 2004). However, four of the first 10 genes are homologues of *parA*, *dnaB*, *ssb* and *topB*; genes known to be associated with plasmid replication (Espinosa *et al.*, 2000; Gerdes *et al.*, 2000). Furthermore, both ICE*Hin1056* found in *H. influenzae* and pKLC102 found in *P. aeruginosa* C appear to exist as extra-chromosomal closed circular plasmid under some conditions (Dimopoulou *et al.*, 1997, Dimopoulou *et al.*, 1992, Klockgether *et al.*, 2004; Leaves *et al.*, 2000). ICE*Hin1056* or other related Haemophilus conjugative elements are found largely in plasmid form in transconjugants immediately following conjugation (Dimopoulou *et al.*, 1992). This is supported by two observations: first, ready isolation of plasmid from transconjugants and not from parent donor clinical isolates (Dimopoulou *et al.*, 1992). Second, the appearance of Southern blot hybridisation patterns consistent with a closed circular form in transconjugants and a pattern indicating chromosomal integration in parent donor strains (Dimopoulou *et al.*, 1992; Dimopoulou *et al.*, 1997; Leaves *et al.*, 2000). The pKLC102 has not been formally shown to replicate autonomously; however, Klockgether *et al.* reported the presence of an extra-chromosomal form and sequences close to the *ssb* gene with features typical of an *oriV* (Klockgether *et al.*, 2004). These data suggest that some, if not all, these GIs may be capable of autonomous replication under some conditions, but this needs to be formally determined experimentally before this can be concluded. Were this to be found, it would be a new, and hitherto unrecognised, dimension to GIs (Dobrindt *et al.*, 2004; Hacker and Carniel, 2001) or integrating and conjugating elements reviewed by Burrus *et al.* (Burrus *et al.*, 2002a). Furthermore, an element that was capable of integrative and excisive

recombination, conjugative transfer and autonomous replication would require finely coordinated regulation of these functions to ensure their timing was appropriate to the relevant state of the GI.

The origin and host range of this family of GI is not apparent from the data available in this investigation. They were only present in organisms belonging to *Proteobacteria*, but the investigation is restricted by the limitations of bio-informatics dependent on algorithms relatively poor at detecting weak homologies across contiguous multi-gene structures necessary for recognizing these related GIs and also the relatively small number of whole genome sequences available for analysis. If this family of GIs is truly restricted to *Proteobacteria*, some insight into the origin of this GI family may come from simple consideration of its average G+C% content. The wide range in the G+C% content from 40% to 70% for the shared core GI genes is not substantially different to the range found for ß- and γ-*Proteobacteria* (Krieg and Holt, 1984). This could be interpreted to suggest a longstanding relationship with this Phylum. This interpretation is based on the proposition, first, that GIs are constrained to this Phylum and, second, that horizontally acquired DNA with a different G+C% equilibrates over time to that of its host's genomic G+C% content through a process referred to as amelioration (Lawrence and Ochman, 1998). It follows that for this family of GIs to have evolved such divergent G+C% contents; it would have necessitated substantial time in hosts with similarly divergent G+C% contents. It would also require limited horizontal transfer and recombination between divergent members of this family of GI, as this would tend to homogenise G+C% contents.

The presence of two GIs with G+C% of 68.2% and 69.6% that is markedly higher than their respective host's genomic G+C%, namely, *R. metallidurans* (63.5%) and *X. axonopodis* (64.7%), is curious. In the case of *X. axonopodis,* this is remarkable as there

are only three regions of the genome with such a highly divergent G+C% content. In addition, the most extensive region is accounted for by the core genes of this GI (28 genes), the others being one or three genes in extent (data are accessible through TIGR http://www.tigr.org/tigr-scripts/CMR2/GCDisplay.spl?asmbl_id=226). One interpretation is that these two high G+C% content GIs have been acquired from hosts with even higher G+C% content than them. This would suggest an origin among organisms with >70% G+C% average content for their genomes. One possible host and organism common to *Proteobacteria* is the environmental organism, *Derxia* spp. a β-*Proteobacteria*, with a genomic G+C% greater than 70% (Becking, 1984). Few, if any, other known *Proteobacteria* have such a high G+C% content (Krieg and Holt, 1984). An alternative and more speculative possibility is that these high G+C% GIs may originate in high G+C% content Gram-positive bacteria such as *Streptomyces* spp. These Gram-positive organisms have a G+C% content as high as 73% (Bentley *et al.*, 2002; Omura *et al.*, 2001). This would then suggest a wider host range than just the *Proteobacteria*. Conclusive evidence of this may become apparent from the growing number of whole genomic sequences available for analysis and improved algorithms for interrogating these databases for large contiguous and weakly homologous sequences. The origin, more precise age and host range of this family of related GIs remains speculative, but future comparative investigations of codon usage and the G+C nucleotides at the third codon position may go a long way to answering these questions.

7.5.2. Antibiotic resistance accessory genes and this family of GIs.

The observation of an evolutionarily related coherent GI provides new insight into how the recent emergence and spread of β-lactamase positive Ap^r and/or tetracycline and chloramphenicol antibiotic resistance in *H. influenzae* has arisen. The phylogenetic

relationship of ICE*Hin1056* to other distantly related GIs indicates that transferable resistance in *H. influenzae* has deep evolutionary origins. The resistance genes, conveyed by transposons (e.g. Tn*10* or Tn*3*) (Elwell *et al.*, 1977, Kaulfers *et al.*, 1978), form clusters of accessory genes on the core element that have apparently evolved stable relationships. No other accessory genes with different properties are apparent from analysis of the whole sequence. The emergence of this resistance element in pathogenic *H. influenzae* only became readily detectable in the early 1970s (Brunton *et al.*, 1986; Elwell *et al.*, 1975; Powell, 1988; Syriopoulou *et al.*, 1976; Syriopoulou *et al.*, 1978). Curiously, resistance then rapidly emerged over the next few years worldwide among pathogenic *H. influenzae* and rose to high prevalence (20% to 30%) in many countries (Powell, 1988; Sentchilo *et al.*, 2003). Evidence indicates that this resistance is accounted for by the appearance of ICEs (GIs) highly related to ICE*Hin1056* (Dimopoulou *et al.*, 1992; Dimopoulou *et al.*, 1997; Dimopoulou *et al.*, 2002; Leaves *et al.*, 2000; Powell, 1988; Powell and Livermore, 1988; Schito *et al.*, 2000). The epidemic expansion of this family of ICEs among pathogenic *H. influenzae* and high prevalence among other commensal haemophili (Leaves *et al.*, 2000; Scheifele and Fussell, 1981b; Scheifele *et al.*, 1982) suggest that this GI is well adapted to these bacterial hosts and offers them a survival advantage under antibiotic exposure. An intriguing question is whether the acquisition of resistance transposons and their intimate and apparent stable relationship with these GIs in haemophili has recently evolved or has deeper origins preceding the antibiotic era. The latter supposition would support the notion that antibiotic resistance is an ancient adaptive response that had pre-evolved in bacteria. Modern day intensive use of antibiotics only serves to change the distribution or organization of these pre-evolved structures.

The association of resistance genes to this family of GIs is not unique to *H. influenzae*. GIs found in both *H. somnus* and *P. aeruginosa* contain accessory genes with homology to resistance genes. In the case of *H. somnus* 2336, homologues of *tetA* and *romA* (multi-drug resistance gene) are present and the strain is tetracycline resistant (minimum inhibitory concentration to tetracycline is 8 μg/ml, T. Inzana, personal communication). In *P. aeruginosa* strain C, the GI, pKLC102, contains *aadB* that encodes tobramycin and gentamicin resistance (Klockgether *et al.*, 2004). This gene is found in a large complex class I transposon named TNCP23 inserted into pKLC102 (Klockgether *et al.*, 2004). As this strain was isolated from a case with cystic fibrosis who regularly received treatments of tobramycin, it was surmised that this variant had been selected by the tobramycin exposures (Klockgether *et al.*, 2004). The relationship between this family of GIs and resistance is unpredictable. For example, the related GI, SPI-7, found in the antibiotic multi-resistant (including tetracycline resistant) strain *S. Typhi* CT 18 does not contain resistance genes, even though tetracycline resistance in this strain is encoded by a Tn*10*-like structure (Parkhill *et al.*, 2001). This Tn*10*-like structure along with other antibiotic resistance genes is found on the large unrelated conjugative plasmid, pHCM1 (Wain *et al.*, 2003). Therefore, this family of related GIs does not simply and predictably acquire resistance genes even when transposons containing resistance genes are present in trans.

7.5.3. Major properties of other accessory genes

The accessory genes conveyed by this family of GI have a wide range of attributes other than antibiotic resistance that contribute to the habitat specific survival of their host organisms. Another and most common associated attribute is Type IV secretion found in five of the GIs, two of which, PAP1 (He *et al.*, 2004) and SPI-7 (Pickard *et al.*, 2003), are

well described pathogenicity islands. The Type IV secretion system in SPI-7 is involved in the adhesion of *S. Typhi* to macrophages and is important to *S. Typhi's* intracellular life cycle and in disease causation (Morris *et al.*, 2003; Tsui *et al.*, 2003; Zhang *et al.*, 2000). The role of genes encoding the Type IV secretion system in PAP1 is less clear. These genes were not reported as contributing to the virulence of *P. aeruginosa* PA14 (He *et al.*, 2004). The Type IV secretion system present in the GI found in *Y. enterocolitica* is probably important in the organism's survival in animal hosts. This can be indirectly inferred from investigations into *Yersinia pseudotuberculosis* (Collyn *et al.*, 2002). The *Y. enterocolitica* Type IV secretion system is highly homologous to the one investigated in *Y. pseudotuberculosis* (comparison performed using ACT, data not shown). The Type IV secretion system found in *Y. pseudotuberculosis* was shown to be crucial in mouse invasion models (Collyn *et al.*, 2002). Furthermore, the distribution of the Type IV secretion system among isolates of *Y. pseudotuberculosis* was restricted to pathogenic strains and its variable presence was regarded as consistent with it being on a mobile element (Collyn *et al.*, 2002), perhaps another of the GI family described here. The data available is incomplete, but it is consistent with the Type IV secretion systems present in this family of GI being involved in adherence of the host organisms to specific targets in their habitat.

Other genes important in virulence are present in these GIs. The Vi antigen encoded by SPI-7 is a well-known capsular antigen characteristic of *S. Typhi* that contributes to the pathogenesis of typhoid fever and is a valuable vaccine antigen (Pickard *et al.*, 2003). PAP1 contains a number of virulence associated accessory genes including two-component regulators that may be involved in regulating pathogenicity activities (He *et al.*, 2004). The island found in *H. ducreyi* contains the cytolethal distending toxin genes. Their role in pathogenesis have been studied and despite *in-vitro*

evidence of causing cell damage and death, these genes do not appear to be essential for causing disease in an animal model or in a human challenge study (Gelfanova *et al.*, 1999; Stevens *et al.*, 1999; Young *et al.*, 2001). Nevertheless it remains an intriguing gene cluster likely involved in bacterial human host interactions.

Accessory genes important in the survival of organisms in environmental habitats include a range of genes encoding metabolic functions. These are found in species that live in potentially nutrient deficient environments such as water or soil and include *P. aeruginosa* SG17M, *Pseudomonas* sp. B13, *B. fungorum* and *R. metallidurans*. Another attribute of accessory genes of some of these environmental species are highly processive degradative enzymes. In particular, the *clc* element found in *Pseudomonas* sp. B13 and *B. fungorum* encodes a number of genes that inactivate toxic chemicals such as chlorocatechols (van der Meer and Sentchilo, 2003). Organisms possessing these enzymes and related DNA elements have become widely distributed, presumably through selection, in environments heavily contaminated with toxic waste (Chatterjee *et al.*, 1981; Chatterjee and Chakrabarty, 1983; van der Meer and Sentchilo, 2003).

7.6. Conclusions

There is a very large and diverse population of GIs present in bacteria (Dobrindt *et al.*, 2004) most of which do not share detectable homology with the family of related GIs described here. However, the methods used here to identify and characterise one family of GI may be used in principle to identify other families of coherent GI among these many apparently unrelated GIs. It seems possible that there will be both examples of mosaics without evidence of a coherent core GI structure and other GIs which have coherent core genes and a common evolutionary origin. This will be increasingly possible to investigate with the improving bio-informatics techniques and the rapidly increasing

sequence data available from complete bacterial genomes. Such a systematic classification of GIs is likely, in turn, to necessitate a commensurate systematic naming scheme.

CHAPTER EIGHT

CONCLUDING REMARKS AND FUTURE RESEARCH

The main objective of this work was to investigate the origin of antibiotic resistance in *H. influenzae*. The principal and unexpected finding was that the ICEs that transfer antibiotic resistance in *H. influenzae* have a common origin with a diverse family of coherent GIs found in β- and γ-*Proteobacteria*. Analysis of the core genes shared between these related GIs has given new insight into how they may conjugate and, thereby, explain how some GIs horizontally transfer between their bacterial hosts. A surprising finding was that these GIs may be capable of autonomous replication given the presence of genes associated with replication and the prediction that an *oriV* is present in some of them.

Future investigation of this family of GIs will focus on experimentally identifying genes encoding conjugation, replication and integrative and excisive recombination. Identifying and understanding the regulatory genes involved in controlling these properties will be particularly interesting and may reveal novel mechanisms of gene control. These studies will, therefore, contribute to our understanding of how GIs propagate and survive in bacterial populations. An intriguing possibility is that these GIs are capable of mobilising whole chromosomes in an Hfr type manner. If this is the case they may contribute to the diversification of bacteria and this may explain how large segments of DNA (e.g. capsular genes) are acquired and exchanged by bacteria in nature.

Another direction of future research will be to determine the extent to which GIs in general are classifiable using the methods described in this thesis. In order to systematically study the large number of GIs present in the many whole genomes now available for analysis, new improved search algorithms will be needed for identifying contiguous sequences with weak homologies. These investigations may not only produce

a systematic classification of GIs, but may give new insight into the host range of the family of related GIs described in this thesis.

The study of the antibiotic resistance ICEs in oropharyngeal haemophili revealed that these elements are highly related. Curiously, the sequences adjacent to and including the integrase gene derived from either of two divergent origins. The explanation for how these divergent regions evolved and how they have recombined with highly conserved modules (replicative and conjugative modules) is unclear. Also, whether this will inform on the recent emergence of antibiotic resistance in *H. influenzae* is unclear. However, the sequence variant that associates with ICE*Hin1056* and ICE*Hin2866* is over represented in *H. influenzae* compared to *H. parainfluenzae* (Dr. Crook, personal communication). One hypothesis is that this variant sequence is better adapted to *H. influenzae* and may have been recently acquired and in part accounts for the successful emergence of transferable resistance in pathogenic *H. influenzae*. In addition, the near identity in sequence found among the Tn*3* sequences from epidemiologically unrelated isolates is consistent with this transposon being relatively recently acquired by this family of ICEs.

By investigating the population structure of these oropharyngeal ICEs, a better understanding of how they have evolved and spread through the population of haemophili will be possible. Methods have been developed in the laboratory to sequence multiple genes conserved in these ICEs and their bacterial hosts. Large collections of isolates from different parts of the world and from people on antibiotics have been assembled. These can be used to investigate whether there has been a recent clonal expansion of a few ICE variants in the population of pathogenic *H. influenzae*. It will also be possible to determine if there is geographical structuring of ICEs, which would be consistent with local evolution rather than rapid, worldwide spread of these ICEs. This experimental approach will allow predictions about the routes of transmission and spread of these ICEs

to be made. This will give insights into the evolution of ICEs in haemophili and how they have acquired resistance transposons and spread to become so prevalent worldwide.

REFERENCES

Alfano, J.R., Charkowski, A.O., Deng, W.L., Badel, J.L., Petnicki-Ocwieja, T., van Dijk, K., and Collmer, A. (2000) The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc Natl Acad Sci U S A* 97: 4856-4861.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.

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: 3389-3402.

Anderson, P., Smith, D.H., Ingram, D.L., Wilkins, J., Wehre, P.F., and Howie, V.M. (1977) Antibody to polyribophosphate of *Haemophilus influenzae* type b in infants and children: effect of immunization with polyribophosphate. *Journal of Infectious Diseases* 136: 57-62.

Aparicio, P., Roman, F., and Campos, J. (1996) [Epidemiological characterization of *Haemophilus influenzae* using molecular markers]. *Enferm Infecc Microbiol Clin* 14: 227-232.

Arnold, L., A, A.P., and Jackson, R. (2003) Pathogenicity and other genomic islands in plant pathogenic bacteria. *Molecular Plant Pathology* 4: 407-420.

