

Investigation of the influence of phase variable restriction modification systems and lipooligosaccharide epitopes on resistance of *Haemophilus influenzae* to infection by bacteriophage

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

by

Christopher Jason Richard Turkington BSc(hons), MSc

Department of Genetics University of Leicester Leicester United Kingdom

September 2016

Abstract

Investigation of the influence of phase variable restriction modification systems and lipooligosaccharide epitopes on resistance of *Haemophilus influenzae* to infection by bacteriophage

Christopher Jason Richard Turkington

The evolution of ON/OFF switching phase variable loci is presumed to have arisen due to the need for bacterial populations to cope with uncertain environments where selection fluctuates between opposing pressures. One such selection is believed to be the presence of bacteriophage. Bacteriophage can influence bacterial evolution by forcing the development of bacteriophage resistance mechanisms within bacterial populations. However, resistance mechanisms can often come at a cost, such as reduced survival against the immune responses. As such, phase variation may circumnavigate this cost by generating heterogeneous populations containing both resistant and sensitive phenotypes.

This study aimed to investigate the two known phase variable bacteriophage resistance genes in Haemophilus influenzae hsdM and lic2A, which encodes the methyltransferase of a type I restriction modification system and a lipooligosacharide biosynthesis associated glycosyltransferase respectively.

Analysis of the diversity of repeat tract lengths and ON/OFF states of these genes across H. influenzae genomes within GenBank revealed that while lic2A was ON in the majority of strains, hsdM was OFF. Thus lic2A may be consistently beneficial while the hsdM benefit is transient. Analysis of samples from patients with COPD, showed that lic2A was ON in the majority of samples, while hsdM may also be ON state in a number of samples.

Although a resistance phenotype could be observed for lic2A, no resistance could be observed for hsdM. Therefore, the dynamics of bacteriophage spread through populations heterogeneous for lic2A was investigated. The heterogeneous populations generated by phase variation reduced bacteriophage dispersal through bacterial populations. This mechanism may also allow bacterial populations to adjust their heterogeneity levels to control bacteriophage densities. The heterogeneity may further create diverse bacteriophage densities across the bacterial macropopulation through resistant populations acting as a barrier.

The results demonstrate the potential for phase variation to aid in the survival of bacterial populations against bacteriophage predation.

Acknowledgements

Blood... Check... Sweat... Check... Tears... *puts on manly face*... No... and... It's done!!! Four years of hard work... Well three years plus one partially wasted one... its been fun!

Anyway, all this wouldn't have been possible without the well wishes and support of friends and family. My mum, Jean, for being the best mum a son could ask for, to me you still know everything. To Douglas, Scott, and all the Rodgers lot. Hannah for making sure I somehow manage to stumble my way through everyday life. Jazz, Tolis, and Anisha for the good times and lots of laughs, I miss having the... nope not going to say three musketeers like the rest of you guys thesis'... damn I suppose I just did... d'oh! (last ones just for you Jazz). Likewise Neda, although you did all that while also working... and making sure I survived the PhD! I'd also like to thanks all the members of the Chris and Marthas labs past and present, Didier, Depesh, Jack, Alex, Jinyu, Peter, Kate, Janet, Ananthi, Ahmed, Guillermo, Ali, Mohammed, Aisha, Thekra, Wafaa, Lamiaa, and the many others that have came and went.

I'd like to thank all the people who have helped me during this work, Dr Andrey Morozov for his role in the mathematical modelling of the experimental data; Dr Mathew Diggle for providing clinical samples that helped me get to grips with *Haemophilus*; Dr Andrew Millard, for sequencing the main strains used in this work; Dr Jason Johnston for allowing me to use the pJJ260 plasmid, and Prof. Michael Apicella for sending this plasmid.

To each and every one of you I say thank you.

I would also like to thank in particular my supervisors Dr Christopher Bayliss and Prof. Martha Clokie for their guidance, support... and for generally putting up with me. Your patience, encouragement, and advice has helped me in more ways than you will ever know. So again thank you.

I would also like to dedicate this thesis to my father Douglas Turkington and my grandmother Jeanie Rodgers. You may be gone but you will never be forgotten.

Contents

1	Intr	oducti	ion	1
	1.1	Haem	ophilus influenzae	1
	1.2	Bacter	riophage	4
		1.2.1	Bacteriophage life cycles and their implications for the evolution of their hosts,	
			and themselves	6
			1.2.1.1 Lytic cycle	6
			1.2.1.2 Lysogenic cycle	11
		1.2.2	The bacteriophage of <i>Haemophilus influenzae</i>	14
	1.3	Bacter	rial resistance mechanisms to bacteriophage	
		infecti	on, and reciprocal bacteriophage adaptations	16
		1.3.1	Restriction modification systems	17
		1.3.2	Surface structures	24
		1.3.3	Other Resistance Mechanisms	27
	1.4	Phase	variation	29
		1.4.1	Mechanisms of phase variation, and its role in survival	30
		1.4.2	Phase variable bacteriophage resistance mechanisms	34
		1.4.3	Phase variation in <i>Haemophilus influenzae</i> : niche adaptation and bacteriophage	
			resistance	40
	1.5	Aims	of Project	50
2	Mat	terials	and Methods	52
	2.1	Bacter	rial strains, bacteriophage, and growth conditions	52
	2.2	Sputu	m supernatants from patients with COPD	54
	2.3	DNA	analysis methods	54
		2.3.1	Polymerase Chain Reaction (PCR)	54
		2.3.2	Restriction endonuclease digestions and ligation reactions	55
		2.3.3	Agarose gel electrophoresis	55
		2.3.4	Purification of DNA fragments	55

		2.3.5	Genescan analysis	56
		2.3.6	DNA sequencing	57
		2.3.7	Splicing by overlap extension-PCR	57
		2.3.8	Ligation reactions	59
		2.3.9	Transformation of <i>E. coli</i> DH5- α	60
			2.3.9.1 Generation of competent <i>E. coli</i> DH5- α cells	60
		2.3.10	Transformation of <i>E. coli</i> DH5- α by heat-shock	61
		2.3.11	Plasmid extraction	61
		2.3.12	Transformation of <i>H. influenzae</i>	61
			2.3.12.1 Generation of competent cells of <i>H. influenzae</i>	61
			2.3.12.2 Transformation of plasmid DNA into <i>H. influenzae</i>	62
		2.3.13	Bioinformatic analysis	62
	2.4	Experi	imental Methods using Bacteriophage	64
		2.4.1	Small Drop Plating Plaque Assay and Double Agar Overlay Plaque Assay Pro-	
			cedures	64
		2.4.2	Bacteriophage HP1c1 One-Step Growth Curve	65
		2.4.3	Bacteriophage Adsorption Assay	65
		2.4.4	Determination of Phase Variable Resistance to Bacteriophage by Haemophilus	
			influenzae	66
		2.4.5	Effect of Various Multiplicities of Infection (MOI) of Bacteriophage HP1c1 on	
			H. influenzae growth	66
		2.4.6	Oscillating prey assay	67
		2.4.7	Multi-directional oscillating prey assay methodology	69
3	Ana	dysis o	of the Phase Variable Type I Restriction	
	Mo	dificati	on System of Haemophilus influenzae	73
	3.1	Abstra	act	73
	3.2	Introd	uction	74
	3.3	Presen	ace of Phase variable hsdM in H. influenzae	75
	3.4	Distrib	oution of phase variable repeats within <i>H. influenzae</i> strains	76
	3.5	Analys	sis of the putative coding states of the Phase Variable $hsdM$ gene of H . influenzae	78
	3.6	Sequer	nce analysis of the $hsdM$ gene of	
		H. infl	luenzae RM118	81
	3.7	Determ	nination of Bacteriophage Resistance Phenotype of RM118 Following Bacterio-	
		phage	HP1c1 Challenge	84
	3.8	Constr	ruction of $RM118_{GAGAC(3x)}$	84
		3.8.1	Generation of $\rm RM118_{GAGAC(3x)}$	85

		3.8.1.1	Construction of the cassette containing plasmid pCT1, and transfor-
			mation into H. influenzae RM118 r ⁻ HinDI m ⁻ HinDI
		3.8.1.2	Construction of the 3x 5'-GAGAC-3' containing plasmid pCT2 and
			transformation into <i>H. influenzae</i> RM118 _{cassette} $\dots \dots \dots$
	3.9	Analysis of SSR	a tract lengths within the HindI-like $hsdM$ gene of 82 COPD samples $\ . \ . \ 90$
		3.9.1 GenBan	k $hsdM$ sequences and disease $\dots \dots 95$
		3.9.2 Analysis	of $hsdS$ gene of the 121 H . influenzae sequences available on GenBank . 98
	3.10	Discussion	
4	Ana	lysis of the ph	ase variable bacteriophage receptor of
	Hae	emophilus influ	<i>enzae</i> and its role in bacteriophage
	resi	stance	115
	4.1	Abstract	
	4.2	Introduction	
	4.3	Presence of pha	se variable <i>lic2A</i> in <i>H. influenzae</i>
		4.3.1 Distribu	tion of phase variable repeats within $lic2A$ gene of
		H. influ	enzae
	4.4	Analysis of the	putative coding states of the phase variable $lic2A$ gene of $H\!\!\!.$ influenzae . 119
	4.5	Association bet	ween $lic2A$ ON/OFF state and the clinical state of the host they were
		isolated from .	
	4.6	Association bet	ween $lic2A$ ON/OFF state and the site of isolation from the host \ldots . 123
	4.7	Analysis of SSR	t tract lengths within the $lic2A$ gene of 65 COPD samples $\ldots \ldots \ldots 126$
	4.8	Sequence analys	sis of the $lic2A$ gene of <i>H. influenzae</i> Rd 30 $\dots \dots \dots$
	4.9	Determination of	of bacteriophage resistance phenotype of Rd 30 following bacteriophage $% \mathcal{A}$
		HP1c1 challeng	e
	4.10	Discussion	
5	Inve	estigation of th	e effect of the Phase Variable
	Bac	teriophage Re	ceptor, <i>lic2A</i> , on Population dynamics
	of E	I. influenzae, a	and Bacteriophage HP1c1 141
	5.1	Abstract	
	5.2	Introduction	
	5.3	One-Step growt	h curve of bacteriophage HP1c1 on
		H. influenzae R	d 30S and Rd 30R $\dots \dots \dots$
	5.4	The effect of Po	pulation Heterogeneity on Bacteriophage HP1c1 Spread
	5.5	Determination	of adsorption rate of bacteriophage HP1c1 to Rd 30S $\ldots \ldots \ldots \ldots 149$
	5.6	Description of t	he mathematical model facilitating simulation of the oscillating prey assay 151

7	Sup	plementary Information 18	31	
	6.1	Potential further work	<i>'</i> 9	
6	Con	cluding discussion 17	$\mathbf{'2}$	
	5.10	Discussion	i5	
		in host survival against bacteriophage HP1c1	52	
	5.9	Determination of whether $lic2A$ phase variation generated population heterogeneity aids		
	5.8	Multi-directional oscillating prey assay	59	
	5.7	Modelling of the effect of phase variation on bacteriophage survival		

List of Tables

1.1	Examples of surface molecules used for adsorption of bacteriopahge
1.2	Phase Variable Bacteriophage Resistance Mechanisms
1.3	Examples of phase variable genes in <i>H. influenzae</i>
2.1	<i>H. influenzae</i> isolates used in this study
2.2	Primers used in genescan reactions
2.3	Primers used in SOE-PCR reactions and screening of transformants
2.4	Ingrediants of media used for chemical competancy of <i>E. coli</i> DH5- α 60
2.5	M-IV Competence Media Component Solutions
3.1	Allele classification from analysis of the 121 strains with an identifiable $hsdM$ gene in
	GenBank
S1	Media/Buffer/Reagents used in this study
S2	Other primers used in this study

List of Figures

1.1	Total number of laboratory-confirmed cases of H . influenzae type b, by age group, in	
	England per year between 1990 and 2015	2
1.2	Total number of laboratory-confirmed cases of H . influenzae, arranged by serotype, in	
	England per year between 1990 and 2015	3
1.3	Representation of the extent of mutations in host chromosomes elicited by co-evolution	
	with bacteriophage	10
1.4	Electron micrograph of three bacteriophage of <i>H. influenzae</i>	16
1.5	Comparison of restriction modification system diversity in four ${\it H.~influenzae}$ strains $% {\it H.~influenzae}$.	18
1.6	Illustration of the impact and pitfalls of restriction modification activity on bacterio-	
	phage infection dynamics	23
1.7	Examples of bacterial resistance mechanisms to bacteriophage infection $\ldots \ldots \ldots$	28
1.8	Streisinger model of slipped strand mispairing	31
1.9	Theorised role of phase variable restriction modification systems in defence against bac-	
	teriophage	39
1.10	Schematic example of an lipooligosaccharide structure for $H.$ influenzae Rd \ldots	46
2.1	Graphic representation for the methodology of the oscillating prey assay $\ldots \ldots \ldots$	68
2.2	Outline of transfer and sampling regime for testing bacteriophage expansion over a fixed	
	area	70
2.3	Layout of which wells received $lic2A$ ON or $lic2A$ OFF cells in the 50 $\%$ ON and 66 $\%$	
	ON population structure simulations	71
3.1	Presence of 5'-GAGAC-3' SSR units in the HindI-like hsdM gene of 121 H. influenzae	
	strains	77
3.2	Sequence analysis of 121 <i>H. influenzae hsdM</i> alleles	79
3.3	Putative reading frames of the $hsdM$ gene of H. influenzae RM118 $_{\rm r^+HinDI\ m^+HinDI},$	
	RM118 $_{\rm r^{-}HinDI}$ $_{\rm m^{-}HinDI},$ and the in silico predicted RM118 $_{\rm GAGAC(3x)}.$	83

3.4	Infectivity assay of the ability of bacteriophage HP1c1 to produce plaques on phase
	variants of <i>H. influenzae</i> RM118
3.5	Products produced during PCR amplification of $hsdM$ for use in constructing pCT0.5
	and pCT1 plasmids
3.6	Restriction profile from screening of transformant colonies of pCT0.5
3.7	Restriction profile from screening of transformant colonies of pCT1
3.8	Electrophoresis image from screening transformed cells of H . influenzae following trans-
	formation with $pCT1 \dots \dots$
3.9	Products produced during PCR amplification of $hsdM$ for use in constructing the pCT2
	plasmid
3.10	Restriction profile from screening of transformant $E. \ coli$ colonies of pCT2 91
3.11	Distribution of 5'-GAGAC-3' repeats within samples from COPD patients 92
3.12	Distribution of 5'-GAGAC-3' repeats within samples from COPD patients 94
3.13	Clinical states associated with the ON/OFF state of $hsdM$ 121 H. influenzae strains
	from GenBank where $hsdM$ was detected $\ldots \ldots 95$
3.14	Putative ON/OFF state associated with the clinical states of the 121 H . influenzae
	strains from GenBank with detectable $hsdM$
3.15	Examples of sequence variability in the three main regions of the HindI-like $hsdS$ gene $\ . \ 99$
3.16	Heatmap showing identity matrix of the level of homology between each domain of the
	hsdS genes, and the entire $hsdS$ gene of the 121 H. influenzae samples available on
	GenBank
3.17	Phylogenetic tree of the $hsdS$ gene of 121 H . influenzae sequences available on GenBank 106
4.1	Presence of 5'-CAAT-3' SSR units in the <i>lic2A</i> gene of 104 <i>H. influenzae</i> strains 118
4.2	Clinical states associated with the hosts from whom the 104 H . influenzae strains with
	sequences in the GenBank database where a $lic2A$ gene was identifiable $\ .\ .\ .\ .\ .\ .$. 121
4.3	Association between each SSR tract length and the ON/OFF state of $lic2A$ in 104
	H. influenzae genomes
4.4	ON/OFF state of the $lic2A$ gene from the 104 H. influenzae strains in the GenBank
	database as a function of the clinical condition of the host from whom they were isolated. 124
4.5	ON/OFF state of the $lic2A$ gene from the 104 H. influenzae strains in the GenBank
	database as a function of the host site from where the bacteria were isolated. \ldots 125
4.6	Peak Scanner output of the 226 bp fragment of $lic2A$ amplified from Rd 30S for screening
	of COPD samples
4.7	Putative $lic2A$ SSR tract lengths within sputum supernatant samples from COPD patients.128
4.8	Nucleotide and amino acid sequences of the $lic2A$ gene of the H. influenzae Rd 30
	variants used in this study

4.9	Infectivity assay of bacteriophage HP1c1 against the $lic2A$ phase variant strains H . in-
	fluenzae Rd30S and Rd30R
5.1	One-step growth curve of bacteriophage HP1c1 on $H.\ influenzaelic2A$ phase variants $\ . \ . 145$
5.2	Oscillation assay, fluctuating host single-cycle infection, of HP1c1 through various host
	population structures for <i>lic2A</i> expression
5.3	Adsorption assay of bacteriophage HP1c1 to H. influenzae Rd 30S $\ldots \ldots \ldots \ldots \ldots 150$
5.4	Examples of the outcome from the model of the three components added at the beginning
	of each cycle
5.5	Examples of bacteriophage densities during simulations from the mathematical model
	of the oscillating prey assay $\ldots \ldots 155$
5.6	Modelled probability of bacteriophage survival following replication through heteroge-
	neous bacterial populations for <i>lic2A</i> expression after 105 cycles
5.7	Multi-directional oscillating prey assay of bacteriophage HP1c1 through $H.$ influenzae
	populations of varying degrees of phenotype mixing for $lic2A$ expression
5.8	Bacterial growth curves in the presence of various bacteriophage HP1c1 multiplicities
	for <i>H. influenzae</i> populations composed of distinct degrees of <i>lic2A</i> phenotypes 163

Abbreviations

\mathbf{Abi}	Abortive infection
ATP	Adenosine triphosphate
BHI	Brain Heart Infusion broth
bl2seq	BLAST 2 Sequences
CDS	Coding DNA Sequence
CFU	Colony forming unit
COPD	Chronic obstructive pulmonary disease
CRISPR-cas	Clustered regulatory inter-spaced short palin- dromic repeat and CRISPR associated proteins
\mathbf{crRNA}	CRISPR RNA
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
Hib	Haemophilus influenzae type b
Kdo	3-Deoxy-D-manno-oct-2-ulosonic acid
LB	Lysogeny broth
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
MLST	Multi-locus sequence typing
MMR	DNA mismatch repair
MOI	Multiplicity of infection
MTase	Methyltransferase
NAD	Nicotinamide adenine dinucleotide
NTHi	Non-typeable Haemophilus influenzae
OD	Optical density
ORF	Open reading frame
PCR	Polymerase Chain Reaction

PBS	Phosphate buffered saline
PFU	Plaque forming units
REase	Restriction endonuclease
RNA	Ribonucleic acid
SAM	S-adenosylmethionine
${ m sBHI}$	BHI supplemented with Hemin and NAD
SCID	Severe combined immune deficiency
sLB	LB supplemented with Hemin and NAD
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
Sie	Superinfection exclusion
\mathbf{SM}	Saline magnesium
SMRT	Single molecule real time
SOB	Super Optimal Broth
SOC	Super Optimal broth with Catabolite repression
SOE	Splicing by overlap extension
\mathbf{SSM}	Slipped strand mispairing
SSR	Simple sequence repeat
T-SDS-PAGE	Tricine-SDS-PAGE
TAE	Tris-acetate-EDTA
TEM	Transmission electron microscopy
UPH ₂ O	Ultra-pure water
UV	Ultra-violet

Disclaimer

I conducted all experimental assays in this work, however the mathematical models in this work were generated by Dr Andrey Morozov (University of Leicester), and are included to provide a complete picture of the results obtained from this study.

1. Introduction

1.1 Haemophilus influenzae

Haemophilus influenzae is a Gram-negative, non-motile, coccobacillus that is a common asymptomatic coloniser of the nasopharynx and which exclusively colonises humans (Finney and Ritchie, 2014), but can act as opportunistic pathogen (Hinz et al., 2015). Carriage rates for *H. influenzae* vary largely for a wide variety of reasons, including age, socio-economic status, allergies, and degree of exposure to children (García-Rodríguez and Fresnadillo Martínez, 2002). The later of these reasons eludes to the most common groups where *H. influenzae* carriage is noted, children. For example, a study into the isolation frequencies across three age groups, 0-6, 7-15, and ≥ 16 in the county of Elfsburg in Sweden indicated that the prevalence of *H. influenzae* carriage in the nasopharynx of healthy individuals was highest in the youngest age group (13.2 % (21/159), 6.1 % (12/198), 2.7 % (7/261) respectively) (Gunnarsson et al., 1998). *H. influenzae* colonisation has been shown to be largely variable though, with colonisation rates reaching as high as around 80 % in young children, however colonisation rates generally decrease as a function of age (García-Rodríguez and Fresnadillo Martínez, 2002).

Although often found as a commensal in the human nasopharynx *H. influenzae* can act as a pathogen, and in particular is noted as a significant cause of symptomatic exacerbations in the respiratory disease chronic obstructive pulmonary disease (COPD), contributing significantly to morbidity and mortality in this condition (Papi et al., 2006; Finney and Ritchie, 2014). In addition *H. influenzae* can cause a number individual conditions, including bacteremia (Laupland et al., 2011); a number of mucosal infections such as otitis media, conjunctivitis, and sinusitis (Van Eldere et al., 2014); meningitis (Ladhani et al., 2010); and pneumonia (Cordero et al., 2000). Like its prevalence in carriage, *H. influenzae* associated disease is more readily observed in the young than in adults (Dworkin et al., 2007). Statistics on the numbers of reported cases of the most severe disease associated serogroup, *H. influenzae* type b, across age ranges in England can be found in Figure 1.1, detailing the number of cases per year from 1990-2015.

H. influenzae isolates can be segregated into two groups, non-encapsulated, or encapsulated strains, with the encapsulated strains segragated into a further six distinct groupings based on their anti-



Figure 1.1. Total number of laboratory-confirmed cases of H. influenzae type b, by age group, in England per year between 1990 and 2015. The year of the Hib vaccines introduction is noted by the dotted line. All statistics were obtained from data published by Public Health England (Public Health England, 2015b). Values for 2014 and 2015 were updated using the most recent statistics on H. influenzae laboratory confirmed cases from Public Health England (Public Health England, 2016).

genically differential polysaccharide capsule (Pittman, 1931; Falla et al., 1994). The group without a polysaccharide capsule, termed non-typeable *H. influenzae* (NTHi), are the current most common cause of laboratory reported invasive *H. influenzae* cases in England (Fig. 1.2) and is also highly prominent in other countries (Van Eldere et al., 2014). The remaining five groups are seperated based on the antigenic properties of their capsule with these groups referred to as serogroups a-f (Falla et al., 1994).

H. influenzae serotype b (Hib) is probably the most well known form of *H. influenzae*, as, in the 1930s, before the introduction of antibiotic therapy, Hib had a mortality rate of close to 100 % in North America (Aubrey and Tang, 2003). After the introduction of antibiotics, but before the introduction of the Hib vaccine, Hib was still the prominent cause of invasive bacterial infections and meningitis in North America with mortality level at around 3-5 % (Agrawal and Murphy, 2011). Despite drops in mortality levels, *H. influenzae* infections were still of significant concern due to sequelae from infection (seen in 20-30 % of individuals after infection), such as loss of hearing or reduced mental capacity (Agrawal and Murphy, 2011). Estimates for the impact of the Hib vaccine on the incidence of Hib associated disease in North America estimate that incidences have reduced from $\sim 85 / 100,000$ per year to $\sim 2 / 100,000$ per year in Canada, and from 88 / 100,000 per year to 1.6 / 100,000 per year in the United States (Peltola, 2000). In the UK, levels of Hib associated disease are estimated to

have been ~ 36 / 100,000 per year while incidences after the Hib vaccine levels were estimated to have dropped to ~ 1 / 100,000 per year (Peltola, 2000). The implementation of the Hib vaccine in 1989, and its application in the UK in 1992, reduced the cases of Hib reported (and *H. influenzae* report numbers overall) significantly (e.g. Figures 1.1 and 1.2) with type b being overtaken by the capsular polysaccharide free group of *H. influenzae* in many countries (Adam et al., 2010; Van Eldere et al., 2014), a trend that has been reflected within the official statistics on *H. influenzae* reports released by Public Health England (Figure 1.2).



Figure 1.2. Total number of laboratory-confirmed cases of *H. influenzae*, arranged by serotype, in England per year between 1990 and 2015. The year of introduction of the introduction of the Hib vaccination is denoted by the dotted line. All statistics were obtained from data published by Public Health England (Public Health England, 2015a), values for 2014 and 2015 were updated using the most recent statistics available from Public Health England for *H. influenzae* (Public Health England, 2016). For 2014 and 2015 values for types a,c,d,e,f (Blue line) data from Public Health England (2016) is only available for types a,e,f and as such 2014 statistics for this group was used based on the partial a,c,d,e,f values of Public Health England (2015a), while for 2015 data can only be said to be representative of number of confirmed cases of types a,e,f.

As well as being an important aetiological agent in disease *H. influenzae* also has a decorated history as a research organism. *H. influenzae* was involved in Alexander Fleming's seminal article describing the discovery of penicillin titled "On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae" (Fleming, 1929). At the time *H. influenzae* was known as *Bacillus influenzae* as well as Pfeiffer's bacillus, due to its morphology, discovery by Richard Pfeiffer, and Pfeiffer's mistaken association of this bacteria as the aetiological agent of influenzae (Taubenberger et al., 2007). *H. influenzae* Rd KW20 served as the organism that lead to the discovery of Type II restriction modification systems (Smith and Wilcox, 1970; Kelly and Smith, 1970), work that would later go towards earning the 1978 Nobel Prize in Physiology or Medicine for Werner Arber, Daniel Nathans, and Hamilton O. Smith (Loenen et al., 2014b; Nobelprize.org, 2014). In 1989 the highly pathogenic *H. influenzae* serotype b served as the basis for the first polysaccharide-conjugate vaccine (Aubrey and Tang, 2003), meanwhile in 1995 *H. influenzae* Rd KW20 was the first free living organism to have its complete genome sequenced (Fleischmann et al., 1995). The organism has also played a key role in the understanding of phase variation, the topic of this current work, and as such, serves as the model organism for the observation of the role of phase variation in the dynamics between bacteria and their viral predator, the bacteriophage.

1.2 Bacteriophage

Bacteriophage are the viral predator of prokaryotes and as such are a key driving forces behind their bacterial hosts evolutionary trajectory (Koskella and Brockhurst, 2014). Interest in these viral parasites has been reignited by the ever increasing rates of bacterial resistance to the current lines of antimicrobial treatments, with the suggestion of bacteriophage as a potential alternative therapeutic treatment (Thiel, 2004). Bacteriophage have been suggested as an alternative therapeutic to current treatment methods for a number of reasons, such as their ability to induce lysis of their host upon replication; their replication in bacterial cells causing an increase in the number of viral cells, effectively resulting in 'auto-dosing'; their high specificity, with a bacteriophage often only targeting one specific host or a small subgroup of hosts; and the relative ease with which new bacteriophage can be discovered, due to their abundance (Loc-Carrillo and Abedon, 2011).

Each of the stated reasons bacteriophage may serve as a successful therapeutic is a reflection of the ability of bacteriophage to shape bacterial populations in natural environments. In nature, bacteriophage shape bacterial host populations across many environments from the rhizosphere, the phyllosphere, oceans, to even shaping the microbiota in living organisms (Gómez and Buckling, 2011; Koskella et al., 2011; Marston et al., 2012; Pommier et al., 2012; Quesada et al., 2012; Koskella, 2013; Mills et al., 2013).

In terms of their lytic action, lysis of their host will result in extinction of the bacterial population without bacterial adaptation, imposing a large selection pressure on the bacterial population to adapt. This will alter the population composition towards new resistant genotypes distinct from the original susceptible forefathers (Lenski and Levin, 1985; Stern and Sorek, 2011; Scanlan et al., 2015). In industry this is a considerable problem as, for instance, contamination with bacteriophage in fermentation cultures for dairy production often can lead to reduced productivity due to lysis of the fermenting

bacteria (Brüssow, 2001).

The auto-dosing nature of bacteriophage replication means that even if the bacteriophage population is present at low numbers, over time, their numbers will increase, eventually reaching a point in which they can cause a reduction in bacterial density; from the therapeutic stand point this is termed 'active treatment' (Cairns et al., 2009; Abedon, 2014). However bacteriophage densities, regardless of whether low or high, diversify bacterial populations in a concentration dependent manner, as mutations implying reduced susceptibility will vary as a function of phage density, where partial resistance phenotype adaptations can suffice in low phage densities, while in higher densities only those conferring absolute resistance are likely to remain viable (Christen et al., 2016).

The high specificity of bacteriophage, and their propensity to often only target a small subsection of a specific species (Flores et al., 2011), means that not all members of the bacterial populations will be targeted by the same phage in the same dose. For example, Golomidova et al. (2007) observed the presence of a large number of *Escherichia coli* bacteriophage with diverse host ranges, the range of bacterial hosts that a bacteriophage can infect, within samples of horse faeces. However, estimations of the proportion of *E. coli* from the same sources these bacteriophage could infect was only 2-4 % of the total strains present in the community (Golomidova et al., 2007). As such in natural settings multiple bacteriophage are necessary to shape the entire bacterial population and the likelihood of these multiple viral challenges eliciting the same genotypic alteration to an already genotypically diverse population is unlikely, especially considering that these viruses are likely to be targeting different surface molecules even within the same species (E.g. Table 1.1).

Bacteriophage discovery is to an extent trivial due to the ubiquitous nature of these parasites, with their numbers estimated to be around 10^{31} , some ten-fold higher than their bacterial counterparts (Clokie et al., 2011; Hatfull and Hendrix, 2011; Suttle, 2007). To highlight their sheer abundance, it has been estimated that a total of roughly 10^{23} infections occur per second in the ocean alone (Suttle, 2007). One creative estimation suggests that if the total number of viruses in the sea were used to create a row of viruses from tip to tail the resulting viral string would extend some sixteen times longer than the diameter of the Milky Way (Weinbauer and Rassoulzadegan, 2004).

Although the development of bacteriophage as a treatment for bacterial infections may be the justification for the recent resurgence in research into these viral predators it is their natural ability to sculpt bacterial populations which serves as the backbone for their application. It is the investigation of the effects on interaction of bacteriophage and their host through one mutational mechanism, phase variation, that may itself have arisen at particular loci through co-evolution with bacteriophage, that will be the focus of this current work.

1.2.1 Bacteriophage life cycles and their implications for the evolution of their hosts, and themselves

Bacteriophage replication is generally segregated into two life cycles, the lytic life cycle, and the lysogenic life cycle (Campbell, 2003). The lytic life cycle is the archetype of the bacteriophage life cycle, as it is this cycle that leads to the eventual lysis of the infected bacterium in order to release viral progeny. In the lysogenic life cycle the bacteriophage integrates its genetic material into the bacterial hosts chromosome (or a plamsid), playing a role in shaping their hosts genetic architecture. As the most commonly studied and well understood viruses are from the order *Caudovirales* (often described as the 'tailed bacteriophage'), and as all known bacteriophage for *H. influenzae* are from this order, when bacteriophage replication is discussed it is in relation to viruses from this order. Regardless of their form, either bacteriophage life cycle can shape the evolutionary pathway for the bacterial host towards the hosts benefit, or even towards its detriment.

1.2.1.1 Lytic cycle

The lytic life cycle is the epitome of the classical explanation of the bacteriophage life cycle. In essence, during this cycle the bacteriophage must first adsorb to a host cell, inject its DNA, use the hosts metabolic machinery to manufacture new viral progeny and lyse the host cell, facilitating dissemination of their progeny (Kutter and Sulakvelidze, 2004). When bacteriophage therapy is discussed, the viruses that are mentioned are those which undergo the lytic life cycle obligately. This is because these viral particles shape bacterial populations through elimination of cells susceptible to infection by the invading bacteriophage. Lytic action upon sensitive populations can have a resounding influence on what members of a bacterial population remain within a given environment, shaping the future direction of bacterial evolution through selecting which members of the bacterial population can pass on their genetic material to future generations.

One such influence is in the control of bacterial densities. The relationship between bacteriophage and host densities was evident even at one of the incarnations of the discover of bacteriophage (bacteriophage are said to have been discovered twice, independently, first by Fredrick Twort in 1915, and then Felix D'Herelle in 1917 (Summers, 2011; Keen, 2015)). In his seminal work D'Herelle (1917) describes a phenomenon of clear zones emerging from the lawn of bacteria derived from the stool samples of patients with dysentery and notes that such clear zones are temporally associated with the recovery of afflicted individuals. In his own words:

"In summary, in certain patients recovering from dysentery, I observed that the disappearance of the dysenteric bacillus coincided with the appearance of an invisible microbe endowed with antagonistic properties toward a pathogenic bacillus. This microbe, which is a veritable "microbe of immunity", is

an obligate bacteriophage. Its parasitism is strictly specific, but although it is limited to one species at a given time, it may in turn have an effect upon diverse germs via acclimatisation. It would thus appear that in bacillary dysentery, along with homologous antitonic immunity emanating directly from the patient's organism, there also exists heterologous antimicrobial immunity produced by an antagonistic microorganism." - D'Herelle 1917 via D'Herelle 2007

What D'Herelle was likely describing is the natural control bacteriophage can elicit on bacterial population numbers, an idea that is reflected in theories on bacteriophage control of bacterial numbers during outbreaks such as that reported for another gastrointestinal pathogen Vibrio cholerae (Faruque et al., 2005b,a). Faruque et al. (2005b) observed that there was a temporal correlation between the estimated number of cases of cholera caused by both the V. cholerae O1 and O139 strains at the International Centre for Diarrhoeal Disease Research hospital in Dhaka, Bangladesh, between 2001 and 2003, and the prevalence of bacteriophage against these pathogenic V. cholerae strains in the aquatic environment of Dhaka. They observed that when numbers of bacteriophage against these V. cholerae strains within water samples were low the number of cholera cases would begin to increase, and when bacteriophage numbers were generally high, the incidence of reported cholera cases would be in low numbers (Faruque et al., 2005b). In their follow up work analysing again the correlation between cholera outbreaks and bacteriophage numbers in Dhaka, this time during an outbreak of antibiotic resistant V. cholerae O1 between August 2004 and December 2004, they again observed this same temporal association, with reported incidences of cholera decreasing as the prevalence of bacteriophage within water samples began to increase (Faruque et al., 2005a). However, it is worth noting that the numbers of cholera cases was extrapolated in their studies from the International Centre for Diarrhoeal Disease Research hospital's 2 % sampling programme, so exact case numbers may not be entirely accurate.

In this study they also estimated the numbers of *V. cholera* within the water samples. They observed a direct correlation between high incidences of *V. cholerae* and low bacteriophage numbers, and vice versa. The estimations of the numbers of bacteria in water environments are after enrichment (necessary for isolation of *V. cholerae* O1) through an 'index of *V. cholerae* concentration'. This is the mean weekly value of *V. cholerae* concentrations divided by the median of all week means in the study period. Thus again the value is not entirely a direct estimation; and additionally the role of environmental factors such as pH, salinity etc cannot be ruled out as contributing factors to alterations in bacterial and viral numbers; however regardless the correlation between bacteriophage numbers and estimated cases due to the bacteriophage target bacteria pertains to an interesting observation. Similar associations between bacteriophage numbers and their hosts abundance have been made in other systems, such as in the control of bacterial population numbers in soil ecosystems (Allen et al., 2010), and bacterial waste-water treatment bio-reactors (Shapiro et al., 2010). Nowhere though is the effect of bacteriophage predation more evident than within the sea, where bacteriophage play significant roles in the regulation of organic carbon cycling within the oligotrophic environment (Wilhelm and Suttle, 1999; Fuhrman, 1999; Suttle, 2007). Marine viruses are estimated to cause lysis of around 20 % of the total prokaryote biomass per day (Suttle, 1994), with this lysis freeing large quantities of organic carbon into the marine environment. Cyanophage, those bacteriophage which infect cyanobacteria, a phylum of photosynthetic organisms ubiquitous to ocean environments that account for a considerable amount of the total biological carbon in the ocean (Wilhelm and Suttle, 1999), contribute significantly to freeing organic carbon. It is estimated that at any one time cyanophage infect 50 % of all cyanobacteria (Clokie et al., 2011), with estimations of the amount of photosynthetically fixed carbon this infection frees to become dissolved organic material to around 25 % of total photosynthetically fixed carbon (Wilhelm and Suttle, 1999). Due to this the lytic action, bacteriophage are believed to be key players in the recycling of global carbon (Suttle, 2005).

Such vast selection pressures on bacterial populations forces adaptation. Within one of the members of the cyanobacteria, *Synechoccus*, such adaptations have been observed. Marston et al. (2012), using chemostat cultures of *Synechoccus* spp. WH7803 and the cyanophage RIM8, observed that the presence of bacteriophage resulted in diversification of the *Synechoccus* populations, while diversification was abscent in the population free from predation over the course of the 6-month experiment. They observed up to three bacterial phenotypes in co-evolved populations at any one time, with the authors conceding that this could be a vast underestimate as only 2-4 host isolates were examined per time point (Marston et al., 2012). These adaptations provided varying degrees of bacteriophage resistance, a topic that will be covered in more detail later (Section. 1.3), but such resistance mutations often result in knock-on phenotypic and genotypic adaptations within bacterial populations, and in *Synechoccus* reduced growth rates between bacteriophage resistant populations and their ancestral populations have been observed (Lennon et al., 2007).

Such knock-on phenotypic and genotypic alterations have been extensively studied within members of the *Pseudomonas* spp. and their cognate bacteriophage, in particular *Pseudomonas fluorescens* SBW25 and bacteriophage Φ 2. It was first observed in *P. fluorescens* SBW25 that predation by bacteriophage can generate a mutator phenotype within the bacterial population that increases bacterial mutation rates (Pal et al., 2007). *P. fluorescens* avoid bacteriophage predation in this way through mutations within the organisms mismatch repair (MMR) genes, *mutS* and *mutL*, allowing global hypermutation across the chromosome (Pal et al., 2007). These hypermutable cells had survived bacteriophage predation due to the increased rates of mutation allowing the increased chance of developing mutations that would confer resistance to bacteriophage attachment. Scanlan et al. (2015) observed the extent such bacteriophage induced global hypermutation can cause alterations in the bacterial chromosome. They observed from analysis of single genotypes within six different lineages of evolved or co-evolved populations (bacteriophage free and bacteriophage present populations respectively) that genotypes from lineages co-evolved with bacteriophage had a greater number of mutations, both synonymous and non-synonymous. The extent of these mutations can be noted in figure 1.3. The \sim 10-fold higher mutation rates noted in 3 of the 5 co-evolved populations displaying a hyper-mutation phenotype were attributed to mutations within one or more DNA repair genes (namely *mutS*, *uvrD*, and *uvrB*). However, evolution with bacteriophage can have detrimental effects on host populations, and Scanlan et al. (2015) observed that populations co-evolved in the presence of bacteriophage had a ~ 25 % reduction in fitness in growth compared to those evolved without bacteriophage under conditions free from bacteriophage predation.

Co-evolution with lytic bacteriophage can even give rise to completely new phenotypes that otherwise may not have arisen in their absence. For example, in a study of the effects of bacteriophage on the diversification of P. fluorescens, Buckling and Rainey (2002b) observed the presence of highly mucoid colonies within lineages evolved with bacteriophage that were absent in the bacteriophage free lineages, implying that bacteriophage selection had resulted in the generation of this morphology. This was later shown to only arise after exposure to bacteriophage, even when multiple environmental conditions (combinations of shaken vs static cultures and high vs low nutrient) were tested, with shaken high nutrient cultures developing this phenotype in 100 % of the bacteriophage exposed lineages (Scanlan and Buckling, 2012). Interestingly this mucoid phenotype can often be observed in *Pseudomonas aeruginosa* in individuals with cystic fibrosis, and is associated with poor patient prognosis (May et al., 1991; Ramsey and Wozniak, 2005). Thus interaction with bacteriophage can potentially facilitate the evolution of clinically relevant characteristics in their bacterial hosts.

The interaction of bacteriophage and their host is not simply a one-sided affair, as the interaction of bacteriophage and a host can cause alterations in the bacteriophage population. For instance, cyanophage are known to harbour some genes that can encode parts of the photosynthetic apparatus that is common in cyanobacteria (Mann et al., 2003; Millard et al., 2004; Lindell et al., 2004). Mann et al. (2003) discovered that the lytic cyanophage S-PM2 contained two genes, *psbA* and *psbD*, which are homologues of cyanobacterial D1 and D2 proteins respectively, part of the cyanobacterial photosystem II. Photosystem II is a crucial component of the the photosynthetic process in cyanobacteria, however, during the photosynthetic process there is a large turnover of D1 molecules, caused by damage to these proteins through reactive oxygen species generated in the photosynthetic process (Baena-González and Aro, 2002). During photosynthesis the D1 protein is constantly generated, but in intense light the rate of D1 degradation is greater than the rate of synthesis, which results in photoinhibition (Aro et al., 1993). However, bacteriophage can overcome this scenario through the expression of the bacteriophage *psb* genes.

The psbA and psbD genes are believed to have been acquired from the host cyanobacteria, *Synechoccus*, at two separate recombinational events, as these genes are separated by two further non-cyanobacteria



Figure 1.3. Representation of the extent of mutations in host chromosomes elicited by co-evolution with bacteriophage. Figure from Scanlan et al. (2015) showing mapped chromosomal mutations of individual genotypes from six populations co-evolved with bacteriophage (a) and evolved in the absence of bacteriophage (b), after ~ 400 generations, onto the circular genome of *P. fluorescens* SBW25. The first and second rings represent the forward and reverse strands respectively, with the remaining rings inward each represent the genotype selected from one of the six replicate populations.

genes (Mann et al., 2003). The psbA and psbD genes, together with other photosynthesis related genes, have similarly been found in bacteriophage that infect other cyanobacteria, such as those infecting *Prochlorococcus* (Lindell et al., 2004), as well as other *Synechoccus* bacteriophage (Millard et al., 2004). For S-PM2 and *Synechoccus* it has been shown that psbA is transcribed after infection of the host, with an initial increase in host psbA transcripts being followed by decreases in host psbAtranscription to below that of the viral psbA (Clokie et al., 2006). A similar observation was also made for *Prochlorococcus* and its cyanophage where host psbA expression decreases overtime, eventually being overtaken by the increasingly expressed bacteriophage psbA (Lindell et al., 2005). The expression of psbA and other photosynthesis related genes by the bacteriophage may be of key importance as in the absence of photosynthesis it is known that there is a reduction in the production of viral particles after infection (Lindell et al., 2005), while viral replication is known to be a function of the growth rate of the bacterial host (Lindell et al., 2007). Thus it has been hypothesised that the presence of bacterial photosynthesis associated genes within bacteriophage may be a possible means by which bacteriophage have adapted to facilitate the generation of energy within their host in order to produce their progeny (Lindell et al., 2007).

1.2.1.2 Lysogenic cycle

The other common life cycle used by bacteriophage is the lysogenic cycle. This life cycle proceeds initially as does the lytic cycle, whereby bacteriophage bind receptors on the host surface and injects its DNA. However, rather than commandeering the hosts metabolic machinery to generate new viral particles, the bacteriophage then integrates its DNA within the chromosome of its bacterial host. These integrated elements, termed prophage, remain dormant replicating together with the cell as part of the chromosome until a point where the prophage element is stimulated to re-enter the lytic cycle, normally through stimulation from a stressor, such as ultra-violet light or antibiotic stress etc, although spontaneous induction often occurs in a small number of the lysogenic population (Kutter and Sulakvelidze, 2004).

Bacteriophage elements are found, to at least some degree, within the genome sequences of a vast number of bacteria sequenced to date (Casjens, 2003). In extreme examples these elements can constitute up to 10-20 % of a bacterial genome in the form of partial remnants (cryptic prophage) or in their complete dormant forms (Casjens, 2003). A good example of the prevalence of prophage carriage within bacterial genomes is *E. coli* 0157:H7 EDL933 which was identified to carry 18 genomic regions related to known bacteriophage, with regions ranging from 7.5 kb to 61.6 kb in size (Perna et al., 2001). Of these bacteriophage elements 17 are cryptic bacteriophage, bacteriophage elements which lack some of the constituent genes required to make a fully functional viral particle (Wang et al., 2010), and one region, BP-933W, encodes a fully functional bacteriophage particle (Perna et al., 2001; Plunkett et al.,

1999).

The prophage elements held within bacterial chromosomes are often maintained due to their provision of new phenotypes to their host that prove beneficial under certain circumstances, occurring classically through a process called 'lysogenic conversion' (Brüssow et al., 2004). These characteristics can be in the form of new means of pathogenicity, or mechanisms that allow the bacterial host to survive in stressful environments (Brüssow et al., 2004; Fortier and Sekulovic, 2013). For example, the prophage elements within the above example of E. coli 0157:H7 strain EDL933 encode both shiga toxin types, stx1 and stx2, encoded on the cryptic prophage CP-933V, and BP-933W respectively (Perna et al., 2001). Shiga toxins are potent virulence factors, with infection by shiga toxin-producing E. coli resulting in more sever conditions than other non-shiga toxin-producing E. coli, such as the condition of haemolytic uraemic syndrome (Tarr et al., 2005). Shiga toxin production requires prophage induction, the process where the bacteriophage begins to re-enter the lytic cycle, with the stx genes present within the late genes of encoding prophage (Karch et al., 1999). Induction results in generation of increased numbers of bacteriophage chromosomes within the cell, and thus copies of the stx gene, increasing toxin production with toxins then released following lysis of the host cell (Muniesa et al., 2012). Induction of these stx encoding prophage can be be initiated by such things as antibiotics (Zhang et al., 2000), immune stressors (Wagner et al., 2001), or other environmental stimuli such as pH changes (Imamovic and Muniesa, 2012). Similarly other toxins are encoded in bacterial species through bacteriophage elements, such as the cholera toxin genes ctxAB within the CTX Φ prophage found in Vibrio cholerae (Heidelberg et al., 2000); the diptheria toxin gene, tox, within a prophage region of Corynebacterium diphtheriae (Cerdeño-Tárraga et al., 2003); the Clostridium botulinum botulinium toxin is encoded in C. botulinium type C and D by the prophage $CE\beta$ and $DE\beta$ respectively (Eklund et al., 1971, 1972); and numerous Staphylococcus aureus toxins are also encoded on prophage elements (Grumann et al., 2014).

Prophage elements do not only encode virulence factors such as toxins; rather their presence can have more understated effects within their hosts. Wang et al. (2010) studied the effect of the removal of cryptic prophage elements from the chromosome of *E. coli* K-12 BW25113, of which there are nine. Wang et al. (2010) found that the mutant lacking these elements, termed $\Delta 9$, had reduced growth under a number of conditions after comparing growth of $\Delta 9$ to the wild-type strain under 1240 different metabolic conditions using phenotype microarrays. Growth reductions for $\Delta 9$ were observed in the presence of six quinolone antibiotics (lomefloxacin, ofloxacin, oxolinic acid, novobiocin, nalidixic acid, and enoxacin), with further testing of nalidixic acid on $\Delta 9$ showing 379-fold growth suppression compared to the wild-type. Exposure to 11 β -lactam antibiotics also resulted in similar growth reductions in $\Delta 9$ (Wang et al., 2010). The growth of prophage cured $\Delta 9$ was also inhibited in high salinity (6% NaCl) and conditions of oxidative stress (presence of potassium tellurite), with retesting of these types of stress in flask cultures (oxidative stress was retested using exposure to 30 mM H_2O_2 for 15 min) quantified as a 787-fold and 245-fold inhibition of growth respectively (Wang et al., 2010). Screening of strains with individual prophage deletions showed that although the reduction in survival against osmotic stress only arose in the $\Delta 9$ strain, oxidative stress survival could be attributed to a single phage deletion (Wang et al., 2010).

Prophage do not exclusively provide beneficial characteristics to their host, but can also confer detrimental effects. Here, deletion of the nine prophage for $\Delta 9$, as well as deletion of eight of the nine prophage, resulted in reduced biofilm formation in comparison to the wild-type. Conversely, deletion of prophage CP4-57 resulted in a ~ 6-fold increase in early biofilm formation (Wang et al., 2010). A similar observation has been noted in *P. aeruginosa* PA14 upon lysogeny by bacteriophage DMS3, 83/86 lysogenised isolates showed a lack of biofilm formation, and cell swarming, compared to the wild-type cells (Zegans et al., 2009).

Work by Martínez-García et al. (2015) linked prophage elements with detrimental effects to hosts survivability. They studied the effects of the removal of the four prophage elements from *Pseu*domonas putida KT2440 again through the use of pheontypic microarrays to determine knock-on phenotypic effects. They observed that a number of traits were enhanced by removal of prophage elements, such as utilisation of certain sole nitrogen sources (e.g. dipeptides) and a higher tolerance for the oxidising 2-hydroxy-1,4-naphtoquinone; conversely some deleterious changes were also noted (Martínez-García et al., 2015). A further example of a beneficial effect of curing the lysogenised P. *putida* KT2440 was in response to UV irradiation, a condition that P. *putida* would be regularly exposed to in the environment, with the four prophage cured isolate showing significantly higher survival than that the wild-type (Martínez-García et al., 2015). This is a significant point as it is a reminder of the largest drawback of lysogeny, that the host can at any time undergo lysis due to induction of the prophage elements, with UV irradiation a well known means of eliciting prophage induction.

The intertwining involvement, and even the cooperation of prophage elements in the regulation of bacterial gene expression is particularly elegantly portrayed through a process described as 'active lysogeny'. Through this process bacteriophage elements can regulate gene expression through integration into or excision from coding DNA sequences on the bacterial host's chromosome (Feiner et al., 2015). There are two forms that active lysogeny can take, either 'reversible active lysogeny' or 'non-reversible active lysogeny', with the former allowing potentially continual gene regulation changes while the latter is a finite process occurring only once per cell (Feiner et al., 2015).

An example of the involvement of active lysogeny in the regulation of bacterial gene expression, and a role in evolution of their bacterial hosts, is the regulation of mutation rates in *Streptococcus pyogenes* by the prophage-like *S. pyogenes* chromosomal island M1 (SpyCIM1). SpyCIM1 (originally termed SF370.4) is located between the MMR genes *mutS* and *mutL*, allowing the prophage element to control expression of *mutL* (Canchaya et al., 2002; Scott et al., 2008). At early exponential growth the prophage

element excised from the chromosome of SF370, remaining present as an episome within the cell. However, as the cell progressed towards, and into, stationary phase SpyCIM1 reintegrated into the SF370 chromosome (Scott et al., 2008). This coincided with initially high levels of the MutL protein within the cell at low population densities, with these levels decreasing with increasing cell density until stationary phase at which point the expression of MutL was undetectable (Scott et al., 2008). Comparison of *S. pyogenes* SF370 with another *S. pyogenes* strain that does not harbour a prophage in this locus, NZ131, revealed that SF370 had a \sim 100-fold higher mutation rate compared to NZ131 (Scott et al., 2008).

The ability of prophage elements to facilitate temporal increases in mutation frequency when nutrient conditions are becoming limited is of clear benefit to a host population as mutations may arise that allow the population to adapt to the stressful environmental conditions. In addition temporal control of mutator phenotypes will reduce the prevalence of deleterious or even toxic mutations within the bacterial populations, a drawback of mutator phenotypes (Palmer and Lipsitch, 2006). It is of note however, that there is a cost to the involvement of SpyCIM1 in controlling *mutL* expression, as SpyCIM1 integration also confers increased sensitivity to UV irradiation (Scott et al., 2008). Active lysogeny plays a key role in other bacterial species including the differentiation of small numbers of cells from populations of *Anabaena* spp. and *Nostoc* spp. to become nitrogen fixing heterocysts in nitrogen limited environments through the excision of multiple prophage remnants from their hosts chromosome (Henson et al., 2011)

In summary, bacteriophage, regardless of life cycle, can shape the evolutionary trajectory of their host organisms, altering phenotypic and genotypic diversity. A key means through which bacteriophage diverge host populations is through bacteriophage resistance. Bacteriophage resistance provides a level of heterogeneity to bacterial hosts whereby some cells may be able to gain the benefits of lysogeny while others are resistant preventing accumulation of negative attributes. While lytic action forces bacterial cells to become resistant as a means of surviving.

1.2.2 The bacteriophage of *Haemophilus influenzae*

While bacteriophage related to other organisms have been extensively isolated and thoroughly characterised the majority of bacteriophage related research involving *H. influenzae* occurred more than 40 years ago. The earliest isolated *H. influenzae* bacteriophage '*Haemophilus* phage 1', or 'HP1', was isolated from the homogenate of throat swabs from 30 patients from Johns Hopkins Hospital, Baltimore, MD, USA some 53 years ago (Harm and Rupert, 1963). This isolate was deemed problematic to work with due to its propensity to produce small, extremely turbid, and hard to view plaques. Therefore 'clear plaque' derivatives were isolated, one of which termed, HP1c1 is the bacteriophage that will be used throughout this body of work. HP1 and its derivatives are linear double stranded DNA bacteriophage from the *Myoviridae* family, with cohesive ends at the 5' ends of each strand of the HP1 chromosome (Harm and Rupert, 1963; Boling et al., 1972). The entire HP1 chromosome is composed of 32,355 bp of double stranded DNA, with the single stranded cohesive ends 7 bp in length and a G+C content almost identical to its *H. influenzae* host strain Rd (\sim 39 % Vs \sim 38 % respectively) (Fleischmann et al., 1995; Esposito et al., 1996). It also shares similarities to, and thus is classified in the same family as, the *E. coli* bacteriophage P2 (Esposito et al., 1996). To date there have only been seven other bacteriophage isolated against *H. influenzae*, HP1 and its clear plaque derivatives (Harm and Rupert, 1963), three types of bacteriophage S2 (A, B, and C) (Bendler and Goodgal, 1968), N3 (Samuels and Clarke, 1969), ϕflu (Morgan et al., 2002), and Bacteriophage HP2 (Williams et al., 2002).

After HP1 another bacteriophage of the family *Myoviridae*, named S2, was the next bacteriophage isolated (Bendler and Goodgal, 1968). Although morphologically similar to HP1c1 (Figure 1.4) and also infecting the same host strain, *H. influenzae* Rd, this bacteriophage is distinct from HP1. Boling et al. (1973) were able to produce plaques, all be it with reduced prevalence, with bacteriophage S2 on isolates of strain Rd lysogenised with HP1c1 and vice versa. This would indicate that the bacteriophage are not identical as lysogeny by one bacteriophage provides superinfection immunity against isogenic bacteriophage (Boling et al., 1973).

Strain Rd itself also harbours inducible prophage together with other *H. influenzae* strains and *Haemophilus aegyptius* (Stachura et al., 1969). However these were deemed as defective as the bacteriophage particles observed by electron microscopy were often structurally incomplete and the bacteriophage capsids appeared empty of DNA following induction (Stachura et al., 1969). It was later observed though that these defective bacteriophage did in fact contain DNA, however no viable infection or binding of these bacteriophage could be observed on the limited number of *H. influenzae* tested (Boling et al., 1973). In addition Boling et al. (1973) observed that the defective prophage within Rd was greatly dissimilar in DNA homology to that of HP1c1 and S2.

Bacteriophage from the *Siphoviridae* family have also been isolated. Samuels and Clarke (1969) isolated three bacteriophage, by UV-irradiation of clinical *H. influenzae* isolates (20 isolates were screened in total). Of these three bacteriophage they state that only one was distinguishable from HP1c1 of Harm and Rupert (1963) which they termed N3. N3 has a different target receptor to that of HP1c1 as bacteriophage HP1c1 is able to cause infection on *H. influenzae* Rd, but N3 cannot. However, extraction of the bacteriophage DNA and transformation into Rd still results in viral particles being produced (Samuels and Clarke, 1969). This would imply then that there is not an internal mechanism preventing N3 infecting its host, such as a restriction modification system, but rather an external means, such as the lack of the cognate receptor on the surface of Rd preventing N3 adsorption.

In terms of the natural ecology of *H. influenzae* bacteriophage, little is known as no studies have



Figure 1.4. Electron micrographs of three bacteriophage of *H. influenzae*. Figure taken from Boling et al. (1973) and shows the bacteriophage (a) HP1c1, (b) S2, (c) the defective prophage of *H. influenzae* Rd

been conducted on their ecology and, as noted above, few have been isolated and characterised. It is known however that DNA reads with homology to H. influenzae bacteriophage have been identified in metagenomic studies of sputum samples from both healthy individuals and those with cystic fibrosis (Willner et al., 2009). Thus these viral predators are present within the only colonisation niche for H. influenzae, and as such H. influenzae would require means to defend itself, of which bacteria have many.

1.3 Bacterial resistance mechanisms to bacteriophage infection, and reciprocal bacteriophage adaptations

Logically with lytic predation killing host cells (and predator numbers increasing with every infection event), and prophage-harbouring strains effectively in a tenuous state due to the ever present chance of lysis at any time, without adaptation bacterial populations can become extinct. As such, bacteria have developed mechanism that prevent, or reduce the prevalence of bacteriophage predation within their communities. However, bacteriophage similarly attempt to develop means to overcome their host resistance, bringing with it a reciprocal state of evolution where host and virus attempt to outdo each other, resulting in mass diversification of both organisms. This ongoing struggle is termed the bacteriophage-host 'arms-race' (Stern and Sorek, 2011). Below the arsenal deployed by bacteria to avert the infection attempts by bacteriophage will be discussed, with mention of means that bacteriophage use to circumnavigate their hosts resistance. The focus however will be largely on two such means, restriction modification systems and surface molecule alterations, the two known means of bacteriophage resistance in *H. influenzae*.

1.3.1 Restriction modification systems

Restriction modification systems are likely the best characterised of the bacterial defence mechanisms against bacteriophage challenge. The importance of restriction modification systems in the host evolution is highlighted by their prevalence, as these systems have been identified in around 90% of all bacterial and archael genomes sequenced, with 80% of genomes containing multiple restriction modification systems, and the number of restriction systems present within a cell increasing with genome size (Vasu and Nagaraja, 2013; Roberts et al., 2015). An example of the prevalence of restriction modification systems in *H. influenzae* can be found in figure 1.5. Their importance stems from their key role as a prokaryotic defence to incoming foreign DNA, such as bacteriophage DNA.

Restriction modification systems can be segregated into four groups based on their complexity, target sequence organisation, and their requirements for certain co-factors to function. These four restriction modification groups are described as 'Types' with the groups subdivided as Type I-IV (Roberts et al., 2003; Tock and Dryden, 2005). Type I-III restriction modification systems function in the classical description of restriction modification systems. That is to say that these enzymes are composed of at least two subunits, a methyltransferase (MTase), which acts to modify host DNA at a specific target sequence denoting this DNA as 'self'; and a restriction endonuclease (REase) that acts on DNA which does not contain the methylation signature of the associated MTase at the target site, resulting in cleavage, or 'restriction' of the incoming unmethylated DNA (Bickle and Krüger, 1993). Type IV restriction modification systems were excluded from the above description of the classical restriction modification system mechanism as this Type functions inversely, acting on DNA which is already methylated, with restriction the fate of incoming modified DNA (Roberts et al., 2003; Tock and Dryden, 2005).

Type I restriction modification systems are the most complex of the restriction modification system Types, consisting of three 'host specificity determinant' subunits, termed HsdM, HsdS, and HsdR. The HsdM and HsdR subunits serve as the MTase and REase subunits respectively, while the additional



Figure 1.5. Comparison of restriction modification system diversity in 4 *H. influenzae* strains. Each coloured triangle originating from the central genome indicates a chromosomal region containing a putative restriction modification system. Brown triangles indicate Type I systems, orange triangles Type II systems, and yellow triangles indicate regions encoding Type III restriction modification systems. Coloured arrows on the outer edge of triangles indicate ORFs encoding restriction modification system subunits. The blue arrows indicate MTase genes, red arrows indicate REase genes, the yellow arrows indicate specificity subunit encoding regions, and grey regions indicate ORFs of other or unknown functions. The restriction modification system diversity is represented for four *H. influenzae* genomes a) Rd KW20, b) F3031 c) 86-028NP d) KR494. Images collated from REBASE (Roberts et al., 2015).

subunit, HsdS, is responsible for the designation of the target loci for methylation/restriction by the system (Murray, 2000). These subunits function as pentomeric proteins composed of 2x HsdM 2x HsdR and 1x HsdS. Although it is this pentameric protein that functions as the complete restriction modification system, a trimer of 2x HsdM and 1x HsdS is still able to function exclusively as a MTase; HsdR cannot function without HsdM and HsdS (Dryden et al., 2001).

HsdS contains two domains, each capable of recognising a different sequence, with each domain responsible for identifying half of the total target sequence (Murray, 2000). These two highly variable sequences are separated by a conserved central region that creates the central space between the recognition sequences (Loenen et al., 2014a). This means that Type I restriction modification systems target typically two, often asymmetric, nucleotide sequences of between 2-4 bp per half, which are separated by spacer sequence of \sim 5-8 bp (Loenen et al., 2014a). Examples of there restriction sites include 5'-AAC(N)₆GTGC-3', 5'-GCA(N)₈GTGG-3', and 5'-GA(N)₇TAY-3' for EcoKI, CfrAI, and MpuII respectively (Kan et al., 1979; Kannan et al., 1989; Clark et al., 2012b; Loenen et al., 2014a).

The MTase and REase activity of Type I restriction modification systems requires co-factors to function. For MTase activity S-adenosylmethionine (SAM) is required to act as a methyl donor (Murray, 2000; Tock and Dryden, 2005). For REase activity SAM, adenosine triphosphate (ATP), and Mg²⁺ are required, as both the MTase and REase subunits are needed for REase activity (Murray, 2000; Tock and Dryden, 2005). For REase activity Mg²⁺ catalyses the reaction, while ATP provides energy for translocation of the REase during the restriction process (Murray, 2000; Loenen et al., 2014a). In the presence of hemi-methylated DNA methyltransferase activity is initiated, while if the target sequence lacks methylation, this sequence is a target of restriction activity (although unmethylated DNA can also undergo methylation). Methylation is believed to occur through inverting the base to be methylated to within the catalytic domain of the MTase. For restriction, two pentomeric proteins are required, with each of the REase proteins pulling the DNA bi-directionally, until the two pentomeric proteins collide, producing a restriction site of variable location between the two recognition sites (Srikhanta et al., 2010).

Type II restriction modification systems are the family of restriction enzymes that have revolutionised the world of molecular biology through there extensive use in the processes of molecular cloning. Type II restriction modification systems are probably the best characterised of the restriction modification Types, largely due to their applications in molecular biology. Type II restriction modification systems traditional consist of a monomeric MTase, and either a monomeric or a homodimeric REase, with these two enzymes normally acting independently of each other (Roberts et al., 2003). The MTase will act on unmethylated DNA at the target sequence to methylate the double stranded DNA, providing either m4C, m5C, or m6A methylation, thus designating it immune to the action of conventional Type II REase activity (Murphy et al., 2013).

The classical view of Type II REases is that their target is an unmethylated symmetric DNA sequences of around 4-8 bp in length, and that cleavage exposes 5'-phosphate group and 3'-hydroxyl group (Pingoud and Jeltsch, 2001). However, Type II restriction modification systems can differ in components, and Type II REases can be divided into 11 subtypes, termed as A, B, C, E, F, G, H, M, P, S, and T based on such features as recognition sequence, form of cleavage, and number of sequences targeted (Roberts et al., 2003). For instance Type IIA and Type IIS REases target asymmetric sequences with Type IIA and Type IIS REases distinguished as Type IIS enzymes cleave at least one DNA strand outside of the target sequence, although most cleave both strands outside of the target sequence, whereas Type IIA cleave asymmetric sequences within the target sequence. Members of the Type Type IIC, Type IIG, Type IIH, and Type IIT REases can target either symmetric or asymmetric sequences. These examples can be distinguished further, whereby: Type IIC REase and MTase are part of the one polypeptide; Type IIG REases are stimulated or inhibited by AdoMet; Type IIH REases are genetically more similar to Type I systems, in that they are composed of three parts (M, S, and R); and Type IIT REases are heterodimers. The Type IIE and Type IIF systems both require two target sequences, however, while Type IIF systems restrict both target sequences, Type IIE REases only cleave one site, with the other activating the restriction of the cleaved site. Type IIM enzymes target only methylated DNA, however, these are separated from Type IV systems as Type IV systems do not cleave at a fixed site. The Type IIB REases cleave DNA at both ends of the target sequence on both strands. Lastly, Type IIP REases are those that encompass the classical description of Type II restriction enzymes, those that target palindromic sequences (thus Type IIP) and cleave within the sequence, or just outside (Pingoud and Jeltsch, 2001; Roberts et al., 2003; Pingoud et al., 2005; Tock and Dryden, 2005; Pingoud et al., 2014; Roberts et al., 2015).

It should be noted that these subgroups of Type II REases, are not mutually exclusive (Roberts et al., 2003). For instance AlfI of *Acinetobacter lwoffi* falls into both the Type IIB and Type IIP groups, as it cleaves DNA at both ends of the target sequence but recognises a palindromic sequence, namely recognising 5'-GCANNNNNTGC-3' and cleaving as (10/12)GCANNNNNTGC(12/10) (Roberts et al., 2015).

The classical interpretation of Type III restriction systems pertains that these systems act similarly to Type I systems, in that they target asymmetric DNA sequences, the restriction activity is facilitated by a complex of the MTase and REase proteins, the co-factors SAM, Mg^{2+} and ATP are required for complete restriction modification activity, and the restriction of DNA found to be unmethylated occurs away from the site of recognition (anywhere from 25-27 bp away from recognition sequence) (Dryden et al., 2001; Loenen et al., 2014b). Examples of such sequences include 5'-AGACC-3', 5'-CAGAG-3', and 5'-CGAAG-3' for EcoP11 (*E. coli*), SbaU (*Salmonella enterica* Subsp. *enterica* serovar Bareilly) and BceS (*Bacillus cereus*) respectively (Bächi et al., 1979; Xu et al., 2012; Pirone-Davies et al., 2015). However, unlike Type I restriction modification systems these enzymes consist of only two subunits, that is Mod and Res, encoding the methyltransferase subunit and restriction endonuclease subunit respectively. Another difference between Type I (also Type II) and Type III is that with Type III systems methylation of the target sequence does not occur on both strands but rather only occurs on one strand.

The Mod protein contains the target recognition domain, and as such for restriction activity a complex containing both the Mod and Res subunits are required in a 2x Mod and 2x Res complex. Restriction occurs through two separate Mod/Res complexes binding to two separate target recognition sequences, with these sequences orientated in opposite directions (i.e. same sequence occurring on opposite strands) (Meisel et al., 1992). From here the two Mod/Res complexes will translocate the DNA in the direction of the other Mod/Res complex, with collision of the two complexes resulting in restriction of the DNA molecule in question (Srikhanta et al., 2010).

Type IV restriction systems (not restriction modification as these systems do not encode methyltransferases), as mentioned, are a group of systems which target incoming DNA that has already undergone modification. Modification can be in the form of methylation, hydroxymethylation and glucosylhydroxymethylation within the target sequence (Tock and Dryden, 2005). Although poorly understood, these systems are often encoded by 1-2 genes, and generally have poorly defined restriction sites with cleavage occurring at variable distance from the recognition sequence (Roberts et al., 2003). Some examples of Type IV systems include SauUS of *Staphylococcus aureus* that recognises 5'-S5mCNGS-3' and cleaves the DNA between 2-18 bp downstream of the recognition site (Xu et al., 2011) and EcoKMcrBC that requires two iterations of a purine followed by a methylated cytosine to cause restriction in sequences can be separated by large distances in the form of $RmC(N)_{40-2000}RmC$ with restriction occurring near one of these sites (Loenen and Raleigh, 2014).

Each of these restriction system Types has been associated with resistance to bacteriophage infection. For instance the Type I system SauI provides resistance against ϕ 75 for *S. aureus* (Waldron and Lindsay, 2006), the Type II system LlaBI provides resistance against bacteriophage jj50, and c2, in *Lactococcus lactis* subsp. *cremoris* (Nyengaard et al., 1996), the Type III system EcoPI can provide resistance against bacteriophage λ for *E. coli* (Linn and Arber, 1968; Rao et al., 2014), and lastly, the *Streptomyces coelicolor* Pgl system is a Type IV system providing resistance against ϕ C31 (Sumby and Smith, 2002). The effects of restriction can reduce bacteriophage infectivity, as measured by efficiency of plating (the number of plaque forming units noted from a lawn of test bacteria compared to its control), from 10⁻¹ to 10⁻⁵ PFUs (Loenen et al., 2014b) and potentially higher. The efficiency of restriction systems on incoming DNA is a function of the number of target sequence sites within the incoming DNA fragment (Krüger et al., 1988).

The observation of a high effectiveness and multiple restriction modification systems per genome suggests that a strong selective pressure exists for evolutionary maintenance of these systems. The role of these systems in preventing access of foreign DNA implies that the most abundant organism on the planet, the bacteriophage, may be a highly significant contributor to this pressure.

However, although bacteria can utilise these systems to avoid bacteriophage predation, the systems themselves are not without fault. Restriction modification systems often act as 'leaky' barriers. That is to say they do not act as an absolute resistance mechanism, but rather serve more as an imperfect resistance mechanism. They are flawed mechanisms due to the potential for inadvertent methylation of bacteriophage DNA as occurs if the hosts methylation enzyme acts at the target site before restriction activity (Kuhnlein and Arber, 1972; Samson et al., 2013). This would designate the bacteriophage DNA as self, permitting its replication, with the progeny of this bacteriophage similarly containing methylation that will provide immunity. This effectively allows the progeny bacteriophage population

to move undetected in host populations carrying restriction systems where the methylation signature is compatible with the self. For example, in *S. aureus* and its bacteriophage, phage K, Kelly et al. (2011) screened a number of *S. aureus* isolates for susceptibility to phage K and isolated 24 examples that were resistant to phage K through restriction modification activity. However, after around 2-3 exposures the phage K population had overcome the restriction modification activity allowing bacteriophage K to clear cultures of the previously resistant host (Kelly et al., 2011).

Bertani and Weigle (1953) was probably the first study to describe the phenomenon of bacteriophage resistance to restriction modification systems through host assigned methylation, meanwhile inadvertently providing evidence for a method for the bacterial host population to negate the bacteriophage populations new countermeasure. Using the restriction proficient *E. coli* K-12 derivative strain S, and the restriction modification deficient *E. coli* strain C, they observed that phage λ progeny produced following infection of strain S, termed λ S, were equally able to infect both strain S, and strain C. Meanwhile, for λ progeny produced from infection of strain C, termed λ C, the bacteriophage was still as efficient at infecting strain C, but a reduction in efficiency of plating was observed on strain S. Furthermore, the progeny of λ C passaged through S are again efficient in infection of strain S (Bertani and Weigle, 1953). A diagram of this dynamic can be found in figure 1.6.

Bacteriophage can also actively utilise other means to avoid the action of restriction modification systems including the use of proteins that mimic DNA, blocking of restriction sites, or the removal/alteration of DNA sequences to avoid restriction (Samson et al., 2013). The T3 and T7 bacteriophage encode a DNA mimicking protein, Ocr (overcoming restriction), that reduces sensitivity to bacterial restriction (Krüger and Bickle, 1983; Murray, 2000). Ocr mimics the structure of a 24 bp bent molecule of the classical right-handed double helix B-DNA, that provides resistance to multiple subgroups of Type I restriction modification systems, regardless of recognition sequence (Walkinshaw et al., 2002). Ocr is encoded by gene 0.3 in T7 in the first most region to be injected into the bacterial cell. Injection occurs in a two step process where a small section of DNA, containing gene 0.3, is injected allowing the production of Ocr while preventing the exposure of restriction site containing bacteriophage DNA to the host restriction enzymes (Krüger and Schroeder, 1981; Hyman and Abedon, 2010). Ocr prevents restriction of bacterial DNA by binding the host restriction modification system with a greater affinity than the bacteriophage DNA, inactivating it, and maintaining this association for a long time period (Atanasiu et al., 2002).

Some bacteriophage are able to mask their restriction sites not by methylation, but rather encoding proteins that bind the injected bacteriophage DNA. The P1 bacteriophage, injects the products of two genes, darA and darB (defence against restriction), which reduce the sensitivity of the bacteriophage to restriction by a number of Type I restriction modification systems by binding to the packaged DNA of the bacteriophage and blocking restriction (Iida et al., 1987).


Figure 1.6. Illustration of the impact and pitfalls of restriction modification activity on bacteriophage infection dynamics. In (a) bacteriophage progeny from infection of a cell devoid of restriction activity proceed to infect another bacterial cell free of restriction modification activity, thus infectivity is uninhibited. However, in (b) bacteriophage progeny from infection of a cell devoid of restriction modification activity attempts to infect a cell containing an active restriction modification system targeting a sequence within the bacteriophage DNA. The unmethylated bacteriophage is therefore susceptible to restriction activity, causing in a reduction in bacteriophage infectivity. For (c) any successfully produced bacteriophage progeny from infection of the restriction modification system containing cell would be methylated at the target sequence. As such, progeny would be able to infect another cell containing this active restriction modification system and would be able to replicate uninhibited due to the methylation of its DNA. Lastly, in (d) methylated bacteriophage produced from the infection of restriction modification containing cells will again be free to infect cells lacking the restriction modification activity. However, the resulting progeny from this infection will lose the methylation signature and be susceptible to downstream action by cells containing restriction modification systems. Tan cells are devoid of restriction modification systems, and pink cells contain active restriction modification systems. Bacteriophage with blue capsids are not methylated at target sequence, while bacteriophage with pink capsids are methylated at the restriction modification systems recognition sequence. In addition to encoding mechanisms that allow the bacteriophage to inhibit the restriction process many bacteriophage have also undergone genome alterations that allow them to evade restriction modification systems. For instance a number of the bacteriophage of *B. subtilis* have lost the thymine residue from their DNA sequence, replacing it with 5-hydroxymethyluracil (Krüger and Bickle, 1983). Such alterations can reduce the effectiveness of restriction modification systems at restricting the foreign DNA (Berkner and Folk, 1979). Additionally despite containing 36 EcoP15 sites within its genome bacteriophage T7 is resistant to restriction by this Type III REase. This is because as mentioned previously Type III REases need to recognise two inversely orientated asymmetric target sites in order to facilitate the collision event that permits restriction of the foreign DNA and all 36 EcoP15 sites within T7 are orientated in the sane direction (Meisel et al., 1992).