Astumian, J.H., Waldman, A.S., and Scocca, J.J. (1989) Site-specific recombination between cloned *attP* and *attB* sites from the *Haemophilus influenzae* bacteriophage HP1 propagated in recombination-deficient *Escherichia coli*. *J Bacteriol* 171: 1747-1750.

Austin, S.J. (1988) Plasmid partition. *Plasmid* 20: 1-9.

Autret, S., Nair, R., and Errington, J. (2001) Genetic analysis of the chromosome segregation protein Spo0J of *Bacillus subtilis*: evidence for separate domains involved in DNA binding and interactions with Soj protein. *Mol Microbiol* 41: 743-755.

Azad, A.K., Coote, J.G., and Parton, R. (1992) Distinct plasmid profiles of *Pasteurella haemolytica* serotypes and the characterization and amplification in *Escherichia coli* of ampicillin-resistance plasmids encoding ROB-1 beta-lactamase. *J Gen Microbiol* 138 (Pt 6): 1185-1196.

Baquero, F. (1996) Trends in antibiotic resistance of respiratory pathogens: an analysis and commentary on a collaborative surveillance study. *J Antimicrob Chemother* 38 Suppl A: 117-132.

Barbour, M.L., Crook, D.W., and Mayon-White, R.T. (1993) An improved antiserum agar method for detecting carriage of *Haemophilus influenzae* type b. *Eur J Clin Microbiol Infect Dis* 12: 215-217.

Barrett, F.F., Taber, L.H., Morris, C.R., Stephenson, W.B., Clark, D.J., and Yow, M.D. (1972) A 12-year review of the antibiotic management of *Haemophilus influenzae* meningitis. Comparison of ampicillin and conventional therapy including chloramphenicol. *J Pediatr* 81: 370-377.

Barrow, G.I., and Feltham, R.K.A. (1993) Characters of Gram-negative bacteria. In *Cowan and Steel's manual for the identification of medical bacteria*. Barrow, G.I. and Feltham, R.K.A. (eds). Cambridge: Cambridge University Press, pp. 150-156.

Barry, A.L. (1986) Procedures and theoretical considerations for testing antimicrobial agents in agar media. In *Antibiotics in laboratory medicine*. Lorian, V. (ed). New York: Williams and Wilkins, pp. 1-14.

Bartosik, D., Wlodarczyk, M., and Thomas, C.M. (1997) Complete nucleotide sequence

of the replicator region of *Paracoccus (Thiobacillus) versutus* pTAV1 plasmid and its correlation to several plasmids of *Agrobacterium* and *Rhizobium* species. *Plasmid* 38: 53-59.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2002) Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J Bacteriol* 184: 4259-4269.

Becking, J.H. (1984) Gram-Negative aerobic rods and cocci. In *Bergey's Manual of Systematic Bacteriology*. Vol. 1. Baltimore, USA: Williams & Wilkins, pp. 140-402.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573-580.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417: 141-147.

Berry, C., O'Neil, S., Ben-Dov, E., Jones, A.F., Murphy, L., Quail, M.A., Holden, M.T., Harris, D., Zaritsky, A., and Parkhill, J. (2002) Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. israelensis. *Appl Environ Microbiol* 68: 5082-5095.

Betz, A.L., and Goldstein, G.W. (1986) Specialized properties and solute transport in brain capillaries. *Annu Rev Physiol* 48: 241-250.

Bonheyo, G., Graham, D., Shoemaker, N.B., and Salyers, A.A. (2001) Transfer region of a bacteriodes conjugative transposon, CTnDOT. *Plasmid* 35: 41-51.

Booy, R., Hodgson, S.A., Slack, M.P., Anderson, E.C., Mayon-White, R.T., and Moxon, E.R. (1993) Invasive *Haemophilus influenzae* type b disease in the Oxford region (1985-91). *Arch Dis Child* 69: 225-228.

Boughton, W.H. (1982) Rapid detection in spinal fluid of beta-lactamase produced by ampicillin-resistant *Haemophilus influenzae*. *J Clin Microbiol* 15: 1167-1168.

Bradley, D.E. (1980) Morphological and serological relationships of conjugative pili. *Plasmid* 4: 155-169.

Bradley, D.E., Taylor, D.E., and Cohen, D.R. (1980) Specification of surface mating systems among conjugative drug resistance plasmids in *Escherichia coli* K-12. *J Bacteriol* 143: 1466-1460.

Bradley, D.E. (1983) Specification of the conjugative pili and surface mating systems of Pseudomonas plasmids. *J Gen Microbiol* 129 (Pt 8): 2545-2556.

Brightman, C.A., Crook, D.W., Kraak, W.A., Dimopoulou, I.D., Anderson, E.C., Nichols, W.W., and Slack, M.P. (1990) Family outbreak of chloramphenicol-ampicillin resistant *Haemophilus influenzae* type b disease. *Lancet* 335: 351-352.

Bruce, K.D., and Jordens, J.Z. (1991) Characterization of noncapsulate *Haemophilus influenzae* by whole-cell polypeptide profiles, restriction endonuclease analysis, and rRNA gene restriction patterns. *J Clin Microbiol* 29: 291-296.

Brunton, J., Clare, D., and Meier, M.A. (1986) Molecular epidemiology of antibiotic resistance plasmids of *Haemophilus* species and *Neisseria gonorrheae*. *Reviews of Infectious Diseases* 8: 713-724.

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002a) Conjugative transposons: the tip of the iceberg. *Mol Microbiol* 46: 601-610.

Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002b) The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* 48: 77-97.

Bush, K. (1989) Classification of beta-lactamases: groups 1, 2a, 2b, and 2b'. *Antimicrob Agents Chemother* 33: 264-270.

Bush, K., and Jacoby, G. (1997) Nomenclature of TEM beta-lactamases. *J Antimicrob Chemother* 39: 1-3.

Bussiere, D.E., and Bastia, D. (1999) Termination of DNA replication of bacterial and plasmid chromosomes. *Mol Microbiol* 31: 1611-1618.

Campos, J., Chanyangam, M., deGroot, R., Smith, A.L., Tenover, F.C., and Reig, R. (1989) Genetic relatedness of antibiotic resistance determinants in multiply resistant *Hemophilus influenzae*. *J Infect Dis* 160: 810-817.

Catlin, B.W. (1975) Iodometric detection of *Haemophilus influenzae* beta-lactamase: rapid presumptive test for ampicillin resistance. *Antimicrob Agents Chemother* 7: 265-270.

Chalmers, R., Sewitz, S., Lipkow, K., and Crellin, P. (2000) Complete nucleotide sequence of Tn10. *J Bacteriol* 182: 2970-2972.

Chandler, M., and Mahillon, J. (2002) Insertion sequences revisited. In *Mobile DNA II*. Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds). Washington, D.C.: ASM Press, pp. 305-366.

Chang, C.C., Gilsdorf, J.R., DiRita, V.J., and Marrs, C.F. (2000) Identification and genetic characterization of *Haemophilus influenzae* genetic island 1. *Infect Immun* 68: 2630-2637.

Chanyangam, M., Smith, A.L., Moseley, S.L., Kuehn, M., and Jenny, P. (1991) Contribution of a 28-kilodalton membrane protein to the virulence of *Haemophilus influenzae*. *Infect Immun* 59: 600-608.

Chatterjee, D.K., Kellogg, S.T., Hamada, S., and Chakrabarty, A.M. (1981) Plasmid specifying total degradation of 3-chlorobenzoate by a modified ortho pathway. *J Bacteriol* 146: 639-646.

Chatterjee, D.K., and Chakrabarty, A.M. (1983) Genetic homology between independently isolated chlorobenzoate-degradative plasmids. *J Bacteriol* 153: 532-534.

Chen, C.Y., and Kado, C.I. (1994) Inhibition of *Agrobacterium tumefaciens* oncogenicity by the osa gene of pSa. *J Bacteriol* 176: 5697-5703.

Chen, E.Y., Schlessinger, D., and Kere, J. (1993) Ordered shotgun sequencing, a strategy for integrated mapping and sequencing of YAC clones. *Genomics* 17: 651-656.

Cheng, Q., Paszkiet, B.J., Shoemaker, N.B., Gardner, J.F., and Salyers, A.A. (2000) Integration and excision of a Bacteroides Conjugative Transposon, CTnDOT. *Journal of Bacteriology* 182: 4035-4043.

Chow, J.C., Young, D.W., Golenbock, D.T., Christ, W.J., and Gusovsky, F. (1999) Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J Biol Chem* 274: 10689-10692.

Cohen, M.L. (1992) Epidemiology of drug resistance: implications for a post-antimicrobial era. *Science* 257: 1050-1055.

Collyn, F., Lety, M.A., Nair, S., Escuyer, V., Ben Younes, A., Simonet, M., and Marceau, M. (2002) Yersinia pseudotuberculosis harbors a Type IV pilus gene cluster that contributes to pathogenicity. *Infect Immun* 70: 6196-6205.

Connor, E.M., and Loeb, M.R. (1983) A hemadsorption method for detection of colonies of *Haemophilus influenzae* type b expressing fimbriae. *J Infect Dis* 148: 855-860.

Cosgrove, S.E., and Carmeli, Y. (2003) The impact of antimicrobial resistance on health and economic outcomes. *Clin Infect Dis* 36: 1433-1437.

Craig, N.L. (1988) The mechanism of conservative site-specific recombination. *Annu Rev Genet* 22: 77-105.

Dahlberg, C., Linberg, C., Torsvik, V.L., and Hermansson, M. (1997) Conjugative plasmids isolated from bacteria in marine environments show various degrees of homology to each other and are not closely related to well-characterized plasmids. *Appl Environ Microbiol* 63: 4692-4697.

Danner, D.B., Deich, R.A., Sisco, K.L., and Smith, H.O. (1980) An eleven-base-pair sequence determines the specificity of DNA uptake in Haemophilus transformation. *Gene* 11: 311-318.

Daubin, V., Gouy, M., and Perriere, G. (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Res* 12: 1080-1090.

Daubin, V., Moran, N.A., and Ochman, H. (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* 301: 829-832.

Daum, R.S., Murphey-Corb, M., Shapira, E., and Dipp, S. (1988) Epidemiology of rob beta-lactamase among ampicillin-resistant *Haemophilus influenzae* isolates in the United States. *J Infect Dis* 157: 450-455.

Davison, J. (1999) Genetic exchange between bacteria in the environment. *Plasmid* 42: 73-91.

de Graaff, J., Elwell, L.P., and Falkow, S. (1976) Molecular nature of two beta-lactamase-specifying plasmids isolated from *Haemophilus influenzae* type b. *J Bacteriol* 126: 439-446.

de Groot, R., Campos, J., Moseley, S.L., and Smith, A.L. (1988) Molecular cloning and mechanism of trimethoprim resistance in *Haemophilus influenzae*. *Antimicrob Agents Chemother* 32: 477-484.

de Groot, R., Dzoljic-Danilovic, G., van Klingeren, B., Goessens, W.H., and Neyens, H.J. (1991) Antibiotic resistance in *Haemophilus influenzae*: mechanisms, clinical importance and consequences for therapy. *Eur J Pediatr* 150: 534-546.

del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M., and Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* 62: 434-464.

Deneer, H.G., Slaney, L., Maclean, I.W., and Albritton, W.L. (1982) Mobilization of nonconjugative antibiotic resistance plasmids in *Haemophilus ducreyi*. *J Bacteriol* 149: 726-732.

Dimopoulou, I.D., Kraak, W.A., Anderson, E.C., Nichols, W.W., Slack, M.P., and Crook, D.W. (1992) Molecular epidemiology of unrelated clusters of multiresistant strains of *Haemophilus influenzae*. *J Infect Dis* 165: 1069-1075.

Dimopoulou, I.D., Jordens, J.Z., Legakis, N.J., and Crook, D.W. (1997) A molecular analysis of Greek and UK *Haemophilus influenzae* conjugative resistance plasmids. *J Antimicrob Chemother* 39: 303-307.

Dimopoulou, I.D., Russell, J.E., Mohd-Zain, Z., Herbert, R., and Crook, D.W. (2002) Site-Specific Recombination with the Chromosomal tRNA(Leu) Gene by the Large Conjugative Haemophilus Resistance Plasmid. *Antimicrob Agents Chemother* 46: 1602-1603.

Dionisio, F., Matic, I., Radman, M., Rodrigues, O.R., and Taddei, F. (2002) Plasmids spread very fast in heterogeneous bacterial communities. *Genetics* 162: 1525-1532.

Dobbins, A.T., George, M., Jr., Basham, D.A., Ford, M.E., Houtz, J.M., Pedulla, M.L., Lawrence, J.G., Hatfull, G.F., and Hendrix, R.W. (2004) Complete genomic sequence

of the virulent *Salmonella* bacteriophage SP6. *J Bacteriol* 186: 1933-1944.
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: 6365-6372.
Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2: 414-424.
Draper, G.C., and Gober, J.W. (2002) Bacterial chromosome segregation. *Annu Rev Microbiol* 56: 567-597.
Edmonson, M.B., Granoff, D.M., Barenkamp, S.J., and Chesney, P.J. (1982) Outer membrane protein subtypes and investigation of recurrent *Haemophilus influenzae* type b disease. *J Pediatr* 100: 202-208.
Elwell, L.P., De Graaff, J., Seibert, D., and Falkow, S. (1975) Plasmid-linked ampicillin resistance in *Haempohilus influenza* type b. *Infect Immun* 12: 404-410.
Elwell, L.P., Saunders, J.R., Richmond, M.H., and Falkow, S. (1977) Relationships among some R plasmids found in *Haemophilus influenzae*. *J Bacteriol* 131: 356-362.
Elwell, L.P. (1994) R Plasmids and antibiotic resistances. In *Molecular Genetics of Bacterial Pathogenesis*. Miller, V.L., Kaper, J.B., Portnoy, D.A., Isberg, R.R (ed): American Society for Microbiology, pp. 17-41.
Eskola, J., Petola, H., Takala, A., Kayhty, H., Hakalinen, M., Karanko, V., Kela, E., Rekola, P., Ronnbery, P., Samuelson, J., Gordon, L., and Mahela, P. (1987) Efficacy of *Haemophilus influenzae* type b polysaccharide-diptheria toxoid conjugate vaccine in infancy. *New England Journal of Medicine* 317: 717-722.
Espinosa, M., Cohen, S., Couturier, M., del Solar, G., R., D.-O., Giraldo, R., Janniere, L., Miller, C., Osborn, M., and Thomas, C.M. (2000) Plasmid replication and copy number control. In *The Horizontal Gene Pool. Bacterial plasmids and gene spread*. Thomas, C.M. (ed). Amsterdam: Harwood Academic Publishers, pp. 1-47.
Esposito, D., and Scocca, J.J. (1994) Identification of an HP1 phage protein required for site-specific excision. *Mol Microbiol* 13: 685-695.
Esposito, D., and Scocca, J.J. (1997a) Purification and characterization of HP1 Cox and definition of its role in controlling the direction of site-specific recombination. *J Biol Chem* 272: 8660-8670.
Esposito, D., and Scocca, J.J. (1997b) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* 25: 3605-3614.
Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8: 186-194.
Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185.
Falla, T.J., Crook, D.W., Brophy, L.N., Maskell, D., Kroll, J.S., and Moxon, E.R. (1994) PCR for capsular typing of *Haemophilus influenzae*. *J Clin Microbiol* 32: 2382-2386.
Fallon, R.J. (1982) Increase in antibiotic resistance in *Haemophilus influenzae* in the United Kingdom since 1977. *British Medical Journal of Clinical Research* 285: 133.
Farley, M.M., and Stephens, D.S. (1992) Pathogenic events during *Haemophilus influenzae* type b infection of human nasopharyngeal mucosa. *J Infect Dis* 165 Suppl 1: S109-110.
Farrar, W.E., and O'Dell, N. (1974) B-Lactamase activity in ampicillin-resistant *Haemophilus influenzae*. *Antimicrobial Agents and Chemotherapy* 6: 625-629.
Firshein, W., and Kim, P. (1997) Plasmid replication and partition in *Escherichia coli*: is the cell membrane the key? *Mol Microbiol* 23: 1-10.
Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage,

A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., Sutton, G., Fits-Hugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, J.F., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., C.M., F., Smith, H.O., and Venter, J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 259: 496-512.

Foster, T.J., Davis, M.A., Roberts, D.E., Takeshita, K., and Kleckner, N. (1981) Genetic organization of transposon Tn10. *Cell* 23: 201-213.

Frank, A.C., and Lobry, J.R. (2000) Oriloc: prediction of replication boundaries in unannotated bacterial chromosomes. *Bioinformatics* 16: 560-561.