1.3.2 Surface structures

Surface structures are the main region of bacteriophage targeting enabling the bacteriophage replication cycle to begin, by facilitating adsorption of the bacteriophage to the host cell. Thus bacteriophage apply a strong selection pressure on their host to alter the presence and/or composition of molecules on their surface. This has become particularly evident since the development of large-scale analysis of bacterial groups through metagenomics. Rodriguez-Valera et al. (2009) discussed the presence of 'metagenomic islands', genomic regions that are under-represented within the metagenome, within sequenced marine bacterial species compared to the total marine metagenome. They observed that genes encoding surface exposed structures were consistently present within the metagenomic islands. These surface related genes included those involved in LPS, pili, and flagella biosynthesis, with their under-representation implying a high degree of divergence across the species investigated (Rodriguez-Valera et al., 2009). Many of the surface structures they identified within their metagenomic islands are targets for bacteriophage adsorption in a number of bacterial species, as can be seen from the examples in table 1.1.

Mutations in surface structure can be observed very soon after exposure to bacteriophage. Scanlan et al. (2015) noted that after co-incubation of *P. fluorescens* SBW25 and bacteriophage Φ 2, mutations in LPS biosynthesis associated genes had arisen at the earliest time point tested of around 14 bacterial generations, with the four earliest time points tested (14, 28, 42, and 56 generations) containing mutations only in LPS biosynthesis genes. Analysis of the LPS structures of bacterial populations after co-evolution with bacteriophage detected variable lengths of O-antigen, while the wild-type and control populations (i.e. those evolved without bacteriophage) had the same O-antigen length (Scanlan et al., 2015).

The degree of bacterial surface variability that allows generation of bacteriophage resistance in bacterial populations is deemed highly problematic for the implementation of bacteriophage as a therapeutic.

The second se		0	
Bacteriophage	Host	Host receptor target	Reference
$\mathbf{phage}\gamma$	$Bacillus\ anthrac is$	GamR	Davison et al. 2005
SPP1	$Bacillus\ subtilis$	YueB	São-José et al. 2004
T4	Escherichia coli	OmpC and LPS	Wilson et al. 1970; Datta et al. 1977; Henning and Jann 1979
Y	Escherichia coli	LamB	Randall-Hazelbauer and Schwartz 1973
c2	Lactococcus lactis	PIP	Geller et al. 1993
$\phi {f K} {f M} {f V}$	Pseudomonas aeruginosa	type IV pilus	Chibeu et al. 2009
JG004	Pseudomonas aeruginosa	LPS	Garbe et al. 2011
MPK6	Pseudomonas aeruginosa	LPS group B polysaccharide	Heo et al. 2009
ES18	Salmonella enterica	FhuA	Killmann et al. 2001
7, ST29, and ST35	Salmonella enterica	TolC	Ricci and Piddock 2010
fsy-1 and Gifsy-2	Salmonella enterica	OmpC	Ho and Slauch 2001
N1S and 6 others	Salmonella enterica	O-antigen of LPS	Shin et al. 2012
N2T and 8 others	Salmonella enterica	FliC flagellin	Shin et al. 2012
N4S and 4 others	Salmonella enterica	Either types (FliC or FljB) of flagellin	Shin et al. 2012
	Salmonella enterica,		
tiple bacteriophage	Escherichia coli, and	ButB	Shin et al. 2012
	Shigella flexineri		
2mw and 5 others	Staphylococcus aureus	wall teichoic acid	Xia et al. 2011

 Table 1.1. Examples of surface molecules used for adsorption of bacteriopahge.

In light of this, bacteriophage therapy relies on bacteriophage 'cocktails', effectively mixes of multiple bacteriophage that should ideally target different surface molecules. Bacteriophage resistance is so problematic that some researchers have noted that even cocktails containing 45 different bacteriophage in one mix, although successful in substantially reducing the concentration of bacteria, do not completely eliminate bacteria over the test period (Andreatti Filho et al., 2007). These observations were attributed to the possibility of bacteriophage resistance (Andreatti Filho et al., 2007).

Bacteriophage are able to utilise a number of host derived structures to facilitate their adsorption, examples of which can be found in table 1.1. For a number of the known bacteriophage receptors their recognition as the bacteriophage receptor stemmed from the natural acquisition of resistance to bacteriophage by the host followed by screening of the genetic differences between the wild-type and resistant derivatives. This was true for Le et al. (2014) who identified a bacteriophage resistance mechanism used by *P. aeruginosa* PA1, that was conferred through the deletion of a large stretch of chromosomal DNA of around 200-300 kb, which always involved the deletion of galU, a gene involved in LPS biosynthesis. The galU gene encodes UDP-D-glucose pyrophosphorylase, an enzyme responsible for the conversion of Glc-1-phosphate to UDP-D-Glc, removal of galU thus resulted in defective LPS (Le et al., 2014). This was confirmed by generation of galU deficient strains, and the observation that the resulting defective LPS significantly reduces bacteriophage PaP1 adsorption, and as such confirmed that it was this gene that conferred bacteriophage resistance (Le et al., 2014).

Although loss of the structures being targeted on the bacteriophage surface is an efficient strategy to prevent bacteriophage infection, it does remove the ability to utilise the physiological role of this receptor from the cells repertoire. As such, bacteriophage resistance can also be achieved while maintaining function through generation of only small alterations to the surface structure. For example, in *E. coli* the TolC protein is used by the TLS bacteriophage as the surface receptor for adsorption (German and Misra, 2001). Amongst other roles TolC serves the bacterial cell as an efflux pump for bactericidal agents, including antibiotics (Fralick, 1996). As bacterial mutants deficient for TolC display an increased sensitivity to antibiotics if bacteriophage resistance arose as a result of complete loss of TolC the bacteriophage resistant cells would now have a high degree of sensitivity to antibiotic stress. When German and Misra (2001) generated isolates of *E. coli* resistant to bacteriophage TLS, they observed that complete loss of TolC was the prefered method of resistance to TLS compared to the structure remaining functional, albeit altered (frequencies were around 10^{-6} to 10^{-9}). Those cells that had lost TolC function were now ~2x more sensitive to novobiocin, while those cells resistant to TLS through various missense mutations in TolC either had the same sensitivity as the wild-type strain or only slightly higher sensitivity (German and Misra, 2001).

Although a bacterial host can alter its cell surface to prevent bacteriophage adsorption, receptor changes likewise necessitate that the bacteriophage itself adapts to these changes to prevent its extinction. One means bacteriophage have at their disposal is altering their anti-receptor components, the bacteriophage structure involved in the interaction with bacterial surface molecules to facilitate adsorption (Samson et al., 2013). Alterations in bacteriophage tail fibre genes are known to cause alterations in host range. For example Duplessis and Moineau (2001) found that they could alter the infectivity of a broad host-range *Streptococcus thermophilus* bacteriophage, DT1, through eliciting recombination of this bacteriophage with the putative anti-receptor encoding region of another *S. thermophilus* phage, MD4, which itself was only able to infect one of the tested stains. The resulting five recombinant bacteriophage isolated DT1.1-DT1.5 were now only able to form plaques on the single *S. thermophilus* strain bacteriophage MD4 could infect. It is worth noting however that MD4 is unable to adsorb to SMQ-301, yet these recombinant DT1 bacteriophage retain the ability to adsorb SMQ-301, but with reduced affinity (Duplessis and Moineau, 2001).

1.3.3 Other Resistance Mechanisms

Although the resistance developed by bacteriophage through alteration of their surface structures and through the use of restriction modification systems are the main focus of this current work, these are not the only means by which bacteria can attempt to avoid predation by bacteriophage. Other bacterial defence mechanisms to bacteriophage infection exist, including superinfection exclusion systems (Sie), CRISPR-cas, and abortive infection (Abi) systems (Labrie et al., 2010).

Sie systems are resistance mechanisms that are often provided to a host by a prophage element already found within the bacterial cell. These systems can prevent further infection, or 'superinfection', by another bacteriophage (Seed, 2015). These mechanisms aid the hosts resistance to incoming bacteriophage by blocking the entry of any foreign bacteriophage DNA into the cell. Such mechanisms include the well characterised Imm protein of bacteriophage T4, that provides resistance for the bacterial host against superinfection by other T-even bacteriophage (Anderson and Eigner, 1971). The Imm protein acts to block new bacteriophage DNA, in the process expelling the released DNA into the cells periplasm (See Figure 1.7 (b)), where it is then degraded by the action of endonuclease I (Anderson and Eigner, 1971).

CRISPR-cas systems are described as a form of bacterial 'adaptive immunity' against bacteriophage infection (Bikard and Marraffini, 2012). This is because CRISPR-cas systems are composed of short sequences that are remnants of a bacteriophage that previously attempted to infect the host cell. CRISPR-cas loci consist of two main regions, the region encoding the cas genes, and the region containing the CRISPR array (Szczepankowska, 2012). The CRISPR array is the region of DNA that contains the short variable spacers that are the remnants of attempted infections, with these variable regions separated by short repetitive sequences of DNA (i.e. the coloured and black boxes, respectively, of figure 1.7 (d)). The cas genes are the effectors of the CRISPR-cas system, with roles such



however, the bacterial host already contains a prophage element that encodes an Sie system, which produces a protein (e.g. Imm of T4) that prevents entry of bacteriophage DNA. In (c) the host encodes a restriction modification system which cleaves the unmethylated bacteriophage DNA preventing successful infection. For (d) a bacteriophage methylated at the restriction target site is able to inject its DNA, however, the cell contains a CRISPR-cas infection process is prevented. Lastly, (e) shows the outcome of infection of a cell with an active Abi system. Recognition of bacteriophage infection Figure 1.7. Examples of bacterial resistance mechanisms to bacteriophage infection. In (a) the host has altered its surface molecule from a previously system with spacer sequence matching a region of the bacteriophage DNA. Through the generation of crRNA targeting the invading virus, a successful infectable form to a new variant unrecognisable to the bacteriophage. For (b) a bacteriophage with the correct receptor specificity is able to bind, results in activation of an ion channel that allows cation passage, reducing the cells membrane potential, and ultimately resulting in cell death. as acquiring new spacers; processing pre-CRISPR RNA (crRNA) formed following the transcription of the CRISPR array into mature crRNA; and the eventual targeting, binding, and degradation of incoming DNA (Richter et al., 2012; Nuñez et al., 2014).

In general the process of CRISPR-cas mediated bacteriophage resistance results from previous exposure of the bacterial host to foreign DNA, from which a small fragment is inserted into the CRISPR-array. The 'remembered' fragments are then converted into pre-crRNA where the entire CRISPR-array, including both the spacers and repeats, is transcribed as a whole. This pre-crRNA is then processed into mature crRNA following endonuclease activity within the repeat regions that results in the release of the mature crRNA containing an individual spacer sequence and a part of the repeat sequence (separated transcribed fragments in figure 1.7 (d)). From here additional cas proteins are recruited that results in the generation of a complex that allows the CRISPR-based surveillance of foreign DNA, and ultimately cleavage (Richter et al., 2012; Terns and Terns, 2011; Sorek et al., 2013).

Abi systems are described as a form of altruistic defence to bacteriophage infections, as the infected cell effectively commits suicide in an attempt to prevent bacteriophage spread to the remainder of the bacterial population. Figure 1.7 (e) illustrates the mechanism of action of the the RexAB system of *E. coli*, encoded by bacteriophage λ . RexAB consists of two main proteins RexA, and RexB. RexA is an intracellular protein that is activated only upon injection by bacteriophage DNA, while RexB is a membrane anchored ion channel that itself is only activated upon interaction with at least two active RexA molecules. Once RexB is activated cations are allowed to leave the cell, reducing membrane potential of the cell and with it reducing intracellular levels of ATP, and ultimately resulting in cell death. Thus little to no bacteriophage molecules will be able to be synthesised and released, and as such will prevent the spread of bacteriophage to neighbouring cells (Labrie et al., 2010; Dy et al., 2014; Seed, 2015).

1.4 Phase variation

Phase variation is the high frequency, stochastic, reversible, alteration in gene expression (van der Woude, 2011; Moxon et al., 2006; Bayliss, 2009). These changes in gene expression can manifest as binary gene expression changes, i.e. ON-OFF and OFF-ON switches; modular changes increasing or decreasing gene expression levels (e.g. full, partial, absent expression); or completely switching between distinct active phenotypic activities (i.e. antigenic switching) (Moxon et al., 2006; Bayliss, 2009). Each of these gene expression changes results in the bacterial populations being able to generate heterogeneous populations for expression of the gene(s) under phase variable control. This allows bacterial populations to create profoundly diverse populations depending on the number of phase variable genes within the organisms chromosome. For example, if a bacterial population contained 8 loci within their chromosome with binary switching ON-OFF/OFF-ON characteristics, a further 2 loci that switch between two alleles, and 4 genes which express high, low, and abolished expression this would mean the population could generate the following number of distinct populations:

 $x = 2^{8+2} \times 3^4 = 82944$ potential phenotypes in the bacterial population

The ability of bacterial species to generate populations composed of a high degree of heterogeneity through phase variation allows the population to be ready for a potential switch in the environmental conditions to that which is unfavourable to the current dominant phasotype, and in-turn facilitate niche adaptation (van der Woude and Bäumler, 2004). The environmental selection pressure can be in the form of factors of the host immune system, predation by bacteriophage, adaptations required for entry into a new niche, or other abiotic stresses (Moxon et al., 2006).

1.4.1 Mechanisms of phase variation, and its role in survival

Phase variation can arise as the result of a number of mechanisms. The two most commonly discussed of mechanisms of phase variable adaptation are the phenomenon of site-specific recombination, and slipped strand mispairing (SSM), although phase variation can also occur via non-mutational means, namely epigenetic regulation (van der Woude and Bäumler, 2004; Moxon et al., 2006; Bayliss, 2009). Phase variation occurs in hypermutable regions of DNA termed 'contingency loci' that mediate the high frequency genotypic switching (Moxon et al., 2006). An example of one such phase variation mechanism are simple sequence repeats (SSRs), these are repetitive elements of DNA (with the repetitive units often consisting of 1-6 nucleotides) that can expand or contract in unit number and in turn alter gene expression (Moxon et al., 2006). SSRs are believed to alter in unit number due to slippage of DNA polymerase on these repetitive DNA sequences during replication through SSM (Streisinger et al., 1966). The key event occurs as the DNA polymerase is generating a new strand and is believed to involve separation of the nascent strand from its template within the SSR tract. When these strands re-anneal, misalignment of the repeat tract would result in either addition or deletion of a repeat unit from the SSR on the newly synthesised strand depending on which strand misalignment occurs (Figure 1.8).

Addition or deletion of repeat units from the SSR tracts can alter gene function though multiple means depending on its location. If the SSR tract is located within the open reading frame (ORF) of a gene then alteration in the SSR tract number will result in a reading frame shift, potentially either introducing a premature stop codon or reinstating gene expression from a non-functional frame. Examples of SSM within gene ORFs are found in a number of bacteria, including *Helicobacter pylori*, *Campylobacter jejuni*, and *Neisseria gonorrhoeae* (Karlyshev et al., 2002; Sanabria-Valentín et al.,



Figure 1.8. Streisinger model of slipped strand mispairing. SSM is believed to arise through a slippage of DNA polymerase during DNA replication resulting in separation of the replicating DNA strands. If, after strand separation, SSR repeats misalign upon re-annealing then addition or deletion of a repeat unit will occur. If this misalignment occurs on the nascent DNA strand (left hand side) then this will result in the addition of a repeat unit to the newly synthesised strand, meanwhile, if this occurs on the template strand (right hand side) the replicating strand will contain one less repeat upon completion of synthesis. Figure is adapted from (Moxon et al., 2006)

2007). Such is the case for tcpH in Vibrio cholerae, where SSM along a poly-G tract (G₉ to G₁₀) within the tcpH ORF prevents expression (Carroll et al., 1997). tcpH is involved in the regulation of genes important for V. cholerae, including the ctxAB and tcpA, involved in the production of cholera toxin and the toxin-co-regulated pilus respectively (Carroll et al., 1997; Beck et al., 2004). Both of the regulated genes are key virulence factors in V. cholerae (Chakraborty et al., 2000), highlighted by the observation that the phase OFF variant of tcpH are avirulent in infant mouse models (Carroll et al., 1997). In Moraxella catarrhalis a poly-A tract controls ON/OFF switching of the uspA2H gene, which encodes a surface structure that can function as an adhesin, and may be involved in resistance of the cell to human sera (Lafontaine et al., 2000). SSM along SSRs is also the main method of phase variation in H. influenzae, where phase variation occurs in this manner mainly through repetitive tetranucleotide repeats (Power et al., 2009).

Neisseria meningitidis serves as a good model organism for the impact of phase variable switching in gene expression mediated by SSM, due to an expansive array of SSR tracts (Bayliss et al., 2001). It also contains well characterised loci that highlight the potential role of SSR containing ORFs in the

control of gene expression in survival against antagonistic selections. For instance, N. meningitidis serogroup B isolates undergo phase variation of their capsule due to the presence of a homopolymeric repeat within the ORF of the capsule biosynthesis gene siaD (Hammerschmidt et al., 1996). Slippage along the poly-C tract from C₇ to C₆ results in a reading frame shift leading to premature truncation of the siaD ORF, and thus an inability to express the capsule protein (Hammerschmidt et al., 1996).

The *N. meningitidis* capsule is a key means of defence to the host immune system, serving to increase resistance to human sera (Vogel et al., 1997), and as such encapsulated strains predominate during invasive disease (Bayliss et al., 2001). However, during carriage, it is the non-encapsulated strains which predominate (Bayliss et al., 2001). This may coincide with the observation that cells expressing the capsule seem to be at a disadvantage during invasion (Hammerschmidt et al., 1996), and that cells expressing the capsule are inhibited in their ability to form biofilms, a characteristic that is believed to be important in their colonisation (Yi et al., 2004). As such the existence of an ON/OFF phase variation of the cells capsule could be of benefit so, during the initial colonisation of the host, non-encapsulated cells can invade and colonise, with populations that then go on to more invasive locations able to survive due to the constant generation of sub-populations that have increased resistance to sera. Thus the majority phasotype in the population will switch from non-encapsulated to encapsulated in the invasive location, allowing the progeny of a serum sensitive population to colonise a niche for which it was originally maladapted.

If the SSR tract is located outside of an ORF it can still effect gene expression (van der Woude and Bäumler, 2004). For instance, if a phase variable SSR tract is located in the promoter region of a gene SSR tract length changes may alter the ability of RNA polymerase to bind to its promoter sequence due to creating a reduced affinity between RNA polymerase and the -10 -35 RNA polymerase binding sites, potentially even impeding its binding completely (Zhou et al., 2014).

In the *porA* gene of *N. meningiditis*, SSM along a poly-G tract within the -10 -35 promoter region mediates modular switching in expression levels of the porin protein PorA (van der Ende et al., 2000). Phase variation from a higher expression state to a lower expression state has been noted to allow bacterial evasion of bactericidal antibodies (Tauseef et al., 2013). Switching from an initial G_{11} SSR tract to G_{10} and G_9 caused increasing survival ~2-fold and ~13-fold respectively, due to progressive reductions in PorA levels (Tauseef et al., 2013). The clinical role of this remains unknown but clinical isolates with low levels of PorA expression have been observed (van der Ende et al., 2000). However, logically, the designation of a loci that is a target for the host immune system as having modular expression levels would be of benefit. During infection, depending on the strength of selection (in this case the strength of the immune response) the bacterial population would be in a position to adjust the dominant phenotype from, for example, one expressing high levels of a protein, to a reduced but still functioning level as the high expression cells are killed off, allowing maintenance of the proteins function while also reducing the negative effect its expression.

Phase variable alterations in gene expression can also occur by site-specific recombination (McCusker et al., 2008; Honarvar et al., 2003; Li et al., 2002). Site-specific recombination involves a region of DNA being flanked by invertible repeat sequences, with recombination at these sequences resulting in the inversion of the stretch of DNA sequence contained between the two invertible repeat sequences (van der Woude, 2011). For instance, if a promoter sequence is held within a region of DNA flanked by invertible repeat sequences, the promoter sequence can become inverted upon recombination at the inverted repeats (van der Woude, 2011). If the promoter was originally facing a coding sequence then altering its orientation would prevent the expression of this coding sequence. If a correctly orientated additional coding sequence is found in the new orientation then the inversion of the promoter may result in transcription of this alternative coding sequence thus allowing switching between two distinct phenotypic activities. However, if the new orientation the promoter is facing is a non-coding region then this will effectively result in the loss of the initial phenotype and thus act as a form of ON/OFF switching (Bayliss, 2009; van der Woude, 2011). Recombination does not only contribute to phase variation through inversion of promoter elements, but can cause reversible genomic rearrangements that give rise to new phenotypes. The elements being rearranged are again flanked by inverted repeats, however the recombination that occurs may invoke insertion of previously non-transcribed alleles to within a genes ORF. The production of such hybrid ORFs thus creates new alleles that a host can express, and allow for creation of a number of new phenotypes to aid in cell survival (Manso et al., 2014; Cerdeño-Tárraga et al., 2005; van der Woude and Bäumler, 2004).

It is not only diversifications in DNA sequences that can produce phase variable expression changes, as phase variation can also occur through epigenetic regulation. Here differential states of methylation of regulatory sequences provide distinctive gene expression states (Bayliss, 2009; van der Woude, 2011). The classical example of this form of epigenetic regulation is the *pap* locus of *E. coli*, which encodes the pyelonephritis-associated (P) pilus (van der Woude and Bäumler, 2004). This occurs due to differential methylation by the Dam methylase within the two sets of three leucine-responsive regulatory protein (Lrp) binding sites of the promotor region of the *papBA* locus. When the proximal Dam sites are unmethylated Lrp will bind this site and prevent the expression of the *papBA* locus. However, when levels of PapB are low in the cell this increases activation of the *papI* locus, increasing the production of PapI. PapI acts to form a complex with Lrp that, following DNA replication, causes Lrp to bind the distal hemimethylated Dam site. The binding of the PapI-Lrp, and dissociation of Lrp to the proximal Dam site, allows the methylation of the proximal Dam site, and expression of the *papBA* locus. Meanwhile, high levels of PapB repress PapB expression and thus trigger ON-OFF switching (van der Woude and Bäumler, 2004; Bayliss, 2009).

As can be seen a number of these phase variable loci discussed encode surface exposed structures.

Although not observed experimentally in these systems, parallels of these systems have been identified as being targets for adsorption of bacteriophage, with common targets of bacteriophage adsorption including capsules, porins, pili, flagella, and LPS epitopes (Bertozzi Silva et al., 2016). Phase variation of such loci would thus allow a transient resistance to bacteriophage predation, similar in fashion to the transient resistance to the host immune system.

1.4.2 Phase variable bacteriophage resistance mechanisms

A number of bacterial species have been identified which contain genes under phase variable control that alter the bacterial hosts resistance/susceptibility to bacteriophage infection (Table 1.2). The presence in bacteria of resistance mechanisms under phase variable control allows the population to be in a constant state of preparation for the invasion of their niche by bacteriophage; provides a largely diverse heterogeneous population that may confer resistance to a large variety of bacteriophage; facilitate the use of any genetic element that would be selected against by bacteriophage presence until the selection pressure becomes too great; and allows members of the population to prevent constant expression of certain genes that are costly to express in the absence of bacteriophage predation.

As mentioned phase variation occurs extensively in surface exposed molecules, and as noted in section 1.3.2 for a number of Gram-negative prokaryotes the LPS is a target of bacteriophage adsorption. Phase variation of LPS associated genes is not uncommon (Zhou et al., 2014), and has been noted in a number of bacterial species (See Table 1.2). The presence of contingency loci within these genes has been both shown, and hypothesised, to be a means of regulating phage resistance.

In V. cholerae, an organism where phage predation is believed to play a key role in decreasing epidemic spread of cholera (Faruque et al., 2005b), phase variation of two LPS biosynthesis related genes has been identified where phase variation of each gene results in distinct levels of bacteriophage resistance. Seed et al. (2012) observed that upon isolation of colonies from the center of plaques formed from infection of the V. cholerae O1 strain E7946 with bacteriophage ICP1 a number of colonies had deletions in the poly-A SSR tracts within two genes, wbeL and manA. These genes are exclusive to the O1 group of V. cholerae with wbeL involved in tetronate acytlation (Stroeher et al., 1995), and manA encoding a type I phosphomannose isomerase (PMI) (Seed et al., 2012).

Phase variation of wbeL from the ON state to the OFF state, denoted in Seed et al. (2012) as $wbeL^*$, occurred by a single nucleotide deletion to A_7 within the normal A_8 tract of the wbeL ORF. This deletion results in a reading frame switch from the production of a 471 amino acid peptide to production of a truncated 42 amino acid product, in the process conferring absolute resistance to bacteriophage ICP1. Phase variation of manA, does not confer complete resistance like the $wbeL^*$ phenotype, but rather confers a reduction in the hosts susceptibility to ICP1 challenge. Quantitatively, following

)		
Organism	PV Gene	Product	Reference
Campylobacter jejuni	cj1421	O methyl phosphoramidate moiety of capsular polysaccharide	Sørensen et al. 2011
Campylobacter jejuni	cj0031c	Type II restriction modification system	Anjum et al. 2016
Campylobacter jejuni	Multiple	DNA inversions that activates a Mu-like prophage element	Scott et al. 2007
Clostridium difficile	cwpV	cell wall protein hypotesised to act as Sie system	Sekulovic et al. 2015
Haemophilus influenzae	lic2A	glycosyltransferase of LOS	Zaleski et al. 2005
Haemophilus influenzae	hsdM	methyltransferase of Type I RM system	Zaleski et al. 2005
Neisseria gonorrhoeae	hsdS	Specificity subunit of Type I RM system	Adamczyk-Poplawska et al. 2011
Pseudomonas aeruginosa	mucA	anti-sigma factor σ^{22}	Latino et al. 2016
Pseudomonas aeruginosa	mzm	O-antigen polymerase	Latino et al. 2016
Pseudomonas aeruginosa	wzz2	regulates O-antigen chain length	Latino et al. 2016
Pseudomonas aeruginosa	wbpL	glycosyltransferase involved in LPS synthesis	Latino et al. 2016
Salmonella enterica	STM2209- $STM2208$	Inner membrane protein that alters O-antigen length	Cota et al. 2012
Salmonella enterica	gtrABC1 cluster	α -1,4-glucosylation of galactose residues on O-antigen	Kim and Ryu 2012
Streptococcus pneumoniae ^a	hsdS	altered specificity of Type I RM system	Manso et al. 2014
Streptomyces coelicolor	pglX	methyltransferase of Type IV RM system	Sumby and Smith 2003
Vibrio cholerae	manA	Phosphomannose isomerase involved in O1-antigen biosynthesis	Seed et al. 2012
Vibrio cholerae	wbeL	involved in O1-antigen biosynthesis via tetronate acytlation	Seed et al. 2012
^a Although not shown in article t	this system can alter bacte	riophage sensitivity (Marco Oggioni, personal communication)	

 Table 1.2. Phase Variable Bacteriophage Resistance Mechanisms

incubation of ICP1 with LPS extracted from wild-type V. cholerae and a manA* strain, free phage were inhibited by Log_{10} 2.3 fold and Log_{10} 1.6 fold respectively, illustrating that ICP1 was unable to bind to the LPS of manA* to the same extent as LPS extracted from the wild-type host. A manA* gene is the product of slippage along either of two poly-A tracts within the manA ORF, both of which are composed of A₉ SSR tracts. A functional manA should produce a peptide of 399 amino acids in length, slippage on the first A₉ tract would produce an incomplete 81 amino acid product and slippage in the second poly-A tract would result in a truncated 207 amino acid peptide. Slippage along either of the A₉ tracts of manA produced the same degree of resistance to ICP1, demonstrating that either tract is capable of providing the reduced susceptibility phenotype (Seed et al., 2012).

The presence of multiple phase variable surface structures is not unusual, and as phase variation is often described as a mechanism that permits a bacterial populations to adapt to sudden environmental changes (Bayliss, 2009), the environment in which bacteriophage invade in itself could be a reason for the requirement of multiple phase variable loci that ultimately cover similar roles. This is demonstrated in Seed et al. (2012) who analysed the resistance phenotypes produced following bacteriophage infection in two distinct environments. In simulated natural environments, where V. cholerae was incubated with ICP1 in pond microcosms, resistance through $wbeL^*$ predominates, while the manA^{*} resistance phenotype was never observed. However, in another environment, an agar plate, where the spread of host and phage are relatively restricted, bacteriophage resistant mutants isolated from the center of plaques had arisen predominately through mechanisms other than phase variation of wbeL or manA. Only $\sim 20\%$ of the resistant isolates selected had undergone phase variation in either wbeL or manA (Seed et al., 2012). Furthermore, of this 20%, $\sim 13\%$ of V. cholerae resistant mutants were the result of phase variation to the $manA^*$ phenotype, and only $\sim 7\%$ through phase variation of wbeL. This points to an association with the spatial restrictions within each environment, emphasising the importance of the presence of multiple phase variable mechanisms in dealing with pressures in different environments. This may imply that phase variation to a $manA^*$ phenotype, providing partial resistance, is sufficient for resistance to bacteriophage in an environment where diffusion is restricted, yet in the less restricted environment of the pond microcosms $wbeL^*$ is required for resistance. Thus phase variation can result in differential adaptation between distinct environments to the same selection pressure.

Although inactivation of certain genes may prove beneficial in environments where bacteriophage are present, bacteria do not frivolously maintain genetic elements within their chromosomes. Normally each gene has a purpose that aids the survival of the bacterial population in at least one environment where they are selected for, otherwise the genes would have been lost or inactivated and eroded over time (Mira et al., 2001). As such, the adaptation of bacteria to avoid predation by bacteriophage can prove detrimental to host survival in phage-free environments as these structures have their own biologically relevant functions (León and Bastías, 2015). In *V. cholerae* for *wbeL* and *manA*, this is also true. In infant mouse colonisation assays OFF state phase variants of both genes show attenuation in their ability to colonise the small intestine (Seed et al., 2012). Similarly in assays involving the antimicrobial peptide polymyxin B, both $wbeL^*$ and $manA^*$ were less adept at survival than their parental wild type strain (Seed et al., 2012). Phase variation can therefore provide a means through which bacterial populations can create heterogeneous populations able to deal with selective pressures associated with selection against gene expression and also pressure requiring expression, effectively providing the population with a contingency plan, should the environment change to an opposing pressure.

The phase variation of molecules on the cell surface would not only facilitate maintenance of a small sub-population expressing the receptor for use of its natural function for the host cells benefit, but rather this would in theory benefit the host through reducing the selection pressure on the co-evolution of adaptations within the bacteriophage population to then modify its specificity to target a different receptor. This would avoid similar results of receptor modifications noted in *E. coli*, where in response to bacteriophage λ predation *E. coli* reduces the expression of its receptor, LamB, and as a response the bacteriophage changes its specificity to now target a completely new molecule OmpF (Meyer et al., 2012).

Although commonly associated with structures that may act as bacteriophage receptors, phase variation has also been noted occurring in other mechanisms that provide bacteriophage resistance. The Sie system of the nosocomial pathogen *Clostridium difficile* is one such system, known as CwpV (Sekulovic et al., 2015). Although the encoding gene cwpV has been found in all isolates of *C. difficile* to date (Reynolds et al., 2011; Sekulovic et al., 2015), with all of these noted as being phase variable (Reynolds et al., 2011), only ~5% of a bacterial population actively express cwpV (Emerson et al., 2009).

When cwp V is in the ON state, the cells display a bacteriophage resistant phenotype, while in the OFF state, cells are significantly more susceptibility to bacteriophage challenge (Sekulovic et al., 2015). Phase variation here occurs through site specific recombination between two inverted repeat elements that alter the orientation of sequences upstream of the start codon of cwp V. The inversion event causes the generation of a terminator 60-70 bp downstream of the cwp V transcriptional start site, prematurely terminating transcription and thus preventing cwp V expression and putting the cell in the OFF state for this gene; inversion in the opposite direction removes this terminator allowing successful expression (Emerson et al., 2009).

The level of resistance conferred by cwpV is dependent on the bacteriophage tested. Testing of five bacteriophage, from two morphologies, three siphoviridae and two myoviridae, showed that when cwpVwas in the ON state complete resistance was observed against all three siphoviridae, yet only a reduced susceptibility following challenge with either myoviridae (Sekulovic et al., 2015). One key distinction between the myoviridae and siphoviridae is the presence of a contractile tail sheath covering the tail tube protein in the former, while the latter has a completely exposed tail tube, and as such CwpV may interact with this molecule to give resistance in some way (Fokine and Rossmann, 2014; Sekulovic et al., 2015).

The existence of a transiently expressed Sie system may serve as a means of reducing the selection pressure on the bacteriophage population to overcome the Sie system. This would reduce the chances of the bacteriophage developing mechanisms that allow circumnavigation of the Sie system, a phenomenon noted in *Streptococcus thermophilus*. In *S. thermophilus* a mechanism of bacteriophage adaptation to an Sie system, Ltp, encoded by bacteriophage TP-J34, has been described (Bebeacua et al., 2013). Ltp provides resistance to not only *S. thermophilus* bacteriophage but also *Lactococcus lactis* bacteriophage (Sun et al., 2006). However, mutants of the *L. lactis* bacteriophage, P008, are able to infect *S. theromophilus* isolates regardless of the presence of its Ltp (Bebeacua et al., 2013) by altering their tail tape measure protein, the molecule which it is believed interacts with the Ltp system (Bebeacua et al., 2013).

As mentioned, phase variation is a means by which bacteria can generate a heterogeneous population. This would mean that a bacterial population could consist of phase variants that are proficient or deficient in restriction and modification if a restriction modification system was under phase variable control. Phase variable contingency loci have been identified within a number of restriction modification systems in a number of organisms, including *Helicobacter pylori* (Salaün et al., 2004), *H. influenzae* (Zaleski et al., 2005; De Bolle et al., 2000), *Streptococcus pneumoniae* (Manso et al., 2014), *Moraxella catarrhalis* (Seib et al., 2002), *Mycoplasma pulmonis* (Dybvig et al., 1998), *Neisseria meningitidis* (Saunders et al., 2000) to name but a few. The presence of phase variable regions within multiple bacterial strains implies that these loci serve an important function as otherwise they would not have been selected for over time in multiple distinct bacterial species.

Although for a number of systems their involvement in bacteriophage resistance is yet to be observed, this remains a highly likely scenario. The constant ongoing battle between viruses and microbes could have been one of the forces that brought about the evolution of such hypermutable regions, and as such a role for phase variable restriction modification systems has been suggested, whereby they may serve as a means of aleviating the aforementioned pitfalls of this leaky barrier (See Section 1.3.1). Bayliss et al. (2006) suggested that the ON/OFF phase variation generating heterogenous population of restriction modification proficient and deficient cells for a Type III restriction modification system in *H. influenzae* may act to counter the known flaw in these systems of accidental methylation of incoming bacteriophage DNA.

As illustrated in figure 1.9 it is hypothesised that a population consisting of solely restriction deficient cells would not be resistant to bacteriophage, thus allowing the spread of bacteriophage throughout the bacterial population in an uncontrolled expansion, and the eventual elimination of the bacterial population (Figure 1.9 (a)). Meanwhile, if a bacterial population consisted of only cells with an active



Figure 1.9. Theorised role of phase variable restriction modification systems in defence against bacteriophage. Each lineage illustrates theoretical outcome of bacteriophage infection of bacterial populations with a (a) non-phase variable fixed OFF restriction modification system, (b) non-phase variable fixed ON restriction modification system, and (c) a population heterogeneous for restriction modification activity due to phase variation. Cells which lack restriction modification activity are represented by green circles and those with functional restriction modification systems are represented in red, while bacteriophage lacking methylation at the target site are depicted in blue, and those which have become accidentally methylated are in red.

restriction modification systems, bacteriophage expansion would be initially reduced by restriction on the majority of the population, yet a small number of bacteriophage could be methylated by the host enzyme providing the bacteriophage progeny with complete resistance to restriction, again eventually causing death of large numbers of the bacterial population (Figure 1.9 (b)).

However, with a phase variable restriction modification system, the bacterial population would consist of a mix of restriction modification ON and OFF phase variants. As such the restriction modification system ON variants would again reduce initial expansion of phage through the population, and again in the process giving rise to small numbers of methylated bacteriophage. However, if these methylated bacteriophage are then to infect a host which has the system in the OFF state the resulting phage progeny would lose their methylation, reverting the bacteriophage progeny to a restriction susceptible phenotype (Figure 1.5 (c)). Phase variable restriction modification expression would thus allow a measure of control on bacteriophage spread through their population by generating this heterogeneous populations. Having a population adept to altering the gene expression phenotype sporadically could ultimately ensure the bacterial populations survival.

Thus phase variable bacteriophage resistance mechanisms within bacterial genomes can play several roles in the interaction of bacteria and their associated viral predators. In this current work however, the focus is mainly on *H. influenzae* a bacterium which contains not only a number of phase variable loci, but also phase variable loci that are involved in bacteriophage resistance.

1.4.3 Phase variation in *Haemophilus influenzae*: niche adaptation and bacteriophage resistance

H. influenzae has a genome that is littered with a plethora of phase variable loci (Hood et al., 1996; Power et al., 2009). Phase variation within *H. influenzae* is mainly through SSM along SSRs, highlighted by the observations of Power et al. (2009) who surveyed the prevalence of SSR tracts within 8 full *H. influenzae* genomes and 8 incomplete genomes. Power et al. (2009) identified a total of 988 putative phase variable SSR tracts within the 16 analysed *H. influenzae* genomes, an average of \sim 62 per chromosome. Such phase variable genes in *H. influenzae* are linked to an array of biological roles from adherence, the acquisition of iron, evasion of the immune response, genes involved in the defence against foreign DNA, and loci involved in gene regulation. Examples of phase variable loci, and their putative functions within *H. influenzae*, can be found in table 1.3.

For instance, phase variable loci can be found within hqbA, hqbB, and hqbC, three genes that all encode haemoglobin binding protein, and thus are involved in iron scavenging from the environment (Ren et al., 1999; Cope et al., 2000). All three of these genes contain tetranucleotide repeats (5'-CCAA-3') within their ORFs, facilitating phase variation by SSM (Ren et al., 1999; Cope et al., 2000). These genes are of particular interest as H. influenzae is unable to complete the biosynthetic process involved in the generation of haem, due to the absence of many of the enzymes involved in the biosynthetic pathway for haem generation, namely those steps involved in the generation of the penultimate substrate of haem generation, protoporphyrin (Panek and O'Brian, 2002). Thus H. influenzae requires exogenous sources of either protoporphyrin or haem for growth during aerobic growth. Haem biosynthesis normally begins from the generation of δ -aminolevulinic acid by one of two means (with no bacteria able to utilise both strategies), either through the C8 pathway, beginning with glutamate as the substrate; or through δ -aminolevulinic acid synthase from glycine and succinyl coenzyme (Panek and O'Brian, 2002). From here the process of haem biosynthesis progresses as a seven-step process concluded by the chelation of protoporphyrin IX with iron to form protohaem (Panek and O'Brian, 2002). White and Granick (1963) noted that incubation of *H. influenzae* with intermediaries of the haem biosynthetic pathway, namely, δ -aminolevulinic acid (starting substrate), porphobilinogen (immediately proceeds δ -aminolevulinic acid), and coproporphyrin (the coproporphyrinogen III isomer is the product of the 4th step in haem biosynthesis) did not result in aerobic growth of H. influenzae, however addition of protoporphyrin did facilitate H. influenzae growth to similar levels achieved by the control population where hemin was added.

In the human body sources of haem are rare, as these are cleared from circulation through binding to haemopexin, albumin, and some lipoproteins followed by transport to the liver (Stojiljkovic and Perkins-Balding, 2002; Schultz et al., 2010); free protoporphyrin meanwhile is not present in large amounts (Morton et al., 2005). Haem can also be obtained from haemoglobin, however, as with free haem, free haemoglobin is cleared from circulation through by binding circulating molecules, in this case haptoglobin (Stojiljkovic and Perkins-Balding, 2002; Schultz et al., 2010). As such the hgp genes encode haemoglobin binding proteins that allow H. influenzae to scavenge haemoglobin or haemoglobin-haptoglobin from the exogenous environment (Cope et al., 2000). However, all three genes encode proteins of similar function and as such the presence of all three genes within the chromosome seems redundant, with the additional aspect of these loci being under phase variable control only further confounding the logic of their consolidated existence within one chromosome. Possible explanations have been suggested for the presence of these phase variable genes within the H. influenzae chromosome, such as ideas related to affinity and specificity switching, combination expression switching in line with nutrient fluctuations, and switching between alleles to avoid immune selection while still allowing iron scavenging (Moxon et al., 2006).

This latter point is of particular interest as surface exposed iron acquisition proteins are not only targets for the immune system but are also targets for bacteriophage predation, e.g. the fhuA receptor of the *E. coli* bacteriophage T5 (Braun et al., 1973). Phase variation within multiple variants in the host chromosome that fulfil similar requirements may therefore exist in *H. influenzae* to function as a form of binary modulation by switching between the alleles allows resistance to bacteriophage predation but maintains the ability to sequester sources of haem. This is an interesting suggestion in light of the recent 'ferrojan horse hypothesis' that speculates that iron within the bacteriophage tail fibres (e.g. what is noted in T4 tail fibres (Bartual et al., 2010)) compete with other iron carrying molecules to gain access to host receptors, that the bacteriophage can then utilise to initiate infection (Bonnain et al., 2016). It is important through to note that the presence of all three of these genes in one chromosome is not a ubiquitous habit amongst *H. influenzae*, with strains known to contain anywhere between 1-4 *hgp*-like genes within their chromosome (Moxon et al., 2006). Thus the theoretical need for modulation may vary between hosts. However, the reason behind the need for multiple genes of apparently overlapping function that are under phase variable control remains a question that has gone unanswered.

Other virulence factors of *H. influenzae* are able to undergo phase variation, such as the appendages of attachment, fimbriae. Here, phase variation is controlled by expansion and contraction of dinucleotide repeats (5'-AT-3') within the shared promoter region of two divergently translationally orientated sets of genes, *hifA* and *hifBCDE* (van Ham et al., 1993; Bayliss et al., 2004). Phase variation here results in three level expression of fimbriae on the cell, abscent, intermediate, and full, corresponding in the study of van Ham et al. (1993), in *H. influenzae* AM20 and AM30, as 9, 11, and 10 5'-AT-3' repeats respectively. The *hifA* gene encodes the major fimbrial subunit, and the *hifBCDE* operon encodes fimbrial chaperone-like proteins and other minor fimbrial subunits (Hardy et al., 2003). Phase variation of these two genes likely occurs due to conflicting selection at different stages of bacterial

Reference	Kroll and Moxon 1988	ein Ren et al. 1999	ein Ren et al. 1999	ein Ren et al. 1999	van Ham et al. 1993	van Ham et al. 1993	Dawid et al. 1999	Dawid et al. 1999	Glover and Piekarowicz n system 1972	Fernaays et al. 2006	Jarosik and Hansen 1994	Hood et al. 1996	thesis Weiser et al. 1989, 1997	is High et al. 1993	Maskell et al. 1991	Erwin et al. 2006	Fox et al. 2005	on system De Bolle et al. 2000
Product	capsular polysaccharide	haemoglobin/haemoglobin-haptoglobin binding prot	haemoglobin/haemoglobin-haptoglobin binding prot	haemoglobin/haemoglobin-haptoglobin binding prot	major fimbrial subunit	fimbrial chaperone-like protein	Adhesin	Adhesin	methyltransferase subunit of Type I restriction modificatio	IgA1 protease	glycosyltransferase involved in LOS biosynthesis	galactosyltransferase involved in LOS biosynthesis	choline kinase of the $lic1$ locus involved in LOS biosyn	eta-galactosyltransferase involved in LOS biosynthesi	sialyltransferase involved in LOS biosynthesis	glycosyltransferase involved in LOS biosynthesis	O-antigen LPS acetylase	methyltransferase subunit of Type III restriction modificati
Mechanism	homologous recombination	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM	SSM
Gene name	cap	hgpA	hgpB	hgpC	$hifA^{\mathbf{a}}$	$hifBCDE^{\mathrm{a}}$	$hmw1A^{ m b}$	$hmw2A^{ m b}$	hsdM	igaB	lex2A	lgtC	licA	lic2A	lic3A	losA1	oafA	modA

^b Phase variation of hmw1A and hmw2A occurs at an SSR between the two separate transcription initiation sites

Table 1.3. Examples of phase variable genes in *H. influenzae*

colonisation within the host. *H. influenzae* isolates obtained from the nasopharynx are more often fimbriated than those from systemic locations (Mason et al., 1985). Fimbriae themselves have further been associated as a means of bacteriophage adsorption (Bertozzi Silva et al., 2016).

A number of phase variable loci are also known in *H. influenzae* to be involved in the resistance of the bacterial population to immune targeting (Clark et al., 2013). In particular *H. influenzae* has a lipooligosaccharide (LOS, similar to LPS but with absence of the O-antigen oligosaccharide chain) that can undergo extensive degrees of phase variation, and thus generate an expansive repertoire of phase variants. The LOS of *H. influenzae* consists of two main regions, lipid A, and the core oligosaccharide region. Lipid A, as in LPS, anchors the core oligosaccharide to the cell membrane (Hood et al., 2001). From here, a triheptosyl region is connected to lipid A through 3-Deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) (Brabetz et al., 2000; Young and Hood, 2013). The triheptosyl region forms the inner core of the core oligosaccharide, and serves as the backbone for oligosaccharides extensions to expand the make up of the outer core (Schweda et al., 2007). It is this outer region of the core oligosaccharide that contains the phase variable epitopes of the LOS (Schweda et al., 2007). A schematic representation of some examples of phase variable loci and their distribution within the LOS can be found in figure 1.10.

In *H. influenzae* a number of these phase variable LOS structures have been associated with enhanced virulence for the host cells, in particular through their association with immune evasion. A number of these loci have been associated with increased survival of the host in the presence of human sera, including *licA*, *lic2A*, and *lgtC*, improving the hosts survivability both as independent structures and cumulatively (Clark et al., 2013).

Although *H. influenzae* has an expansive assemblage of potential LOS structures, with a number having expression regulated through phase variation, its close relative, *Haemophilus parainfluenzae*, lacks much of this collection and the ability to undergo phase variation of its LOS genes, despite at times containing homologues of those known to undergo phase variation in *H. influenzae* (Young and Hood, 2013). With a number of the phase variable LOS transferase loci in *H. influenzae* associated with virulence attributes the observation of the lack of phase variation and as wide an LOS repertoire in *H. parainfluenzae* was suggested to be a potential explanation for its far lesser involvement in disease than its related counterpart (Schweda et al., 2007; Young and Hood, 2013). In fact in another close relative of *H. influenzae*, *Haemophilus haemolyticus*, McCrea et al. (2008) screened for the presence of three known *H. influenzae* phase variable LOS loci, believed to play a role in host virulence, in 109 *H. haemolyticus* isolates. The three loci, *licA* (also known as *lic1A*; part of the *lic1* loci, and the gene within this region that contains the phase variable SSR tract (Weiser et al., 1989)), *lic2A*, and *lgtC*, were found in 95.5, 90, and 100 % respectively of the 88 *H. influenzae* isolates screened but only 42.2, 15.6, and 1.8 % of the *H. haemolyticus isolates*. Similar to *H. parainfluenzae*, *H. haemolyticus* is traditionally noted as rarely associated with disease, however disease attributed to this organism have been documented (Anderson et al., 2012).

The *licA* gene, which is a component of the *lic1* locus that is responsible for the addition of phosphorylcholine to the *H. influenzae* LOS, is under phase variable control through the presence of a 5'-CAAT-3' repeat tract within the *licA* ORF (Weiser et al., 1989, 1997). The *lic1* locus itself consists of four genes in total, *licA*, *licB*, *licC*, and *licD*, which encode a choline kinase, choline permease, pyrophosphorylase, and diphosphonucleoside choline transferase respectively (Weiser et al., 1997). Using the components of the *lic1* locus, *H. influenzae* can incorporate exogenous choline into its LOS, with the heptose which receives the addition of this molecule dependant on *licD* (Weiser et al., 1997; Lysenko et al., 2000a).

The addition of phosphorylcholine to the *H. influenzae* LOS has been shown to be important in bacterial colonisation and persistence in animal models. (Weiser et al., 1998). For instance in chinchilla models it has been noted that populations that switched from *lic1A* OFF to *lic1A* ON would regularly be found in higher densities following intranasal inoculation in chinchillas (Tong et al., 2000). Furthermore expressing phosphorylcholine on *H. influenzae* LOS has been shown in chinchillas to be involved in disease. After 3-10 days following intranasal introduction of *lic1A* OFF cells into 35 chinchillas Tong et al. (2000) noted that 26 % (9/35) of the chinchillas were now colonised with H. influenzae populations composed of > 90 % lic1A ON cells. Of these lic1A ON colonised chinchillas 78 % (7/9) then went on to suffer from severe otitis media, while of the 26 chinchillas that were still colonised by majority *lic1A* OFF cells only 1 chinchilla contracted severe otitis media (Tong et al., 2000). Furthermore, when a mixture of *lic1* ON/OFF cells (~30 % ON) were introduced intra-nasally into BALB/c mice and BALB/c mice with severe combined immune deficiency (SCID), after three days, the H. influenzae population recovered by nasal lavage had now become ~ 80 % ON in the BALB/c mice, while the H. influenzae population recovered from the BALB/c SCID mice were till majority lic1A OFF $(\sim 20 \% \text{ ON})$ (Clark et al., 2012a). Thus as the BALB/c SCID mice lack an adaptive immune system it appears that the adaptive immune system selects for phosphorylcholine expressing *lic1A* ON cells.

Phosphorylcholine expressing populations have been observed to predominate in both healthy and diseases human populations. In human colonisation studies using *H. influenzae*, Poole et al. (2013) observed that *licA* expression may be important in early colonisation of the nasopharynx as, from an initial population composition of ~ 2 % ON, by the end of the sixth day of the experiment the proportion of *H. influenzae* in the *lic1A* ON state increased to ~ 25 %. Meanwhile, the expression of *lic1A* in respiratory tract colonisation was emphasised by the observation that in both asymptomatic patients, and patients with *H. influenzae* associated pneumonia, *lic1A* was constitutively found in the ON state (Weiser and Pan, 1998).

Mechanistically, phosphorylcholine on the H. influenzae is believed to be important in the adherence of H. influenzae to, and invasion of, host cells (Swords et al., 2000). Phosphorylcholine has also been implicated in resistance to pulmonary clearance, with this moiety on the LOS surface mimicking the human platelet activation factor (Swords et al., 2000; Harvey et al., 2001; Pang et al., 2008). Phosphorylcholine expression can reduce sensitivity to the antimicrobial peptide LL-37/hCAP18, which is expressed and secreted by cells in the respiratory tract (Lysenko et al., 2000b). Furthermore it has been shown to reduce binding of IgG to the cell surface of *H. influenzae* (Clark et al., 2012a), allowing this LOS extension to play both an independant, and collective role, in the survivability of *H. influenzae* against the host immune system (Clark et al., 2013).

However, ON expression of licA does have its drawbacks. Weiser et al. (1997) observed that after exposure of *H. influenzae* to human sera, cells expressing phosphorylcholine on their LOS were some 200-fold more sensitive to the bactericidal action of human sera as compared to cells free of phosphorylcholine. In fact they even conceded that the majority of the isolated survivors from the phosphorylcholine expressing populations were spontaneous revertants that had lost their phosphorylcholine through phase variation. This was later determined to be through the immune targeting of phosphorylcholine by C-reactive protein, allowing initiation of the classical pathway of complement mediated bactericidal action, with the extent of its killing action a function of where in the heptose backbone the phosphorylcholine is added (Weiser et al., 1998; Lysenko et al., 2000a). As such the expression of phosphorylcholine is not beneficial to the host in areas where C-reactive protein are in high concentrations such as the blood stream, evident through the observations of Hosking et al. (1999) that consistently, when rats were exposed to majority *lic1A* ON cells, *H. influenzae* cells recovered from the blood were now more abundant for cells that had switched to the *licA* OFF state.

Thus through phase variation of *licA*, the host cells can utilise the expression of phosphorylcholine to its advantage to allow the population to initially colonise and migrate to other niches inside the host. However, should the cells become localised in an area that contains high levels of C reactive protein (e.g. blood), then through the constant shedding of cells devoid of phosphochline and selection by C reactive protein, the population will rapidly evolve into one that is less susceptible to this immune factor. Although this would mean the population is now again more susceptible to antibodies, such as IgG, *lic1* is only one of multiple phase variable loci present within the LOS repertoire of *H. influenzae*. For instance the expression of *lic2A* or a combination of *lic2A* and *lgtC* can increase resistance of *lic1* expressing strains to serum through generation of an LOS with a Galp-(1-4)- β -Gal- β -D-Glcp by *lgtC* (Clark et al., 2013). Expression of *lic2A* without *lgtC* provides only the short extension, but is still able to provide an intermediate level of resistance compared to whole expression of the α -Gal-(1-4)- β -Gal once both *lic2A* and *lgtC* are in the ON state in *H. influenzae* R2846 (Clark et al., 2013). Other phase variable LOS loci have also been shown to provide serum resistance to bacterial populations such as *oafA* and *losA1* (Fox et al., 2005; Erwin et al., 2006).



Figure 1.10. Schematic example of a lipooligosaccharide structure for *H. infleunzae* Rd. Short hand for saccharides are as follows: Kdo = 3-Deoxy-D-*manno*-oct-2-ulosonic acid; PPEtn = pyrophosphoethanolamine; Hep = L-glycero-D-manno-heptose; Glc = D-glucose; PCho = phosphorylcholine; PEtn = phosphoethanolamine; Gal = D-galactose; Neu5Ac = N-acetylneuraminic acid (Sialic acid); and GalNAc = N-acetylogalactosamine. Phase variable loci are represented by green lines, with the phase variable bacteriophage receptor of Zaleski et al. (2005) indicated in blue. Biosynthesis associated genes for each step are presented in italics. Structure is adapted from Hood et al. (2001) and Young and Hood (2013).

From the perspective of the current work, the most interesting of these phase variable loci is that of lic2A, a gene which encodes a glycosyltransferase that has been shown to be involved in bacteriophage resistance in *H. influenzae* Rd30 (Zaleski et al., 2005). lic2A, together with another phase variable LOS transferase, lgtC, is responsible for the expression of the α -Gal-(1-4)- β -Gal extension of the *H. influenzae* oligosaccharide core (High et al., 1993) (See figure 1.10). The expression of this α -Gal-(1-4)- β -Gal extension serves as a means of immune evasion, with similar extensions to the α -Gal-(1-4)- β -Gal extension present on the human P blood group antigens (Virji et al., 1990; Bitzan et al., 1994; Weiser and Pan, 1998). The expression of the α -Gal-(1-4)- β -Gal is responsible for reduced sensitivity to human sera (Weiser and Pan, 1998; Clark et al., 2013).

In terms of clinical significance, Weiser and Pan (1998) found that when screening isolates of *H. in-fluenzae* from five asymptomatic patients and five expressing symptoms of *H. influenzae* associated pneumonia, expression of *lic2A*, and *lgtC* together, and thus α -Gal-(1-4)- β -Gal, was sporadic in asymptomatic individuals, only occurring in one patient, meanwhile every patient suffering from pneumonia had both *lic2A* and *lgtC* in frame. This role in disease, but not in colonisation has been further noted in the observations from the human colonisation study of Poole et al. (2013) that there is no selection for the *lic2A* ON state (nor *lgtC*) during colonisation. Furthermore, an involvement of *lic2A* in

other diseases, such as otitis media, is non-critical, as inconsistent expression states have been noted in samples from such patients (Fox et al., 2014). A role in more invasive blood stream infections was suggested from infections in the infant rat model where the α -Gal-(1-4)- β -Gal epitope provided an advantage to the cells (Griffin et al., 2005).

Phase variation of *lic2A* occurs in a similar way as *licA*, through SSM on a tetranuclotide 5'-CAAT-3' repeat tract within the ORF of the *lic2A* gene (High et al., 1993). This repeat region appears to have no structurally relevant roles, but rather appears to have arisen exclusively as a mechanism to regulate the expression of *lic2A* through phase variation (High et al., 1996). Phase variation of *lic2A* has been shown to potentially result in the display of three levels of expression in a study by Dixon et al. (2007) through use of a *lacZ* reporter. The *lacZ* reporter illustrated that a *lic2A* gene with 22x 5'-CAAT-3' repeats undergoes phase variation at a frequency of ~ 1.1 x 10⁻⁴ for ON-OFF switching and ~ 1.9 x 10⁻⁴ for OFF-ON switching (Dixon et al., 2007). Because the ON-OFF:OFF-ON ratio was ~ 0.6, it was concluded that *lic2A* could be expressed from two reading frames (for single frame expression this ratio should be closer to 2, as for every one ON frame there will thus be two OFF frames). However, although a weak ON phenotype could be observed by the LacZ reporter, testing with monoclonal antibodies targeting the digalactoside epitope associated with *lic2A* expression showed no reactivity to the weak ON reading frame and thus it was concluded this frame may not provide a level of *lic2A* expression sufficient to add the *lic2A* encoded galactose to the LOS structure (Dixon et al., 2007).

lic2A expression is responsible for the addition of the proximal galactose residue to the α -Gal-(1-4)- β -Gal LOS extension when in combination with *lgtC*, while on its own *lic2A* forms Galp-(1-4)- β -D-Glcp. It is this galactose that has been identified as the cell surface receptor for bacteriophage HP1c1 (Zaleski et al., 2005). With *lic2A* in the ON state, *H. influenzae* is sensitive to bacteriophage HP1c1 challenge, however, alterations in the length of the SSR tract to produce a frame shift that leads to the production of a premature stop codon would remove the bacteriophage receptor and thus confer bacteriophage resistance to the host. Zaleski et al. (2005) observed that following exposure to bacteriophage HP1c1, colonies of *H. influenzae* Rd30 could be isolated that displayed complete resistance to further bacteriophage challenge, even when the level of exposure was as high as ~1 x 10⁹ PFU ml⁻¹.

Zaleski et al. (2005) generated knock-outs of three transferase encoding genes, lgtF, lgtC, and lic2Aand utilised these mutants to determine that lic2A was responsible for the bacteriophage-sensitivity switching. The $\Delta lgtF$ mutants remained sensitive to bacteriophage HP1c1, indicating that the extension from HepI was not necessary for bacteriophage infection (see Fig. 1.10). A $\Delta lic2A$ mutant however, was immune to bacteriophage HP1c1 infection, while the $\Delta lgtC$ mutant remained sensitive to infection. Thus the bacteriophage receptor was localised to the Galp-(1-4)- β -D-Glcp extension encoded by lic2A rather than a requirement for the digalactose residue produced by lgtC and lic2A. They then determined the number of 5'-CAAT-3' repeats present within the ORF of lic2A, and observed that all sensitive isolates contained 22x 5'-CAAT-3', while those *H. influenzae* isolates that had developed complete resistance to bacteriophage HP1c1 now had 21x 5'-CAAT-3' repeats in the *lic2A* ORF, correlating to complete and truncated reading frames respectively (Zaleski et al., 2005). Alterations in the repeat tract of *lgtC* were observed in both resistant and sensitive cells further refuting its involvement in bacteriophage resistance (Zaleski et al., 2005).

Phase variation of lic2A expression may have thus developed as a means of dealing with conflicting selection, one occurring in the colonised environment where there is selection for cells that maintain either Gal*p*-(1-4)- β -D-Glc*p* or α Gal(1-4) β Gal expression for immune evasion, and another selection pressure that forces the host to lose this extension to survive bacteriophage selection. Without a bacteriophage receptor to target, bacteriophage will be unable to replicate and their numbers will decrease at a rate of decay functional to criteria such as environmental clearance, bacteriophage inactivation, DNA damage, and physical damage. Thus phase variation could lower bacteriophage numbers to a point where *lic2A* ON variants could facilitate colonisation and immune evasion.

It is not only molecules on the bacterial surface however, that can undergo phase variation. For instance, *H. influenzae* is known to have two phase variable restriction modification systems, one Type I system, HindI (Zaleski et al., 2005), and a Type III restriction modification system, HindIII (De Bolle et al., 2000).

The Type III system of *H. influenzae* undergoes phase variation due to the presence of a SSR tract within the *modA* ORF (De Bolle et al., 2000). Although two reading frames contain start codons for *modA*, the distal start codon, encoding a ~ 86 kDa protein, is the one which is responsible for production of the greatest level of *modA* activity (Srikhanta et al., 2005). The repeat tract within the *modA* ORF differs between strains and has been noted as occurring in iterations of tetranucleotide repeats of either 5'-AGTC-3' or 5'-AGCC-3', with isolates of *H. influenzae* type d containing 5'-AGTC-3', and NTHi and the remaining capsular types (a, b, c, e, and f) containing 5'-AGCC-3' repeats (Fox et al., 2007). However, not all isolates of *H. influenzae* contain an SSR tract within their *modA* ORF, with this dependent on which version of the *modA* allele is held by the isolate (Fox et al., 2007, 2014; Atack et al., 2015). However, in the majority of *H. influenzae* isolates it appears that those alleles which carry the phase variable form predominate (Fox et al., 2007; Atack et al., 2015). It was through study of the phase variation of *modA* that the correlation with increasing repeat tract length and phase variation frequency was discovered, with the frequency of switching increasing from 1.02 x 10⁻³ with 18 repeats to 2.65 x 10⁻³ with 37 repeats for OFF to ON switching (De Bolle et al., 2000).

Phase variation of restriction modification systems, alters the methylation pattern within the bacterial genome, termed the phasevarion, and this was first observed for the Type III system of H. influenzae

(Srikhanta et al., 2005, 2010). With at least 25 alleles associated with the modA locus across H. influenzae isolates, and each allele containing a different target recognition domain, H. influenzae has the potential to methlyate a wide range of sequences (Atack et al., 2015). Atack et al. (2015), when surveying the distribution of modA alleles from four H. influenzae strain collections noted that the 5 most common alleles to be modA2, 4, 5, 9 and modA10. Through generation of knockouts of strains harbouring each of these five alleles, and through use of single molecule real-time (SMRT) sequencing, the methylation sequence of each of these five modA alleles was determined. Four of these alleles were noted to target five base pair sequences, with only modA4 targeting a four base sequence; all alleles methylated sequences through m6A methylation. The phenotypic alterations associated with modA methylation include:- alterations in antibiotic susceptibility; outer membrane protein expression, including those involved in iron savaging; and biofilm expression. Phase variation of genes in H. influenzae similarly may exist to offset the costs associated with a particular gene expression phenotype. For example VanWagoner et al. (2016) observed that in *H. influenzae* R2866, inactivation of the phase variable modA10 gene resulted in a ~10-fold and ~100-fold increase in the ability of R2866 to adhere and invade human middle ear epithelial cells (HMEEC) and normal human bronchial epithelial (NHBE) cells respectively. They also noted that modA10 altered expression of olpA2, a gene which encodes an outer membrane adhesin, with a knock-out of olpA2 significantly reduced in adhesion to eukaryotic cells (VanWagoner et al., 2016).

Using modA2 for further analysis in *in vivo* chinchilla otitis media models, Atack et al. (2015) noted that introduction of an OFF phase variant into the chinchilla would consistently result in switching to the ON variant. Meanwhile, introduction of ON variants resulted in maintenance of modA in this state. The importance of modA in *H. influenzae* was also implied in previous mouse models, with transposon mutants without a functioning modA unable to colonise their host (Herbert et al., 2002).

Furthermore, the phase variation of restriction modification system, as mentioned in section 1.4.2 has been suggested as a means of avoiding the consequences of accidental methylation of bacteriophage DNA (Bayliss et al., 2006). The Type I restriction modification system of *H. influenzae* is one such system that has been identified to alter the resistance of *H. influenzae* to bacteriophage infection (Zaleski et al., 2005). This is again due to SSM of an SSR within the ORF of a gene encoding a methyltransferase, hsdM. This gene contains a pentanucleotide repeat tract, 5'-GAGAC-3', with slippage from 4 repeats to 5 repeats (and thus presumably 3 repeats) noted as removing the reduced susceptibility to bacteriophage HP1c1 observed when hsdM is in the ON state.

Little work has been conducted into the effects of the expression of hsdM. However, hsdM does not appear to play a significant role in at least the early stages of colonisation, as in the human colonisation study of Poole et al. (2013) the proportion of cells with hsdM in the ON state (from an already majority off innoculum) did not increase across the length of the six day experiment. However, as with the expression of modA, this may be dependent on the specificity of the sequence which is being targeted, which in this case is due to the hsdS gene that encodes the target sequence. Research into the degree of heterogeneity in hsdS alleles for this Type I system has not been conducted either. Finally as the HindI restriction modification system can reduce susceptibility to bacteriophage HP1c1, and bacteriophage against *H. influenzae* are known to be present within the respiratory tract of both healthy individuals and those with chronic respiratory conditions (Willner et al., 2009), phase variation of hsdM may serve as a means of allowing bacterial colonisation (the OFF state), while providing colonising bacteria with a means of resistance against the sudden appearance of bacteriophage.

1.5 Aims of Project

The presence of phase variable loci within bacteria are believed to be the result of a bacterial cell facing conflicting selection pressures, resulting in the evolution of hyper-variable loci that can rapidly, stochastically, and reversibly alter gene expression as a means of improving fitness and for survival. These putative adaptive traits have been suggested, and for some systems illustrated, as beneficial to bacteria for adaptation to bacteriophage predation.

In *H. influenzae* phase variable loci have been observed in a large number of genes, however, two of these genes have been associated as alterations in resistance/susceptibility to infection by bacteriophage HP1c1. These two genes, hsdM and lic2A, encode the MTase of a Type I restriction modification system and a glycosyltransferase involved in LOS biosynthesis respectively. As such the current project aimed to explore the dynamics of populations that are phase variable for these loci, as well as determining the repercussions of a bacterial host demonstrating phase variable resistance mechanisms on the bacteriophage population. Information on the dynamics between the presence of phase variable loci on the survival of both the bacterial and bacteriophage populations may allow insight into the origins of such a molecular hyper-variability mechanism, as well as potential to shape natural population structures.

The aim of the project therefore, was to determine the effect of phase variable resistance mechanisms on both the population dynamics of H. *influenzae* and the challenging bacteriophage, HP1c1. To achieve this the following specific objectives were outlined:

- 1. Investigate prevalence of the phase variable hsdM and lic2A loci within other isolates of H. influenzae, using data from:
 - (a) Sequenced genomes from GenBank
 - (b) Data obtainable from sputum samples from patients suffering from COPD

- 2. Determine if phase variation state for *hsdM* and *lic2A* is associated with clinical characteristics (i.e. disease state of host/site of isolation)
- 3. Validate the bacteriophage resistance phenotypes of hsdM and lic2A phase variants
- 4. Determine infection characteristics of bacteriophage HP1c1
- 5. Develop an assay to simulate bacteriophage spread through bacterial populations composed of differing degrees of bacterial resistance : susceptibility ratios, allowing generation of hypothetical degrees of phase variation induced population heterogeneity
- 6. Generate a mathematical model of bacteriophage propagation dynamics, from the spread simulation experiment
- 7. Create a multi-directional spread experiment to determine phase variation induced population heterogeneity can affect bacteriophage spread across a fixed area
- 8. Determine how heterogeneous populations affect the chances of bacterial population survival

2. Materials and Methods

2.1 Bacterial strains, bacteriophage, and growth conditions

H. influenzae strains used in this study are detailed in table 2.1. *H. influenzae* isolates were obtained from Dr. Christopher D. Bayliss, sourced originally from the culture collection of Prof. E. Richard Moxon (University of Oxford, UK). The *H. influenzae* strain harbouring bacteriophage HP1c1 as a prophage, RM118L _{r⁻HinDI} m⁻HinDI, was generated by Dr. Jaspreet Sahota (University of Leicester, UK).

The compositions of all media can be found in supplementary table S1. Unless otherwise stated, all *H. influenzae* isolates were cultured on supplemented brain-heart infusion (sBHI) 1% agar plates, supplemented with 10 % Levinthals and 2 μ g nicotinamide adenine dinucleotide (NAD) ml⁻¹. For liquid culture *H. influenzae* strains were routinely cultured in sBHI broth, supplemented with 10 μ g hemin ml⁻¹ and 2 μ g NAD ml⁻¹, shaken at 100 revolutions per minute (rpm). The above, unless otherwise stated, were incubated at 37°C overnight.

E. coli DH5- α (Lucigen) cells were grown on Lysogeny Broth (LB) 1% agar, and if liquid culture was required LB broth was used. Both plates and broth cultures were incubated at 37°C, with broth cultures again agitated at 100 rpm.

Bacteriophage HP1c1, a 'clear-plaque' derivative of bacteriophage HP1 (Harm and Rupert, 1963), was used throughout this investigation. Bacteriophage HP1c1 stocks were stored at -80°C in the lysogenic state, integrated into the chromosome of *H. influenzae* RM118 $_{r^{-}HinDI} m^{-}HinDI$ (this lysogenic strain was termed RM118L $_{r^{-}HinDI} m^{-}HinDI$). As such, before being used in experimental protocols, HP1c1 first required induction from its host. Herein, 100 μ l of an overnight culture of RM118L $_{r^{-}HinDI} m^{-}HinDI$ was added to 10ml sBHI broth and incubated at 37°C until growth reached OD₆₀₀ 0.1. From this 1 ml of culture was transferred to 9 ml fresh broth, with Mitomycin C (Sigma) added to the culture to a final concentration of 300 ng ml⁻¹. This was then incubated at 37°C, shaken at 100 rpm, for 5 hours. After this time the culture was then centrifuged at 4696 x g, and the resulting supernatant filtered using a 0.22 μ m syringe filter (Millipore), before being stored at 4°C. In all cases, following mitomycin C induction the bacteriophage released from the cell underwent one further round of propagation before

H. influenzae strain name	Notes	Source
$\rm RM118 \ _{r^{+}HinDI \ m^{+}HinDI}$	4x 5'-GAGAC-3' isolate	Zaleski et al. 2005
RM118 _{r⁻HinDI m⁻HinDI}	5x 5'-GAGAC-3' isolate	Zaleski et al. 2005
m RM118- $hsdM$ - $ m Cass$	RM118 $_{r$ -HinDI m-HinDI containing the tetR- $sacB/nptII$ cassette in $hsdM$	This study
RM118L _r -HinDI m ⁻ HinDI	HP1c1 containing lysogen of RM118 _r -HinDI m-HinDI	Dr. Jaspreet Sahota
Rd 30S	Bacteriophage HP1c1 sensitive isolate	Zaleski et al. 2005
Rd 30R	Bacteriophage HP1c1 resistant isolate	Zaleski et al. 2005

 Table 2.1.
 Bacterial Isolates used in this study

application in bacteriophage experiments.

For bacteriophage HP1c1 propagation, *H. influenzae* Rd 30S was used routinely as a propagating host, while for the investigation of restriction modification associated resistance RM118 $_{r^-HinDI} m^-HinDI$ was used (this host does not have a functional *hsdM* gene and as such all bacteriophage progeny would be unmethylated). For propagation, 100 μ l of an overnight Rd 30S (or RM118 $_{r^-HinDI} m^-HinDI$) culture was inoculated into 10 ml sBHI broth, and allowed to grow to OD₆₀₀ 0.1 (~1 x 10⁸ CFU ml⁻¹). From this culture, 100 μ l was then transferred to 10 ml sBHI to maintain the bacteria at exponential growth for a longer period, at which point bacteriophage were then added at a multiplicity of infection (MOI, the ratio of bacteria to bacteriophage) of 0.01. Bacteriophage/host suspension was then left to incubate at 37°C, shaken at 100 rpm, for 5 hours. After which, the bacteriophage-host mix was centrifuged at 4696 x g for 10 mins and filtered through a 0.22 μ M filter. The resulting lysate was then stored at 4°C until required for bacteriophage based assays.

2.2 Sputum supernatants from patients with COPD

During the course of this study the opportunity arose to conduct PCR analysis on sputum supernatants isolated from individuals with COPD. The 174 sputum samples were originally obtained from the COPD study group of Prof. Christopher Brightling (University of Leicester), with sputum supernatants to be used in PCR reactions prepared by Koirobi Haldar. Samples were from a total of 89 patients, with samples taken at points including stable return visits, exacerbations and/or post exacerbations. All samples were handled and destroyed in accordance with the Human Tissue Act.

2.3 DNA analysis methods

2.3.1 Polymerase Chain Reaction (PCR)

Amplification of DNA products was conducted through use of polymerase chain reaction (PCR). For cloning reactions Phusion high-fidelity polymerase (New England Biolabs) was used to generate amplicons, while Biotaq polymerase (Bioline) was used for reactions involving the screening of DNA samples, GeneScan, and sequencing reactions. All reactions and compositions were dependent on the purpose of the assay being carried out; further information on the PCR reactions used can be found in the appropriate section.

2.3.2 Restriction endonuclease digestions and ligation reactions

Restriction digests were conducted to facilitate ligation of insert DNA fragments into the appropriate vector during construction of plasmids for allelic exchanges, or to screen plasmids extracted from transformants to ensure fragments had inserted correctly. Reactions were carried out in 20 μ l volumes, consisting of 1 μ l of the appropriate restriction enzyme(s) per μ g DNA to be digested, 2 μ l of the appropriate 10x buffer, 1-2 μ g of DNA, with the remaining volume made to 20 μ l with upH₂O. All enzymes used were purchased from New England Biolabs.