Freeman, J.M. (1998) Patterns of genome organization in Bacteria. *Science* 279: 1827a.

Galli, D.M., Chen, J., Novak, K.F., and Leblanc, D.J. (2001) Nucleotide sequence and analysis of conjugative plasmid pVT745. *J Bacteriol* 183: 1585-1594.

Garner, D., and Weston, V. (2003) Effectiveness of vaccination for *Haemophilus influenzae* type b. *Lancet* 361: 395-396.

Gelfanova, V., Hansen, E.J., and Spinola, S.M. (1999) Cytolethal distending toxin of *Haemophilus ducreyi* induces apoptotic death of Jurkat T cells. *Infect Immun* 67: 6394-6402.

Gerdes, K., Ayora, S., Canosa, I., Ceglowski, P., Diaz-Orejas, R., Franch, T., Gultyaev, A.P., Bugge Jensen, R., Kobayashi, I., Macpherson, C., Summers, D., Thomas, C.M., and Zielenkiewicz, U. (2000) Plasmid Maintenance System. In *The Horizontal Gene Pool*. Thomas, C.M. (ed). Amsterdam: Harwood Academic Publishers, pp. 49-85.

Gessner, B.D. (2002) Worldwide variation in the incidence of *Haemophilus influenzae* type b meningitis and its association with ampicillin resistance. *Eur J Clin Microbiol Infect Dis* 21: 79-87.

Gil, R., Silva, F.J., Zientz, E., Delmotte, F., Gonzalez-Candelas, F., Latorre, A., Rausell, C., Kamerbeek, J., Gadau, J., Holldobler, B., van Ham, R.C., Gross, R., and Moya, A. (2003) The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc Natl Acad Sci U S A* 100: 9388-9393.

Gilsdorf, J.R., McCrea, K.W., and Marrs, C.F. (1997) Role of pili in *Haemophilus influenzae* adherence and colonization. *Infect Immun* 65: 2997-3002.

Gomez-Lus, R. (1998) Evolution of bacterial resistance to antibiotics during the last three decades. *Int. Microbiology* 1: 279-284.

Gonzalez Lopez, M., del Pino De La Fuente, A., Garcia Martin, F.J., Calbo Torrecillas, F., Obando Santaella, I., Campos, J., and Martinez Valverde, A. (2000) [Meningitis due to *Haemophilus influenzae* type f]. *An Esp Pediatr* 53: 369-371.

Goodgal, S.H., and Mitchell, M.A. (1990) Sequence and uptake specificity of cloned sonicated fragments of *Haemophilus influenzae* DNA. *J Bacteriol* 172: 5924-5928.

Grindley, N.D.F. (2002) The movement of Tn3-like elements: Tranposition and cointegrate resolution. In *Mobile DNA II*. Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds). Washington, DC: ASM Press, pp. 272-305.

Grocock, R.J., and Sharp, P.M. (2002) Synonymous codon usage in Pseudomonas aeruginosa PA01. *Gene* 289: 131-139.

Groeneveld, K., van Alphen, L., Eijk, P.P., Jansen, H.M., and Zanen, H.C. (1988) Changes in outer membrane proteins of nontypable *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease. *J Infect Dis* 158: 360-365.

Groisman, E.A., Saier, M.H., Jr., and Ochman, H. (1992) Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. *Embo J* 11: 1309-1316.

Guerina, N.G., Langermann, S., Clegg, H.W., Kessler, T.W., Goldman, D.A., and Gilsdorf, J.R. (1982) Adherence of piliated *Haemophilus influenzae* type b to human oropharyngeal cells. *J Infect Dis* 146: 564.

Guillemot, D. (1999) Antibiotic use in humans and bacterial resistance. *Current Opinion in Microbiology* 2: 494-498.

Gutmann, L., Williamson, R., Collatz, E., and Acar, J.F. (1988) Mechanisms of beta-lactam resistance in *Haemophilus influenzae*. *Eur J Clin Microbiol Infect Dis* 7: 610-615.

Hacker, J., Blum-Oehler, G., Muhldorfer, I., and Tschape, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* 23: 1089-1097.

Hacker, J., and Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54: 641-679.

Hacker, J., and Carniel, E. (2001) Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2: 376-381.

Hacker, J., and Kaper, J. (2002) *Pathogenicity islands and the evolution of pathogenic microbes.* Berlin: Springer-Verlag.

Hardy, K. (1986) *Bacterial plasmids.* Berkshire, England: Van Nostrad Reinhold (International) Co. Ltd. Berkshire, England.

Hasegawa, K., Yamamoto, K., Chiba, N., Kobayashi, R., Nagai, K., Jacobs, M.R., Appelbaum, P.C., Sunakawa, K., and Ubukata, K. (2003) Diversity of ampicillin-resistance genes in *Haemophilus influenzae* in Japan and the United States. *Microb Drug Resist* 9: 39-46.

He, J., Baldini, R.L., Deziel, E., Saucier, M., Zhang, Q., Liberati, N.T., Lee, D., Urbach, J., Goodman, H.M., and Rahme, L.G. (2004) The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc Natl Acad Sci U S A* 101: 2530-2535.

Heath, P.T., and McVernon, J. (2002) The UK Hib vaccine experience. *Arch Dis Child* 86: 396-399.

Heffron, F., Sublett, R., Hedges, R.W., Jacob, A., and Falkow, S. (1975) Origin of the TEM-beta-lactamase gene found on plasmids. *J Bacteriol* 122: 250-256.

Hentschel, U., and Hacker, J. (2001) Pathogenicity islands: the tip of the iceberg. *Microbes Infect* 3: 545-548.

Hillen, W., and Berens, C. (1994) Mechanisms underlying expression of Tn*10* encoded tetracycline resistance. *Annu Rev Microbiol* 48: 345-369.

Hochhut, B., and Waldor, M.K. (1999) Site-specific integration of the conjugal Vibrio cholerae SXT element into prfC. *Mol Microbiol* 32: 99-110.

Holt, J.G., Krieg, N.R., Sneath, P.H.A., Staley, J.T., and Williams, S.T. (1994) Facultatively anaerobic Gram-neative rods. In *Bergey's Manual of Determinative Bacteriology*. Hensyl, W.R. (ed). Maryland: Williams & Wilkins, pp. 175-201.

Howard, A.J., Dunkin, K.T., and Millar, G.W. (1988) Nasopharyngeal carriage and antibiotic resistance of *Haemophilus influenzae* in healthy children. *Epidemiol Infect* 100: 193-203.

Jahn, G., Laufs, R., Kaulfers, P.M., and Kolenda, H. (1979) Molecular nature of two *Haemophilus influenzae* R factors containing resistances and the multiple integration of drug resistance transposons. *Journal of Bacteriology* 138: 584-897.

Jordens, J.Z., Leaves, N.I., Anderson, E.C., and Slack, M.P. (1993) Polymerase chain reaction-based strain characterization of noncapsulate *Haemophilus influenzae*. *J Clin Microbiol* 31: 2981-2987.

Jordens, J.Z. (1998) Characterisation of non-capsulate *Haemophilus influenzae* by repetitive extragenic palindromic (REP)-PCR. *J Med Microbiol* 47: 1031-1034.

Jorgensen, J.H., Lee, J.C., and Alexander, G.A. (1977) Rapid penicillinase paper strip test for detection of beta-lactamase-producing *Haemophilus influenzae* and Neisseria gonorrhoeae. *Antimicrob Agents Chemother* 11: 1087-1088.

Jorgensen, J.H. (1992) Update on mechanisms and prevalence of antimicrobial resistance in *Haemophilus influenzae*. *Clin Infect Dis* 14: 1119-1123.

Juteau, J.M., Sirois, M., Medeiros, A.A., and Levesque, R.C. (1991) Molecular distribution of ROB-1 beta-lactamase in *Actinobacillus pleuropneumoniae*. *Antimicrob Agents Chemother* 35: 1397-1402.

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: 6592-6601.

Karlin, S. (2001) Detecting anomalous gene clusters and pathogenicity islands in diverse bacterial genomes. *Trends Microbiol* 9: 335-343.

Katsura, K., Suzuki, F., Miyashita, S.I., Nishi, T., Hirochika, H., and Hashiba, T. (1997) The complete nucleotide sequence and characterization of the linear DNA plasmid pRS64-2 from the plant pathogenic fungus *Rhizoctonia solani*. *Curr Genet* 32: 431-435.

Kaulfers, P.M., Laufs, R., and Jahn, G. (1978) Molecular properties of transmissible R factors of *Haemophilus influenzae* determing tetracycline resistance. *J Gen Microbiol* 105: 243-252.

Kerr, A.R., Peden, J.F., and Sharp, P.M. (1997) Systematic base composition variation around the genome of *Mycoplasma genitalium*, but not *Mycoplasma pneumoniae*. *Mol Microbiol* 25: 1177-1179.

Khan, W., Ross, S., and Rodrigues, W. (1974) *Haemophilus influenzae* type B resistant to ampicillin. A report of two cases. *Journal of American Medical Association* 229: 229-301.

Killian, M. (1976) A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. *J Gen Microbiol* 93: 9-62.

Killian, M., Frederiksen, W., and Biberstein, E.L. (1981) *Haemophilus, Pasteurella and Actinobacillus*. London: Academic Press.

Kinmonth, A.L., Storrs, C.N., and Mitchell, R.G. (1978) Meningitis due to chloramphenicol-resistant *Haemophilus influenzae* type b. *Br Med J* 1: 694.

Kleckner, N. (1989) Transposon Tn*10*. In *Mobile DNA*. Berg, D.E. and Howe, M.M. (eds). Washington, D.C: American Society for Microbiology, pp. 229-262.

Klockgether, J., Reva, O., Larbig, K., and Tummler, B. (2004) Sequence analysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. *J Bacteriol* 186: 518-534.

Konieczny, I. (2003) Strategies for helicase recruitment and loading in bacteria. *European Molecular Biology Organization* 4: 37-41.

Kostman, J.R., Sherry, B.L., Fligner, C.L., Egaas, S., Sheeran, P., Baken, L., Bauwens, J.E., Clausen, C., Sherer, D.M., Plorde, J.J., and et al. (1993) Invasive *Haemophilus influenzae* infections in older children and adults in Seattle. *Clin Infect Dis* 17: 389-396.

Krieg, N.R., and Holt, J.G. (1984) *Bergey's Manual of Systemic Bacteriology*. Baltimore, USA: William & Wilkins.

Kroll, J.S., and Moxon, E.R. (1988) Capsulation and gene copy number at the cap locus of *Haemophilus influenzae* type b. *J Bacteriol* 170: 859-864.

Kroll, J.S., Loynds, B., Brophy, L.N., and Moxon, E.R. (1990) The bex locus in

encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. *Mol Microbiol* 4: 1853-1862.

Kroll, J.S., Loynds, B.M., and Moxon, E.R. (1991) The *Haemophilus influenzae* capsulation gene cluster: a compound transposon. *Mol Microbiol* 5: 1549-1560.

Kroll, J.S. (1992) The genetics of encapsulation in *Haemophilus influenzae*. *J Infect Dis* 165 Suppl 1: S93-96.

Kroll, J.S., Farrant, J.L., Tyler, S., Coulthart, M.B., and Langford, P.R. (2002) Characterisation and genetic organisation of a 24-MDa plasmid from the Brazilian Purpuric Fever clone of *Haemophilus influenzae* biogroup aegyptius. *Plasmid* 48: 38-48.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.

Lanka, E., and Wilkins, B.M. (1995) DNA processing reactions in bacterial conjugation. *Annu Rev Biochem* 64: 141-169.

Larbig, K.D., Christmann, A., Johann, A., Klockgether, J., Hartsch, T., Merkl, R., Wiehlmann, L., Fritz, H.J., and Tummler, B. (2002) Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol* 184: 6665-6680.

Laufs, R., and Kaulfers, P.-M. (1977) Molecular characterization of a plasmid specifying ampicillin resistance and its relationship to other R factors from *Haemophilus influenzae*. *Journal of General Microbiology* 103: 277-286.

Laufs, R., Kaulfers, P.M., Jahn, G., and Teschner, U. (1979) Molecular characterization of a small *Haemophilus influenzae* plasmid specifying beta-lactamase and its relationship to R factors from *Neisseria gonorrhoeae*. *Journal of General Microbiology* 111: 223-231.

Laufs, R., Riess, F.C., Jahn, G., Fock, R., and Kaulfers, P.M. (1981) Origin of *Haemophilus influenzae* R factors. *J Bacteriol* 147: 563-568.

Lawley, T.D., Burland, V., and Taylor, D.E. (2000) Analysis of the complete nucleotide sequence of the tetracycline-resistance transposon Tn10. *Plasmid* 43: 235-239.

Lawrence, J.G., and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci U S A* 95: 9413-9417.

Lazou Ahren, I., Bjartell, A., Egesten, A., and Riesbeck, K. (2001) Lipopolysaccharide-binding protein increases toll-like receptor 4-dependent activation by nontypeable *Haemophilus influenzae*. *J Infect Dis* 184: 926-930.

Leaves, N.I., and Jordens, J.Z. (1994) Development of a ribotyping scheme for *Haemophilus influenzae* type b. *Eur J Clin Microbiol Infect Dis* 13: 1038-1045.

Leaves, N.I., Dimopoulou, I., Hayes, I., Kerridge, S., Falla, T., Secka, O., Adegbola, R.A., Slack, M.P., Peto, T.E., and Crook, D.W. (2000) Epidemiological studies of large resistance plasmids in Haemophilus. *J Antimicrob Chemother* 45: 599-604.

Lee, P.S., and Lee, K.H. (2000) Genomic analysis. *Curr Opin Biotechnol* 11: 171-175.

Lerman, S.J., Kucera, J.C., and Brunken, J.M. (1979) Nasopharyngeal carriage of antibiotic resistant *Haemophilus influenzae* in healthy children. *Pediatrics* 64: 287-291.

Levy, J., Verhaegen, G., De Mol, P., Couturier, M., Dekegel, D., and Butzler, J.P. (1993) Molecular characterization of resistance plasmids in epidemiologically unrelated strains of multiresistant *Haemophilus influenzae*. *J Infect Dis* 168: 177-187.

Levy, S.B., Buu-Hoi, A., and Marshall, B. (1984) Transposon Tn10-like tetracycline resistance determinants in *Haemophilus parainfluenzae*. *J Bacteriol* 160: 87-94.

Lewin, B. (1990) *Genes IV.* New York: Oxford University Press.