Restriction digests for linearising plasmids before transformation into *H. influenzae* was conducted similarly as above, however with one modification. Reactions were as above, however, 2 μ g of plasmid DNA was consistently used in the digestion.

Digests of samples were ran on agarose gel as per section 2.3.3, and the resulting band required extracted from the gel and purified as per section 2.3.4. This was true for all restriction reactions except for those involved in the linearisation of plasmids to be used in the transformation of *H. influenzae* as here, a 5 μ l aliquot was used in electrophoresis, with the remainder of the sample being purified from solution as per the PCR reaction purification method of section 2.3.4.

2.3.3 Agarose gel electrophoresis

DNA fragments produced by PCR, or following restriction digestion were interpreted using agarose gel electrophoresis through a 1% agarose gel. Agarose gels consisted of 1% (w/v) agarose in TAE (Tris-acetate-EDTA) buffer (composition can be found in supplementary table S1), and 0.5 μ g/ml ethidium bromide for visualisation of DNA. Each sample was loaded using 6x DNA loading dye (New England Biolabs), with the resulting DNA product bands compared to a Generuler 1kb DNA ladder (Thermo-Scientific).

Electrophoresis was conducted at 100 V for 60 minutes unless otherwise specified, with products >6 kb ran for 90 minutes. Gels were then visualised using a Syngene Gbox gel viewer.

2.3.4 Purification of DNA fragments

For PCR products where a single band of DNA was expected, 5 μ l of PCR reaction to be tested was ran on an agarose gel (Section 2.3.3) to confirm only a single band was present. If this were the case, the remainder of the PCR reaction was directly purified using the E.Z.N.A Cycle Pure Kit (Omega biotek) following the purification protocol of the manufacturer. For restriction digests, or PCR reactions expected to produce more than one product, the whole reaction was ran on an electrophoresis gel. The required band was then cut from the agarose gel, and purified using the Zymoclean Gel DNA recovery kit (Zymo Research) by following the manufacturers instructions for purification of gel products.

2.3.5 Genescan analysis

For determination of repeat tract size of SSRs within amplicons of phase variable genes the fragment size of an isolate of known repeat tract length (with repeat tract of the known isolate determined by DNA sequencing [See Section 2.3.6]) was compared to the sample in question by genescan analysis. The known isolate was *H. influenzae* Rd 30S. For genescan analysis each amplicon to be analysed, including the control amplicon, were amplified using primers containing a fluorescent dye, 6-FAM, modification at the 5' end of the forward primer of the primer set used. The products were amplified by PCR in 15 μ l volumes, consisting of 0.3 μ l of each of the 10 μ M primers (*hsdM* was amplified using HsdM-CT-F2 and HsdM-GS-R2, *lic2A* was amplified using Lic2A-GS-F and Lic2A-GS-R; for sequences see table 2.2), 0.45 μ l 50mM MgCl₂, 1.5 μ l 10x reaction buffer, 0.15 μ l Biotaq polymerase, 0.3 μ l 10mM dNTPs, 2 μ l of boiled lysate DNA, made to 15 μ l with upH₂O. This reaction was then amplified in a Sensoquest Gradient Thermal Labcycler (Geneflow) under the following conditions: an initial denaturation step of 95°C for 5 minutes; followed by 30 cycles of 95°C for 1 minute, 47°C for 1 minute, and 72°C for 1 minute; with a final step of 72°C for 5 minutes.

Primer Name	Sequence	Origin
HsdM-CT-F2	5'-CGGCATTCTACTGGAAATCATTG-3'	This study
HsdM-GS-R2	5'-ATAGCGGCATCAAGTTGTT-3'	This study
Lic2A-GS-F	5'-CGGGCAACTGAACGTCGCAA-3'	This study
Lic2A-GS-R	5'-CGCTCTCCTTTAGTAAGTATTC-3'	This study

Table 2.2. Primers used in genescan reactions

Following this initial amplification reaction, as Taq polymerase produces overhangs of adenine residues on amplified products, A-tailing was conducted to ensure that all fragements produced during the reaction had these overhangs. Herein, each of the 15 μ l reactions recieved a further 0.18 μ l MgCl₂, 0.6 μ l 10x reaction buffer, 0.06 μ l Biotaq polymerase, and 5.16 μ l upH₂O. Each sample was then incubated at 72°C for 45 minutes.

Each reaction was then subsequently diluted 10-fold using upH₂O. From the diluted product 0.5 μ l was then added to 9.25 μ l deionised formamide (AMRESCO) and 0.25 μ l Genescan 500-LIZ size standard (Applied Biosystems). Genescan analysis was conducted using an ABI 3730 DNA analyser (Applied Biosystems). Results were then uploaded to Peak Scanner 1.0, where fragment size information was

exported onto Microsoft Excel for analysis.

2.3.6 DNA sequencing

DNA sequencing of PCR products was conducted by GATC biotech. For sequencing, PCR or plasmids first required purification as per sections 2.3.4 and 2.3.11 respectively. 30 μ l of ~30 ng μ l⁻¹ DNA to be sequenced was sent, together with 30 μ l of the primers from where sequencing was to occur. GATC biotech conduct sequencing reactions using an ABI 3730xl sequencer.

For bacterial genome sequencing, bacterial DNA was isolated from overnight cultures of *H. influenzae* using the Wizard Genomic DNA Purification Kit (Promega) according to the manufacturers instructions. Once bacterial DNA was isolated, genomic DNA was tested for purity using the Nano-Drop 1000 (Thermo-Scientific). As a further test of purity of DNA, three-fold diluted samples of genomic DNA were ran on a 1% agarose gel. The samples were then further screened to confirm genomic DNA was *H. influenzae* using the *recA* primer set from that recommended for use for *H. influenzae* MLST (Meats et al., 2003). Once all tests were passed, the DNA was quantified on a Qubit 3.0 flourometer (Thermo-Scientific) and 10 ng μ l⁻¹ DNA of each bacterial genome was sent to Dr Andrew Millard at the University of Warwick to be sequenced. There, the genomes were sequenced using an Illumina MiSeq.

2.3.7 Splicing by overlap extension-PCR

For generation of constructs for mutation of *H. influenzae* splicing by overlap extension-PCR (SOE-PCR) was used. Primers used for all SOE-PCRs can be found in table 2.3. The SOE-PCR reactions were conducted to ultimately generate a *H. influenzae* strain with a 3x 5'-GAGAC-3' tract within its *hsdM* gene. Mutations were achieved in a two-part allelic exchange using the tetR-*sacB/nptII* cassette of plasmid pJJ260 (Johnston, 2012), graciously provided by Dr Jason Johnston (University of Kentucky, USA) and Prof. Michael A. Apicella (University of Iowa, USA).

For the two step allelic exchange two inserts were generated using SOE-PCR. The first, eventually ligated into pUC19 to form pCT0.5, involved the amplification of two regions, one upstream of the 5'-GAGAC-3' SSR tract of hsdM and one downstream of this tract. These regions were amplified using HsdM_5'_F_1 and

The 5' and 3' regions upstream and downstream of the 5'-GAGAC-3' SSR tract was amplified using primers HsdM_5'_F_1 and HsdM_5'_cass_R_2, and HsdM_3'_cass_F_3 and HsdM_3'_R_4 respectively. For amplification of each fragment, reagents were added as follows: 2.5 μ l of each of the associated two primers (stocks = 10 μ M), 1 μ l 10mM dNTPs (Bioline), 0.5 μ l Phusion high-fidelity polymerase

seduences			
Primer Name	Sequence	Restriction site	Origin
HsdM_5'_F_1	5'-TAT <u>GAGCTC</u> AAAGGCAATGAAGTTATGCG-3'	SacI	This study
HsdM_5 '_cass_R_2	5'-AAAATACGTTAT <u>CCCGGG</u> TTGCATAAATTATGCCCTCTG-3'	SmaI	This study
HsdM_5'_fix_R_2	5'-GTCTCGTCTCGTCTCGCTT-3'	N/A	This study
HsdM_3 '_cass_F_3	5'-TAATTTATGCAA <u>CCCGGG</u> ATAACGTATTTTGGGTGCCA-3'	SmaI	This study
HsdM_3' fix_F_3	5'-GCGAGACGAGACGAGACGAC-3'	N/A	This study
$HsdM_3'R_4$	5'-TAT <u>GGATCC</u> TCTGATAAGGCTTTTCATTTCCC-3'	BamHI	This study
$pCT1_screen_F$	5'-TTCCACATCAGAAACCG-3'	N/A	This study
$pCT1_screen_R$	5'-TAGGTTGCGTATTGGAAG-3'	N/A	This study
$pCT2_screen_F$	5'-TGAACTTTGGGCATCTGCTG-3'	N/A	This study
$pCT2_screen_R$	5'-TGATCACGGTTTTCTAGTTCTGC-3'	N/A	This study
(New England biolabs), 10 μ l 5x Phusion reaction buffer (New England biolabs), 33 μ l upH₂O, and 0.5 μ l of *H. influenzae* RM118 _{r⁻HinDI} m⁻HinDI DNA (stock concentration ~1 μ g ml⁻¹. The initial PCR reaction for the 5' fragment involved the following cycling conditions: an initial denaturation step of 98°C for 5 minutes; followed by 30 cycles of 98°C for 10 seconds, 64°C for 30 seconds, and 72°C for 1.5 minutes; with a final step of 72°C for 5 minutes. The resulting product was then purified as per section 2.3.4.

Once these fragments had been produced and purified, ~50 ng of each purified fragment was added to one PCR tube to be used as a template in the SOE-PCR reaction. To this tube, 1 μ l dNTPs, 10 μ l 5x Phusion reaction buffer, 0.5 ml Phusion high fidelity polymerase, and 32.7 μ l upH²O. No primers were added at this point as this first reaction aims to allow the annealing of the two fragments, with extension of the annealed fragments occurring in the subsequent step. For this initial annealing step thermocycling occurred as follows: initial denaturation at 98°C for 5 minutes; 10 cycles of 98°C for 10 seconds, 66°C for 30 seconds, and 72°C for 3 minutes; followed by a final extension at 72°C for 5 minutes. Following this 2.5 μ l of the outer primers (HsdM_5'_F_1 and HsdM_3'_R_4, 10 μ M stocks) were then added to the reaction. The samples then underwent a further thermocycling reaction of: initial denaturation at 98°C for 5 minutes; 20 cycles of 98°C for 10 seconds, 64°C for 30 seconds, and 72°C for 3 minutes; followed by a final extension step at 72°C for 5 minutes.

The resulting product was then viewed on a 1 % agarose gel with the the 3677 bp product and extracted from the gel and purified as per section 2.3.4. This product was then used to generate pCT0.5 by ligation into pUC19 as per section 2.3.8.

A similar process was conducted for the generation of the PCR fragment containing the 3x 5'-GAGAC-3' SSR. For this reaction 5' and 3' fragments were generated using the primer pairs HsdM_5'_F_1 and HsdM_5'_fix_R_2, and HsdM_3'_fix_F_3 and HsdM_3'_R_4 respectively. Each of the inner primers (i.e. HsdM_5'_fix_R_2 and HsdM_3'_fix_F_3) contain an overlapping sequence of homology with the opposing inner primer that contains a 3x 5'-GAGAC-3' SSR tract. Thus annealing of these fragments will generate one fragment containing a 3x 5'-GAGAC-3' SSR. PCR reaction for this primer set were as the same as for the previous primer set, as were the cycling conditions to generate each fragment. From this point onward all steps and reactions were conducted as per the previous set of SOE PCRs. Except that this SOE-PCR reaction would go on to form the insert for pCT2.

2.3.8 Ligation reactions

Ligation reactions were conducted using the purified final products of section 2.3.7, following digestion with the appropriate restriction enzymes (SacI and BamHI for generation of pCT 0.5 and pCT2, and SmaI for pCT1). For ligation of fragments with both blunt and cohesive ends into their vectors, vector and insert were mixed in 3:1 ratios (total DNA concentration of ~10 ng μ l) in a 20 μ l reaction. The remainder of the reaction volume consisted of 1 μ l T4 ligase (New England Biolabs) and 2 μ l 10x T4 ligase buffer (New England Biolabs), made to 20 ml with upH₂O. Each reaction was incubated at 16°C overnight, with the resulting products then used in the transformation of *E. coli* DH5- α as per section 2.3.9.

2.3.9 Transformation of *E. coli* DH5- α

2.3.9.1 Generation of competent E. coli DH5- α cells

Transformation of DNA into *E. coli* was facilitated through use of chemically competent cells of *E. coli* DH5- α . Competent cells of *E. coli* were prepared using RbCl, based on the method of Hanahan (1985). Herein, a colony of *E. coli* DH5- α was inoculated into 5 ml LB broth, and incubated overnight at 37°C, shaken at 100 rpm. The entire 5 ml culture was then inoculated into 200ml of fresh LB broth containing 20 mM MgSO₄, and allowed to grow at 37°C agitated at 100 rpm until reaching OD₆₀₀ 0.3. Once this density was reached cells were transferred to 50 ml centrifuge tubes and incubated on ice for 15 minutes. The tubes were then centrifuged at 4500 x g for 15 minutes at 4°C. The supernatant was then discarded and the resulting pellet resuspended in 0.4 volume of pre-chilled RF1 (for composition see table 2.4) solution and incubated on ice for 5 minutes. After this the cells were again centrifuged at 4500 xg for 5 minutes at 4°C, the supernatant discarded, and the pellet resuspended in 1/25 volumes of RF2 solution (table 2.4), and further incubated on ice for 15 minutes. The resulting cells were then aliquoted in 100 μ l volumes, and each aliquot was then snap frozen in liquid nitrogen and stored until use at -80°C until required.

Solution name	${f Ingredient}$	
RF1	RbCl (Acros)	0.48 g
	$MnCl_2$ (Sigma)	$0.40~{\rm g}$
	$1M CH_3 COOK$ (Sigma)	$1.2 \ \mathrm{ml}$
	$CaCl_2 \cdot 2H_2O$ (Acros)	$0.06~{\rm g}$
	Glycerol (Acros)	$6 \mathrm{ml}$
	Make to 40 ml in distilled water and adjust to pH 5.8	
	Sterilise by filtration	
$\mathbf{RF2}$	RbCl (Acros)	$0.02~{\rm g}$
	$0.5\mathrm{M}$ 3-(N-Morpholino) propanesulfonic acid (MOPS) (Sigma)	$0.4 \ \mathrm{ml}$

Table 2.4. Ingrediants of media used for chemical competancy of E. coli DH5- α

Primer Name	Ingredient	Origin
	$CaCl_2 \cdot 2H_2O$ (Acros)	$0.22 \mathrm{~g}$
	Glycerol (Acros)	$3 \mathrm{ml}$
	Make to 20 ml in distilled water and adjust to pH 6.8	
	Sterilise by filtration	

Table 2.4. – Continued from overleaf

2.3.10 Transformation of *E. coli* DH5- α by heat-shock

For transformation of the competent *E. coli* DH5- α cells with plasmid DNA heat-shock was conducted. Here, *E. coli* DH5- α cells were thawed on ice for around 15 minutes. During which time the 3 μ l of ligation reaction was transferred into 0.5 ml microcentrifuge tubes and placed on ice to chill. Once the cells had thawed, 30 μ l of competent cells was transferred to the 0.5 ml microcentrifuge tube containing the ligation reaction, and pipetted once to mix. The cell-ligation mix was then left to incubate on ice for 15 minutes. After this point the cells underwent heat-shock at 42°C by transfer to a water bath set to 42°C, incubation at this temperature for 30 seconds, followed by immediate transfer back to ice. Once transferred the tubes were then incubated for a further 2 minutes on ice. After this, 370 μ l of LB broth was then added to the cells followed by incubation at 37°C, agitated at 100 rpm, for 1 hour.

The resulting cell suspension was then screened for transformants by spreading 100 μ l of this suspension onto LB 1% agar containing the appropriate antibiotic (100 μ g ml⁻¹ carboncillin disodium salt [Melford] for pCT0.5 and pCT2 transformations, and 15 μ g ml⁻¹ kanamycin sulphate [Thermo-scientific] for screening for pCT1 transformants). Plates were incubated at 37°C overnight.

2.3.11 Plasmid extraction

Where necessary all plasmid extractions were conducted using the GenElute Plasmid Miniprep Kit (Sigma) as per the manufacturers instructions.

2.3.12 Transformation of *H. influenzae*

2.3.12.1 Generation of competent cells of *H. influenzae*

Transformation of DNA into *H. influenzae* was achieved through use of competent cells of *H. influenzae*, obtained using the protocol of Poje and Redfield (2003). *H. influenzae* RM118 $_{r^-HinDI}$ m⁻HinDI was used to generate the initial competent cells for use in the first step of the allelic exchange, with the second

allelic exchange facilitated using the positive transformants of the first allelic exchange (RM118-*hsdM*-Cass). Briefly the competence method of Poje and Redfield (2003) is as follows, 350 μ l of overnight culture was inoculated into 35 ml sBHI broth, and allowed to grow at 37°C shaken at 100 rpm until reaching OD₆₀₀ 0.2-0.25. Of the 35 ml, 20 ml was passed through a 500 ml 0.2 μ M bottle top filter (Sarstedt) to collect the bacterial cells on the filter paper. Once the 20 ml culture had passed through the filter, 20 ml M-IV media (for M-IV media add 0.5 ml of solutions 22, 23, 24, and 40 to 50 ml solution 21, table 2.5) was then passed through the same filter to wash the cells. Following this the filter paper was then removed and placed in a 200 ml Erlenmeyer flask containing 20 ml M-IV media. The culture was then left at 37°C, shaken at 100 rpm, for exactly 100 minutes. To this culture, 5 ml of 80% glycerol was then added, and mixed, with the culture then separated into 1 ml aliquots and frozen at -80°C until required.

2.3.12.2 Transformation of plasmid DNA into H. influenzae

When required for transformation with DNA, 4 sets of frozen cells were allowed to thaw on ice for ~10 minutes, after which they were centrifuged at 5000 x g for 5 minutes, and the supernatant discarded and the pellet resuspended in 1 ml fresh M-IV media. To each tube, linearised plasmid DNA (linearised in all cases with SacI) was then added, and the cells allowed to incubate at 37°C for 15 minutes. Each tube recieved one dilution of a series of 10-fold dilutions of linearised plasmid DNA, ranging from 1 μ g to 1 ng. The cells were then plated onto sBHI 1% agar plates, containing 100 μ g ml⁻¹ carboncillin disodium salt (Melford) for the first allelic exchange using pCT1, while transformants for the second allelic exchange were plated on sBHI 1% agar containing 10 % sucrose (Fischer) and 0.5 μ g ml⁻¹ chlortetracycline hydrochloride (Sigma). Plates were incubated at 37°C for ~24 hours.

2.3.13 Bioinformatic analysis

All Genome visualisation was conducted using Artemis v15.0.0 (Rutherford et al., 2000). Genome screening for gene homologues was conducted using Blast (Altschul et al., 1990). Gene sequence visualisation, alignments, and phlogenetic trees were achieved using MEGA v6.0 (Tamura et al., 2013). Restriction site planning for cloning was conducted using APE v2.0.45 (ApE, 2016).

olution Name	Component	
Solution 21	L-Aspartic acid (Sigma)	4 g
	L-Glutamic acid (Sigma)	$0.2~{ m g}$
	Furmaric acid (Sigma)	1.0 g
	NaCl (Fisher)	4.7 g
	K_2HPO_4 (Fisons)	$0.87~{ m g}$
	$\rm KH_2PO_4$ (Fisher)	$0.67~{ m g}$
	Tween-80 $(Acros)$	$0.2 \ \mathrm{ml}$
	Distilled water	$850 \mathrm{~ml}$
pH adjusted to 7.4 us	sing 4 N NaOH (Fisher), and then make to 1 L using	g Distilled water
	Sterilise by autoclaving	
Solution 22	L-Cystine [*] (Sigma)	$0.04~{ m g}$
	L-Tyrosine [*] (Sigma)	$0.1~{ m g}$
*Dissolve at $37^{\circ}C$	in 10 ml of 1 N HCl, then bring to 100 ml Distilled	water and add:
	L-Citruline (Sigma)	$0.06~{ m g}$
	L-Phenylalanine (Alfa Aesar)	0.2 g
	L-Serine (Sigma)	$0.3~{ m g}$
	L-Alanine (Sigma)	$0.2~{ m g}$
	Sterilise by filtration	
Solution 23	$CaCl_2$ (Acros)	0.1 M
	Sterilise by autoclaving	
Solution 24	$MgSO_4$ (Fisher)	$0.1 \ \mathrm{M}$
	Sterilise by autoclaving	
Solution 40	Vitamin-free Casamino acids (Difco)	5% (w/v)
	Make to 100 ml with Distilled water	
	Sterilise by autoclaying	

 Table 2.5.
 M-IV Competence Media Component Solutions

2.4 Experimental Methods using Bacteriophage

2.4.1 Small Drop Plating Plaque Assay and Double Agar Overlay Plaque Assay Procedures

To enumerate bacteriophage numbers, either double agar overlay or small drop plating plaque assays were utilised. Small drop plating plaque assays were conducted when large numbers of bacteriophage samples were to be screened, while double agar overlay plaque assays were utilised where either low numbers of samples were to be screened, or for samples expected to contain low bacteriophage numbers.

Small drop plating plaque assay were conducted in a similar to the method of Mazzocco et al. (2009b). Herein, an sBHI 1% agar plate was overlayed with 3ml of sBHI 0.3% agar innoculated with 150 μ l of indicator bacteria (*H. influenzae* Rd 30S unless otherwise stated). This indicator bacterial inoculum had been subcultured in sBHI broth from an overnight culture, wherein 100 μ l of overnight culture was added into 10 ml fresh broth and incubated at 37°C, agitated at 100 rpm, for ~2 hours, until the OD₆₀₀ of the culture was ~0.1. The agar overlay was allowed to set at room temperature for around 5 minutes, after which point 10 μ l of ten-fold serial dilutions of the bacteriophage to be screened was place onto the agar surface, until all dilutions had been spotted. These spots were then left to dry for around 10 minutes, after which point the plates were then incubated overnight at 37°C. The following day the number of visible single plaques were counted and used to determine the concentration of bacteriophage present within the test solution (termed plaque forming units [PFU] ml⁻¹).

Bacteriophage concentration in PFU ml⁻¹, ϕ , is calculated as follows:

$$\phi \ = \ \frac{100N}{D}$$

Where N is the number of plaques present in the 10 μ l spot of the appropriate dilution, and D is the dilution factor at which plaques where the single plaques were visible.

For double agar overlay plaque assays, a method similar to Mazzocco et al. (2009a) was used. Bacterial cultures to impregnate the sBHI 0.3% agar were made as above, however, before addition of the agar/bacteria mix to the sBHI 1% agar plates, 100 μ l of bacteriophage sample was added to sBHI 0.3% agar/bacteria overlay. This mixture was then homogenised by inversion and poured onto the sBHI 1% agar plates. This was again allowed to set for around 5 minutes, after which point plates were incubated at 37°C and enumerated as above, except that the N is now multiplied by 10 rather than 100 to adjust for the volume of bacteriophage sample tested now being 100 μ l rather than 10 μ l used in the small drop assay.

2.4.2 Bacteriophage HP1c1 One-Step Growth Curve

In order to determine the time required for the bacteriophage to undergo one complete replicative cycle a bacteriophage one-step growth curve was conducted. Herein, 400 μ l of an overnight culture of either Rd 30S or Rd 30R was inoculated into 40 ml sBHI broth and grown at 37°C until reaching OD_{600} 0.1. To this culture bacteriophage HP1c1 to a final concentration of ~1 x 10⁵ PFU ml⁻¹ then added (MOI 0.001). This was then mixed by inversion, with 1 ml then taken and filtered with a $0.22 \ \mu m$ filter to determine the actual starting bacteriophage concentration. To allow an initial period of binding, the bacteriophage-host mix was then incubated at 37°C for 10 minutes. Following this the suspension was then centrifuged at 4696 x q for 5 minutes at 4°C to pellet the bacterial cells and bound bacteriophage. The supernatant was then removed and filtered through a 0.22 μ m filter to measure the concentration of unbound HP1c1. The pelleted bacterial cells were then resuspended in 40 ml fresh sBHI pre-chilled to 4°C. The cells were then again centrifuged at 4696 x q for 5 minutes at 4°C, with the unbound bacteriophage containing supernatant then filtered through 0.22 μ m filter for determination of the amount of unbound bacteriophage. The pelleted cells and bound was then again resuspended in 40 ml chilled sBHI and the incubation and centrifugation process repeated, and the supernatant again collected. Following centrifugation the bacteria and bound bacteriophage containing pellet was then resuspended in 40 ml room temperature sBHI broth. At this point the 40 ml suspension was then separated into 13x 2 ml aliquots in 7 ml bijous, one bijou per time point. These bijous were then incubated at 37°C, shaken at 100 rpm. Samples were taken at 5 minute intervals from 0-60 minutes. At each time point 1 ml was removed from the corresponding bijou, filtered through a 0.22 μ m filter for measuring bacteriophage concentration. Bacteriophage titres were measured using the small drop plaque assay of section 2.4.1.

2.4.3 Bacteriophage Adsorption Assay

In order to determine the rate at which bacteriophage HP1c1 was able to adsorb to host cells adsorption assays were conducted using the method of Kropinski (2009). Briefly, 10 ml of *H. influenzae* Rd30S was grown in sBHI from an overnight culture to 0.1 OD₆₀₀. From this culture, 9 ml was aliquoted into a 50 ml tube. The remaining 1 ml of bacterial culture was then placed on ice to determine the bacterial density used upon conclusion of the experiment, by spreading 100 μ l of 10-fold serial dilutions of cell suspension on sBHI 1% agar plates.

The 9ml Rd 30S tube being used for adsorption rate determination then received 2 μ l of 30 mg ml⁻¹ chloramphenicol (Sigma). This tube was then placed into a water bath at 37°C, which also allows agitation at 100 rpm, and was left for 5 minutes to allow the media to adjust to this temperature. After this point 1 ml of bacteriophage HP1c1, at a concentration of ~1-3 x 10⁷ PFU ml⁻¹, prewarmed

to 37°C, was added to the tube. From 0-10 minutes after the addition of HP1c1, in 1 minute intervals, 50 μ l was then removed from this tube and added to chilled tubes containing 950 μ l sBHI broth plus three drops of chloroform. These chloroform containing tubes were then briefly vortexed and placed back on ice. This was repeated until all time points had been completed. After this point plaque assays were conducted as per section 2.4.1, using 100 μ l from each of the chloroform containing tubes.

Bacteriophage adsorption constant, in ml min⁻¹, k, was determined using the equation:

$$k = \frac{2.3}{Bt} \log \frac{Pa}{P}$$

Where B is the concentration of bacterial cells and t is the time required for bacteriophage titre to fall from the original value (Po) to the final titre (P) at that point.

2.4.4 Determination of Phase Variable Resistance to Bacteriophage by Haemophilus influenzae

In order to determine the degree to which resistance to bacteriophage HP1c1 is conferred by the phase variable resistance mechanisms, efficiency of plating assays were conducted. These assays were conducted using the small drop plating plaque assay method as per Section 2.4.1, using a bacterial lawn of the appropriate hosts to test resistance. Herein, to determine the resistance attributed to phase variation of *lic2A*, a stock of 1×10^9 PFU ml⁻¹ bacteriophage HP1c1 was spotted onto cultures of *H. influenzae* Rd 30S, and Rd 30R, with the difference between the number of plaques formed on each lawn serving as a measure of resistance.

This was also conducted to determine the resistance characteristics attributed to phase variation of hsdM. Here, the difference in plaque number formation from the 1×10^9 PFU ml⁻¹ bacteriophage HP1c1 stock on lawns of RM118 $_{r^+HinDI}$ m⁺HinDI and RM118 $_{r^-HinDI}$ m⁻HinDI was used. Importantly, the bacteriophage HP1c1 stock had originated as the product of replication of HP1c1 in *H. influenzae* RM118 $_{r^-HinDI}$ m⁻HinDI m⁻HinDI, as this strain does not contain an active methyltransferase, and as such the bacteriophage used in the assay have unmethylated DNA at the HindI recognition sites.

2.4.5 Effect of Various Multiplicities of Infection (MOI) of Bacteriophage HP1c1 on *H. influenzae* growth

For investigation of the effect that various degrees of selection pressure elicited by bacteriophage has on the composition of phase variable H. influenzae populations, bacterial growth curves in the presence of a range of bacteriophage concentrations were conducted. For this, bacteriophage HP1c1 was incubated with *H. influenzae* isolates at MOIs of 0, 0.01, 0.1, 1, 10, and 100. MOIs were achieved through the addition of HP1c1 at a concentration of $\sim 1 \ge 10^7$ PFU ml⁻¹ (MOI 100) to a well of a 96-well with 10-fold serial dilutions then made of this in subsequent wells until $\sim 1 \ge 10^3$ PFU ml⁻¹. These were then made to 198 µl with fresh sBHI broth, with 2 µl of bacteria added to each well.

The bacteria that was added to each well was composed to varying degrees of *lic2A* ON and OFF cells i.e. ratios of Rd 30S:Rd 30R. The ratios tested were 100 % *lic2A* ON, 75 % ON, 66 % ON, 50 % ON/OFF, 66 % OFF, 75 % OFF, and 100 % OFF. These ratios were made by mixing *H. influenzae* Rd 30S and Rd 30R population at different ratios. For this, 100 μ l from an overnight culture of each strain was inoculated into separate tubes of fresh sBHI broth and allowed to grow to an OD₆₀₀ 0.1, corresponding to ~1 x 10⁸ CFU ml⁻¹. These two phasotypes were then mixed proportionally to give one inoculum containing both cells. For example in the 50 % ON/OFF population 50 μ l of the 0.1 OD₆₀₀ Rd 30S and 50 μ l Rd 30R cultures were then added to 900 μ l sBHI broth (giving ~1 x 10⁷ CFU ml⁻¹ of *H. influenzae* cells total, but ~5 x 10⁶ CFU ml⁻¹ of Rd 30S and Rd 30R). Thus following inoculation of the wells of the microtitre plate with 2 μ l of this suspension will give a final concentration of ~1 x 10⁵ CFU ml⁻¹ *H. influenzae* composed of both *lic2A* ON and *lic2A* OFF cells. This same procedure was carried out for the remaining population structures to be tested, with proportions of Rd 30S:Rd 30R cells adjusted accordingly.

The microtitre plates were then placed into a Multiskan GO Microplate Spectrophotometer (Thermo-Scientific) for measurement of bacterial density (OD_{600}) over a 25 hour period. Optical density was measured every 10 minutes over the 25 hour time course. Plates were shaken at the Multiskan's 'low' setting and incubated at 37°C throughout. Results were obtained from the Thermo-Scientific SkanIt Software, and exported to Microsoft Excel for analysis.

2.4.6 Oscillating prey assay

For testing the effect of host population phase heterogeneity on the migration of bacteriophage through H. influenzae populations an assay involving the transfer of bacteriophage HP1c1 to prey bacterial populations with varying lic2A expression states was derived.

Here 200 μ l of overnight culture of either *H. influenzae* Rd 30S and Rd 30R were inoculated into separate tubes of 20 ml sBHI followed by incubation at 37°C until reaching an OD₆₀₀ 0.1. The OD₆₀₀ 0.1 cultures were then split into 5 ml aliquots depending on the requirement of the cycling pattern. These aliquots were then inoculated with an MOI of ~0.01 bacteriophage HP1c1 (i.e. ~1 x 10⁸ CFU ml⁻¹ and ~1 x 10⁶ PFU ml⁻¹), with new sBHI added to give a total volume of 6 ml in each tube. From the 6 ml suspensions, 1 ml was then removed and filtered through a 0.22 μ m filter to determine starting PFU count of the cycle. The remaining 5 ml was then incubated for 50 mins at 37°C shaking at ~100





rpm. Incubation for 50 minutes allows only one bacterial replication cycle to be completed. At the end of the incubation the 5 ml culture was then filtered through a 0.22 μ m filter. From the filtrate 600 μ l was then passaged into a new tube containing 5 ml of OD₆₀₀ 0.1 Rd 30S or Rd 30R depending on the cycling pattern. This again was then adjusted to a total volume of 6 ml with sBHI, with 1 ml subsequently removed following mixing by inversion for filtration and measurement of bacteriophage titre. The above process was then repeated for a total of 20 cycles for each population composition tested.

Bacterial population compositions tested were: 100 % *lic2A* ON (cycling pattern = ON-ON-ON...), 66 % ON (cycling pattern = ON-ON-OFF...), 50 % ON (cycling pattern = ON-OFF-ON-OFF...), 50 % OFF (cycling pattern = OFF-ON-OFF-ON...), 66 % OFF (cycling pattern = OFF-OFF-ON...), and 100 % OFF (cycling pattern = OFF-OFF-OFF...). A graphic representation of this methodology can be found in figure 2.1. Bacteriophage titres at each time point were obtained by the small drop plating assay method of Section 2.4.1.

2.4.7 Multi-directional oscillating prey assay methodology

To estimate how heterogeneous populations would effect bacteriophage expansion, a miniaturised oscillating prey assay of section 2.4.6 was conducted over multiple 96-well plates. Here 100 μ l of overnight culture of *H. influenzae* Rd 30S and Rd 30R were inoculated into separate tubes of 10 ml sBHI followed by incubation at 37°C until reaching an OD₆₀₀ 0.1. At which point 250 μ l was added to a single well of a 96 well plate. Whether this first well contained *H. influenzae* Rd 30S or Rd 30R was dependant on the population structure and the predetermined random distribution of ON and OFF wells. The population structures simulated were the same as in section 2.1. Bacteriophage HP1c1 was then added to this well to give a final concentration of $\sim 1 \ge 10^6$ PFU ml⁻¹ per 300 μ l, with the bacteria/bacteriophage suspension then made to 300 μ l with sBHI broth. From this 100 μ l was removed, centrifuged at 1500 x g for 4 minutes and the supernatant extracted to determine initial bacteriophage titre by the small drop plating assay method of Section 2.4.1.

The remaining 200 μ l suspension was then incubated at 37°C shaking at ~100 rpm for 70 minutes. 70 minutes was used in this assay rather than the 50 minutes of the original oscillating prey assay as during the miniaturisation of the experiment bacteriophage replication was noted to require extra time to occur. It is important to note that this 70 minute period in this miniaturised assay was simply the equivalent of the 50 minute incubation of the larger assay (i.e. the end of only one viral replicative cycle). After this the 96 well plate was then centrifuged at 1500 x g for 4 minutes, to pellet bacterial cells with 20 μ l of the bacteriophage containing supernatant passed into the wells in the sequential manner shown in figure 2.2 (b). The remainder of the supernatant was then extracted from select wells (for sampling locations see figure 2.2 (a)) for determining bacteriophage titres again by the small drop







the 50 % ON population structure, while (b) shows the 66 % ON population structure. In both the examples shown the blue/purple wells represent *lic2A* ON cells, while grey cells represent the *lic2A* OFF cells. The 50 % *lic2A* OFF and 66 % OFF populations were structured as shown, however Figure 2.3. Layout of which wells received *lic2A* ON or *lic2A* OFF cells in the 50 % ON and 66 % ON population structure simulations. (a) shows the blue/purple wells in these examples contained *lic2A* OFF cells and the grey cells contained *lic2A* ON cells. The well outlined in red is the one which received the initial inoculum of HP1c1 in all cases.

plating assay method of Section 2.4.1.

The new well the 20 μ l of bacteriophage containing supernatant was transferred into contained 167 μ l of OD⁶⁰⁰ 0.1 of the appropriate *lic2A* expression state host. Following introduction of bacteriophage this well was then adjusted to 200 μ l using fresh sBHI broth. The plate was then incubated again at 37°C for 70 minutes after which the plates were again centrifuged at 1500 x g for 4 minutes with 20 μ l then transferred to the next set of wells. This process was repeated for a total of 10 cycles in the left, right, and downward directions, and 20 cycles in the upward direction (see figure 2.2 (a)).

For the 100 % *lic2A* ON or 100 % OFF populations all wells either contained Rd 30S or Rd 30R respectively. For the mixed populations each well was designated with a number of 1-631. First, to determine what wells should contain either the *lic2A* ON or OFF hosts, a random set of 316 numbers was requested, from between 1-631, using R (Team, 2016). The numbers produced were then applied to the corresponding wells as ON wells, with the remainder as OFF wells. A diagram of this can be seen in figure 2.3 (a). In the case of the 50 % OFF wells the inoculum was reversed, whereby the wells designated to contain *lic2A* ON cells, now contained *lic2A* OFF, and vice versa. For the 66 % *lic2A* ON population structure an additional random set of 100 distinct numbers from the 50 % OFF populations was then obtained from R, and applied to the map again (as seen in figure 2.3 (b)). For the 66 % OFF population structure the designations of ON and OFF *lic2A* hosts was reversed as was the case for the 50 % OFF population.