Ling, J.M., Khin-Thi-Oo, H., Hui, Y.W., and French, G.L. (1989) Antimicrobial

susceptibilities of *Haemophilus* species in Hong Kong. *J Infect* 19: 135-142.
Livrelli, V.O., Darfeuille-Richaud, A., Rich, C.D., Joly, B.H., and Martel, J.L. (1988) Genetic determinant of the ROB-1 beta-lactamase in bovine and porcine *Pasteurella* strains. *Antimicrob Agents Chemother* 32: 1282-1284.
Lobry, J.R. (1996a) Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* 13: 660-665.
Lobry, J.R. (1996b) A simple vectorial representation of DNA sequences for the detection of replication origins in bacteria. *Biochimie* 78: 323-326.
Louis, P., and Galinski, E.A. (1997) Identification of plasmids in the genus *Marinococcus* and complete nucleotide sequence of plasmid pPL1 from *Marinococcus halophilus*. *Plasmid* 38: 107-114.
Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., and Spratt, B.G. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95: 3140-3145.
Makela, P.H. (1988) Unencapsulated *Haemophilus influenzae*--what kind of pathogen? *Eur J Clin Microbiol Infect Dis* 7: 606-609.
Manchak, J., Anthony, K.G., and Frost, L.S. (2002) Mutational analysis of F-pilin reveals domains for pilus assembly, phage infection and DNA transfer. *Mol Microbiol* 43: 195-205.
Maneewannakul, K., Maneewannakul, S., and Ippen-Ihler, K. (1995) Characterization of *traX*, the F plasmid locus required for acetylation of F-pilin subunits. *J Bacteriol* 177: 2957-2964.
Mantri, Y., and Williams, K.P. (2004) Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. *Nucleic Acids Res* 32 Database issue: D55-58.
Marshall, B., Roberts, M., Smith, A., and Levy, S.B. (1984) Homogeneity of transferable tetracycline-resistance determinants in *Haemophilus* species. *J Infect Dis* 149: 1028-1029.
Martin, P.R., Shea, R.J., and Mulks, M.H. (2001) Identification of a plasmid-encoded gene from *Haemophilus ducreyi* which confers NAD independence. *J Bacteriol* 183: 1168-1174.
Massung, R.F., Esposito, J.J., Liu, L.I., Qi, J., Utterback, T.R., Knight, J.C., Aubin, L., Yuran, T.E., Parsons, J.M., Loparev, V.N., and et al. (1993) Potential virulence determinants in terminal regions of variola smallpox virus genome. *Nature* 366: 748-751.
Mathies, A.W. (1972) Penicillins in the treatment of bacterial meningitis. *Journal Royal College of Physician of London* 6: 584-588.
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K. (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413: 852-856.
McMurry, L., Petrucci, R.E., Jr., and Levy, S.B. (1980) Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proc Natl Acad Sci U S A* 77: 3974-3977.
McVernon, J., Andrews, N., Slack, M.P., and Ramsay, M.E. (2003) Risk of vaccine failure after *Haemophilus influenzae* type b (Hib) combination vaccines with acellular

pertussis. *Lancet* 361: 1521-1523.
Meats, E., Feil, E.J., Stringer, S., Cody, A.J., Goldstein, R., Kroll, J.S., Popovic, T., and Spratt, B.G. (2003) Characterization of encapsulated and noncapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *J Clin Microbiol* 41: 1623-1636.
Medeiros, A., and O'Brien, T.F. (1975) Ampicillin-resistant *Haemophilus influenzae* type B possessing a TEM-type enzyme but little permeability barrier to ampicillin. *Lancet*: 716-719.
Medeiros, A., Levesque, R., and Jacoby, G. (1986) An animal source for the ROB-1 b-lactamase of *Haemophilus influenzae* type b. *Antimicrobial Agents and Chemotherapy* 29: 212-215.
Medzhitov, R., Preston-Hurlburt, P., and Janeway, C.A., Jr. (1997) A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* 388: 394-397.
Mendelman, P.M., Doroshow, C.A., Gandy, S.L., Syriopoulou, V., Weigen, C.P., and Smith, A.L. (1984) Plasmid-mediated resistance in multiply resistant *Haemophilus influenzae* type b causing meningitis: molecular characterization of one strain and review of the literature. *J Infect Dis* 150: 30-39.
Mendelman, P.M., Syriopoulou, V.P., Gandy, S.L., Ward, J.I., and Smith, A.L. (1985) Molecular epidemiology of plasmid-mediated ampicillin resistance in *Haemophilus influenzae* type b isolates from Alaska. *J Infect Dis* 151: 1061-1072.
Messer, W., and Weigel, C. (1997) DnaA initiator--also a transcription factor. *Mol Microbiol* 24: 1-6.
Mhlanga-Mutangadura, T., Morlin, G., Smith, A.L., Eisenstark, A., and Golomb, M. (1998) Evolution of the major pilus gene cluster of *Haemophilus influenzae*. *J Bacteriol* 180: 4693-4703.
Miller, R.B., Lein, D.H., McEntee, K.E., Hall, C.E., and Shin, S. (1983) *Haemophilus somnus* infection of the reproductive tract of cattle: a review. *J Am Vet Med Assoc* 182: 1390-1392.
Mills, D.A., Phister, T.G., Dunny, G.M., and McKay, L.L. (1998) An origin of transfer (oriT) on the conjugative element pRS01 from *Lactococcus lactis* subsp. lactis ML3. *Appl Environ Microbiol* 64: 1541-1544.
Mitsuda, T., Kuroki, H., Ishikawa, N., Imagawa, T., Ito, S., Miyamae, T., Mori, M., Uehara, S., and Yokota, S. (1999) Molecular epidemiological study of *Haemophilus influenzae* serotype b strains obtained from children with meningitis in Japan. *J Clin Microbiol* 37: 2548-2552.
Morris, C., Yip, C.M., Tsui, I.S., Wong, D.K., and Hackett, J. (2003) The shufflon of *Salmonella enterica* serovar Typhi regulates Type IVB pilus-mediated bacterial self-association. *Infect Immun* 71: 1141-1146.
Moxon, E.R., Glode, M.P., Sutton, A., and Robbins, J.B. (1977) The infant rat as a model of bacterial meningitis. *J Infect Dis* 136 Suppl: S186-190.
Moxon, E.R., and Murphy, P.A. (1978) *Haemophilus influenzae* bacteremia and meningitis resulting from survival of a single organism. *Proc Natl Acad Sci U S A* 75: 1534-1536.
Moxon, E.R., and Maskell, D. (1992) *Haemophilus influenzae*: the biochemistry and biology of a virulence factor. In *Molecular Biology of Bacterial Infection: Current Status and Future Perspectives*. Harmaeche, C.E., Penn, C.W. and Smyth, C.J. (eds). Cambridge, UK: Cambridge University.
Munson, R.S., Jr., and Granoff, D.M. (1985) Purification and partial characterization of outer membrane proteins P5 and P6 from *Haemophilus influenzae* type b. *Infect*

Immun 49: 544-549.

Munson, R.S., Jr., Grass, S., and West, R. (1993) Molecular cloning and sequence of the gene for outer membrane protein P5 of *Haemophilus influenzae*. *Infect Immun* 61: 4017-4020.

Murata, T., Ohnishi, M., Ara, T., Kaneko, J., Han, C.G., Li, Y.F., Takashima, K., Nojima, H., Nakayama, K., Kaji, A., Kamio, Y., Miki, T., Mori, H., Ohtsubo, E., Terawaki, Y., and Hayashi, T. (2002) Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. *J Bacteriol* 184: 3194-3202.

Murphey-Corb, M., Nolan-Willard, M., and Daum, R.S. (1984) Integration of plasmid DNA coding for beta-lactamase production in the *Haemophilus influenzae* chromosome. *J Bacteriol* 160: 815-817.

Murphy, T.F., and Apicella, M.A. (1987) Non-typable *Haemophilus influenzae*: a review of clinical aspects, surface antigens and the human immune response to infection. *Reviews of Infectious Diseases* 9: 1-15.

Murphy, T.F., and Bartos, L.C. (1988) Purification and analysis with monoclonal antibodies of P2, the major outer membrane protein of nontypable *Haemophilus influenzae*. *Infect Immun* 56: 1084-1089.

Murphy, T.F., Nelson, M.B., and Apicella, M.A. (1992) The P6 outer membrane protein of nontypeable *Haemophilus influenzae* as a vaccine antigen. *J Infect Dis* 165 Suppl 1: S203-205.

Musser, J.M., Granoff, D.M., Pattison, P.E., and Selander, R.K. (1985) A population genetic framework for the study of invasive diseases caused by serotype b strains of *Haemophilus influenzae*. *Proc Natl Acad Sci U S A* 82: 5078-5082.

Musser, J.M., Barenkamp, S.J., Granoff, D.M., and Selander, R.K. (1986) Genetic relationships of serologically nontypable and serotype b strains of *Haemophilus influenzae*. *Infect Immun* 52: 183-191.

Musser, J.M., Kroll, J.S., Granoff, D.M., Moxon, E.R., Brodeur, B.R., Campos, J., Dabernat, H., Frederiksen, W., Hamel, J., Hammond, G., Holby, E.A., Jonsdottir, K.E., Kabber, M., Kallings, I., Khan, W.N., Killian, M., Knowles, K., Koornhof, H.J., Law, B., Li, K.I., Montgomery, J., Pattison, P.E., Piffaretti, J.C., Takala, A.K., Thong, M.L., Wall, R.A., Ward, J.I., and Selander, R.K. (1990) Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. *Reviews of Infectious Diseases* 12: 75-111.

Musser, J.M. (1996) Molecular population genetic analysis of emerged bacterial pathogens: selected insights. *Emerg Infect Dis* 2: 1-17.

Nelson, M.B., Murphy, T.F., van Keulen, H., Rekosh, D., and Apicella, M.A. (1988) Studies on P6, an important outer-membrane protein antigen of *Haemophilus influenzae*. *Rev Infect Dis* 10 Suppl 2: S331-336.

Noel, G.J., Hoiseth, S.K., and Edelson, P.J. (1992) Type b capsule inhibits ingestion of *Haemophilus influenzae* by murine macrophages: studies with isogenic encapsulated and unencapsulated strains. *J Infect Dis* 166: 178-182.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405: 299-304.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and *et al.* (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol* 223: 1-7.

Okeke, I.N., and Edelman, R. (2001) Dissemination of antibiotic-resistant bacteria across geographic borders. *Clin Infect Dis* 33: 364-369.

Omikunle, A., Takahashi, S., Ogilvie, C.L., Wang, Y., Rodriguez, C.A., St Geme, J.W.,

3rd, and Adderson, E.E. (2002) Limited genetic diversity of recent invasive isolates of non-serotype b encapsulated *Haemophilus influenzae*. *J Clin Microbiol* 40: 1264-1270.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., and Hattori, M. (2001) Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A* 98: 12215-12220.

Osborn, A.M., and Boltner, D. (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* 48: 202-212.

Overturf, G.D., Cable, D., and Ward, J. (1987) Ampicillin-chloramphenicol-resistant *Haemophilus influenzae*: plasmid-mediated resistance in bacterial meningitis. *Pediatr Res* 22: 438-441.

Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12: 357-358.

Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrell, B.G. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413: 848-852.

Paterson, A.J., Macsween, K.F., and Pennington, T.H. (1987) *Haemophilus influenzae* subtyping by SDS-PAGE of whole-cell polypeptides. *Epidemiol Infect* 99: 179-189.

Patrick, D., Betts, J., Frey, E.A., Prameya, R., Dorovini-Zis, K., and Finlay, B.B. (1992) *Haemophilus influenzae* lipopolysaccharide disrupts confluent monolayers of bovine brain endothelial cells via a serum-dependent cytotoxic pathway. *J Infect Dis* 165: 865-872.

Pearson, W.R., and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444-2448.

Pearson, W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* 183: 63-98.

Peltola, H., Kayhty, H., Sivonen, A., and Makela, H. (1977) *Haemophilus influenzae* type b capsular polysaccharide vaccine in children: a double-blind field study of 100,000 vaccinees 3 months to 5 years of age in Finland. *Pediatrics* 60: 730-737.

Peltola, H. (2000) Worldwide *Haemophilus influenzae* type b disease at the beginning of the 21st century: global analysis of the disease burden 25 years after the use of the polysaccharide vaccine and a decade after the advent of conjugates. *Clin Microbiol Rev* 13: 302-317.

Philpott-Howard, J., and Williams, J.D. (1982) Increase in antibiotic resistance in *Haemophilus influenzae* in the United Kingdom since 1977: report of study group. *Br Med J (Clin Res Ed)* 284: 1597-1599.

PHLS (2002) Antimicrobial Resistance in 2000: England and Wales. London: Public Health Laboratory Services.

Picardeau, M., Lobry, J.R., and Hinnebusch, B.J. (1999) Physical mapping of an origin of bidirectional replication at the centre of the Borrelia burgdorferi linear chromosome. *Mol Microbiol* 32: 437-445.

Pichichero, M.E., Loeb, M., Anderson, and Smith, D.H. (1982) Do pili play a role in

pathogenicity of *Haemophilus influenzae* type B? *Lancet* 2: 960-962.
Pickard, D., Wain, J., Baker, S., Line, A., Chohan, S., Fookes, M., Barron, A., Gaora, P.O., Chabalgoity, J.A., Thanky, N., Scholes, C., Thomson, N., Quail, M., Parkhill, J., and Dougan, G. (2003) Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. *J Bacteriol* 185: 5055-5065.
Pittman, M. (1930) The 'S' and 'R' forms of *Haemophilus influenzae*. *Proceedings of Society of Experimental Biological Medicine* 27: 299.
Pittman, M. (1931) Variation and type specifity in the bacterial species *Haemophilus influenzae*. *Journal of Experimental Medicine* 53: 471-493.
Plaut, A.G. (1978) Microbial IgA proteases. *N Engl J Med* 298: 1459-1463.
Powell, M., Koutsia-Carouzou, C., Voutsinas, D., Seymour, A., and Williams, J.D. (1987) Resistance of clinical isolates of *Haemophilus influenzae* in United Kingdom 1986. *Br Med J (Clin Res Ed)* 295: 176-179.
Powell, M. (1988) Antimicrobial resistance in *Haemophilus influenzae*. *J Med Microbiol* 27: 81-87.
Powell, M., and Livermore, D.M. (1988) Mechanisms of chloramphenicol resistance in *Haemophilus influenzae* in the United Kingdom. *J Med Microbiol* 27: 89-93.
Powell, M., Yeo, S.F., Seymour, A., Yuan, M., Williams, J.D., and Fah, Y.S. (1992) Antimicrobial resistance in *Haemophilus influenzae* from England and Scotland in 1991. *J Antimicrob Chemother* 29: 547-554.
Pushparajah, K., Ramnarayan, P., Maniyar, A., Paget, R., and Britto, J. (2003) Continued threat of *Haemophilus influenzae* type B disease in the UK. *Lancet* 361: 90.
Quintiliani, R., and Courvalin, P. (1995) Mechanisms of resistance to antimicrobial agents. In *Manual of Clinical Microbiology*. Murray, P.R., Baron, E.J., Pfaller, M.A., Tenover, F.C. and Yolken, R.H. (eds). Washington, D.C.: American Society for Microbiology, pp. 1308-1326.
Rafii, F., and Crawford, D.L. (1988) Transfer of conjugative plasmids and mobilization of a nonconjugative plasmid between Streptomyces strains on agar and in soil. *Appl Environ Microbiol* 54: 1334-1340.
Ramsay, M.E., McVernon, J., Andrews, N.J., Heath, P.T., and Slack, M.P. (2003) Estimating *Haemophilus influenzae* type b vaccine effectiveness in England and Wales by use of the screening method. *J Infect Dis* 188: 481-485.
Ravatn, R., Studer, S., Zehnder, A.J., and van der Meer, J.R. (1998) Int-B13, an unusual site-specific recombinase of the bacteriophage P4 integrase family, is responsible for chromosomal insertion of the 105-kilobase clc element of Pseudomonas sp. Strain B13. *J Bacteriol* 180: 5505-5514.
Reinholdt, J., and Kilian, M. (1997) Comparative analysis of immunoglobulin A1 protease activity among bacteria representing different genera, species, and strains. *Infect Immun* 65: 4452-4459.
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: 1907-1914.
Richards, J.C., Cox, A.D., Schweda, E.K., Martin, A., Hood, D.W., and Moxon, E.R. (2001) Structure and functional genomics of lipopolysaccharide expression in *Haemophilus influenzae*. *Adv Exp Med Biol* 491: 515-524.
Roberts, M., Stull, T.L., and Smith, A.L. (1981) Comparative virulence of *Haemophilus influenzae* with a type b or type d capsule. *Infect Immun* 32: 518-524.
Roberts, M.C., and Smith, A.L. (1980) Molecular characterization of "plasmid-free" antibiotic-resistant *Haemophilus influenzae*. *J Bacteriol* 144: 476-479.
Roberts, M.C., Swenson, C.D., Owens, L.M., and Smith, A.L. (1980) Characterization of

chloramphenicol-resistant *Haemophilus influenzae*. *Antimicrob Agents Chemother* 18: 610-615.

Roszczyk, E., and Goodgal, S. (1975) Methylase activities from *Haemophilus influenzae* that protect *Haemophilus parainfluenzae* transforming deoxyribonucleic acid from inactivation by *Haemophilus influenzae* endonuclease R. *J Bacteriol* 123: 287-293.

Rowe-Magnus, D.A., and Mazel, D. (1999) Resistance gene capture. *Curr Opin Microbiol* 2: 483-488.

Roy, P.H. (1999) Horizontal transfer of genes in bacteria. *Microbiology Today* 26: 168-170.

Rubin, L.G., Medeiros, A.A., Yolken, R.H., and Moxon, E.R. (1981) Ampicillin treatment failure of apparently beta-lactamase-negative *Haemophilus influenzae* type b meningitis due to novel beta-lactamase. *Lancet* 2: 1008-1010.

Rubin, L.G., and Moxon, E.R. (1983) Pathogenesis of bloodstream invasion with *Haemophilus influenzae* type b. *Infect Immun* 41: 280-284.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944-945.

Sahm, D.F., Jones, M.E., Hickey, M.L., Diakun, D.R., Mani, S.V., and Thornsberry, C. (2000) Resistance surveillance of Streptococcus pneumoniae, *Haemophilus influenzae* and Moraxella catarrhalis isolated in Asia and Europe, 1997-1998. *J Antimicrob Chemother* 45: 457-466.

Sakai, D., and Komano, T. (2000) The *pilL* and *pilN* genes of IncI1 plasmids R64 and ColIb-P9 encode outer membrane lipoproteins responsible for thin pilus biogenesis. *Plasmid* 43: 149-152.

Salyers, A.A., Shoemaker, N.B., Stevens, A.M., and Li, L.Y. (1995) Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev* 59: 579-590.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual.* New York: Cold Spring Harbor Laboratory Press.