3. Analysis of the Phase Variable Type I Restriction Modification System of *Haemophilus influenzae*

3.1 Abstract

H. influenzae encodes a phase variable Type I restriction modification system, that switched between ON and OFF expression based on expansion or contraction of a pentanucleotide repeat tract (5'-GAGAC-3') within the hsdM gene. In this chapter the previous work of Zaleski et al. (2005) is examined in an attempt to validate their observations that 4x 5'-GAGAC-3' within *H. influen*zae $_{r^+HinDI m^+HinDI}$ gives a bacteriophage resistant phenotype. Their observations of a bacteriophage resistance phenotype at 4 repeats could not be replicated. Further *in silico* investigation indicated that 4 repeats would produce an OFF phenotype in RM118 $_{r^+HinDI m^+HinDI}$, however 3 repeats would (in addition the reading frame associated with 3 repeats would also confer an ON hsdM in the majority of *H. influenzae*). Although construction of a 3x 5'-GAGAC-3' hsdM RM118 was attempted, efforts did not prove successful, thus a resistance phenotype for 3 repeats remains to be tested.

Correlations between host disease state and hsdM ON/OFF state were investigated using the genome sequences available in genBank but no associations were observed. A putative role for phase variation of hsdM in the colonisation of patients with COPD was observed by analysing sputum supernatant samples that may warrant further investigation.

Finally, hsdS alleles associated with HindI-like systems were catalogued, with one allele noted as of further interest. hsdS32 allele was constitutively in the ON state across strains encompassing a number of MLST types and as such may play a key role in the physiology of *H. influenzae* strains containing this allele.

3.2 Introduction

Zaleski et al. (2005), found that phase variation within a type I restriction modification system, HindI, of a *H. influenzae* RdKW20 derivative, RM118 (Hood et al., 1996), altered the resistance/susceptibility to bacteriophage infection. Phase variation of HindI is the product of SSM along a pentanucleotide (5'-GAGAC-3') SSR tract, within the ORF of the *hsdM* gene of *H. influenzae* RM118 (Zaleski et al., 2005). *hsdM* encodes for the methyltransferase subunit of the HindI restriction modification system (Smith and Wilcox, 1970). With four 5'-GAGAC-3' repeats *H. influenzae* RM118 displayed partial resistance upon challenge with bacteriophage HP1c1 causing a $\sim 10^{-2}$ to 10^{-3} reduction in plaque formation, while, with three or five 5'-GAGAC-3' repeats this noticeable resistance was eliminated (Zaleski et al., 2005).

It has been previously hypothesised that the presence of phase variable contingency loci within restriction modification systems, may serve as a means to prevent bacteriophage adaptation to host restriction modification systems (Bayliss et al., 2006). The hypothesised mechanism for which is outlined in section 1.4.2. Briefly, as restriction modification systems are imperfect mechanisms of bacteriophage resistance, with inadvertent methylation of incoming bacteriophage DNA providing a means of circumnavigating the restriction activity of the system and providing resistance to resulting viral progeny (González et al., 1994; Lobocka et al., 2012). Any cell not expressing a restriction modification system is unable to provide such an avoidance strategy to the incoming bacteriophage and as such, in death, will nullify the methylation of the bacteriophage achieved within a previous host. Thus the constant generation of heterogeneous populations by phase variation of a restriction modification system could potentially provide a means to prevent bacteriophage circumnavigating the restriction modification systems via accidental methylation. Restriction modification system active populations can reduce bacteriophage numbers, with those escaping the system through methylation re-sensitised upon infection of a restriction modification system deficient host. From here the cycle can continue, either maintaining a small subpopulation of bacteriophage as a reservoir for horizontal gene transfer when conditions require; reducing the bacteriophage numbers to levels where the bacterial host is no longer under a great deal of selection; or reducing the overall selection pressure on the bacteriophage population to find another means of circumnavigating the system. A diagram of the above outlined mechanism can be found in Figure 1.9.

Thus the observations of Zaleski et al. (2005) makes the phase variable hsdM of H. influenzae an ideal system to study the effects of a phase variable restriction modification system on bacteriophage-host interactions. However, before further work was undertaken the observations of Zaleski et al. (2005) had to be further analysed and confirmed.

3.3 Presence of Phase variable hsdM in H. influenzae

To begin analysing the phase variable hsdM gene of H. influenzae RM118, the presence of homologues to this gene within other isolates of H. influenzae was investigated using the hsdM sequence data of the parental strain of RM118, Rd KW20, as a basis for screening the H. influenzae genome sequences available in GenBank. As of 06-09-16 there are 126 H. influenzae sequences with DNA sequence data available in GenBank, 17 as complete genome sequences, and the remainder as various levels of assembly in contiguous sequences. However, of these 126, two are the same sequence present as different assembly stages, and as such there are a total of 125 genomes that were used in analysis.

From these sequences each genome was screened for hsdM homologues using BLAST 2 Sequences (bl2seq) comparison against the HindI hsdM gene sequence of Rd KW20, using the redefined start codon of the gene suggested by (Zaleski et al., 2005). This resulted in homologues of the Rd KW20 hsdM gene being identified in 121 of these H. influenzae strains. The four strains where no hsdM data was obtained were all from strains where the sequence data was available only as contigs. Of these, the genomes of two strains (H. influenzae 159_HINF and 167_HINF), showed no regions with homology to the Rd KW20 HindI hsdM following pairwise comparison. The genome of one strain, 839_HINF contained a 109bp region at the 3' end of contig 11 corresponding to 94% DNA homology to the Rd KW20. The final excluded strain, HK1212, consists of 1262 contigs, with homology to the Rd KW20. The final excluded strain, HK1212, consists of 1262 contigs, with homology to the Rd KW20 HindI hsdM found across numerous contigs, making isolation of a complete sequence impractical. Therefore it cannot be excluded that some of these genomes may in fact contain a HindI hsdM homologue but due to potential incomplete coverage and assemblies these were not picked up by the screening method.

Even if the four excluded strains are taken as truly absent of homology to the Rd KW20 HindI hsdM, this would still mean 96.8% of the *H. influenzae* strains with sequence data available on GenBank contained a homologue of the phase variable Rd KW20 HindI hsdM. This would imply that this hsdM gene is highly conserved amongst *H. influenzae* strains, and therefore that this gene may be of importance in the long-term growth/survival of the cell. However, although the hsdM gene is conserved, the conserved nature of the phase variable 5'-GAGAC-3' repeat tract must also be elucidated.

3.4 Distribution of phase variable repeats within *H. influenzae* strains

Although the Rd KW20 HindI hsdM gene has been noted as being phase variable, the degree, or even presence, of phase variable contingency loci within other strains of *H. influenzae* has not previously been investigated. Therefore, the presence of SSRs within the Rd KW20 HindI hsdM homologues was explored using the hsdM sequence data obtained from the 121 strains from GenBank. The whole length of the hsdM genes of all strains were aligned and found to be highly conserved (81.3 %-100 % nucleotide identity). However, a high degree of nucleotide variability was found in the region within the first 10-30 bp of the gene, the region which correlates to the repeat tract location in Rd KW20 (High et al., 1993). Of the 121 hsdM sequences that were collected, 116 (95.9 %) were found to contain at least one instance of 5'-GAGAC-3' within the initial 16bp of the 5' end of the coding region, the location of repeats found in Rd KW20 (Figure. 3.1 (a)). Of the five strains where no 5'-GAGAC-3' units were found in the immediate 5'-end of the coding region, namely H. influenzae strains MiHi64, 137_HINF, 841_HINF, 781, and 156, only the two latter isolates lacked repetitive DNA motifs. The other three isolates all contained distinct repeat sequences to that of Rd KW20 within the same region of the hsdM ORF as Rd KW20's 5'-GAGAC-3' SSR. The SSRs of all three of these strains were pentanucleotide repeats; strains 137_HINF and 841_HINF both contain 5'-ACGAC-3' repeat tracts of 7 and 11 repeats respectively, while MiHi64 contains nine repeats of 5'-AAACG-3'. The difference in repeat tract length between 137_HINF and 841_HINF points to the occurrence of SSM on these SSRs within the hsdM gene of these stains, while the large length of the SSR of MiHi64 also indicates this tract as a potential locale for phase variable expression.

The length of SSR tracts within the hsdM ORFs of the 121 strains of H. influenzae can be noted in Figure 3.1 (b). Across the 121 hsdM sequences containing a 5'-GAGAC-3' SSR tract or alternative SSR, the SSR tracts ranged from only one repeat unit to a maximum length of 13 incarnations within a single ORF. However, the majority of SSR lengths were found to cluster between 1-5 units, with 91.7 % of isolates containing one of these SSR tract length. All of these units are 5'-GAGAC-3' units as the 3 isolates with alternative SSR tract units contained 8, 9, and 11 iterations of their SSR sequence. This may imply a low level of phase variation within this gene as phase variation rate is known to increase with repeat tract length (De Bolle et al., 2000). The occurrence of only one 5'-GAGAC-3' within the ORF of hsdM was noted as predominating across the strains, occurring in 38.8 % (47/121) of the sequences. This is interesting as a single 5'-GAGAC-3' would not be phase variable by SSM due to a lack of consecutive repeats for a slippage event to occur. This would again imply that a high phase variation rate for hsdM may not be required, possibly due to a lack of rapidly altering conflicting selection pressures to drive the need for expansion of the SSR tract. The presence of two 5'-GAGAC-3'



Figure 3.1. Presence of 5'-GAGAC-3' SSR units in the Hindl-like hsdM gene of 121 H. influenzae strains. (a) Example of 5'-GAGAC-3' repeat tract diversity in nucleotide sequences of H. influenzae hsdM gene. Nucleotide sequences used for illustration are derived from sequence data of nine H. influenzae strains, available on Genbank. The 5'-GAGAC-3' SSRs are highlighted in yellow. (b) Shows the distribution of the number of SSR repeat units within the ORF of all 121 H. influenzae hsdM sequences analysed in this study. SSR numbers include those strains that did not contain a 5'-GAGAC-3' sequence, whether that was due to having no repeat at all, or a unique repeat sequences in the same region as the 5'-GAGAC-3'.

units had the second highest occurrence across the strains at 20.7% (25/121), further implying a low phase variation rate for *hsdM* across *H. influenzae* species.

The presence of four repeats within the hsdM ORF, as noted as the ON state for hsdM by Zaleski et al. (2005) in the Rd KW20 derivative RM118 $_{r^{+}HinDI m^{+}HinDI}$, was present in 12.4 % (15/121) of *H. influenzae* isolates. While the two OFF states of their study were also reasonably abundant. Three 5'-GAGAC-3' repeats were present in 13 isolates (10.7 %), while five repeats were found in 11 of the isolates (9.1%). The diversity noted in SSR tract length allows for observation to what extent phase variation potentially occurs across the *H. influenzae* strains analysed; however the state of expression cannot simply be attributed based on repeat number, and hence examination of the selection for different phase variation states was the next parameter investigated.

3.5 Analysis of the putative coding states of the Phase Variable *hsdM* gene of *H. influenzae*

In order to determine the expression states associated with the SSR unit numbers found within the 121 HindI-like hsdM genes, hsdM sequences were translated from the putative initiation codon suggested by Zaleski et al. (2005). Figure 3.2 (a) shows putative expression states associated with the SSRs identified in the 121 H. influenzae strains based on whether the amino acid translation produced a putative full-length peptide sequence uninterrupted by stop codons. Across all strains it was observed that hsdM was most often found to be in the OFF state, with 69.4 % of the sequences coding for truncated products (Fig 3.2 (a)). This may imply that the presence of hsdM is detrimental to the cell under normal conditions, as the majority of strains do not encode a complete peptide. It was noted that the presence of one or two 5'-GAGAC-3' units within the SSR tract, the two most common tract lengths, produce both full and truncated putative amino acid sequences. The ratio of ON:OFF for one and two 5'-GAGAC-3' units were 0.52:1 and 1.78:1 respectively, implying that presence of one 5'-GAGAC-3' is more frequently associated with an OFF state than with an ON state, but that this is inverted for two repeats of 5'-GAGAC-3'. One and two 5'-GAGAC-3' units within the SSR of hsdM were also responsible for the largest total number of ON isolates, with 16 isolates each. However, the tract length that gave the greatest number of OFF isolates was the presence of also one 5'-GAGAC-3' unit, followed by four 5'-GAGAC-3' units, the repeat length Zaleski et al. (2005) found was associated with ON restriction modification activity in RM118.

While attempting to determine the SSR number that would produce a reading frame with a complete polypeptide for each hsdM sequence, it was observed that 18 of the 121 sequences were unable to produce complete amino acid sequences regardless of reading frame, thus these can be said to be fixed OFF in regard to phase variation. These included the two GenBank sequences containing no



Figure 3.2. Sequence analysis of 121 H. influenzae hsdM alleles. (a) and (b) show the distribution of SSR repeat tracts amongst H. influenzae strains shows the distribution and ON/OFF states for all 121 sequences analysed in this study (b) is the analysis of the remaining isolates after exclusion of sequences which had produce stop codons in all reading frames (fixed OFF). (c) Repeat number associated with expression of hsdM in the remaining 103 isolates. That is the closest number of repeats to produce a full length amino acid sequence after in silico expansion or contraction of the repeat and their associated ON (grey) or OFF (black) state, as well as the overall proportion of these states across the H. influenzae strains analysed. (a) units.

repetitive motif, together with a further 16 isolates containing repetitive motifs ranging from 3-13 repeats in length. As these genes would be putatively unable to express hsdM regardless of phase variable alterations in the active reading frame the inclusion of these sequences within the analysis of phase variation was thought to be somewhat extraneous, and as such these sequences were removed from further analysis, with the analysis of ON/OFF state in relation to repeat tract length repeated (Fig 3.2 (b)). The reduced total of 103 isolates showed a similar pattern with 61.2 % having the hsdM gene in an OFF state (Fig. 3.2 (b)).

This further proved to make no difference to the ON ratio of sequences consisting of one or two 5'-GAGAC-3' units, with both SSR number ON:OFF ratios maintained at 0.52:1 and 1.78:1 respectively. The greatest alteration came for strains containing three and five 5'-GAGAC-3' repeats, both of which were suggested OFF states by Zaleski et al. (2005). For the strains containing three repeats of 5'-GAGAC-3', before exclusion of the fixed OFF strains, 7 strains were ON for hsdM, while 6 where OFF. However, 5 OFF strains were excluded, leaving 87.5 % as having an ON state for hsdM with three repeats (Figure 3.2b). The observations for the change following the exclusion of fixed OFF examples with three repeats is interesting in light of Zaleski et al. (2005)'s observation of RM118 showing no restriction activity with three repeats of 5'-GAGAC-3', concluded to be due to hsdM being in the OFF state with this number of repeats. For five repeats no ON isolates were noted, while for the OFF state 11 isolates were initially found. Following exclusion of the 5 fixed OFF isolates this number therefore reduced to 6 examples. For four 5'-GAGAC-3' units, only 1 of the 15 sequences with this SSR number was found to contain a sequence which could encode the full length amino acid sequence for hsdMbefore exclusion of the fixed OFF isolates, with the exclusion removing only 2 OFF isolates. That is to say that in 93.3 % of cases hsdM was not in frame with this number was only slightly reduced upon exclusion of fixed OFF isolates, reducing to 80 %. Crucially however, the ON isolate with four repeats was not the parental strain of RM118, Rd KW20, but rather, H. influenzae HI1408. Rd KW20 on the other hand was noted as producing a truncated reading frame with four repeats.

This raises questions in relation to the observations of Zaleski et al. (2005), as they defined hsdM ON with four repeats of 5'-GAGAC-3' for RM118, yet its parental strain Rd KW20 is in the OFF state with this repeat number. As RM118 is derived from Rd KW20, it is highly unlikely that RM118 would be ON with this repeat number. This then raised the question of how many repeats were required for production of uninterrupted amino acid sequences in Rd KW20, as well as the remainder of the 103 *H. influenzae* strains capable of phase variation to the ON state.

To determine the number of repeats required to convert OFF isolates of H. influenzae to putative ON expression, repeat units were added or removed in silico to give the closest SSR tract that produces ON expression. The results of in silico adjustments for the GenBank isolates to produce ON expression via phase variation, are shown in Figure 3.2 (c). This shows the closest number of repeats required for

the production of an uninterrupted amino acid sequence.

The presence of two SSR units within the hsdM ORF was found to code for the putative ON state in the greatest number of *H. influenzae* strains (38/103, 36.9 %). In strains where two pentanucleotide repeats would facilitate expression of hsdM, the other ON states would include five, eight, eleven repeats etc. However, no ON isolates were observed for any of these repeat numbers. Furthermore, with two repeats the phase variation rate would be very low as smaller reapeat tracts have lower phase variation rates than larger tracts (De Bolle et al., 2000). This, combined with the abundance of strains with this SSR number, could potentially imply that this number of repeats may correlate to fixed ON expression, as isolates ON with this reading reading frame are centred on two repeats.

The reading frame which produced the highest number of putative ON isolates was the one associated with 0, 3, 6, and 9 repeats with 47/103 (45.6 %) isolates. Within this repeat set zero repeats was associated with ON expression in 11 examples, three repeats in 28 isolates, six repeats in 6 isolates, and nine repeats in 2 isolates (Fig 3.2 (c)). This is of interest as it is this frame that was found to produce a putative full length peptide sequence in Rd KW20, and additionally was a repeat number that was determined phenotypically by Zaleski et al. (2005) to be in the OFF phase.

The final set of repeats that share a reading frame, associated with one, four, seven, and ten repeats, was found to produce putative full length amino acid sequences in the lowest number of isolates with 18/103 (17.5 %) *H. influenzae* strains (Fig 3.2 (c)). This set had examples of ON expression across a range of repeat numbers, albeit in low numbers, however, as a number of these are longer tracts (i.e. 4, 7, 10 repeats) it is highly likely that phase variation may occur as a means of altering gene expression.

As Rd KW20 produced an ON state with the removal of one 5'-GAGAC-3' repeat, giving three 5'-GAGAC-3' repeats, the observations of the current study contrast to the findings of Zaleski et al. (2005). In order to determine if additional mutations exist within RM118 that allow phase variation into an ON state, as occurs in strains containing a different nucleotide composition to Rd KW20, the degree of similarity between the HindI hsdM sequences of Rd KW20 and its RM118 derivatives obtained from Zaleski et al. (2005) was next investigated.

3.6 Sequence analysis of the *hsdM* gene of *H. influenzae* RM118

The observation that the repeat tract within the phase variable hsdM gene of Rd KW20 was in the OFF state with four repeats is contradictory to the findings of Zaleski et al. (2005) for the Rd KW20 derivative, RM118. However, although RM118 is derived from Rd KW20 the sequences may not be identical and mutations acquired during *in vitro* growth of RM118 may result in expression with four

5'-GAGAC-3' repeats. For this reason the sequence of the hsdM gene of two of the RM118 isolates from Zaleski et al. (2005) were analysed to determine the level of similarity of these derivatives to the parental strain, Rd KW20.

The hsdM gene of the two RM118 isolates, RM118 $_{r^{+}HinDI} _{m^{+}HinDI}$ and RM118 $_{r^{-}HinDI} _{m^{-}HinDI}$, are noted to contain four and five repeats respectively by Zaleski et al. (2005) and are termed '+' and '-' based on the phenotypic observations of Zaleski et al. (2005). Here an opportunity arose to sequence the genomes of these isolates (conducted by Dr Andrew Millard, University of Warwick), and as such this sequence data was used to obtain the sequence of the full hsdM locus in each isolate.

Analysis of the sequence data confirmed that RM118 $_{r^+HinDI m^+HinDI}$ and RM118 $_{r^-HinDI m^-HinDI}$ contain four and five 5'-GAGAC-3' repeats respectively. However, pairwise comparison of the *hsdM* gene of RM118 $_{r^+HinDI m^+HinDI}$ and Rd KW20, which contains the same number of 5'-GAGAC-3' showed that these two *hsdM* genes have 99.94 % homology to each other. A single nucleotide difference was due to, a sequencing error in the Rd KW20 genome sequence, which is located 80 bp into the Rd KW20 *hsdM* ORF, correlating to a W, denoting the presence of either an adenine or thymine residue at this position. In both RM118 $_{r^+HinDI m^+HinDI}$ and RM118 $_{r^-HinDI m^-HinDI}$ this base is an adenine. Comparison of the sequence error region of Rd KW20 to another Rd KW20 derivative available on GenBank, RdAW, revealed that the Rd KW20 is likely to also contain this adenine as this same residue is also found in RdAW. Regardless, a single nucleotide substitution in this locus would not produce a stop codon within any of the three reading frames.

These sequencing results indicate that the *hsdM* gene of Rd KW20 is 100% identical to RM118 $_{r^+HinDI m^+HinDI}$. This signifies that RM118 $_{r^+HinDI m^+HinDI}$ would not produce a full length amino acid sequence for *hsdM* with four 5'-GAGAC-3'. Review of the amino acid sequence coded by RM118 $_{r^+HinDI m^+HinDI}$ in figure 3.3 (b) shows that translation would produce an interrupted product of only 24 amino acids, instead of the expected ~560 amino acids. RM118 $_{r^+HinDI m^+HinDI}$ which contains five repeats, was identical to both RM118 $_{r^+HinDI m^+HinDI}$ and Rd KW20 except for possessing one additional repeat unit. RM118 $_{r^+HinDI m^+HinDI}$ can be noted in figure 3.3 (a) to produce an even shorter length polypeptide of 16 amino acids before the interruption by a stop codon. When an *in silico* deletion of one 5'-GAGAC-3' repeat in RM118 $_{r^+HinDI m^+HinDI}$ (Fig 3.3c) was made, an uninterrupted peptide sequence of ~560 amino acids was produced, as noted for Rd KW20.

The above points raise further queries of the observation of Zaleski et al. (2005). If four 5'-GAGAC-3' repeats does not theoretically produce a full amino acid sequence in RM118 $_{r^+HinDI m^+HinDI}$ then a phenotype attributed to this SSR length would be unlikely. As such the phenotypes for RM118 $_{r^+HinDI m^+HinDI}$ and $_{r^-HinDI m^-HinDI}$ need to be further determined to clarify this genotype to phenotype correlation.

(A) RM118 r⁻HinDI m⁻HinDI

M Q A R R D E T R R D D Q D C R # K S H I T C K R D E T R R D E T T K I V D K N H T L Q A S E T R R D E T R R P R L S I K I T H Y N

ACAAGCATTTTTTAAATGAATTAGACGAAAAAACTTTGGGCATCTGCTGATAAACTCCGCAAACAACT

T S I F K * I R R K T L G I C * # T P Q T T Q A F L N E L D E K L W A S A D K L R K Q L K H F # M N + T K N F G H L L I N S A N N L

TGATGCCGCTAACTATAAACATATCGTTCTTGGTCTTATCTTCTTAAAATACATTTCCGATAGTTT

* C R # L # T Y R S W S Y L L K I H F R + F D A A N Y K H I V L G L I F L K Y I S D S F M P L T I N I S F L V L S S # N T F P I V S

(B) RM118 r⁺HinDI m⁺HinDI

M Q A R R D E T R R P R L S I K I T H Y N K C K R D E T R R D D Q D C R # K S H I T T S A S E T R R D E T T K I V D K N H T L Q Q A

CATTTTTAAATGAATTAGACGAAAAACTTTGGGCATCTGCTGATAAACTCCGCAAACAACTTGATG

H F # M N + T K N F G H L L I N S A N N L M I F K * I R R K T L G I C * # T P Q T T * C F L N E L D E K L W A S A D K L R K Q L D A

CCGCTAACTATAAACATATCGTTCTTGGTCTTATCTTCTTAAAATACATTTCCGATAGTTTCACCC

PLTINISFLVLSS#NTFPIVSP R#L#TYRSWSYLLKIHFR+FHP ANYKHIVLGLIFLKYISDSFTH

(C) <u>RM118^{GAGAC(3x)}</u>

M Q A R R D E T T K I V D K N H T L Q Q A F C K R D E T R R P R L S I K I T H Y N K H F A S E T R R D D Q D C R # K S H I T T S I F

TTAAATGAATTAGACGAAAAACTTTGGGCATCTGCTGATAAACTCCGCAAACAACTTGATGCCGCT

L N E L D E K L W A S A D K L R K Q L D A A # M N + T K N F G H L L I N S A N N L M P L K * I R R K T L G I C * # T P Q T T * C R #

AACTATAAACATATCGTTCTTGGTCTTATCTTCTTAAAATACATTTCCGATAGTTTCACCCATCAA

N Y K H I V L G L I F L K Y I S D S F T H Q T I N I S F L V L S S # N T F P I V S P I N L # T Y R S W S Y L L K I H F R + F H P S T

Figure 3.3. Putative reading frames of the hsdM gene of H. influenzae RM118 $_{r^+HinDI m^+HinDI}$, RM118 $_{r^-HinDI m^-HinDI}$, and the *in silico* predicted RM118^{GAGAC(3x)}.(a) and (b) show partial sequence data for the RM118 isolates of Zaleski et al. (2005). (a) is the isolate noted to be restriction modification deficient by Zaleski et al. (2005), consisting of four repeats of 5'-GAGAC-3', while (b) shows the sequence data of the isolate which they found to display restriction modification activity, consisting of five 5'-GAGAC-3' units. (c) shows the *in silico* created putative partial sequence of a RM118 isolate consisting of three repeats of 5'-GAGAC-3', termed RM118_{GACGA(3x)}. Data shown displays the DNA sequence (top lines), and their corresponding amino acid sequences. Amino acid sequences for all three reading frames are displayed below the DNA sequence, with letters indicating amino acids, and symbols (#,+,*) putative stop codons. The repeat regions of each sequence are underlined and highlighted in bold.

3.7 Determination of Bacteriophage Resistance Phenotype of RM118 Following Bacteriophage HP1c1 Challenge

With RM118 $_{r^+HinDI m^+HinDI}$ found to share the same reading frame and putative expressive/nonexpressive states as Rd KW20, the phenotypic state of the two RM118 isolates of Zaleski et al. (2005) was then tested to determine the restriction modification activity of each of the RM118 isolates. If the observations of Zaleski et al. (2005) hold true, bacteriophage HP1c1 should have a reduced ability to form plaques on a lawn of RM118 $_{r^+HinDI m^+HinDI}$, with four repeats of 5'-GAGAC-3' within its ORF. Infection of RM118 $_{r^-HinDI m^-HinDI}$ meanwhile, should produce no reduction in the plaque number as this strain, containing five repeats, should be devoid of restriction modification activity. If on the other hand the analysis of the sequences holds true then no reduction in the ability of bacteriophage to form plaques on these hosts should be observed.

To determine the phenotype of these isolates, ten-fold serial dilution of a $\sim 1 \ge 10^9$ PFU ml⁻¹ bacteriophage HP1c1 were spotted onto a lawn of each of the RM118 isolates, as any reduction in plaque formation on either strain would indicate bacteriophage resistance. No difference in bacteriophage plaque production was noted between RM118 r⁺_{HinDI} m⁺_{HinDI} and RM118 r⁻_{HinDI} m⁻_{HinDI} upon exposure to bacteriophage HP1c1 (two sample *t*-test: t = 0.1718, P = 0.8719; mean \pm SEM PFU ml⁻¹ values for bacteriophage HP1c1 infectivity on RM118 r⁺_{HinDI} m⁺_{HinDI} = $1.23 \pm 0.38 \ge 10^9$; RM118 r⁻_{HinDI} m⁻_{HinDI} m⁻_H

This would then imply that in terms of bacterial defence to HP1c1 that both cells are equipped with the same resistance level or resistance mechanism, or rather lack there of. As such this is consistent with the observation that hsdM could not be in the ON state with four or five repeats, as found in RM118 $_{r^+HinDI m^+HinDI}$ and RM118 $_{r^-HinDI m^-HinDI}$ respectively. According to the sequence analysis however, if hsdM contained three repeat units this would putatively facilitate the generation of a complete amino acid sequence. The observations here again raise questions about the observations of Zaleski et al. (2005), due to the abscence of any reduction in plaque formation when RM118 $_{r^+HinDI m^+HinDI}$ is exposed to bacteriophage HP1c1. One key piece of information remains to be obtained however, that of whether three 5'-GAGAC-3' does in fact provide restriction modification activity.

3.8 Construction of RM118_{GAGAC(3x)}

To determine whether three repeats of 5'-GAGAC-3' within hsdM provides a resistance phenotype to bacteriophage HP1c1 infection, a *H. influenzae* RM118 isolate containing three 5'-GAGAC-3' repeats



Figure 3.4. Infectivity assay of the ability of bacteriophage HP1c1 to produce plaques on phase variants of *H. influenzae* RM118. Results show the plaque production by bacteriophage HP1c1 on lawns of two phase variants of RM118. RM118 $_{r^{+}HinDI} m^{-}HinDI$ contains five repeats within its SSR (black bar) and was found to show no restriction modification activity in the study of Zaleski et al. (2005). RM118 $_{r^{+}HinDI} m^{+}HinDI$ (grey bar) contains four repeats and was found to show restriction modification activity in the same study. N = 3, Error bars = mean ± SEM.

within its ORF was required. To facilitate this RM118 was subject to mutagenesis in order to generate this three 5'-GAGAC-3' repeat strain, termed RM118_{GAGAC(3x)}. This three repeat mutant was generated in *H. influenzae* RM118 $_{r^{-}HinDI}$ as a background as this variant does not produce a resistance phenotype here, nor in Zaleski et al. (2005) (Figure 3.3).

3.8.1 Generation of RM118_{GAGAC(3x)}

For generation of RM118_{GAGAC(3x)}, a counter-selection cassette was used allowing two-step allelic exchange to produce an unmarked mutation of RM118 $_{r^{-}HinDI}$ m⁻HinDI. This counter-selection cassette has been demonstrated previously to allow the successful deletion of the *H. influenzae cyaA* gene by the same process (Johnston, 2012). The plasmid containing this counter-selection cassette, pJJ260, was graciously provided by Dr Jason W. Johnston (University of Kentucky, USA) and Prof. Michael Apicella (University of Iowa, USA). This tetR-*sacB*/*nptII* cassette allows initial positive selection through the presence of *nptII* conferring kanamycin resistance, while providing secondary counter selection through the presence of an inducible *sacB* under the control of a tetracycline repressor; *sacB* originates from *Bacillus subtilis* and encodes levansucrase an enzyme which converts sucrose to levans, with the accumulation of these levans toxic in Gram negative bacteria (Gay et al., 1983; Cianciotto et al., 1988; Pelicic et al., 1996; Fan et al., 2003; Johnston, 2012). The expression of sacB is under control of tetR as the presence of sacB in this cassette was initially noted as toxic to *E. coli*, even in the absence of sucrose, during the creation and application of the cassette (Johnston, 2012). For the two step allelic exchange two plasmids were constructed, with one carrying the counter selection cassette placed within the hsdM ORF, pCT1, and one carrying a hsdM allele carrying a 3x 5'-GAGAC-3' SSR, pCT2.

3.8.1.1 Construction of the cassette containing plasmid pCT1, and transformation into H. influenzae RM118 $_{r-HinDI}$ m⁻HinDI

pCT1 was generated in a two step process, first by SOE-PCR whereby two fragments, one upstream and one downstream of the hsdM SSR, were fused to create a product that would allow the insertion of the tetR-sacB/nptII cassette between two regions of homology in the second step following restriction and ligation. The 5' fragment, amplified using the primers HsdM_5'_F_1 and HsdM_5'_cass_R_2.2, is a 1515 bp product (Fig. 3.5) encompassing the region immediately 5' to the hsdM gene and the first 6 bp of the ORF. The 3' fragment was amplified using HsdM_3'_cass_F_3 and HsdM_3'_R_4.2, and is 2192 bp in length (Fig. 3.5). The 3' fragment begins ~ 300 bp downstream of the 5'-GAGAC-3' SSR extending for ~ 800 bp after the stop codon of hsdM.



Figure 3.5. Products produced during PCR amplification of *hsdM* for use in constructing pCT0.5 and pCT1 plasmids. In (a) products are the regions of homology that are eventually contained with within pCT1, while (b) shows the products that were used to create the 3x GAGAC SSR. Both image sets show the 5' and 3' amplicons respectively

Both internal primers, HsdM_5'_cass_R_2.2 and HsdM_3'_cass_F_3, contain a 5' tail that consists of a 12 bp region of homology at the 5' most end and an SmaI restriction site. The homologous tail region of each primer is the reverse complement of the first 12 bp after the SmaI site of the other

internal primer. This allowed for use of SOE PCR to link the two flanking regions by annealing of the homologous sequences followed by amplification with the two outermost primers, HsdM_5'_F_1 and HsdM_3'_R_4.2 which contain SacI and BamHI sites at their 5' ends respectively. This process resulted in the generation of a 3677 bp product (Fig. 3.5) that was then ligated into pUC19 following restriction digest of the fragments with SacI and BamHI to produce pCT0.5.

Positive transformants for the 5' and 3' fragments within pUC19 were screened for on LB agar plates containing 100 μ g ml⁻¹ carbenicillin, overlayed with X-gal and IPTG for blue/white screening. Eight resulting white colonies were then screened by restriction digest for successful ligation. Restriction digest was conducted using XbaI and PstI which, if ligation of the fragments was successful would produce four fragments of 12 bp, 965 bp, 1777 bp, and 3582 bp, with the latter three fragments visible as bands following gel electrophoresis. All colonies screened produced bands of the expected sizes and patterns and as such could now serve as the backbone for the introduction of the tetR-*sacB/nptII* cassette (Fig. 3.6).



Figure 3.6. Restriction profile from screening of transformant colonies of pCT0.5. Figure shows the profiles produced following restriction digest of ~ 1 μ g of plasmid extracted from colonies after blue white screening on LB agar with 100 μ g ml⁻¹ carbenicillin using XbaI and PstI. Each of the first eight wells are the digests of the selected transformant colonies. The final lane shows the digestion of pUC19.

To produce pCT1, the 5' and 3' flanking homology containing plasmid, pCT0.5 required the introduction of the tetR-sacB/nptII cassette. For this, the tetR-sacB/nptII cassette was released from pJJ260 using SmaI, and the pCT0.5 plasmid was also digested with SmaI to open the plasmid at the region between the 5' and 3' flanking sequences of hsdM. The tetR-sacB/nptII cassette and pCT0.5 were then ligated, transformed into DH5 α and transformants screened for on LB agar containing 15 μ g ml⁻¹ kanamycin. Colonies that grew on these plates were then screened for the insertion, and insertion of the correct orientation of the tetR-sacB/nptII cassette between the 5' and 3' homologous regions by restriction digest using NotI and PstI. If the tetR-sacB/nptII cassette has inserted in the correct direction then the tetR gene would be located at the proximal end to the 5' flank, and nptII would be located proximal to the 3' end. In this case, digestion with NotI and PstI will cleave the DNA into three fragments. PstI cleaves twice within pCT1, once in the multiple cloning site of the pUC19 backbone at the distal end of the 3' flanking fragment and another time within the 3' flanking region ~ 1 kb upstream of the other cleavage site. The second fragment is generated between the latter PstI site and the sole NotI restriction site which is found ~45 bp downstream of the *sacB* gene. This fragment is the main marker for discrimination between correct and incorrect orientation of the tetR-*sacB/nptII* cassette, as in the correct orientation this fragment will be 2427 bp in length, while in the opposite orientation will be 3520 bp. The remainder of the plasmid forms the final band, and would be 6453 bp if the cassette is in the correct orientation and 5360 bp in the incorrect orientation.



Figure 3.7. Restriction profile from screening of transformant colonies of pCT1. Figure shows the band profiles produced following restriction digest of ~ 1 μ g of plasmid extracted from colonies screened on LB agar with 15 μ g ml⁻¹ kanamycin using NotI and PstI. Each set of five lanes interrupted by a 1 kb Generuler ladder are three sets of five individual colonies picked from the kanamycin plates. The final lone sample is pCT0.5 digested with NotI and PstI (note PstI should cut twice within pCT0.5, giving a ~ 1 kb band and a ~ 6 kb band. No NotI sites are present in pCT0.5)

The results of this restriction screening of 15 transformants can be noted in figure 3.7. Here it was observed that eight colonies produced the expected banding pattern (Lanes 1, 4, 5, and 7-11). Plasmid DNA extracted from three of the positive transformants were used for transformation of *H. influen*zae RM118 $_{r^-HinDI}$ m⁻HinDI. The pCT1 plasmid was first linearised before transformation into *H. in*fluenzae using SacI, and transformation was carried out by the M-IV method outlined in Poje and Redfield (2003).

Cells that had successfully undergone the transformation process were selected for on sBHI 1 % agar plates containing 15 μ g ml⁻¹ kanamycin. Eight resulting positive colonies were then screened by PCR using primers pCT1_screen_F and pCT1_screen_R, which amplifies a product beginning in the 5' flanking region, within the *rbfA* gene upstream of the *hsdM* start codon, extending to within the tetR-*sacB/nptII* cassette, ending ~ 200 bp within tetR. This amplification should therefore produce a product of 554 bp. The results of PCR screening of the eight selected colonies can be seen in figure 3.8. All eight colonies were then screened on sBHI plates containing 10 % sucrose and 0.5 μ g ml⁻¹ chlortetracycline by plating serial dilutions of OD 0.1 cultures on these plates. Cells containing the tetR-*sacB/nptII* cassette should be more sensitive to sucrose than the wild-type RM118 r⁻HinDI m⁻HinDI. All transformant colonies showed sensitivity to sucrose compared to the wild-type and as such were ready for use in transformation with pCT2.



Figure 3.8. Electrophoresis image from screening transformed cells of *H. influenzae* following transformation with pCT1. The first 8 sample containing wells are the 8 colonies picked from the kanamycin containing sBHI plates, with the next single sample a positive control of extracted pCT1 DNA, with the final well as a negative control.

3.8.1.2 Construction of the 3x 5'-GAGAC-3' containing plasmid pCT2 and transformation into H. influenzae RM118_{cassette}

The 3x 5'-GAGAC-3' hsdM containing plasmid pCT2 was also generated through use of SOE-PCR. Here the 5' fragment was amplified using the primers HsdM_5'_F_1 and HsdM_5'_fix_R_2.1 which produces a 1514 bp product, and the 3' fragment was amplified using HsdM_3'_fix_F_3.1 and HsdM_3'_R_4.2, which gives a 2499 bp product (Fig. 3.9). These products together cover the entire length of the hsdM gene plus the areas upstream and downstream of hsdM. The inner most primers used in generating these fragments, HsdM_5'_fix_R_2.1 and HsdM_3'_fix_F_3.1 and HsdM_3'_R_4.2, each contain a region of homology at their 5' ends to the other internal primer, as was the case with the generation of pCT0.5. In this case the overlapping region contains a 3x 5'-GAGAC-3' tract that can be used to correct the reading frame of hsdM.

Fusing of the two fragments resulted in the generation of a 3996 bp product (Fig. 3.5 (b)) that contained a *hsdM* gene in the correct frame due to the presence of the 3x 5'-GAGAC-3' SSR. The fragment was then ligated into pUC19 following restriction digest of the fragments with SacI and BamHI, (the two outermost primers, HsdM_5'_F_1 and HsdM_3'_R_4.2, contain SacI and BamHI sites at their 5' ends respectively) and successful transformants were again screened for on LB agar plates containing 100 μ g ml⁻¹ carbenicillin, overlayed with X-gal and IPTG for blue/white screening.

Eight of the white colonies that arose from the screening process were then picked and subject to restriction digest with XbaI and PstI. Four fragments of DNA were expected with sizes of 12 bp, 965 bp, 2096 bp, and a 3582 bp fragment. The results of the restriction digest can be noted in figure 3.10, showing that all colonies screened produced the expected banding pattern.



Figure 3.9. Products produced during PCR amplification of hsdM for use in constructing the pCT2 plasmid. In (a) products are the regions of homology that contain a 3x 5'-GAGAC-3' tract at their 3' and 5' ends respectively. (b) shows the products produced following SOE-PCR that produces a ~4 kb fragment that contains a hsdM gene with 3x 5'-GAGAC-3' in its SSR tract.

The pCT2 plasmid was linearised before transformation into one of the *H. influenzae* colonies successfully transformed with pCT1 in section 3.8.1.1 using the M-IV method of Poje and Redfield (2003). The resulting transformants were then screened by plating of serial dilutions of the transformation mixture on sBHI plates containing 10 % sucrose and 0.5 μ g ml⁻¹ chlortetracycline. Following incubation, colonies were observed on the sBHI plates until the 10⁻³ dilution, however, following screening of these colonies no positive transformants were observed. Screening was conducted using primers pCT2_screen_F and pCT2_screen_R, which amplifies a 237 bp product within the region of *hsdM* deleted by the introduction of the tetR-*sacB/nptII* cassette. Additional attempts to transform these cells with pCT2 lead to similar results. Thus a 3x 5'-GAGAC-3' *H. influenzae* RM118 derivative remains to be generated.

3.9 Analysis of SSR tract lengths within the HindI-like *hsdM* gene of 82 COPD samples

Chronic obstructive pulmonary disease (COPD) is a debilitating respiratory condition, most commonly associated with smoking, characterised by symptoms such as a chronic cough, increased sputum production, an inability to undergo high to moderate physical activity, and chronic dyspnea (laboured breathing) (Moghaddam et al., 2011; Otczyk et al., 2011; Alikhan and Lee, 2014). Additionally patients with COPD are often subject to increased bacterial colonisation, particularly during exacerbations of the disease The most commonly isolated bacteria in cases of stable COPD, and COPD exacerbations,



Figure 3.10. Restriction profile from screening of transformant *E. coli* colonies of pCT2. Figure shows the band profiles produced following restriction digest of ~ 1 μ g of plasmid extracted from 8 colonies screened on LB agar with 100 μ g ml⁻¹ carbenicillin using XbaI and PstI. Each lane shows the result of digestion of the plasmid extracted a colony that grew on the LB agar. The penultimate lane contains a 1 kb Generuler ladder. The final lane sample is pUC19 digested with XbaI and PstI

is non-typeable *H. influenzae* (NTHi) (Bandi et al., 2001, 2003; Groenewegen and Wouters, 2003). In one study NTHi was found in 33% of stable COPD patients, and 87% of those with exacerbated symptoms (Bandi et al., 2003). The study of samples from patients with this condition is highly likely to provide information on the *in vivo* characteristics of the state of the HindI-like hsdM gene and of other phase-variable genes of *H. influenzae* during exposure to the inflammatory environment associated with COPD.

A total of 174 sputum supernatant samples from COPD patients were screened by PCR with primers amplifying a *H. influenzae* specific 280 bp region of DNA containing the 5'-GAGAC-3' repeat region of the HindI-like hsdM gene. These primers were designed within regions conserved across the *H. influenzae* species, allowing amplification if any *H. influenzae* is present that encodes for the HindIlike hsdM. The 174 samples originated from patients of varying disease state, that is to say some patients were suffering from exacerbations of the condition, while others were in a stable state or post-exacerbation. Additionally a small number of these samples originated from the same patients taken at unique visits. Upon screening of the 174 samples, 87 tested positive for hsdM by PCR (50%) using the screening primers. These positive products were then analysed by capillary electrophoresis, a method for estimating the exact size of a DNA fragment, and allowing for detection of expansion and contraction of the SSRs by comparison to a sample of known size. This analysis was successful for 82 of the samples. The loss of signal from some samples may be due to degradation of the DNA samples, as a number of the samples screened were positive for *H. influenzae* DNA by qPCR when originally processed (Marco Oggioni, personal communication, 2015).

The putative SSRs within the *H. influenzae* HindI-like hsdM genes of the COPD samples were found to have a similar repeat number distribution to strains analysed from the GenBank database (i.e. the majority containing less than 5 repeats). Interestingly, the most common putative SSR repeat



Figure 3.11. Distribution of 5'-GAGAC-3' repeats within samples from COPD patients. (a) shows the presence of insertion and deletion mutations within the hsdM gene of isolates containing a single 5'-GAGAC-3' repeat unit within their ORF to illustrate the potential diversity in sequence length around the repeat tract of the hsdM gene. (b) shows the distribution of putative 5'-GAGAC-3' repeat number found within the hsdM gene of 82 COPD samples. These data are based on capillary electrophoresis product sizes after PCR across the repeat tract of the hsdM gene. These are then correlated to a control product of known repeat length size, with repeat tract unit values extrapolated by comparison. Decimal values correlate to presence of additional insertion/deletion mutations within the product, such as those presented in (a); each 0.2 value correlates to 1bp.

number amongst the COPD samples was three repeats within hsdM (40.2%). This is of note as, based on the analysis of section 3.5, this implies that the hsdM may be mainly in an ON state, which is contradictory to the observations made for the most common SSR tracts of the GenBank isolates. This would imply that *H. influenzae* may benefit from the presence of their methyltransferase/whole restriction modification system during colonization of individuals with COPD. For the remainder of the COPD samples, due to the presence of insertion/deletion mutations within the regions immediately surrounding the hsdM 5'-GAGAC-3' repeat tract (Figure 3.11 (a)), further interpretation is even more speculative without direct nucleotide sequencing data for the complete gene. This was not possible due to limitations with use of these samples and cost. However the prevalence of ON hsdM phase-variants in COPD isolates does highlight a region of potential future study in the characterization of the phase variable HindI-like restriction modification system, and a putative role in virulence.

To further investigate the relationships between restriction modification repeat tracts and the disease state of COPD the above results were correlated to clinically relevant details, where available. Firstly, the repeat tract unit numbers were correlated to disease state at time of sampling. Herein, the repeat tract states where plotted according to the state of the patient during sampling, whether their COPD was stable, undergoing an exacerbation, or if sampling occurred a number of weeks post exacerbation. Data on disease state was available for only 43 of the 82 samples (Figure 3.12a). For patients with stable COPD, three repeats of 5'-GAGAC-3' was most prevalent (7/20, 35% of stable samples) followed by 2.2 repeats (4/20, 20% of stable samples). This would imply that stable COPD isolates may be predominantly ON if the hsdM sequences are consistent with the other observations on SSR number and expression state. For samples from patients suffering from exacerbations, 2.2 repeats was present in 14.3% of isolates and three repeats in 42.9%. A similar number of post exacerbation samples had three repeats (33%) but interestingly none of these isolates contained 2.2 repeats. This may point to immunological clearing of these bacteria from the host, however as only nine samples in total came from patients post exacerbation this remains highly speculative.

COPD is a chronic condition that occurs in differing degrees of severity between patients. This range in disease severity has been classified into four levels, termed the 'GOLD stages', by the Global Initiative for Chronic Obstructive Lung Disease (GOLD). Here, disease severity increases with GOLD stage, ranked mild, moderate, severe, and very severe by the numbers 1, 2, 3, and 4 respectively. Data on disease severity was available for 74 of the 82 COPD samples, and the correlations between disease severity and the repeat tract data can be observed in Figure 3.12b. Only four COPD samples were from patients with GOLD 1, or mild COPD, 50% of which were found to contain 2.2 putative 5'-GAGAC-3' repeats within their HindI-like hsdM, while the others contained three repeats. The majority of COPD samples here were either from patients with GOLD values of 2 (30/74, 40.5% of all COPD samples) or 3 (31/74, 41.9% of all COPD samples), meaning moderate and severe disease respectively. For GOLD 2 patients, a spread of SSR unit numbers was found from 1.4 to 6.8 repeats, with 2.2 and 3 repeats



Figure 3.12. Correlations of 5'-GAGAC-3' repeat tract length within the HindI-like hsdM gene of the 83 COPD samples to clinically significant attributes. (a) The putative repeat tract length within hsdM in relation to disease state for samples where data was available. These disease states include, stable COPD, samples from during an exacerbation, and examples after an exacerbation. (b) shows the distribution of 5'-GAGAC-3' unit number in relation to disease severity, based on clinical diagnosis using the GOLD scale for COPD disease severity; the greater the GOLD stage the greater the disease severity.
accounting for 46.7% (14/30) of the samples. However, for GOLD 3 samples this number was higher, with these two repeat tract lengths being found in 71% of patients. Only 11% (9/82) of the total COPD samples came from patients with GOLD 4 COPD, here, three putative 5'-GAGAC-3' repeats was also found in the majority of patients (4/9, 44.4%); whilst no very severe COPD samples contained 2.2 putative 5'-GAGAC-3' repeats. This points to the potential importance of isolates containing 2.2 repeats, and 3 repeats in colonisation of patients suffering from COPD. The 'true' effect of these putative repeat tracts could not be elucidated as these samples were not sequenced, and as such, this present work highlights only a potential role of phase variation in COPD which needs to be studied further.

3.9.1 GenBank *hsdM* sequences and disease

The observation that the repeat numbers in the COPD samples may correlate with COPD disease state suggests that there may be correlations between the ON/OFF state of the hsdM gene in disease. Therefore these correlations were investigated for each of the 121 *H. influenzae* sequences and the clinical state of the patients they were isolated from.



Figure 3.13. Clinical states associated with the ON/OFF state of hsdM 121 *H. influenzae* strains from GenBank where hsdM was detected. Pie chart shows the frequency of each clinical state across all 121 strains from GenBank that had detectable hsdM gene. Colours represent the different clinical states, with size of section proportional to the number of strains that were associated with that clinical state. Key shows colour to state association as well as the percentage of the total that strain sequences were available for.

As can be seen in figure 3.13 the *H. influenzae* strains with genome sequences in GenBank represent a wide range of clinical conditions, with the most prevalent condition being meningitis (41.32 %), followed

by otitis media (18.18 %). However, there remains a large proportion of isolates where there was no metadata available on the clinical condition of the patient (14.88 %). Rarer conditions associated with sequence strains on the other hand include those less commonly studied conditions or those not regularly attributed to *H. influenzae*, including Brazilian purpuric fever, conjunctivitis, necrotizing myositis, and septic arthritis.

To determine if there was any association between the putative ON/OFF states of each strains hsdM gene, and thus the HindI-like restriction modification system, the frequency of ON/OFF states of strains with an identifiable hsdM gene was plotted as a function of the disease state of the host from whom the strain was isolated. The resulting figure, figure 3.14, shows the distribution of each of the hsdM isolated from a particular clinical condition together with their putative expression state for hsdM. From this it can be seen that the largest number of sequenced strains come from patients with meningitis and that, for this condition, hsdM in the OFF state outnumber the ON strains with a ratio of 2.57:1.

Asymptomatic carriage shows a similar trend with an OFF:ON ratio of 5:1. This is consistent with the human colonisation studies of Poole et al. (2013), where the introduction of a bacterial population that was 7.4 % ON for hsdM did not increase beyond this level, with the frequency of ON isolates dropping as low as 1.6 % during the course of their experiment. Interestingly this was also the case for isolates from the sequenced strains that had been initially isolated from individuals with COPD. Here again, OFF examples considerably outnumbered ON isolates by 5:1. However, for these two clinical conditions there was a very low number, 6 sequences. In addition it should also be pointed out that the for COPD, three of the OFF isolates (strains 411, 584, and 1104) all come from the same patient.

The only clinical condition in which the hsdM ON state was greater than the OFF state was in individuals with bacteraemia where the ratio was 1.5:1. However, only 10 sequenced strains have been isolated from patients with bacteraemia. Contrastingly 22 isolates were obtained from individuals with otitis media, and as can be seen in figure 3.14 these strains are equally divided between hsdM ON and OFF, with 11 examples for each state. However, a number of these isolates are believed to be in the fixed ON or fixed OFF states due to a low number of pentanucleotide repeat units in their SSR tract.

Statistical analysis of the above examples showed no statistically significant associations with ON/OFF state compared to all other conditions in relation to an association with ON/OFF state (i.e. p > 0.05, χ^2 test, for association between ON/OFF state of each individual condition vs. all others). That being said, a weak, non-significant, association between otitis media and ON/OFF state was observed (p = 0.062, χ^2 test, for association between ON/OFF state of each individual condition vs. all other conditions).

The lack of significant associations between ON/OFF state and host disease state may be in part due to the limited number of samples available from each of the disease states restricting identification





of any trends. However, in addition restriction modification systems do not simply have one target sequence, but rather exhibit major variations due to allelic differences in the target recognition domain of the restriction modification system. These specificities can be diverse even within a bacterial species, and if two strains are targeting different sequences for methylation/restriction these systems may have differing fitness advantages in differing environments. As such this could mean that specificity of the sequences associated with the HindI-like restriction modification system may reveal a correlation between either hsdM ON/OFF state, the clinical state associated with these isolates, or even a combination of the two.

3.9.2 Analysis of *hsdS* gene of the 121 *H. influenzae* sequences available on GenBank

Sequence specificity in the Type I restriction modification system of H. influenzae, as with other Type I systems, is controlled by the hsdS gene. This gene encodes the specificity subunit that, when paired with two MTase subunits or two MTases and two REases, allows targeting of DNA resulting in methylation or restriction respectively. It is possible that there are correlations between hsdS sequence and ON/OFF expression of hsdM, and thus the sequence targeted by the restriction modification system. This would bring with it alterations in sites of methylation within the host genome, and thus altered gene expression throughout the bacterial genome. Alternatively these sequences may allow for the targeting of distinct bacteriophage populations and as such may not be required unless a specific family of bacteriophage are present in the environment.

Each of the samples where a hsdM gene was observed was screened for a hsdS gene, as these genes should occur nearby the hsdM gene. In all 121 cases where a hsdM gene was identified a corresponding hsdS counterpart was identified just downstream of the hsdM gene. The organisation of all of the H. influenzae HindI-like systems are of the form hsdM-hsdS-hsdR, with occasional additional ORFs occurring between. The hsdS genes of all 121 H. influenzae sequences were analysed for their homology. From the alignments it was observed that the HindI-like hsdS gene is a composite of conserved and variable regions (Fig 3.15). In Type I restriction modification systems variable regions are normally found that flank a central conserved region. Each of these variable regions are responsible for targeting one half of the target recognition sequence. The central conserved region acts as a spacer that alters the necessary gap between the target recognition domains. This organisation produces an asymmetric recognition sequence separated by a central spacer such as seen in the 5'-AAC(N)₆GTGC-3' recognition sequence of EcoKI (Kan et al., 1979; Bickle and Krüger, 1993).

In the *H. influenzae* HindI-like hsdS the first variable region begins immediately following the start codon of the gene (Fig. 3.15 (a)). This variable region, herein referred to as V1, extends for



sections of the HindI-like hsdS gene of 20 *H. influenzae* strains. In (a) part of the first variable region, V1, which starts just after the start codon can be seen. In (b) the central conserved region can be noted. (c) shows part of the second and final variable domain closest to the end of the hsdS gene, V2.

 \sim 220 amino acid residues, until it reaches the central conserved region shown in figure 3.15 (b). The central conserved region itself extends for \sim 90 amino acids, terminating in the second and final variable domain, V2 (3.15 (c)). V2 continues for around \sim 200 amino acids until reaching the end of the gene.

The degree of homology found across the entire length of the hsdS gene was investigated. Following alignment, and construction of an identity matrix, the percentage amino acid identity between each of the *H. influenzae* HindI-like hsdS sequences ranged from 26.1-100 %. This is indicative of the high degree of variability across the V1 and V2 domains. If taken independently the level of conservation between V1, central conserved, and V2 domains was between 17.4-100 %, 91.1-100 %, and 6.5-100 % respectively.

A phylogenetic tree was constructed using the full length of the hsdS gene. Heatmaps of the level of identity between each of the regions were then plotted against the tree of the relationships between each of the three main regions grouped as a function of their overall sequence relationships can be found in figure 3.16. From the heatmaps the high levels of variance in the levels of sequence identities for the V1 and V2 domains can be seen (Fig. 3.16 (a) and (c)), whereas there is a high level of conservation of the central domain (Fig. 3.16 (c)).

The V1 domain consisted of 28 potential peptide sequences. However, based on amino acid identity a number of these sequences have a high degree of homology. For instance, the V1 alleles 4, 7, and 8 have an amino acid sequence homology of ~ 96 %, and the V1 alleles 12 and 23, 18 and 19, and 27 and 28, each of which share homologies of ~ 99 , 99.5, and 99 % respectively. For 12 and 23 there is a difference of 2 amino acids near the end of the V1 domain, for 18 and 19 there is a single amino acid switch (lysine vs arginine) near the middle of V1, and for 27 and 28 the variance is a two amino acid switch, one at the N-terminal end and towards the end of V1.

The most common V1 domain was V1 allele 1, found in 17/121 isolates (14 %). Both allele 23 and 28 the next most common, both present in 14/121 sequences (11.6 %). Meanwhile the V1 alleles 4, 5, 8, 13, 14, and 24 were far rarer being only present in 1 of the 121 sequences each (0.8 % of the total number of sequences).

For the V2 domain there were observed to be a collective of 32 potential alleles. For seven groups of V2 domains, namely the V2 alleles 1 and 2; 3 and 4; 8, 9, 10, and 11; 13, 14, 16, 17, 18, 19, 21, and 22; 26 and 27; 28, 29, and 30; and 31 and 32, a high level of homology was observed. The levels of homology differ in these groups but all are \geq 89.9 %. The largest group with similar V2 domains, alleles 13, 14, 16, 17, 18, 19, 21, and 22 share between 89.9-99.6 %, V2 domains 8, 9, 10, and 11 have between 92.7-99.6 % homology, and the V2 domains 28, 29, and 30 are between 96.1-98.2 % identical. V2 domains 1 and 2 are 96.3 % identical with a small number of altered amino acids throughout the length of the V2 domain. For the V2 domains 3 and 4, 98.6 % homology is shared, due two nearby



Figure 3.16. Heatmap showing identity matrix of the level of homology between each domain of the hsdS genes, and the entire hsdS gene of the 121 *H. influenzae* samples available on GenBank. Samples are ordered according to phylogeny of figure 3.17 (Clockwise from 2842STDY5882085 to R2846; left to right on horizontal axis, and from top to bottom on vertical axis) which was constructed using the full length hsdS gene. Identity values are based on amino acid homology following alignment. (a)-(c) show the V1, central conserved, and V2 domains respectively, while (d) shows a heatmap of the entire hsdS gene

amino acids within the middle of V2 differing between the two sets of domains. The V2 domains 26 and 27, and 31 and 32 are extremely similar with 99.8 % homology between 26 and 27, and between 31 and 32. For both sets of sequence the sole difference is a premature stop codon at the proximal end of the V2 domain. HI1722 is the sole representative of the V2 allele 27, while HI2192 is the sole representative of V2 allele 32. In both of these domains a premature stop codon is found within the V2 domain and as such, while it is unclear whether the level of amino acid differentiation in the other aforementioned groups of V2 domains are enough to alter this domains target sequence, the premature truncation in HI1722 likely would be, irrespective of the otherwise highly similar V2 domain.

The most commonly found V2 domain are the V2 domains 1, 26, and 31. Each of these domains are found in 10.74 % of the *H. influenzae* sequences analysed. Of all the V2 domains 13 were observed in only one isolate, as mentioned some of these are due to only varying slightly from other V2 domains.

The central conserved region across all the hsdS sequences was largely conserved in the region of with a minimum of 91.1 % homology. However, if one looks at absolute homology a total of 34 central conserved domains can be noted. Although any differences in amino acid sequences within the central conserved region were taken into account when it came to designating alleles to each of the hsdS sequences, their relevance in alterations in target recognitions sequence specificity may be less relevant. Cowan et al. (1989) noted that after generating hybrid hsdS sequences by combining domains of the hsdS gene of two S. typhimurium Type I restriction modification systems, StySP and SB, that hybrids fused at their central conserved region had the same specificity as those which contained none of the StySP central region and only the SB central conserved region. Therefore some of the hsdSallele numbers suggested here may have the same specificity as they are homologous in every other region except small differences in the central conserved domain.

Overall allele designations were then made for each of the *H. influenzae* sequences analysed resulting in sub-division into 40 hsdS alleles. It is herein proposed that these alleles be referred to in a similar way to that of the modA allele of *H. influenzae* (Fox et al., 2007), wherein each allele is suffixed with their associated allele number. As such these alleles will from hereon be referred to as hsdS1-40. A list of the associated domains for each of these 40 hsdS alleles can be found in table 3.1, together with the example strains for each allele number and the proportion of *H. influenzae* sequences checked that contain that hsdS allele.

It is worth noting however, that some of these hsdS alleles differ within the central conserved domain and are likely to have overlapping, or identical, recognition sequences. For example the hsdS alleles hsdS21 and hsdS22; hsdS29 and 30; and hsdS38 and hsdS39 are all classed as distinct alleles only due to differences in the peptide composition of the central conserved domain.

The most prominent hsdS alleles were hsdS1 and hsdS23, both found in 13/121 isolates (10.74 %). However, 19 of the hsdS alleles are present only in a single *H. influenzae* strain, with the majority

Strain	V1 domain	Conserved V2 domain domain		hsdS allele	% of total isolates			
10810	1	1	1	1	10.74			
Hi359	1	4	4	2	3.31			
137_HINF	2	2	2	3	1.65			
HI1980	3	3	3	4	3.31			
CCUG 26214	4	4	4	0.83				
F3031	5	5	5	6	0.83			
KR494	6	6	6	7	3.31			
CGSHiCZ412602	7	7	7	8	0.83			
PittGG	7	7	8	9	0.83			
60294N1	8	8	9	10	0.83			
492_HINF	9	9	10	11	0.83			
22.4-21	9	27	24	12	0.83			
R2866	10	10	11	13	3.31			
F3047	11	11	12	14	0.83			
1057_HINF	11	13	14	15	3.31			
HI1417	12	12	13	16	4.13			
HI1426	13	14	15	17	0.83			
R3021	14	15	16	18	0.83			
NT127	15	16	17	19	0.83			
411	15	30	28	20	2.48			
Hi322	16	17	18	21	0.83			
1209	16	18	18	22	2.48			
PittHH	16	19	19	23	0.83			
RdKW20	17	20	20	24	1.65			
2842STDY5882045	18	21	21	25	1.65			
615_HINF	19	21	21	26	0.83			
Hi361	19	21	22	27	5.79			
477	20	22	23	28	2.48			
HI2007	21	23	23	29	0.83			

Table 3.1. Allele classification from analysis of the 121 strains with an identifiable hsdM gene in GenBank. Table also shows the allele classification for each od the three main domains associated with the overall hsdS allele classification.

Strain	V1 domain	Conserved V2 domain domain		hsdS allele	% of total isolates		
2842STDY5882061	21	26	23	30	1.65		
156_HINF	22	28	25	31	1.65		
723	23	29	26	32	10.74		
HI1722	23	29	27	33	0.83		
HI1408	24	30	28	34	0.83		
2842STDY5882080	25	31	29	35	1.65		
PittAA	26	32	30	36	4.13		
2019	27	32	30	37	4.96		
C486	28	33	31	38	9.92		
HI1988	28	34	31	39	0.83		
HI2192	28	33	32	40	0.83		

Table 3.1. – Continued from overleaf

containing differences in either of the two variable domains. For example hsdS8 and hsdS9 have the same V1 and central region but differ in their V2 domain. As no recognition sequences for any *H. influenzae* Type I restriction modification systems is currently known it is not possible to infer any sequence specificities based on the sequence analyses.

The relationships between the ON/OFF states of hsdM and each of the hsdS alleles were then investigated. Within this phylogenetic tree (Fig. 3.17), one group of hsdS sequences clearly stands out, the arm of the tree incorporating HI2114-1123HINF.

This arm of figure 3.17 has hsdM in an ON state across all samples. However, this extension is associated with two sets of hsdS alleles, hsdS32 and hsdS33. These alleles differ only in the V2 domain, with the hsdS33 allele having a premature stop codon within the V2 domain. As such this would be expected to considerably alter the 3' recognition sequence of this system. Interestingly though, each of the hsdM genes associated with these systems contain 2x 5'-GAGAC-3', and as discussed in sections 3.4 and 3.5 it may be that with this number of repeats the *HindI-like* restriction modification system will be in a fixed ON state. Thus the hsdS alleles may aim the restriction modification system to sequences that play significant roles in the cells survival.

This could be because the MTase component and the specificity subunit allow for epigenetic regulation of host genes. However, the HindI-like REases of each of these strains does appear to encode a full length product and as such restriction activity, together with the methylation activity, will be constitutively expressed possibly reflecting the need for restriction of a common bacteriophage.

It is further worth noting that the 13 hsdS32 do not all represent the same MLST type, but rather these 13 strains are members of at least 5 MLST groups, 3, 14, 33, 136, and 367. HI1722, which has the hsdS33 allele meanwhile is a member of MLST group 329. Each of the strains with either hsdS32 or 33 were isolated from diseased individuals, including those suffering from bacteremia, otitis media, and meningitis etc, although there is not an association with a particular disease state or site of isolation. The continual expression of the restriction modification system with hsdS32 (10.74 % compared to a mean frequency of 2.5 %), and the prevelance of this hsdS allele across multiple MLST types may signify that this hsdS allele is of benefit to H. influenzae survival.

Other notable groups of hsdS alleles include hsdS1 and hsdS38, prevalent in 10.74 % and 9.92 % of the sequences screened respectively. All strains with the hsdS1 allele were OFF for hsdM, while the hsdS38 had mixed expression of hsdM, with the OFF state favoured. The strains with the hsdS1 allele are largely from the same MLST group, with the hsdS1 allele of 11/13 strains from MLST group 6 (the other two are from group 190). For hsdS38, the strains with this allele were from a more diverse range of MLST groups, including 34, 36, 107, 155, 156, 244, 396, and one isolate of unknown MLST type. That being said there was a significant link between MLST type and hsdS allele (p < 0.05, χ^2 test, for association between MLST types and hsdS allele). Namely that although hsdS alleles could



Figure 3.17. Phylogenetic tree of the hsdS gene of 121 *H. influenzae* sequences available on GenBank. Labels display the *H. influenzae* strain the hsdS sequence belongs to, with labels in green indicative of strains where the hsdM gene was in an ON state based on the sequence analysis of section 3.5, and red labels indicating the hsdS sequences associated with an OFF hsdM. \blacklozenge indicates sequences where hsdM is appears fixed OFF. The four most common hsdS groups are annotated. Tree was constructed based on an amino acid homology, following alignment, using the Maximum-Likelihood method using the Whelan and Goldman model. Values indicate boot strap values following 1000 replications.

be associated with multiple MLST types in all but one case MLST types were associated with only one hsdS allele.

In terms of the SSR tract lengths, the hsdS1 strains all contained a sole 5'-GAGAC-3' unit and thus are in a permanently fixed OFF state. For hsdS38 the hsdM genes contain between 2-13 repeats with three strains having a broken hsdM that would be unable to turn ON; two isolates in the ON state; and seven in the OFF state. This suggests expression of HindI-like with the specificity of hsdS38 may provide transient benefits to the bacterial population. Through phase variable control then, bacterial strains expressing hsdS38 can gain the advantage associated with the expression of a restriction modification system with this target sequence, whilst retaining the potential to access OFF variants that and thus do not incur any costs associated with restriction modification system expression.

Other hsdS alleles were also associated with a mix of hsdM ON/OFF examples, and thus implying transient benefit through phase variation include hsdS2, hsdS36, and hsdS37. The remaining hsdS alleles were either in only one state (either ON or OFF), or have the same repeat length but differing expression (i.e. the system is active through a means other than phase variable expansion and contraction of the SSR units).

3.10 Discussion

Mechanisms for the adaptation of an organism to constantly changing environments are crucial for their successful survival (Moxon et al., 1994). Phase variation provides one such mechanism for the adaptation of bacterial species to readily changing environmental pressures (Bayliss, 2009; Moxon et al., 2006; van der Woude and Bäumler, 2004). Environmental pressures that force population changes include, immune targeting of the bacterial cells, changes in nutrient availability, relocalisation of bacteria to new niches, or, the main focus of this current work, the predation by viral predators such as bacteriophage (Koskella and Meaden, 2013; Bayliss et al., 2006; Williams, 2013). Phase variable bacteriophage defence systems have been described (Hoskisson and Smith, 2007), and in *H. influenzae* one of the examples that have been described is the HindI-like restriction modification system of *H. influenzae* RM118 (Zaleski et al., 2005).

Phase variation of HindI is caused by a pentanucleotide SSR tract within the MTase encoding hsdM gene (Zaleski et al., 2005). ON phase variants of hsdM of *H. influenzae* RM118 were previously shown to have a reduced susceptibility to bacteriophage HP1c1, in the order of 10^{-2} - 10^{-3} . The reduced infectivity, was previously associated with a restriction modification system of *H. influenzae* (Glover and Piekarowicz, 1972). Zaleski et al. (2005) then went on to correlate this partial resistance phenotype to alterations in the 5'-GAGAC-3' repeat number of the hsdM gene. Isolates with the reduced susceptibility phenotype were observed to contain four 5'-GAGAC-3' repeats whereas isolates with complete

sensitivity to bacteriophage HP1c1 contained three or five repeats. Due to these observations this system seemed like the ideal model for study of the effects of phase variation of a restriction modification system on the dynamic interactions between bacteriophage and their hosts.

Surprisingly, upon doing background sequence analysis of the hsdM genes of the sequenced strains available in GenBank it appeared that the repeat tracts of the parental strain of RM118 r⁺HipDI m⁺HipDI and RM118 $_{r-HinDI}$ m⁻HinDI, Rd KW20, would encode a truncated hsdM gene based on the starting position as suggested by Zaleski et al. (2005). This was a troubling observation as the hsdM gene of Rd KW20 contains four 5'-GAGAC-3' repeats within its reading frame, as does the supposedly restriction proficient RM118 $_{r^+HinDI} m^+HinDI$. Thus if RM118 $_{r^+HinDI} m^+HinDI$ and the strain it is derived from share an absolutely homologous hsdM gene it would be therefore be highly unlikely that RM118 $_{r^+HinDI m^+HinDI}$ would express an active restriction modification system. As such, the hsdM genes of each of the hsdM ON and OFF RM118 isolates was sequenced and compared to the hsdM sequence of the parental Rd KW20 strain. From this it was confirmed that the hsdM genes of $RM118_{r^+HinDI m^+HinDI}$ and Rd KW20 are identical. This would therefore mean that if hsdM begins at the starting location suggested by Zaleski et al. (2005) then with four repeats the hsdM gene of RM118 $_{\rm r^+HinDI\ m^+HinDI}$ would produce a truncated peptide product. Lack of restriction modification activity in RM118 $_{\rm r^+HinDI\ m^+HinDI}$ (and RM118 $_{\rm r^-HinDI\ m^-HinDI})$ was confirmed upon testing sensitivity of the cells to bacteriophage HP1c1 infection, where both RM118 $_{r^+HinDI}$ m⁺HinDI and RM118 r-HinDI m-HinDI had the same sensitivity to bacteriophage HP1c1 challenge.

A number of *H. influenzae* isolates were analysed and found to contain three, four, and five 5'-GAGAC-3' repeats within hsdM, as observed by (Zaleski et al., 2005). Five 5'-GAGAC-3' repeats did not code for a complete amino acid sequence, as observed by Zaleski et al. (2005). Crucially, four repeats within the hsdM gene also did not give rise to a complete amino acid sequence in the majority of H. influenzae isolates (Figure 3.2a) including for Rd KW20, the strain from which RM118 is derived. Conversely, three repeats was shown to produce a complete peptide in the majority of isolates with this repeat number, contrary to Zaleski et al. (2005). Exclusion of strains with additional mutations that prevented full translation of the peptide in any reading frame showed that putatively ON hsdMstrains were almost invariantly associated with three repeats of 5'-GAGAC-3' within the SSR. This implies that the reading frame associated with three 5'-GAGAC-3' is commonly associated with ON expression. Furthermore, modification of all sequences in silico to their closest ON reading frame following addition of 5'-GAGAC-3' units, showed that all isolates with three, four and five 5'-GAGAC-3' repeats in their SSR would switch to ON isolates with contraction to three 5'-GAGAC-3' repeats within the hsdM SSR. One isolate, HI1408, was observed to be putatively ON for hsdM with four 5'-GAGAC-3' repeats, and hence it was necessary to sequence the RM118 gene to check for small indels around the SSR.

The hsdM gene of two of the RM118 phase variants discussed in Zaleski et al. (2005) were sequenced. It was determined that neither the five repeat RM118 rHinDI m-HinDI, nor the four repeat RM118 $_{r^+HinDI m^+HinDI}$ were able to encode for a full length peptide sequence. However, *in silico* manipulation of the 5'-GAGAC-3' repeat tract showed that a putative full peptide sequence could be produced with three 5'-GAGAC-3' repeats, as observed from the genomic data available for a number of isolates in GenBank, as well as for Rd KW20. Whether three repeats can mediate phage resistance remains to be empirically confirmed through generation of a three 5'-GAGAC-3' repeat *hsdM* derivative of RM118 and is required before a definitive interpretation of the Zaleski et al. (2005) data can be attempted.

The explanation of the observations of Zaleski et al. (2005) remain to be elucidated. However, during the course of the current investigation it was noted that clear and reliable measurement of plaque formation between H. influenzae and bacteriophage HP1c1 occurs only under very specific conditions. Herein, utilisation of older media, or using cultures outside of early exponential growth, can result in the production of smaller plaques making enumeration difficult, or potentially resulting in the production of highly turbid plaques, near indistinguishable from the surrounding bacterial lawn. This may make false reductions being noted more common but this does not explain the reliable demonstration that isolates with four 5'-GAGAC-3' show the reduced susceptibility characteristics in the work of Zaleski et al. (2005). With 4x 5'-GAGAC-3' Rd KW20 does contain a large uninterrupted region present in the same locus as the hsdM, extending from ~ 70 bp before the start codon associated with the 3x 5'-GAGAC-3' frame, and extending until the same stop codon used by the 3x 5'-GAGAC-3' frame. That being said, the first methionine is not present in this frame until ~ 700 bp downstream of the start codon of the 3x 5'-GAGAC-3' ORF. Thus, should this methionine act as the start for hsdM it would produce a truncated product of 313 amino acids (opposed to the amino acid product of the 3x 5'-GAGAC-3' start position). The original annotation of the Rd KW20 (before the amendment of Zaleski et al. (2005)) states that the hsdM ORF starts at the a value start codon 351 bp upstream of the 3x 5'-GAGAC-3' start site. Zaleski et al. (2005) disputed this as a start codon based on the region preceding this value containing a high level of homology to the MTase subunit of other Type I restriction modification systems. This led Zaleski et al. (2005) to redefine the start codon to be the methionine that, in contrast to their results, requires 3x 5'-GAGAC-3' to produce a complete polypeptide sequence.