Sanders, J.D., Cope, L.D., and Hansen, E.J. (1994) Identification of a locus involved in the utilization of iron by *Haemophilus influenzae*. *Infect Immun* 62: 4515-4525.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977a) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265: 687-695.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977b) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265: 687-695.

Saunders, J.R., and Sykes, R.B. (1977) Transfer of a plasmid-specified beta-lactamase gene from *Haemophilus influenzae*. *Antimicrob Agents Chemother* 11: 339-344.

Saveson, C.J., and Lovett, S.T. (1999) Tandem repeat recombination induced by replication fork defects in *Escherichia coli* requires a novel factor, RadC. *Genetics* 152: 5-13.

Scheifele, D.W., Syriopoulou, V.P., Harding, A.L., Emerson, B.B., and Smith, A.L. (1976) Evaluation of a rapid beta-lactamase test for detecting ampicillin-resistant strains of *Hemophilus influenzae* type b. *Pediatrics* 58: 382-387.

Scheifele, D.W., and Fussell, S.J. (1981a) Frequency of ampicillin-resistant *Haemophilus parainfluenzae* in children. *Journal of Infectious Diseases* 143: 495-498.

Scheifele, D.W., and Fussell, S.J. (1981b) Ampicillin-resistant *Haemophilus influenzae* colonizing ambulatory children. *Am J Dis Child* 135: 406-409.

Scheifele, D.W., Fussell, S.J., and Roberts, M.C. (1982) Characterization of ampicillin-

resistant *Haemophilus parainfluenzae*. *Antimicrob Agents Chemother* 21: 734-739.
Scheifele, D.W., and Fussell, S.J. (1986) Molecular epidemiology of antibiotic resistance plasmids of *Haemophilus* species and *Neisseria gonorrhoeae*. *The Journal of Infectious Diseases* 143: 495-498.
Schell, M.A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.C., Desiere, F., Bork, P., Delley, M., Pridmore, R.D., and Arigoni, F. (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* 99: 14422-14427.
Schito, G.C., Debbia, E.A., and Marchese, A. (2000) The evolving threat of antibiotic resistance in Europe: new data from the Alexander Project. *J Antimicrob Chemother* 46 Suppl T1: 3-9.
Schwarz, F.V., Perreten, V., and Teuber, M. (2001) Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from *Enterococcus faecalis* RE25. *Plasmid* 46: 170-187.
Scriver, S.R., Walmsley, S.L., Kau, C.L., Hoban, D.J., Brunton, J., McGeer, A., Moore, T.C., and Witwicki, E. (1994) Determination of antimicrobial susceptibilities of Canadian isolates of *Haemophilus influenzae* and characterization of their beta-lactamases. Canadian Haemophilus Study Group. *Antimicrob Agents Chemother* 38: 1678-1680.
Sentchilo, V., Zehnder, A.J., and van der Meer, J.R. (2003) Characterization of two alternative promoters for integrase expression in the *clc* genomic island of *Pseudomonas* sp. strain B13. *Mol Microbiol* 49: 93-104.
Setlow, J.K., Boling, M.E., Beattie, K.L., and Kimball, W.A.G. (1972) A complex of recombination and repair gene in *Haemophilus influenzae*. *Journal of Molecular Biology* 68: 361-378.
Shah, D.S., Cross, M.A., Porter, D., and Thomas, C.M. (1995) Dissection of the core and auxiliary sequences in the vegetative replication origin of promiscuous plasmid RK2. *J Mol Biol* 254: 608-622.
Sharp, P.M., Shields, D.C., Wolfe, K.H., and Li, W.H. (1989) Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 246: 808-810.
Silverman, P.M. (1997) Towards a structural biology of bacterial conjugation. *Mol Microbiol* 23: 423-429.
Simasathien, S., Duangmani, C., and Echeverria, P. (1980) *Haemophilus influenzae* type B resistant to ampicillin and chloramphenicol in an orphanage in Thailand. *Lancet* 2: 1214-1217.
Smith, H.O., Gwinn, M.L., and Salzberg, S.L. (1999) DNA uptake signal sequences in naturally transformable bacteria. *Res Microbiol* 150: 603-616.
Smith, M.D., and Kelsey, M.C. (1989) Demonstration of a functional variant of chloramphenicol acetyltransferase in *Haemophilus influenzae*. *J Med Microbiol* 29: 263-268.
Smith-Vaughan, H.C., Sriprakash, K.S., Mathews, J.D., and Kemp, D.J. (1995) Long PCR-ribotyping of nontypeable *Haemophilus influenzae*. *J Clin Microbiol* 33: 1192-1195.
Sonnhammer, E.L., and Durbin, R. (1994) A workbench for large-scale sequence homology analysis. *Comput Appl Biosci* 10: 301-307.
Speer, B.S., Shoemaker, N.B., and Salyers, A.A. (1992) Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. *Clin Microbiol Rev* 5: 387-399.
Spencer, D.H., Kas, A., Smith, E.E., Raymond, C.K., Sims, E.H., Hastings, M., Burns, J.L., Kaul, R., and Olson, M.V. (2003) Whole-genome sequence variation among

multiple isolates of *Pseudomonas aeruginosa*. *J Bacteriol* 185: 1316-1325.
Spratt, B.G. (1999) Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. *Curr Opin Microbiol* 2: 312-316.
Srimanote, P., Paton, A.W., and Paton, J.C. (2002) Characterization of a novel Type IV pilus locus encoded on the large plasmid of locus of enterocyte effacement-negative Shiga-toxigenic *Escherichia coli* strains that are virulent for humans. *Infect Immun* 70: 3094-3100.
St.-Geme-III, J.W. (1994) Early events in the pathogenesis of *Haemophilus influenzae* disease. In *Molecular Genetics of Bacterial Pathogenesis*. Miller, V.L., Kaper, J.B., Portnoy, D.A., Isberg, R.R (ed). Washington, D.C: ASM Press, pp. 157-172.
Staden, R. (1996) The Staden sequence analysis package. *Mol Biotechnol* 5: 233-241.
Stark, W.M., Boocock, M.R., and Sherratt, D.J. (1992) Catalysis by site-specific recombinases. *Trends Genet* 8: 432-439.
Stevens, M.K., Latimer, J.L., Lumbley, S.R., Ward, C.K., Cope, L.D., Lagergard, T., and Hansen, E.J. (1999) Characterization of a *Haemophilus ducreyi* mutant deficient in expression of cytolethal distending toxin. *Infect Immun* 67: 3900-3908.
Stuy, J.H. (1979) Plasmid transfer in *Haemophilus influenzae*. *J Bacteriol* 139: 520-529.
Stuy, J.H. (1980a) Mechanism of additive genetic transformation in *Haemophilus influenzae*. *J Bacteriol* 144: 999-1002.
Stuy, J.H. (1980b) Chromosomally integrated conjugative plasmids are common in antibiotic-resistant *Haemophilus influenzae*. *J Bacteriol* 142: 925-930.
Subramanya, H.S., Bird, L.E., Brannigan, J.A., and Wigley, D.B. (1996) Crystal structure of a DExx box DNA helicase. *Nature* 384: 379-383.
Sullivan, J.T., Trzebiatowski, J.R., Cruickshank, R.W., Gouzy, J., Brown, S.D., Elliot, R.M., Fleetwood, D.J., McCallum, N.G., Rossbach, U., Stuart, G.S., Weaver, J.E., Webby, R.J., De Bruijn, F.J., and Ronson, C.W. (2002) Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J Bacteriol* 184: 3086-3095.
Syriopoulou, V., Scheifele, D., Howie, V., Ploussard, J., Sloyer, J., and Smith, A.L. (1976) Incidence of ampicillin-resistant *Hemophilus influenzae* in otitis media. *J Pediatr* 89: 839-841.
Syriopoulou, V., Scheifele, D., Smith, A.L., Perry, P.M., and Howie, V. (1978) Increasing incidence of ampicillin resistance in *Hemophilus influenzae*. *J Pediatr* 92: 889-892.
Takeya, T., Nomiyama, H., Miyoshi, J., Shimada, K., and Takagi, Y. (1979) DNA sequences of the integration sites and inverted repeated structure of transposon Tn*3*. *Nucleic Acids Res* 6: 1831-1841.
Tenover, F.C., Huang, M.B., Rasheed, J.K., and Persing, D.H. (1994) Development of PCR assays to detect ampicillin resistance genes in cerebrospinal fluid samples containing *Haemophilus influenzae*. *J Clin Microbiol* 32: 2729-2737.
Terry, T.D., Zalucki, Y.M., Walsh, S.L., Blackall, P.J., and Jennings, M.P. (2003) Genetic analysis of a plasmid encoding haemocin production in *Haemophilus paragallinarum*. *Microbiology* 149: 3177-3184.
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., and Venter, J.C. (2000) Complete genome sequence of *Neisseria men*ingitidis serogroup B strain MC58.

Science 287: 1809-1815.
Thomas, C.M. (2000) Paradigms of plasmid organization. *Mol Microbiol* 37: 485-491.
Thomas, N.S., Lusher, M., Storey, C.C., and Clarke, I.N. (1997) Plasmid diversity in Chlamydia. *Microbiology* 143 (Pt 6): 1847-1854.
Thorne, G.M., and Farrar, W.E., Jr. (1975) Transfer of ampicillin resistance between strains of *Haemophilus influenzae* type B. *J Infect Dis* 132: 276-281.
Toussaint, A., and Merlin, C. (2002) Mobile elements as a combination of functional modules. *Plasmid* 47: 26-35.
Tsui, I.S., Yip, C.M., Hackett, J., and Morris, C. (2003) The Type IVB pili of *Salmonella enterica* serovar Typhi bind to the cystic fibrosis transmembrane conductance regulator. *Infect Immun* 71: 6049-6050.
Tu, Q., and Ding, D. (2003) Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. *FEMS Microbiol Lett* 221: 269-275.
Turk, D.C. (1982) Clinical importance of *Haemophilus influenzae* - 1981. In *Haemophilus influenzae - Epidemiology, immunology and prevention of disease*. Sell, S.H. and Wright, P.F. (eds). New York: Elsevier Science Publishing Co., Inc, pp. 3-9.
Turk, D.C. (1984) The pathogenicity of *Haemophilus influenzae*. *J Med Microbiol* 18: 1-16.
Turner, S.A., Luck, S.N., Sakellaris, H., Rajakumar, K., and Adler, B. (2001) Nested deletions of the SRL pathogenicity island of *Shigella flexneri* 2a. *J Bacteriol* 183: 5535-5543.
Uchiyama, N., Greene, G.R., Kitts, D.B., and Thrupp, L.D. (1980) Meningitis due to *Haemophilus influenzae* type b resistant to ampicillin and chloramphenicol. *J Pediatr* 97: 421-424.
Uduman, S.A., Adeyemi, E., El-Khadir, A., Jose, K., Benedict, S., and Bener, A. (2000) *Haemophilus influenzae* type b still remains a leading cause of meningitis among unvaccinated children--a prospective CSF analysis study. *J Trop Pediatr* 46: 331-334.
Uno, R., Nakayama, Y., Arakawa, K., and Tomita, M. (2000) The orientation bias of Chi sequences is a general tendency of G-rich oligomers. *Gene* 259: 207-215.
Urwin, G., Krohn, J.A., Deaver-Robinson, K., Wenger, J.D., and Farley, M.M. (1996) Invasive disease due to *Haemophilus influenzae* serotype f: clinical and epidemiologic characteristics in the H. influenzae serotype b vaccine era. The *Haemophilus influenzae* Study Group. *Clin Infect Dis* 22: 1069-1076.
Van, A.D., Bieth, G., and Bouanchaud, D.H. (1975) [Tetracycline resistance in *Haemophilus influenzae* is plasmid-linked]. *C R Acad Sci Hebd Seances Acad Sci D* 280: 1321-1323.
van Alphen, L., van den Berghe, N., and Geelen-van den Broek, L. (1988) Interaction of *Haemophilus influenzae* with human erythrocytes and oropharyngeal epithelial cells is mediated by a common fimbrial epitope. *Infect Immun* 56: 1800-1806.
van Belkum, A., Duim, B., Regelink, A., Moller, L., Quint, W., and van Alphen, L. (1994) Genomic DNA fingerprinting of clinical *Haemophilus influenzae* isolates by polymerase chain reaction amplification: comparison with major outer-membrane protein and restriction fragment length polymorphism analysis. *J Med Microbiol* 41: 63-68.
van der Meer, J.R., and Sentchilo, V. (2003) Genomic islands and the evolution of catabolic pathways in bacteria. *Curr Opin Biotechnol* 14: 248-254.
van der Zwan, J.C., Dankert, J., de Vries, K., Orie, N.G., and Kauffman, H.F. (1978) Purification of specific precipitinogen and extraction of endotoxin from *Haemophilus influenzae*. *J Clin Pathol* 31: 370-377.
van Elsas, J., Fry, J., Hirsch, P., and Molin, S. (2000) Ecology of plasmid transfer and

spread. In *The horizontal gene pool. Bacterial plasmids and gene spread.* Thomas, C.M. (ed). Netherlands: Harwood Academic Publishers, pp. 175-199.

van Ham, S.M., van Alphen, L., Mooi, F.R., and van Putten, J.P. (1994) The fimbrial gene cluster of *Haemophilus influenzae* type b. *Mol Microbiol* 13: 673-684.

van Klingeren, B., van Embden, J.D., and Dessens-Kroon, M. (1977) Plasmid-mediated chloramphenicol resistance in *Haemophilus influenzae*. *Antimicrob Agents Chemother* 11: 383-387.

Venter, J.C., Smith, H.O., and Hood, L. (1996) A new strategy for genome sequencing. *Nature* 381: 364-366.

Virkola, R., Lahteenmaki, K., Eberhard, T., Kuusela, P., van Alphen, L., Ullberg, M., and Korhonen, T.K. (1996) Interaction of *Haemophilus influenzae* with the mammalian extracellular matrix. *J Infect Dis* 173: 1137-1147.

Wain, J., Diem Nga, L.T., Kidgell, C., James, K., Fortune, S., Song Diep, T., Ali, T., P, O.G., Parry, C., Parkhill, J., Farrar, J., White, N.J., and Dougan, G. (2003) Molecular analysis of *incHI1* antimicrobial resistance plasmids from *Salmonella* serovar Typhi strains associated with typhoid fever. *Antimicrob Agents Chemother* 47: 2732-2739.

Wang, X., Moser, C., Louboutin, J.P., Lysenko, E.S., Weiner, D.J., Weiser, J.N., and Wilson, J.M. (2002) Toll-like receptor 4 mediates innate immune responses to *Haemophilus influenzae* infection in mouse lung. *J Immunol* 168: 810-815.

Waters, V.L. (1999) Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance. *Front Biosci* 4: D433-456.

Watson, R., Dowell, S., and Jayaraman, M. (1999) Antimicrobial use for pediatric upper respiratory infections: reported practice, actual practice, and parent beliefs. *Pediatrics* 104: 1251-1257.

Weber, A., Harris, K., Lohrke, S., Forney, L., and Smith, A.L. (1991) Inability to express fimbriae results in impaired ability of *Haemophilus influenzae* b to colonize the nasopharynx. *Infect Immun* 59: 4724-4728.

Weber, J.L., and Myers, E.W. (1997) Human whole-genome shotgun sequencing. *Genome Res* 7: 401-409.

Weiser, J.N., Goldberg, J.B., Pan, N., Wilson, L., and Virji, M. (1998) The phosphorylcholine epitope undergoes phase variation on a 43-kilodalton protein in Pseudomonas aeruginosa and on pili of *Neisseria meningitidis* and Neisseria gonorrhoeae. *Infect Immun* 66: 4263-4267.

Wenger, J.D., Pierce, R., Deaver, K., Franklin, R., Bosley, G., Pigott, N., and Broome, C.V. (1992) Invasive *Haemophilus influenzae* disease: a population-based evaluation of the role of capsular polysaccharide serotype. *Haemophilus influenzae* Study Group. *J Infect Dis* 165 Suppl 1: S34-35.

West, N., and Tang, C. (2000) Mobile genes. *Microbiology Today* 27: 178-180.

Williams, D.R., and Thomas, C.M. (1992) Active partitioning of bacterial plasmids. *J Gen Microbiol* 138 (Pt 1): 1-16.

Williams, J.D., and Moosdeen, F. (1986) Antibiotic resistance in *Haemophilus influenzae*: epidemiology, mechanisms, and therapeutic possibilities. *Rev Infect Dis* 8 Suppl 5: S555-561.

Wright, C.L., Strugnell, R.A., and Hodgson, A.L. (1997) Characterization of a *Pasteurella multocida* plasmid and its use to express recombinant proteins in *P. multocida*. *Plasmid* 37: 65-79.

Young, R.S., Fortney, K.R., Gelfanova, V., Phillips, C.L., Katz, B.P., Hood, A.F., Latimer, J.L., Munson, R.S., Jr., Hansen, E.J., and Spinola, S.M. (2001) Expression of cytolethal distending toxin and hemolysin is not required for pustule formation by *Haemophilus ducreyi* in human volunteers. *Infect Immun* 69: 1938-1942.

Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C.M., Wilkins, B.M., and Zatyka, M. (2000) Conjugative-DNA Transfer Processes. In *The Horizontal Gene Pool*. Thomas, C.M. (ed). The Netherlands: Harwood Academic Publishers, pp. 87-155.

Zhang, H.L., Malpure, S., and DiGate, R.J. (1995) *Escherichia coli* DNA topoisomerase III is a site-specific DNA binding protein that binds asymmetrically to its cleavage site. *J Biol Chem* 270: 23700-23705.

Zhang, X.L., Tsui, I.S., Yip, C.M., Fung, A.W., Wong, D.K., Dai, X., Yang, Y., Hackett, J., and Morris, C. (2000) *Salmonella enterica* serovar typhi uses Type IVB pili to enter human intestinal epithelial cells. *Infect Immun* 68: 3067-3073.

Zhong, Z., Caspi, R., Mincer, T., Helinski, D., Knauf, V., Boardman, K., Wilkinson, J.E., Shea, T., DeLoughery, C., and Toukdarian, A. (2002) A 50-kb plasmid rich in mobile gene sequences isolated from a marine micrococcus. *Plasmid* 47: 1-9.

APPENDIX A

MEDIA AND ANTIBIOTICS

Composition of Media

Brain heart infusion (BHI) broth

To prepare a litre of the broth, 37 g of brain heart infusion (BDH/Merck, Poole, UK) was added to 1 L of deionised water. The broth was mixed and autoclaved for 20 minutes at 121°C. When the broth has cooled to 55°C, 3.0 ml of stock haemin (5 mg/ml) and 1.5 ml of stock NAD (10 mg/ml) was added and dispensed into Petri-dishes.

Chocolate agar

Chocolate agar was prepared by adding 40 g of blood agar base No. 2 (Oxoid, UK) to 1 L of distilled water, and mixed until completely dissolved. Following autoclaving for 15 minutes at 121°C, the medium was cooled to 55°C before the adding 100 ml of horse blood (previously heated to 75°C).

Hib agar

Hib agar was made by adding 39 g of Columbia base (Oxoid, UK) and 5 g of yeast extract (Oxoid, UK) in 1 L of distilled water. It was autoclaved for 15 minutes at 121°C. After the medium had cooled to 55°C, 1.5 ml of stock NAD (10 mg/ml) and 3.0 ml of stock haemin (5 mg/ml) were added before dispensing into Petri-dishes.

Luria Bertani (LB) broth and agar

To prepare 1 L volume of LB broth, 10 g of typtone, 5 g of yeast extract and 10 g of NaCl were dissolved in 950 ml deionised water. The pH of the solution was adjusted to 7.0 with NaOH and brought up to 1 L volume. It was autoclaved for 20 minutes at 121°C.

For preparation of LB agar, 15 g/L of agar was added to the LB medium before autoclaving for 20 minutes at 121°C. Antibiotic/s was/were added when the agar had cooled to 55°C, and dispensed into Petri-dishes.

Muller Hinton Broth

For the preparation of a litre of the media, 35 g of Muller Hinton dehydrated medium (BioMerieux, USA) and 5 g of yeast extract were suspended in 1 L of deionised distilled water and autoclaved for 20 min at 115°C. When it had cooled to 55°C, 3.0 ml of stock haemin and 1.5 ml of stock NAD (10 mg/ml) were added.

For the preparation of Muller Hinton agar, 1.5% of agar was added to the medium before autoclaving. Haemin and NAD were added when the agar had cooled at 55°C prior to dispensing.

SOC Medium

This medium was prepared by adding 20 g of tryptone, 5 g of yeast extract and 0.5 g of NaCl to 1 L of distilled water, before autoclaving for 20 minutes at 121°C.

Tryptone soy broth (TSB)

To 1L of distilled water, 30 g of a dehydrated tryptone soya broth media (Oxoid, UK) was added and mixed before sterilising by autoclaving at 121°C for 15 minutes.

Additives to culture media

IPTG

For the preparation of a 100 mM stock solution, 238 mg of isopropylthio-β-D-galactosidase (IPTG) was dissolved in 10 ml deionised distilled water. The solution was filter-sterilised and stored in 1 ml aliquots at –20°C. To prepare agar with IPTG, a volume of 40 µl of IPTG stock solution was pipetted onto the surface of a LB plate and smeared with a sterile spreader. The solution was allowed to diffuse into the plate by drying at 37°C for 10-15 minutes.

X-Gal

To prepare a 40 mg/ml X-Gal stock solution, 400 mg 5-Bromo-4-chloro-3-indolyl-β-D-galactosidase (X-Gal) was dissolved in 10 ml dimethylformamide (Sigma). The solution was stored in a brown bottle at –20°C to protect from light. To a previously made LB agar plate, 40 µl of the stock solution was pipetted onto the surface of the agar and evenly spread with a glass spreader and dried for 15 minutes.

Growth supplements

Table 1. Growth supplements for the culture of *H. influenzae*

Compound	Concentration		Role
	Stock (mg/ml)	Final (µg/ml)	
Haemin in DMSO	5	10	Provides V-factor for growth of Haemophilus
Nicotinamide adenine dinucleotide (NAD)	10	10	Supplement for growth of *H. influenzae*

Antibiotics solutions

Table 2. List of antibiotics used

Antibiotic	Concentration	
	Stock (mg/ml)	Final (μg/ml)
Ampicillin	50	50
Chloramphenicol in ethanol	34	20 -50
Rifampicin	10	Various
Streptomycin	100	50
Tetracycline	5	50

Composition of stock solution and buffers

Table 3. List of buffers used

Solution	Composition
Phosphate buffered saline (PBS)	150 mM NaCl, 60 mM $NaPO_2$, pH 7.0
x10 Alkaline phosphatase buffer (MBI Fermentas)	500 mM Tris-HCl (pH7.9), 1 M NaCl, 100 mM $MgCl_2$, 10 mM DTT
x10 Ligation buffer (MBI Fermentas)	500 mM Tris-HCl (pH7.4), 100 mM $MgCl_2$, 10 mM DTT, 10 mM ATP, 250 μg/ml BSA
10x PCR buffer without $MgCl_2$ (MBI Fermentas)	100 mM Tris-HCl, pH 8.3, 500 mM KCl

Table 4. List of chemicals and compositions used

Solution	Composition
Tris-borate EDTA, TBE buffer (10x)	89 mM Tris base, 89 mM boric acid, 2 mM EDTA
Tris EDTA, TE buffer (pH 7.4, 7.6 or 8.0)	10 mM Tris-HCl (pH 7.4, 7.6 or 8.0), 1 mM EDTA (pH8.0)
6x Loading buffer, type III (Sigma, Dorset, England)	Bromophenol Blue 0.25% (w/v), Xylene cyanol FF 0.25% (w/v), sucrose in water 40% (w/v)
2.5x Sequencing buffer	200 mM Tris-HCl, 5 mM MgCl (pH9.0)
Ethidium bromide, EtBr (10 mg/ml in water)	Final concentration 0.5 μg/ml

APPENDIX B

PRIMER SEQUENCES

Table 1. List of primers used for PCR and sequencing.

Primer	Sequence (from 5′ to 3′)	Annealing Temp °C	Length (bases)
M13Fs*	GTA AAA CGA CGG CCA GT	52.8	17
M13Rs	GGA AAC AGC TAT GAC CAT G	55.4	19
M13Fp*	GCC AGG GTT TTC CCA GTC ACG AC	66.0	23
M13Rp	GAG CGG ATA ACA ATT TCA CAC AGG	61.0	24
R1	GGG AGC TTC TTC AGT TAA AG	62.4	20
R1a	AGC ATA ATA AGA GCT TAA GGT TGG	57.6	24
R2	CGT GTA CCT AGC CAA TAC AC	63.0	20
F1	TTA TCG TAA TGT TTG TAT CAG CCT	50.5	24
F2	CGA TCG TAG TGT ATT GGC TAG GTA C	63.0	25
X1	GTC AGC AGT AGT GCT GAT TAG GT	60.6	23
X2	CGC AAA TGT TAG CAC TAG GCG G	62.1	22
X5	GAT ACC ACA ATG ACC ATA AGC CG	60.6	23
X7	TTA TCT TAC TTT ATA TCG CCG C	54.7	22
Y2	CAA ACT TTA GAT GAA GAT GGC GAT G	59.7	25
Y5	TGC ACA ACT ACG GGC TTA CTT ATT	61.3	24
Y7	CAA ACT TTA GAT GAA GAT GGC GAT	60.6	24
Y8	CTC AAC GAA AGT AAC TCA TTG G	58.0	22

* p & s refer to M13 primers for PCR and sequencing, respectively

Table 2. Primers sequences for plasmid sequencing

Primer	Sequence (from 5'to 3')	Annealing Temp °C	Length (bases)
JR1	CCA AGA CCA TCT CAT TGA TGA A	57.4	22
JR10	CGT TGT AAT GAT GAC CAA CCG	58.7	21
JR11	CCT GAT GGT GCA ATT CGA	55.1	18
JR12	TGC ATT CCC TCA AAG AGA GGT	58.8	21
JR13	CGT TGG TAT CCC TAG CCA AG	57.5	20
JR14	CTA AAC CAT GAA CGG ACA GTC	54.6	21
JR15	AAC CCT AGC ACA CTG GTT AAA	54.6	21
JR16	TTG GCT TCA AAT TAC CGC CT	60.1	20
JR17	ACT AGC ATT CCA GCT TTC CC	56.4	20
JR18	AGT TGG TTG CAG CTG TAT GTC T	56.9	22
JR19	ATC TGC AGG AAT TCG ATA TCA A	56.9	22
JR2	CGT AGT ATT CTT TCG CTT GCT GA	59.8	23
JR20	CTT ATC ATC ATT TTG GGC ACT G	57.7	22
JR21	TTT ACT GCG CTC TGA TGA TTT	55.7	21
JR22	GAT AGT GAT GGC ACG ACT AAG C	57.0	22
JR23	CAT CGC ATA CAG AGC GTA GTT G	58.9	22
JR24	ACA GCG GTA AGA TCC TTG AGA G	58.1	22
JR25	CCC GTC GTG TAG ATA ACT ACG A	57.4	22
JR26	GGC CTC AGA TGC TGG TAA TAA T	58.1	22
JR27	TGC TTG AAT ACG AGT ATT CCC A	57.9	22
JR28	TGG GAA TAC TCG TAT TCA AGC A	57.9	22

Primer	Sequence (from 5'to 3')	Annealing Temp °C	Length (bases)
JR29	GAT CGG CAT TAA TAG AGG ACG T	58.1	22
JR3	GAT AAG GGG ATC AAT TTC CCA C	58.4	22
JR30	ACG TCC TCT ATT AAT GCC GAT C	58.1	22
JR31	AAC GGG CCA TTA CGC TGC TTA C	65.6	22
JR32	CCC TTG GAA AGG TTA CCT TTA T	56.8	22
JR33	CGA AGT ATT GCG GGG AGA TAT T	60.4	22
JR34	CGT AGA GGG AAA GAA ATG ACA G	56.2	22
JR35	CAG TTT CAC AAA GTG TTA GCC A	55.9	22
JR36	GCC CGA AGT AGC CTG GGA AAA T	65.8	22
JR37	AGT TCC AAT AAG GAA CTC AAT G	53.4	22
JR38	ATC ACG TAG AAG ACA ATT GGT T	53.5	22
JR39	CAG AGG ATG AAG AGT TTG ATG A	54.3	22
JR4	CTT GAT ATC GAA TTC CTG CAG C	59.5	22
JR40	TTT GAT GTG ATC ATA CTC ATG C	53.4	22
JR41	CAA ATA CCC AAT CAA CTG AGA C	54.1	22
JR42	CAA CTT AAC GTA ACG CAC TAG G	55.5	22
JR43	CGG TGT ACC ACA GAT GTG TTA T	55.0	22
JR44	TGT TGG TCC AAG AAA AAC GTC	55.9	21
JR45	CAT AAG CGT AAA TAG AGA TAA AAT	52.5	24
JR46	ATT TGC AAG TAA TTC TGA CAA CA	53.5	23
JR5	GAC TGT AGC AAG GGA TTC AGC T	58.1	22
JR6	CAG CTG AGA AAG AAC GCA TTT T	59.4	22
JR7	CAG CGT AAC TCT CGC TTT AGG	57.7	21
JR8	GAG ATT AAC CAT TCT CCA GGC	55.7	21
JR9	TAC TGG CGC TTG TAC TAA AGG C	59.7	22
JRF1	TTA TCG TAA TGT TTG TAT CAG CCT	56.3	24
M13Fp	GCC AGG GTT TTC CCA GTC ACG AC	66.0	23
M13Fs	GTA AAA CGA CGG CCA GT	52.8	17
M13Rp	GAG CGG ATA ACA ATT TCA CAC AGG	61.0	24
M13Rs	GGA AAC AGC TAT GAC CAT G	54.5	19
P038F	TCA GAT ATT GAA GCT CGT ATT GGA AG	60.1	26
P059F	CGA CGT GCT AAA TCA CGC TCT	59.8	21
P111F	GTG GGA TCG CTT TAC CTG CA	59.4	20
P112F	ATG CGG TTG CTA TCA TCC AAG TG	60.6	23
P177R	CTA AAT ACA GCA TCC AAA GCG CAC	61.0	24
P290F	GTC AAA GTA TGC AAG CCC ATT ATC	59.3	24
P302R	CAT CAG AAG GGC ACT GG	55.2	17
P310R	GCT TTC TGA TAG TCT AAC GAG AGG	61.0	24
P336F	GAT AAT ATT CGT AAG CAG CCA G	56.5	22
P370R	CCA AAA GAG CTG GCT GAA GG	59.4	20
PBCR	GCG AGC TCG CAT CCA GTC	60.5	18
PC038F	AAA CGG GAC AAA CAA CAG GT	55.3	20
PC058R	TCA ATG GTT CAA GCC TAC CC	57.3	20
PC182F	CCT TCT GTA GCT AGA CTT TGA AAC	59.3	24
PC182R	CCC TTC TTG TTG AGC TTT CGC	59.8	21
PC200F	CGA TTC CGA CCT CAT TAA GCA G	60.3	22
PC200R	TCA ACA TCT CAA TGG CTA AGG CG	60.6	23
PC200RA	TCT AAC ATC TCA ATG GCT AAG GCG	64.0	24
PC200RB	CAG AAG CTA GGT GTA GAG C	56.7	19
PC217F	CTG TAC GTA ATG AGC AAG GTC	57.9	21
PC240F	ATG TTT TAT GAC CGA AGG GAC	55.9	21

Primer	Sequence (from 5'to 3')	Annealing Temp °C	Length (bases)
PC240FA	TTA TCG TAT CCT CTA TCA ACT CTG C	59.7	25
PC240R	GCA GTT ACA TTA TTT GCA CCG	55.9	21
PC251F	TGA TGA AGC CGG ACG ATA TG	57.3	20
PC260F	CTC GTT TTG CTG AAG TGC TTG	57.9	21
PC260R	CAC GTT TTG CTT CTG CAT ACA	55.9	21
PC294F	TCA TCT AGT GCG GAC AAC	53.7	18
PC352F	CAC TTC AGA TGA TGT GTG ATA AC	57.1	23
PC472F	CAA TGG CTG GTT TAT GCA TAT CGC	61.0	24
PC472R	CCA TGG CTG GTT TAT GCA TAT CGC	64.0	24
PC472RC	CGT GCT GAG TTA ACC AGT GAT TG	60.6	23
PC532F	CGA ATG GTG AGG CAG TTG C	58.8	19
PC532R	CTC AAG AAG GGA GCT CAA GTC TAT	61.0	24
PC543R	GAG TCG TGT AAA ATA TCG AGT TCG	59.3	24
PC558RF	AAT CGA ATT CGA CCA TCA GG	55.3	20
PC702F	AGA AGA TCC CGA TGC TCA AGT G	60.3	22
PC702R	CAC AAT AGT AAC AGT ATT GAA GAG CC	60.1	26
PC717F	GTC GTA ATA ATG GCG GCA TAC	57.9	21
PC717R	CTT AAC GGC AAC TTC GCC AG	59.4	20
PC802F	GTC ACG TCG AAG CAG AAT AC	57.3	20
PC802FA	TAG ACT ACC CAT GCA GTA TAG	55.9	21
PC804R	ACC TTG AAC ACG CCA ATG	53.7	18
PC804RB	TGC CAT GGC ATC ACA CCT TG	59.4	20
PC821R(FC)	GCT CGT TGA ATA ACC ATT GCA GG	60.6	23
PC847RR	AAA TTC CTG TGC ACC TTG GT	55.3	20
PE109R	GAT TTA CTG GGA TCA GAC CGA T	58.4	22
PE15R	CGG CAG TTC AAT ACC ATT TGC	57.9	21
PE21F	CAG ATA TTG AAG CTC GTA TTG G	56.5	22
PE33R	CCG TCC AAA GTG AGA TGC C	58.8	19
PE43R	GCC GAT AAT AAC TTG GTG TAA C	56.5	22
PE63R	GGA GTT CCG GAT GGC AAT G	58.8	19
PE73F	GAC CTG AAG ATC AAG AGA CTG	57.9	21
PE75F	CAG ACC GCT ACC GCC AG	64.4	17
PF132F	GGT GGC AAT CTG CAT TTT CTG	57.9	21
PF33FA	GAA TTG CTA TCG GTG CAT TGC	57.9	21
PF45R	GAA TAG CCG AGC GGT CAG	58.2	18
PF60F	TTG ACC GCC TCT CTC CTA TG	59.4	20
PF74RA	TAA AGG TTT CCC TTG CTG TAG	55.9	21
PF7R	TCG ACC GCC AAC ATA GC	55.2	17
PF85F	CTT TGT GTG GTG ATT GAT AAT CTC	57.6	24
PF90R	GAT AAT AGA CCC CAG GAG TC	57.3	20
PG004F	CCT GGT GTC GGG TGA TAG G	61.0	19
PG013R	GTA ACA CTT CTG TAC GTT CAG G	58.4	22
PG154F	CAT TTC CTC AGC GTG TAT TTA C	56.5	22
PG164R	GTC GTA AAT GGC GCA TAG CC	59.4	20
PG23F	AGA TGG GTT ATT GAC GGT GG	57.3	20
PG23R	TGA GTA GTT CAT TTA CCA GCA TG	57.1	23
PG23RA	CAT TTG ATT TTG GTT ATT AGA TGA GT	55.3	26
PG25F	CGA GCT GTT TGG CAA TGC G	62.3	19
PG26F	CAG CGT ATA TTG AAC ATT G	50.2	19
PG273F	CAA CAG AAT CAG AGA ATG AAT ATA G	56.4	25

Primer	Sequence (from 5'to 3')	Annealing Temp °C	Length (bases)
PG284R	CTC CCC ATC ATT TTA TCT ATT AG	55.3	23
PG289F	AAT ATT GTC TAA TCA CTG CTG CC	57.1	23
PG296R	CAT TAC CCT TAA TCC CGT TAG G	58.4	22
PG305F	GCT GAT CAT AAA GAT AAA GCT GG	57.1	23
PG309R	TTT AGG TGG GCA CCA GCC	58.2	18
PG33F	GGC ATT GAT ACC ATT ACC AC	55.3	20
PG33FA	GAA TTG CTA TCG GTG CAT TGC	57.9	21
PG33FB	CAC CAA TGA CGG CAT CTG C	58.8	19
PG33FC	CGA ATT GCT ATC GGT GCA TTG	57.9	21
PG33R	TGG AGA ATG GCG AGT AAG	53.7	18
PG33RA	GAT AAC CCG CCA TAC ACA CT	57.3	20
PG33RB	GCA GAT GCC GTC ATT GGT G	58.8	19
PG392F	GAG GTG TAG CAA ACC ATA TTG TA	57.1	23
PG395R	TTG ATT TTA CTG AGT TCC ATT TTG G	56.4	25
PG41F	GAT GAC TCT GCT CTA CCA TTA C	58.4	22
PG421F	TGA AAT TTG CGG CCA ATA TCT C	56.5	22
PG428R	TTG GGG CAA AGG AGA ATG AGG	59.8	21
PG42R	GTT GCA GGC AGT GGA TAC G	58.8	19
PG48F	GAC ATT AGC CCC TCT TAT CCC	62.6	21
PG50R	GAA GCC ATA ACC CTG TTC C	56.7	19
PG57R	GGT GTA AAG GAG TTT CCT ATT TTG	57.6	24
PG59F	CAC TTA TAA CAT CCT CAA TCA ATG	55.9	24
PG5F	ACA ACA AGA TGT GCG AAC TCG	57.9	21
PG5FA	GGT TGG CAT TCC CAA TAA CTG	57.9	21
PG5R	GAC AAC AAG ATG GGC GAA CTC G	62.1	22
PG62F	CTC CGG CCA AAA TTG GAC TC	59.4	20
PG64R	CTG GTT AAG AAG CTC GAC CG	59.4	20
PG76F	TGG CGC TAT AAG TGC TCA AG	57.3	20
PG81F	GCA GCG TAT ATT GAA CAT TGG	55.9	21
PG81R	CTG CAT TGC CTG TGC GA	55.2	17
PG81RA	ATT ATG CTG AAA TGT TGC TGC	54.0	21
PG81RB	TCG CAC AGG CAA TGC AG	55.2	17
PH104F	GAC GGT CAG CTT ACA GTT AC	57.3	20
PH113R	GTT TAG CGC AGA TAG TAC ACG	60.6	21
PH117F	ATG GCG TTA GAT TTA GGT GAT G	56.5	22
PH118F	GCT GAT CGT GTA CTG CAA TAT G	58.4	22
PH123R	GAG GAA TCA GTT AAG TGC GG	57.3	20
PH139R	GTT TTT TAT CTC ACG GTC ACG	58.7	21
PH143F	CAA GGT CGA TTC TAT GCA GC	57.3	20
PH149R	GAC CAG CCA TTG ATA CCA AAT C	58.4	22
PH209F	CTG CCG ATA TGA TCC AAC TG	57.3	20
PH209F	GAT TTG CTG AGG ATG AAG GTG	60.6	21
PH223F	GAT TTA GCT CTG TAT CAT CAT TG	55.3	23
PH223R	CAA TGA TGA TAC AGA GCT AAA TC	55.3	23
PH226R	ATC AGC TGG CAT TTC TTG ACT G	58.4	22
PH237R	GCA GAT GCC GTC ATT GGT G	58.8	19
PH247F	CAT CCT ATT CTG CCC TAT TCT TC	58.9	23
PH24R	CAC TCA GAC GAA CTA CAG AC	57.3	20
PH250R	CTT CGG CCA GCT TAT AAT GAG	57.9	21
PH269R	TTG TTG CTC ATC GGC TGT TTG	57.9	21

Primer	Sequence (from 5'to 3')	Annealing Temp °C	Length (bases)
PH286R	ATG TGG ATC AGA TAA CCA GCC	57.9	21
PH287F	GGA TTA GCA ATA CTT AGC TGA C	58.7	22
PH293R	CCC TTA TGA ACA CAA CCC TC	57.3	20
PH30F	GCA GAC TGG TTC CGT CAG	62.2	18
PH334AF	GAA GCA TGC CGA ATG GTG AAG	59.8	21
PH334F	GCG CAA GTT TAT ATT GAA GGA C	56.5	22
PH342R	CCT CTG TTT TAT CAT CTG CTT C	58.9	22
PH346R	GAG CTT GCA AGC CAG TGT G	58.2	19
PH346RA	CAC TAT CAG TGT TCG CCC AG	62.5	20
PH348R	CAA GGA TCT TAC CGC TGT TG	60.4	20
PH354F	GGC ATA CAG CAA CAA ATC	60.6	18
PH364R	GAG TAT TGA CTA CGC TAG GTT G	58.4	22
PH364RA	GAA ACG GTT GGT CAT CAT TAC	58.7	21
PH377R	CCT ATT CAA CCA TTG ACA CCG	57.9	21
PH377RA	CGC TGT TAT CTC TGC TGT TTG	60.6	21
PH384R	GCC GCA TAG CTT GAT TTA TCC	57.9	21
PH385F	CTG AGT ACC AAT ATA TAC ACC TC	58.7	22
PH410R	GAT TCC CGT CAT CAT CTA CTG	57.9	21
PH416F	TAT TGT ATA GAT GAT GAA TCT GAG C	56.4	25
PH421F	CTC TGT GAG GTC AAG ATG TC	57.3	20
PH428F	CAC GGG AGG AAG CAT TAA GTG	62.6	21
PH42F	CAT CAC TGA AGC AGA TCG TAG	60.6	21
PH440F	CAT CAT AAG CAT AGA TTG GTT TAG	57.7	24
PH44F	GAT AAC AAC AAG ATA ACG CAA GC	57.1	23
PH46R	CAA ACA AAG TCC ATC AGG ATC AC	58.9	23
PH487R	CTG TTG TAG TTG TTG TCG TTG	58.7	21
PJR24R	GGA TAA TAC CGC ACC ACA TAG	57.9	21
PJR26R	GAG AGG ATG TGA TAT TAA GGG	55.9	21
PJR39R	GCC AGG AGT AAT GCC AAA TTG	57.9	21
PP112F	GGA TTC ATC AGC ACT TTG ATT G	56.5	22
PP310F	GTT AGC GGA CCC TCG TG	57.6	17
PP370R	GCA CCC ATA ATG ACA ACA TAG A	57.9	22
PPC702F	TCT GAA CCA TAT CCA GTG CG	60.4	20
PPF7F	TCA GCC AAT GCC ACA ACA AG	57.3	20
PPF7R	CTT GTT GTG GCA TTG GCT GA	57.3	20
PPG164F	GGC GTG ACG TAA GCT AAA CAA	57.3	21
PPG284R	CAA TAA CCC ATC TAC TCG GC	60.4	20
PPG421R	GTA CAT AAC ATT TGC ATT AGA TCA	54.2	24

APPENDIX C

PAIRWISE ALIGNMENT OF GENES IN *HAEMOPHILUS* SPP. ELEMENTS

Table 1. Pairwise distance of 12 genes in *Haemophilus* spp. calculated using MEGA.

soj	ICE*Hin1056*	ICE*Hin299*	ICE*Hpa8F*
ICE*Hin1056*			
ICE*Hin299*	0.006		
ICE*Hpa8F*	0.008	0.002	
ICE*Hin2866*	0.006	0.000	0.002
dnaB			
ICE*Hin1056*			
ICE*Hin299*	0.010		
ICE*Hpa8F*	0.010	0.001	
ICE*Hin2866*	0.010	0.000	0.001
ssb			
ICE*Hin1056*			
ICE*Hin299*	0.024		
ICE*Hpa8F*	0.024	0.000	
ICE*Hin2866*	0.024	0.000	0.000
topB			
ICE*Hin1056*			
ICE*Hin299*	0.020		
ICE*Hpa8F*	0.016	0.008	
ICE*Hin2866*	0.000	0.000	0.008

pilL	ICE*Hin1056*	ICE*Hin299*	ICE*Hpa8F*
ICE*Hin1056*			
ICE*Hin299*	0.010		
ICE*Hpa8F*	0.010	0.010	
ICE*Hin2866*	0.010	0.000	0.010
traD			
ICE*Hin1056*			
ICE*Hin299*	0.017		
ICE*Hpa8F*	0.034	0.033	
ICE*Hin2866*	0.017	0.000	0.032
Gene 48			
ICE*Hin1056*			
ICE*Hin299*	0.002		
ICE*Hpa8F*	0.000	0.002	
ICE*Hin2866*	0.000	0.002	0.000
Gene 52			
ICE*Hin1056*			
ICE*Hin299*	0.001		
ICE*Hpa8F*	0.005	0.004	
ICE*Hin2866*	0.002	0.001	0.005

virB4	ICE*Hin1056*	ICE*Hin299*	ICE*Hpa8F*
ICE*Hin1056*			
ICE*Hin299*	0.006		
ICE*Hpa8F*	0.006	0.001	
ICE*Hin2866*	0.006	0.000	0.001
traC			
ICE*Hin1056*			
ICE*Hin299*	0.096		
ICE*Hpa8F*	0.099	0.013	
ICEHin2866	0.099	0.013	0.000
traI			
ICE*Hin1056*			
ICE*Hin299*	0.227		
ICE*Hpa8F*	0.250	0.037	
ICEHin2866	0.001	0.228	0.251
int			
ICE*Hin1056*			
ICE*Hin299*	0.345		
ICE*Hpa8F*	0.340	0.011	
ICEHin2866	0.001	0.343	0.338

Table 2. Pairwise distance of ampicillin genes in *Haemophilus* spp. calculated using MEGA

tnpA	ICE*Hin1056*	ICE*Hin299*	ICE*Hpa8F*	ICE*Hin* 2866	*E.coli* V00613
ICE*Hin299*					
ICE*Hpa8F*	*nd	0.000			
ICEHin2866	nd	0.000	0.000		
E.coli V00613	nd	0.022	0.022	0.022	
E. coli pIGAL1	nd	0.022	0.022	0.022	0.000
tnpR					
ICE*Hin1056*					
ICE*Hin299*	0.002				
ICE*Hpa8F*	0.002	0.000			
ICE*Hin2866*	0.002	0.000	0.000		
E.coli V00613	0.056	0.058	0.058	0.058	
E. coli pIGAL1	0.056	0.058	0.058	0.058	0.000
bla					
ICE*Hin1056*					
ICE*Hin299*	0				
ICE*Hpa8F*	0	0.000			
ICE*Hin2866*	0	0.000	0.000		
E.coli V00613	0.003	0.003	0.003	0.003	
E. coli pIGAL1	0.003	0.003	0.003	0.003	0.000

*nd: not done

APPENDIX D

GENOTYPIC CHARACTERISTICS OF GIs

Table 1. Detailed functions of accessory genes in GIs of various bacteria

Pathogens/strain	Island	Attributes of accessory genes		
		Region 1	Region 2	Region 3
H. influenzae 1056 *H. somnus* 129PT	ICE*Hin1056*	*YbeB, YbeA, YbdA, tetA, tetC, tetD, tnp* Antirestriction protein *ardC*, ATP binding protein for virulence, histone acetyltransferase	*bla, tnpR*	
H. somnus 2336		Integrase/recombinase, NADPH quinone reductase, transcriptional regulator, Co/Zn/Cd efflux system component, *romA*, quartenary ammonium compound resistance protein, tetracycline resistance efflux protein, tetracycline resistance repressor protein, exporter protein, haloacid dehydrogenase, putative periplasmid oxidoreductase, putative metal-binding protein, aspartyl-tRNA synthetase	*ngrC* phosphopantheinyl transferase, type I restriction-modification system	
H. ducreyi 35000HP		Antirestriction protein, type I restriction enzyme M subunit, *HicB* protein, transposon gamma-delta resolvase, *IstB* homologue, *cdt* protein		
Pseudomonas sp. B13	*clc*	3'-phosphoadenine 5'-phosphosulfate sulfotransferase (PAPs reductase)/FAD synthetase and related enzymes	Pyruvate 2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogene component, acyl-coA synthetases, aminophenol repressor, ferredoxin-like protein	Chlorocatechol 1,2, dioxygenase, threonine efflux protein, *clcR, clcA*, chloromuconate cycloisomerase, maleyacetate reductase, aromatic dioxygenase, oxidoreductase

Pathogens	Island	Attributes of accessory genes		
		Region 1	Region 2	Region 3
P. luminescence TT01		Type IV secretion: *pilN, pilO, pilP, pilQ, pilR, pilS, pilU, pilV*, transposase	Sugar fermentation, lactylglutathione, aminotransferase, L-mandelate and L-lactate dehydrogenase, *RetA* protein, peptide synthetase, transposase, peptide synthetase	
P. aeruginosa PA14	PAP1	Putative phage protein, pyocin, Type IV secretion: *pilN, pilO, pilQ, pilV2*, 70-kb heat shock protein	*FimA*, *fimD*, kinase sensor *RcsC*, putative response regulator *PvrR*	ATP-dependent endonuclease of the OLD family, pathogen-related protein, deoxycytidine triphosphate deaminase *dcd*
P. aeruginosa C	PKLC102	Type IV secretion: *pill, pilN, pilO, pilP, pilQ, pilR, pilS, pilU, pilV and pilM*	Virulence protein	
P. aeruginosa SG17M	PAG1-3	*LysR, oriT* homologue, transposase, *mloA* type I restriction modification, DNA repair protein	Cyclopropane-fatty-acyl-phospholipid synthesis, yddR, transposase, putative adenosylmethionine, putative alcohol dehydrogenase, putative glutamine amidtransferase, uroporphyrinogen decarboxylase	
P. fluorescens PfO-1		*RcsC* kinase, Vic response regulator, *RcsB* kinase, *fimA*, pilin chaperone, *fimD* methyltransferase, epoxide hydrolase, oxidoreductase		
P. fluorescens SBW25		*Rpfn* protein, carbohydrate selective porin, permease of major facilitator superfamily/multidrug-efflux transporter, histidinol dehydrogenase, nodulation protein, glyoxalase family protein, hydrolases, succinate-semi-aldehyde dehydrogenase, ornithine cyclodeaminase	Zn-dependent peptidase	

Pathogens	Island	Attributes of accessory genes		
		Region 1	Region 2	Region 3
S. Typhi CT18	SPI-7	Type IV secretion: *pilN, pilO, pilQ, pilV*	*IbrA*, DNA-binding prophage protein, type I restriction-modification, UV protection, phage-related protein, Vi biosynthesis, shuflon-specific DNA recombinase	
R. metallidurans CH34		Cation transport ATPase, *tnpA* transposase, plasmid-related protein, superfamily II DNA/RNA helicases	Protein disulphide isomerase, putative DNA repair protein *radC*	Putative pathogenesis-related protein
X. axonopodis 306		Cytosine-specific DNA methyltransferase, phage related protein, permease of drug/metabolite transporter	Plasmid-related protein, oxidoreductase	Haemolysin, haemolysin-type calcium binding protein
Y. enterocolitica 8081		Type IV secretion: *pilN, pilO, pilQ, pilR, pilS, pilV2*	Transposase, arsenical pump, arsenical pump-driving ATPase, arsenical resistance operon repressor	Putative DNA-binding prophage protein, putative haemagglutinin-related protein

APPENDIX E

PHYLOGENETIC RELATIONSHIP BETWEEN GENES IN 12 GIs

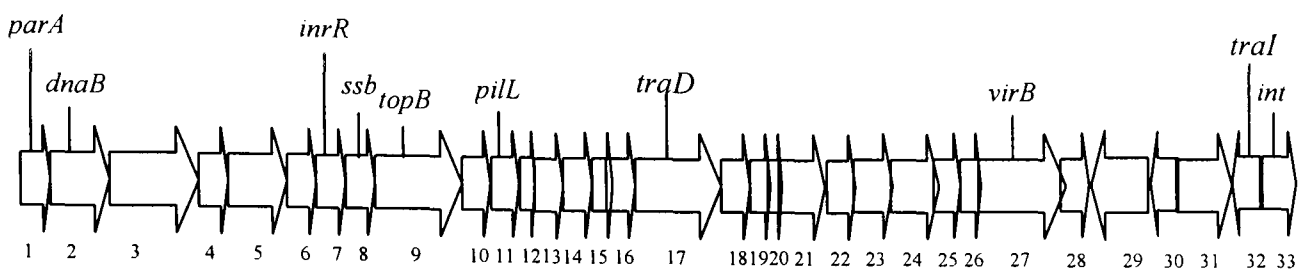


Fig. 1. Virtual core genes

A SplitsTree (Fig. 2 - 5) representation of the ClustalX analysis of each of the 4 genes: 1, 6, 18 and 27 (numbered according to the coherent virtual element), and common to all the GIs. The amino acid sequences of each of the predicted genes alone were aligned using ClustalX. There are only minor differences in the topology of the SplitTrees between each of the genes or the concatenated sequences of the 15 shared genes indicating congruent structure.

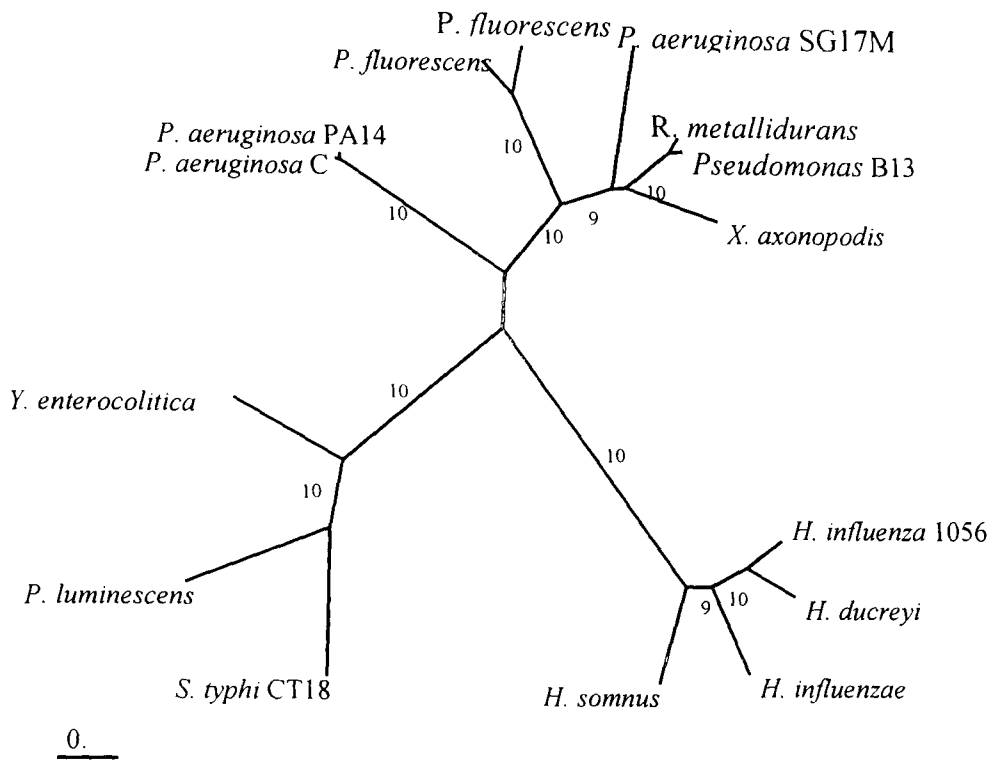


Fig. 2. Gene 1(*soj/parA*)

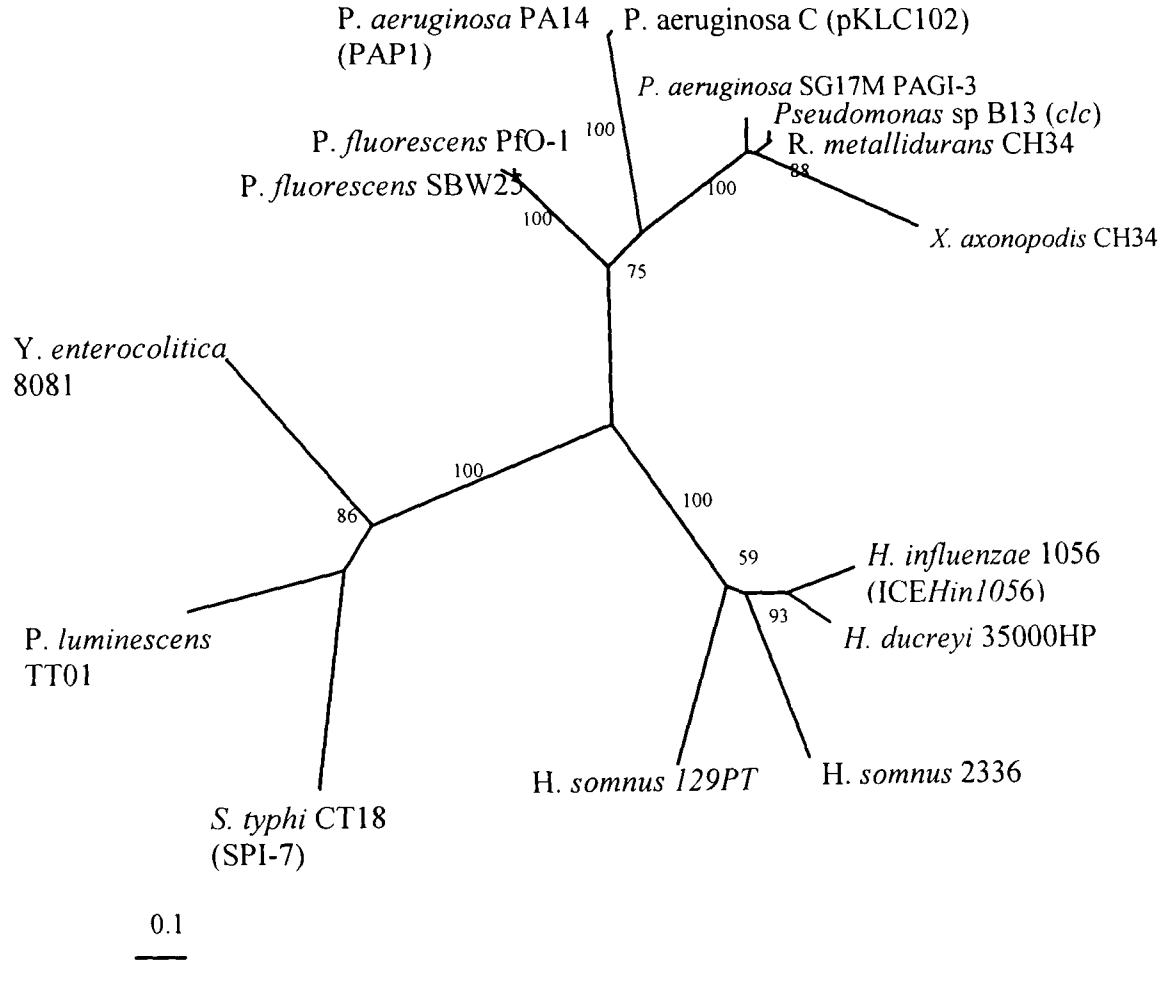


Fig. 3. Gene 6

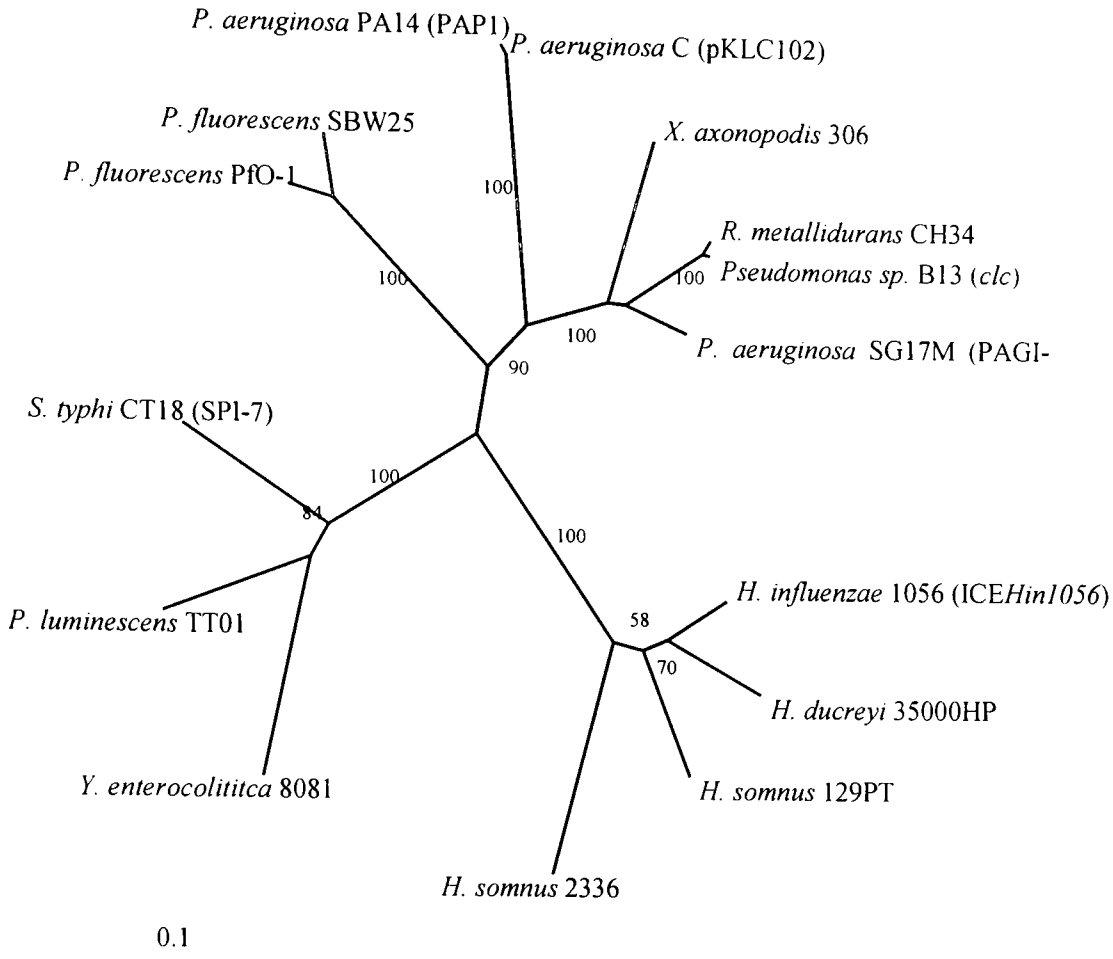


Fig. 4. Gene 18

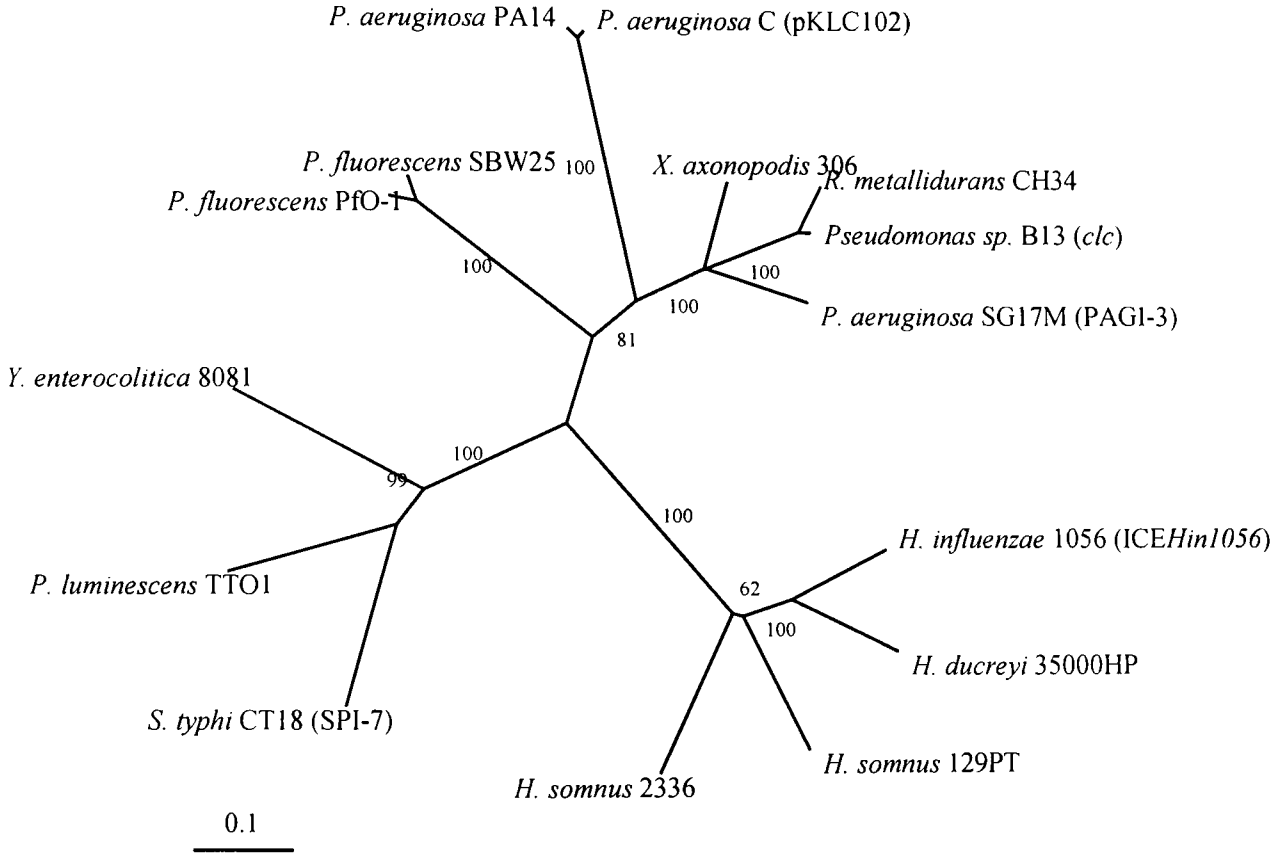


Fig. 5. Gene 27