Attempts in the current study to generate a 3x 5'-GAGAC-3' strain were unsuccessful. Although plasmids containing the desired alleles to conduct the allelic exchange were generated, and the counterselection cassette was successfully inserted into *H. influenzae*, an isolate that had successfully exchanged the cassette for the 3x 5'-GAGAC-3' allele was not obtained. Sucrose sensitivity was observed, however following screening of colonies that grew on the sucrose containing plates no colonies contained the 3x 5'-GAGAC-3' allele. Spontaneous inactivation of *sacB* is often observed, at a frequency of ~10⁻⁵ (Blomfield et al., 1991) and as such may explain the colonies identified in this current work. Attempts to generate the 3x 5'-GAGAC-3' RM118 strain is currently ongoing.

Overall, the investigation of the 125 H. influenzae strains within the GenBank database revealed that 121 of these isolates contained a homologue to the hsdM gene of the HindI-like restriction modification system allowing for further investigation of the putative translation state of hsdM and SSR unit numbers. For the 4 genome sequences where no full hsdM could be identified the genomes are in contiguous sequences, and as such, due to potentially missing coverage, the presence of hsdM homologue in these strains cannot be excluded. This would imply that the hsdM gene of the HindI-like restriction modification system is conserved within the majority, if not all, H. influenzae strains. This is not necessarily surprising with similar observations noted for other Type I restriction modification systems (Waldron and Lindsay, 2006). Furthermore, unlike Type II restriction modification systems where deactivation of the methyltransferase activity can result in the death of the cell, providing the notion that such systems are 'selfish' elements (Naito et al., 1995), whereby their continued presence in their host chromosomes is maintained by REase activity on any unmethylated chromosomal DNA, Type I systems are not as abhorent to subunit inactivation (Kulik and Bickle, 1996). As such the continued presence of the HindI-like type I restriction system points to a beneficial role of the system in the physiology of H. influenzae. However, interestingly, due to hsdM being under phase variable control, conservation does not necessarily mean the systems are active, and upon further analysis of the ON/OFF states of these hsdM genes it was determined that in the majority of the sequenced strains available for analysis, the hsdM gene, and thus the HindI-like restriction modification system, was in the OFF state.

The fact that hsdM is in the OFF state in the majority of the *H. influenzae* genomes, if reflective of their state *in vitro* and not simply an artefact of adaptation to plate culturing, would possibly imply that its benefit to the host cell is transient. This observation provides underpinning evidence to ideas that their involvement may be in ways such as the temporal activation of restriction modification systems lowering selection pressure on bacteriophage to adapt to these systems, the proposed means of nullifying the 'leaky' nature of restriction modification systems in host defence to bacteriophage, or coinciding with potential roles of such systems in gene regulation through epigenetic activity etc (Bayliss et al., 2005; Hoskisson and Smith, 2007). Add into this that the majority of the repeat tracts within the hsdM genes were of 1-5 repeat units in length, would point to a low level of phase variation frequency as the rate of switching has been shown, at least for tetranucleotide repeat tracts, to be proportional to the length of the repetitive sequence (De Bolle et al., 2000).

In terms of the association between the overwhelming number of H. influenzae strains in the OFF state for hsdM and its role in bacteriophage resistance, this would imply that there is not sufficient pressure being exerted on the bacterial host to switch ON its hsdM gene due to scenarios including the presence of bacteriophage that are immune to targeting by HindI-like due to lack of target sequence in the bacteriophage genome; bacteriophage present only at low numbers; the complete lack of bacteriophage that can infect the *H. influenzae* strains in the environment; or even absence of selection for this system to switch ON due to complete resistance being provided by an alternative resistance mechanism; or selection for OFF due to a cost with ON expression. However, again its benefit may be transient, with selection to turn ON occurring in the presence of a bacteriophage sensitive to restriction, being antagonised by a negative cost associated with the expression of the system, selecting for the OFF state. This negative selection pressure could be the role of restriction modification systems in the epigenetic regulation of other host genes.

An example of a phase variable restriction modification being detrimental to the bacterium's ability to colonise a host has been observed in *Helicobacter pylori*. Gauntlett et al. (2014) noted that 2 phase variable restriction modification systems in *H. pylori* OND79 when in the ON state reduces the overall colonisation ability of the bacterial population. These phase variable genes, HP1522 and HP0464, encode a the MTase subunit of a Type II system and the REase of a type I restriction modification system respectively. In the wild-type *H. pylori* OND79 strain both of these systems are found in the OFF state. Upon alteration of these genes to produce functional products a reduction in the ability of mutants for each of these genes to colonise the mouse stomach was noted; with reductions of $\sim 10^2$ noted upon activation of HP1522, and $\sim 10^3$ for activation of HP0464. The reason for the reduction in colonisation upon activation of HP0464 is unknown. The authors however did suggest that the previous association between the OFF state of HP1522 and up-regulation of a *H. pylori* adhesin HopG in *H. pylori* P12 (Srikhanta et al., 2011). Currently no association between *hsdM* activity and alterations in bacterial phenotype (out-with bacteriophage resistance) has been attributed to *hsdM* or the HindI-like restriction modification system.

During this work an opportunity arose to study sputum supernatant from individuals with COPD. Following GeneScan analysis of 83 sample which produced a product following PCR screening of the hsdM repeat tract the putative repeat tracts for each of the samples was compiled. From this it was observed that the repeat tract sizes were similar in distribution to that noted for the 121 GenBank sequences analysed. That is to say that the majority of the samples contained hsdM genes with a repeat tract lengths of between 1-5 repeats. Interestingly though, unlike the data from the 121 GenBank samples, where the majority of isolates were observed to contain 2x 5'-GAGAC-3' repeats, the GeneScan analysis of the 83 COPD sputum supernatants appeared to be centred around containing 3 repeats in their pentanucleotide SSR tract. This is interesting as based on the results from correlating repeat tract numbers to putative ON/OFF states (Section 3.5), this number of repeats is likely to be a number of repeats which will result in ON expression of hsdM. That being said this suggestion is highly speculative and the observations from this current study can serve only as interesting preliminary data to be expanded on in future work as the region either side of the hsdM SSR tract contains a large number of indels, and as such the observations of repeat tract number cannot be guaranteed in every case.

One further important observation is the associations between clinical state of the host the hsdM ON/OFF state of the 121 strains with an identifiable hsdM gene, only 1/5 strains from a host with COPD was in the ON state. This though is only a small number of samples in comparison to the 83 COPD sputum supernatant samples analysed in this current work. Furthermore, for all of the OFF isolates from the GenBank database isolated from patients with COPD *in silico* alteration of their repeat tracts to the nearest ON state resulted in the ON state being noted with either 3 SSR repeats (3/4) or 6 SSR repeats (1/4); the sole ON isolate also had 3 repeats within hsdM. As such, and regardless of the above stated limitations, the association between the hsdM ON state and COPD may prove an interesting point for further study. In particular due to the strong association between *H. influenzae* and COPD across disease states and severities (Leanord and Williams, 2002; Patel et al., 2002; Zhang et al., 2010; Wang et al., 2016).

In terms of a putative association between COPD and bacteriophage it is probable that the environment within the respiratory tract of individuals with COPD may suit bacteriophage maintenance. The respiratory tract of individuals with COPD is characterised by the prevalence of high levels of mucus in the airways (Fahy and Dickey, 2010). Mucus has been shown to facilitate the retention of bacteriophage through the binding of mucin glycoproteins and Ig-like domains on the bacteriophage capsid (Barr et al., 2013). The presence of bacteriophage in mucus has been termed as a form of 'non-host-derived immunity', with the mucosal environment allowing for sub-diffusive movement of bacteriophage within the mucus that increases the number of bacteriophage/host interactions (Barr et al., 2013, 2015). Although no direct studies into bacteriophage prevalence in individuals with COPD has been conducted a similar highly mucous respiratory disease, cystic fibrosis, bacteriophage for H. influenzae have been shown to be prevalent in individuals with this condition through metagenomic analysis of the host viral communities (Willner et al., 2009). As such, the prevalence of *H. influenzae* across individuals, disease states, and severities of the highly mucous condition of COPD, combined with the potential presence of bacteriophage in the mucosal layer against this bacterium would thus mean that within the respiratory tract of individuals with COPD there may be a need for access to an active restriction modification system to defend against bacteriophage predation. Again however, this would require further investigation.

Analysis of the repeat tracts and putative expression of hsdM across sequenced isolates from the GenBank database showed that in the majority of these isolates hsdM genes were in the OFF phase. This is consistent with data from a 6-day human nasopharyngeal colonisation study by Poole et al. (2013) involving colonisation of human subjects with *H. influenzae* strain NTHi 2019Str^R1, with strain 2019 being one of the sequenced isolates analysed in this current study. In their study, Poole et al. (2013) started with an initial inoculum containing 7.4% of the population ON for hsdM, and observed no increase in the proportion of the *H. influenzae* population with hsdM in the ON state. Expression was actually noted to decrease slightly across the initial five days (2.6% day 3, 2.5% day 4, 1.6% day 5) before increasing slightly on the final day, but still below the initial inoculum (6.3% day 6). This may imply that expression of hsdM is only required in specific conditions and that otherwise expression comes with a fitness cost. This possibility was tested by investigating the metadata on each of the GenBank strains to determine potential clinical associations with hsdM expression.

Alterations to bacterial physiology can, as mentioned, alter the bacterium's ability to colonise their host in some manner. For pathogenic bacteria, alterations in bacterial physiology that boost colonisation would be noted in individuals due to alterations in the patients clinical condition. From the examination of the ON/OFF states of the 121 sequenced strains with an identifiable hsdM gene no overwhelming association between disease and isolates where hsdM was in the ON state could be observed.

The only conditions where ON outnumbered or equalled those isolates where hsdM was in the OFF state were those isolated from patients with bacteraemia and otitis media. However, these associations are tentative for two separate reasons. In the case of those strains isolated from patients with bacteraemia, the total number of strains from this condition is below that which would allow a link to be confidently suggested. While in the case of otitis media a previous study into the switching of phase variation states in a number of H. influenzae strains presented data that showed lack of phase ON hsdM in both samples obtained from the nasopharynx, and where the strain had migrated into the inner ear during otitis media (Fox et al., 2014). There is however a caveat to this latter result as the authors of the otitis media study based their assumptions of ON/OFF status on the repeat tract size following GeneScan analysis of PCR fragments from a reaction spanning from either sides of the traditionally 5'-GAGAC-3' repeat tract. This, as conceded in section 3.9 from the GeneScan analysis of PCR fragments encompassing the pentanucleotide tract of hsdM, may not be 100 % accurate due to the presence of indels in the region immediately flanking the SSR tract. Furthermore in the otitis media study ON/OFF state classifications were made based on the presence of repeat tracts of lengths being multiples of 3x 5'-GAGAC-3' repeats in length (i.e. 3 repeats, 6 repeats, 9 repeats etc), however in section 3.5 it was noted that strains with even 1x or 2x 5'-GAGAC-3' could be in the ON state, dependant on the presence of those indels mentioned above. Therefore although no confident correlation could be noted based solely on ON/OFF state for hsdM for otitis media, this is an avenue for further study to gain a clearer picture of whether a specific phase variable state for hsdM has any association with otitis media.

Any effect attributed to restriction modification on bacterial physiology through epigenetic regulation is dependent on the sequence targeted for methylation. As such the diversity of the specificity subunit of the HindI-like restriction modification system, hsdS, was analysed. The presence between the association of ON/OFF state and the disease state of the host from whom the strain was isolated from was also investigated. From this it was observed that for the 121 *H. influenzae* strains where a copy of hsdM was identified could be subdivided into 40 groups based on variants in the peptide sequence of their associated hsdS gene. This is a greater number of hsdS alleles than has been described in the *H. influenzae* Type III restriction modification system. Across all strains that have previously been analysed 21 alleles for modA, the gene which contains the specificity and MTase activity of the system, have been identified (Atack et al., 2015). The hsdS allele subdivisions were made based on absolute amino acid homology, and does not necessarily mean that these hsdS genes will target distinct nucleotide sequences as in some cases the distinguishing features are amino acid substitutions in the the central conserved region, previously noted as having no impact on Type I specificity in *S. typhimurium* (Cowan et al., 1989). Meanwhile others contain only small numbers of alterations in a single variable domain and as such it is unclear whether these small differences could be sufficient to cause the restriction modification system to target a new DNA sequence.

Attempts to determine any correlation between the hsdS alleles and disease state were unfruitful, largely due to the small numbers of strains containing for each of the hsdS alleles. As such no disease specific correlations between hsdS allele and disease state could be made. However, one trend was noted. It was observed that the hsdS32 allele that was present in a reasonable number of strains was constitutively in the ON state for hsdM. This allele was isolated from a number of conditions including meningitis, bactaeremia, and otitis media. Interestingly the hsdS32 allele contained two repeats and is unlikely to undergo phase variation, as found for short tetranucleotide tracts (De Bolle et al., 2000). Thus this allele may be fixed ON. It can be postulated that this allele has some role in regulation and that this phenotype may be of benefit to some strains of H. influenzae.

In summary, the observations of Zaleski et al. (2005) could not be repeated in this study. This is partly because *in silico* predictions indicate that a hsdM gene would not produce a full length peptide with the presence of four 5'-GAGAC-3' repeats, but also because phage resistance was never observed in the experiments with this phase variant. Therefore, the role of phase variation of hsdM in bacteriophage-host interactions remains to be elucidated. Additionally, although no confident correlations between disease state and hsdM state could be made base on the sequenced strains available in the GenBank database, a putative role for phase variation of hsdM in the colonisation of patients with COPD has been observed and may warrant further investigation. Lastly, through the comparison of hsdS alleles associated with the phase variable HindI-like hsdM gene it appears that one allele may be of future interest, with the hsdS32 allele found to be constitutively in the ON state across strains encompassing a number of MLST types. As such the relevance of this allele to the physiology of *H. influenzae* may be of key interest.

4. Analysis of the phase variable bacteriophage receptor of *Haemophilus influenzae* and its role in bacteriophage resistance

4.1 Abstract

As in the previous chapter the observations of Zaleski et al. (2005) for a bacteriophage resistance phenotype associated with the Type I restriction modification system of *H. influenzae* could not be replicated, their observations of a resistance phenotype associated with phase changes of the UDPgalactose-LOS-galactosyltransferases encoding *lic2A* gene was investigated. *lic2A* phase varies through expansion or contraction of a tetranucleotide (5'-CAAT-3') repeat tract within its ORF, with ON expression (22x 5'-CAAT-3') facilitating bacteriophage infection and the OFF state (21x 5'-CAAT-3') preventing infection. Although sequence analysis of our phase variants showed that they had repeat lengths of 22x for the *lic2A* ON Rd30S, and 26x for the *lic2A* OFF Rd30R a resistance phenotype was observed in line with the observations of Zaleski et al. (2005) following exposure to bacteriophage HP1c1. These observations of putative ON/OFF states and SSR lengths were further strengthened by sequence analysis of sequences available in genbank that showed in nearly all cases the reading frame associated with 22x repeats would encode a full length polypeptide, while those frames associated with the 21x repeat of Zaleski et al. (2005) and this study's 26x SSR tract would both produce truncated peptides.

Analysis of the ON/OFF state of *lic2A* in all the *H. influenzae* genomes available in genbank showed that *lic2A* is ON in the majority of *H. influenzae* isolates. Correlation of ON/OFF state to a disease phenotype showed no significant correlations to a particular disease state although this may be limited by the overall limited number of samples associated with a number of disease states. Investigation of SSR tract lengths in sputum supernatants from patients with COPD showed that 22x 5'-CAAT-3' were the most prominent tract length, potentially indicating that the ON state is preferred in colonisation of individuals with COPD.

4.2 Introduction

H. influenzae can generate a diverse assemblage of LOS structures, with a high degree of this diversity due to a large number of phase variable loci encoding LOS biosynthesis associated transferases (Power et al., 2009). In *H. influenzae* the LOS is composed of three main regions, lipid that anchors the LOS to the cell membrane; the inner core, which is a triheptosyl region and serves as the backbone for oligosaccharides extensions to expand; and the outer core, which is composed of these oligosaccharide extensions. In *H. influenzae* a number of phase variable genes encode for transferases involved in the biosynthesis of these extensions. In *H. influenzae* Rd these include *lic1A*, which is responsible for the addition of phosphorylcholine to the glucose extension from the first heptose on the triheptosyl backbone; and *lic2A* and *lgtC*, UDP-galactose-LOS-galactosyltransferases, which each together add a α -Gal-(1-4)- β -Gal to the third heptose of the backbone (Hood et al., 2001) (see figure 1.10 for the organisation of the *H. influenzae* Rd LOS).

In their work Zaleski et al. (2005), suggested that the galactose residue added by lic2A, forming a Galp-(1-4)- β -D-Glcp extension, may act as the receptor for bacteriophage HP1c1, with lic2A knock outs showing complete resistant to bacteriophage infection. Phase variation of surface exposed molecules is believed to have developed as a means of allowing the bacterial cell to utilise the function of the gene being encoded, whether that be cross membrane transport, adherence, motility, or immune evasion, in a temporal manner, likely due to the fluctuating presence/strength of an opposing selection pressure (Moxon et al., 2006). For lic2A its function is believed to be in assisting the bacterial cell in evading the immune system. For instance, Clark et al. (2013) observed that passage of a majority lic2A OFF *H. influenzae* Rd populations through human sera resulted in increased levels of survival when the population had switched to a lic2A ON phenotype.

Within *lic2A* phase variation occurs through SSM along a tetranucleotide repeat tract roughly midway through the genes ORF. This repeat tract consists of repetitive units of 5'-CAAT-3', and phase varies at a rate of ~ 1.88×10^{-4} for ON-OFF switching and ~ 0.95×10^{-4} for OFF-ON switching for repeat tracts of 22 and 21 repeats respectively (High et al., 1993, 1996; Dixon et al., 2007). That being said this rate of switching will depend on the number of repeats, with larger repeat tracts likely showing higher switching rates and lower rates when fewer repeats are present (De Bolle et al., 2000). Zaleski et al. (2005) showed that a mutant lacking a functional *lic2A* gene was resistant to bacteriophage HP1c1 infection. They observed that bacteriophage resistant colonies had an alteration in the number of repeats, switching from 22x 5'-CAAT-3' repeats with a bacteriophage sensitive phenotype to 21x 5'-CAAT-3' repeats with a bacteriophage resistant phenotype. The resistance phenotype conferred by the switch to 21x 5'-CAAT-3' provides absolute resistance to bacteriophage infection, allowing bacterial survival even in the presence of 10⁹ PFU ml⁻¹ bacteriophage. As attempts to replicate the observations of Zaleski et al. (2005) for the HindI restriction modification system of *H. influenzae* were unsuccessful similar analyses were conducted to validate these previous observations and expand the putative role of this phase variable gene in providing the population a means of survival against bacteriophage HP1c1 infection.

4.3 Presence of phase variable *lic2A* in *H. influenzae*

As with the earlier analysis of the 125 *H. influenzae* genome sequence for the hsdM gene of *H. influenzae*, the sequence of the Rd KW20 *lic2A* gene was utilised to screen for this gene in all the *H. influenzae* genomes available in GenBank to determine the prevalence of the *lic2A* gene across the sequenced isolates of *H. influenzae*. Of the 125 genomes of *H. influenzae* in the GenBank genome on 06-09-16 homologs of the *lic2A* gene of Rd KW20 were identified in 121 genomes (96.8 %) following bl2seq comparisons.

However, a number of these homologous sequences were located at the start or end of contigs and were excluded from the analysis as full ORFs were thus not available to confirm ON/OFF state. Furthermore, others such as 137_HINF, 167_HINF, and 841_HINF all show homology to the 3' end of the Rd KW20 *lic2A* but contain a distinct 5' region, no repeat sequence, and were not located between the *ksgA* and *apaH* genes which is the normal location for *lic2A* sequences. When the *ksgA* and *apaH* genes were identified in these three strains, *lic2A* was not present. In its place now was an ORF that shares \sim 72 % homology to a hypothetical gene of *Aggregatibacter actinomycetemcomitans*. As no *lic2A* could be identified these genes were excluded from further analysis.

Only 4 genomes lacked a sequence homologous to lic2A (156_HINF, 159_HINF, 781_HINF, and 839_HINF). These sequences contained the ksgA and apaH genes but these genes are not seperated by lic2A, instead being separated by only a ~10 bp stretch of DNA. Thus, these genomes were excluded from the analysis leaving 104 lic2A genes to be further analysed.

4.3.1 Distribution of phase variable repeats within *lic2A* gene of *H. influenzae*

Previous studies into the presence of phase variable repeat tracts within the *lic2A* gene of *H. influenzae* have been conducted (e.g. High et al. 1996 and Power et al. 2009). However, none of these studies have analysed > 30 strains of *H. influenzae*. This study was significantly larger due to the use of the sequence data from the 104 *lic2A* available within the GenBank database.

Within the 104 sequence it was noted that the *lic2A* gene ranged from between 822 bp to 1125 bp in length (strains HI1998 and 552_HINF respectively). If the repeat tracts are excluded from analysis



CAAACTCCATATTACATAATATAGAGGAATCTAGAA CAAACTCCATATTACATATATAGAGGAATCTAGAA	CAAACTCCATATTACATATATAGAGAAATCTAGAA 	CANANCTCCATATTACATATACAGGGAAATCTAGAA	CAAACTCCATATTACATAATATAGAGAAATCTAGAA	CAAACLICCAIAIIACAIAAIAIAGAGGAAICIAGAA CAAACTCCAIAIIACAIAAIAIAIAGAGGAAICIAGAA	CAAACTCCATATTACATAATATAGAGGAATCTAGAA	CAAACTCCATATTACATATATAGAGGAATCTAGAA	CAAACTCCATATTACATATATAGAGAAATCTAGAA	CAAACTCCATATTACATATATAGAGAAAATCTAGAA	CAAACTCCATATTACATATATAGAGAAATCTAGAA	CAAACTCCATATTACATATATAGAGAAATCTAGAA	CAAACTCTATATTACATATATAGAGAAATCTAGAA	CAAACTCCATATTACATATATAGAGAAATCTAGAA	CAAACTCCATATTACATATATAGAGGAATCTAGAA	AT CAAACTCCATATTACATAATATAGAGGAATCTAGAA	
					CAAT	CAATCAAT	CAATCAATCAAT	CAATCAATCAATCAAT	CAATCAATCAATCAATCAAT	CAATCAATCAATCAATCAATCAAT	CAATCAATCAATCAATCAATCAATCAATCAAT	CAATCAATCAATCAATCAATCAATCAATCAATCAAT	CAATCAATCAATCAATCAATCAATCAATCAATCAAT-	CARTCARTCARTCARTCARTCARTCARTCARTCARTCA	
		AATCAAT	ATCANTCANTCANT	AALCAALCAALCAALCAALCAALCAAL	CARTCAATCAATCAATCAATCAAT	CARTCAATCAATCAATCAATCAATCAAT	CANTCAATCAATCAATCAATCAATCAAT	CANTCAATCAATCAATCAATCAATCAAT	CANTCAATCAATCAATCAATCAATCAAT	CARTCAATCAATCAATCAATCAATCAAT	CARTCAATCAATCAATCAATCAATCAAT	CANTCAATCAATCAATCAATCAATCAAT	CANTCAATCAATCAATCAATCAATCAAT	CARTCARTCARTCARTCARTCART	
ACTTAT CAATCAATCAAT	ACTTAT CAATCAATCAAT	ACTTAT CAATCAATCAATCAATCAATC	ACTIAT CAATCAATCAATCAATCAATC	ACTIAL CARLEAST CARLEAST CAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTATCAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CAATCAATCAATCAATCAATC	ACTTAT CANTCANTCANT CANTON	
AAACTTTCATTTAGTTTCTTTAATGCGTAC AAACTTTCATTTAGTTTCTTTAATGCGTAC	NAACTITCATITAGTITCITTAATGCGTAC: DEECTTTCETTTEETTCETEC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAAACIIICAIIIAGIIICIIIAAIGCGIAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	AAACTTTCATTTAGTTTCTTTAATGCGTAC	
CGSHiCZ412602 3 CCUG 26214 A	CCUG 54503 2 HT1408	28423TDY5882018	615_HINF 2	28428TDY5882036	KR 494	F3031	7P49H1	H1359 A	2019	477	RMH193	28423TDY5882046	22.4-21	BIGRN20	



in the H. influenzae lic2A gene. Nucleotide sequences used for illustration are derived from sequence data of 18 H. influenzae strains, available on Genbank. The 5'-CAAT-3' SSRs are highlighted in yellow, with the reference strain Rd KW20 the last sequence shown. (b) Shows the distribution of Figure 4.1. Presence of 5'-CAAT-3' SSR units in the *lic2A* gene of 104 *H. influenzae* strains. (a) Example of 5'-CAAT-3' repeat tract length diversity the number of SSR repeat units within the ORF of all 104 H. influenzae lic2A sequences analysed in this study.

the lengths range from 806 bp to 821 bp (HI1998 being the shortest, while 91 of the 104 strains would have a *lic2A* gene of 821 bp should the repeat tract be deleted). Despite the disparity in length of *lic2A* between each of the strains the 5'-CAAT-3' repeat tract was found to begin 121 bp into the ORF in all cases. Examples of the 5'-CAAT-3' repeat tract within *lic2A* can be found in figure 4.1 (a). *lic2A* itself is well conserved across all the 104 *H. influenzae* strains, with nucleotide homology ranging from 71.3-100 % identity. The majority of the differences in nucleotide identity are found within the 5'-CAAT-3' repeat, exclusion of this SSR results in an increase in homology to between 95-100 %.

The number of 5'-CAAT-3' repeats was noted to range from between 1 to 77 iterations (Fig 4.1 (b)). The majority of the isolates containing ≤ 26 repeats (96.2 %), with a mean repeat tract length of 21.9. Low SSR unit numbers ($\leq 5x$ 5'-CAAT-3') were found in 12/104 strains (11.5 %).

As seen in figure 4.1 (b), the most common repeat number was 13x 5'-CAAT-3' (13/104, 12.5 %), followed by 22x 5'-CAAT-3' (12/104, 11.5 %), and 16x 5'-CAAT-3' (9/104, 8.7 %). A number of repeat lengths were found in only one isolate, namely 1x, 9x, 20x, 23x, 29x, 33x, 34x, and 77x 5'-CAAT-3' SSR units. The 77x 5'-CAAT-3' repeats of 552_HINF is of interest as this is the longest ever noted SSR tract within *H. influenzae*.

Zaleski et al. (2005) denoted the ON state of *lic2A* in Rd 30 (Rd 30S) to be with 22x 5'-CAAT-3', while the OFF state they identified was with 21x 5'-CAAT-3' repeats (thus also implying an OFF state with 23x 5'-CAAT-3'). Strains with the putative ON state of 22x 5'-CAAT-3' were the most prevalent with 12/104 isolates being identified with this SSR number.

4.4 Analysis of the putative coding states of the phase variable lic2A gene of H. influenzae

The next step was to analyse the putative ON/OFF states associated with each of these repeat numbers. In figure 4.3 (a) all ON/OFF associations are based on the production of an uninterrupted peptide sequence following analysis of the amino acid sequences produces following translation of the nucleotide sequences of each *lic2A* gene.

Across the sequences the ON phase predominated (Fig 4.3). Of the 104 isolates 65 were in the ON state (62.5 %), while 39 sequences were out of frame for lic2A (37.5 %). It was also noted that the ON/OFF states associated with each of the repeat tract was generally more predictable than that noted for hsdM in section 3.5. The ON/OFF states correlated well with the reading frames predicted by Zaleski et al. (2005) for the bacteriophage sensitive and bacteriophage resistant Rd 30 strains. The numbers of repeats associated with ON expression in this frame are, 1, 4, 7, 10, 13, 16, 19, 22, and 25 etc. This pattern gave the majority of the ON isolates from the 104 sequences tested, with 54 of the

65 ON isolates being associated with this frame (83.1 % of the *lic2A* ON strains). If each OFF repeat tract is adjusted to the closest repeat number that will produce an ON state it is observed that 93/104 (89.4 %) *lic2A* sequences were associated with this frame (figure 4.1 (b)).

Across the 26 SSR tract lengths, only four repeat tract lengths (13, 14, 19, and 25) showed examples of both ON and OFF expression. These SSR tract lengths with mixed ON/OFF states, and the examples of ON strains at repeat numbers outside the set associated with 22 repeats are the result of either a motif deleted in some strains at the 3' end of the *lic2A* SSR tract, or a mutation within the SSR sequence. For the the examples with a mutation at the 3' end of the SSR, a 5'-CAAA-3' motif has been deleted in a number of isolates, located immediately following the 5'-CAAT-3' repeat tract. *H. influenzae* HI1413 is considered in this analysis to contain 14x 5'-CAAT-3' repeats, however there are in fact 15x 5'-CAAT-3' units. The excluded 5'-CAAT-3' unit is separated from the rest of the SSR by a 5'-CAAG-3' motif. Similarly, strain HI1426 is donated here as containing an SSR tract 13x 5'-CAAT-3' units in length, however after the thirteenth 5'-CAAT-3' unit is a 5'-CA-3' motif followed by a further 6x CAAT, giving a region of 5'-(CAAT)_{x13}CA(CAAT)_{x6}-3'.

The imposition of lic2A being more readily in the ON state than the OFF state implies that there is selection for the expression of lic2A. This is likely a function of the association between the need for lic2A expression in immune avoidance, and as such the association between the phase state of lic2Aand the clinical state of the host was the next point of investigation.

4.5 Association between *lic2A* ON/OFF state and the clinical state of the host they were isolated from

The known involvement of lic2A in immune evasion (Clark et al., 2013), and a putative role in disease (Weiser and Pan, 1998), suggested that there could be an association between lic2A ON/OFF state and disease state. The range of clinical conditions associated with the hosts from whom the 104 strains used in the analysis of lic2A can be observed in figure 4.2.

As can be seen in figure 4.2 the *H. influenzae* strains with genome sequences in GenBank represent a wide range of clinical conditions similar to that for the investigation into hsdM (Section 3.9.1) despite encompassing a lower number of *H. influenzae* strains. As with the hsdM analysis the largest number of samples were isolated from patients with meningitis (47.12 %), with otitis media the next most prominent (19.23 %). Again a relatively high number of isolates do not have clinical meta-data associated with them (9.62 %).

To determine if there is any putative association between the ON/OFF state of lic2A and the clinical state of the host the phase of the lic2A gene in the 104 *H. influenzae* strains was plotted against the



Figure 4.2. Clinical states associated with the hosts from whom the 104 *H. influenzae* strains with sequences in the GenBank database where a *lic2A* gene was identifiable. Pie chart shows the frequency of each clinical state across all 104 strains from GenBank that had detectable *lic2A* gene. Colours represent the different clinical states, with size of section proportional to the number of strains that were associated with that clinical state. Key shows colour to state association as well as the percentage of the total that strain sequences were available for.



Figure 4.3. Association between each SSR tract length and the ON/OFF state of *lic2A* in 104 *H. influenzae* genomes. (a) shows the ON/OFF state associated with each SSR tract length and the the overall proportion of ON Vs. OFF strains. (b) shows the result of in silico alteration of each of the OFF SSR tracts to a number of 5'-CAAT-3' repeats that will give a full peptide product. Note that in (b) the repeat tract number data for 103 strains is present as one strain cannot produce a complete reading frame regardless of SSR number.

clinical states (fig. 4.4). Four clinical states, where observed to have a greater number of examples with *lic2A* in the ON phase than the OFF phase. Meningitis exhibited a ratio of 1.9:1, ON:OFF. Similarly the ratio for otitis media was 1.5:1 (12 ON Vs. 8 OFF). The largest difference was in bacteraemia, where a ratio of 6:1 (ON:OFF) was observed. Lastly, in *H. influenzae* strains isolated from individuals with COPD ON isolates outnumbered OFF isolates 2:1 (4 examples in the ON state compared to 2 strains with *lic2A* in the OFF state). However, that being said none of the host disease states had a direct correlation to ON/OFF state (i.e. p > 0.05, χ^2 test, for association between ON/OFF state of each individual condition vs. all others).

Across the collection of clinical conditions, no clinical condition had a greater number of OFF isolates than ON. In fact, within figure 4.4 this only occurred for the samples where no meta data was available. The only other case where more than one example was available was for strains that were carriage/commensal isolates. Here there was an equal number of ON/OFF examples (2 examples in each case).

The clinical conditions of bacteraemia, miningitis, and otitis media are caused by the invasion of bacteria to within the otherwise sterile blood, CSF, and middle ear fluid. Meanwhile conditions such as bronchitis, COPD, and cystic fibrosis are all distinct clinical conditions where bacterial colonisation can occur due to the disease state rather than the cause of these clinical condition. However, what they do share are sites of colonisation (in the case of the previous examples the respiratory tract) as such the site of isolation of the bacteria in relation to their ON/OFF phase for *lic2A* was next investigated.

4.6 Association between *lic2A* ON/OFF state and the site of isolation from the host

Further to examining any putative association between lic2A state and disease, the localisation of *H. influenzae* within the body can also likely cause changes in lic2A state. For instance, due to the role of lic2A in evading the host immune response (Clark et al., 2013) it would be beneficial to have a functional lic2A to express the Galp-(1-4)- β -D-Glcp moiety in the more invasive sites where an immune response would be more readily active, such as blood. This was examined by looking for associations between the locale where each of the 104 *H. influenzae* strains were isolated and the putative expression state of their lic2A gene, the results of which can be observed in figure 4.5.

The site of isolation was able to be identified in the majority of samples (99 samples, 95.2 % of the total number of samples). For the majority of the sites of isolation there was found to generally equal numbers of isolates with *lic2A* in the ON and OFF state. For example, for samples isolated from the nasopharynx and from sputum samples exactly 50 % were in the ON or OFF phase (ON:OFF =









2:2 and 5:5 isolates respectively). While for samples from bronchoalveolar lavage OFF only slightly outnumbered ON (2:3, ON:OFF isolates).

However, for bacterial isolates from the CSF and bloodstream there was a larger number of *lic2A* ON strains compared to their OFF counterparts. CSF showed the most substantial difference, with 20 strains isolated from CSF being ON for *lic2A*, while 8 were in the OFF phase (2.5:1, ON:OFF). For blood isolates, the ratio was 1.73:1 ON:OFF, with 19 *lic2A* ON isolates, and 11 *lic2A* OFF isolates identified in GenBank. From the above it appears that the *lic2A* ON phase is important in these two otherwise sterile invasive locations. Again however, no statistically significant link between location of isolation and ON/OFF state was observed (p > 0.05, χ^2 test, for association between ON/OFF state of each isolation site vs. all others)

4.7 Analysis of SSR tract lengths within the *lic2A* gene of 65 COPD samples

With the ON phase of the lic2A gene noted in a previous study to be putatively associated with the respiratory condition of pneumonia (Weiser and Pan, 1998), it was of interest to analyse the phase state of lic2A in further respiratory conditions. As mentioned in section 3.9, sputum supernatants from individuals with COPD became available for study during the course of this work. These samples provided an opportunity to investigate the distribution of SSR tract lengths within *H. influenzae* populations colonising individuals with an important respiratory condition.

A set of 174 sputum supernatant samples were screened with primers targeting the conserved regions that flank the *H. influenzae lic2A* repeat tract. Of these, 65 samples produced a PCR fragment that was successfully sized following capillary electrophoresis. The PCR fragments were compared to a *lic2A* fragment amplified from *H. influenzae* Rd 30S, which contains 22x 5'-CAAT-3' repeats and produces a product of 226 bp (Fig. 4.6).

The most common fragment size correlated to an SSR tract 22x tetranucleotide repeats in length (Fig. 4.7, 16/65 samples, 24.6% of total samples). This is the SSR number found in Rd 30S, and thus would putatively encode for a functional *lic2A* product. In contrast, only 36.9 % of the COPD samples produced a fragment correlating to the size of an ON number of repeats in Rd 30S. This is interesting as it seems unusual for the majority of samples to be in the OFF state, while the greatest number of samples contain a repeat number that would be in the ON phase. There are potentially three explanations for this. Firstly, as seen in figure 4.7, some samples contain X.25 or X.75 repeats, this would mean that these samples contain X repeats +1 bp, or -1 bp respectively. A number of these examples surround *lic2A* tract lengths associated with a *lic2A* ON state. This may mean that



Figure 4.6. Peak Scanner output of the 226 bp fragment of *lic2A* amplified from Rd 30S (blue peak) for screening of COPD samples. Y-axis indicates the strength of the fluorescent signal from the 6-FAM labelled fragment, while X-axis indicates the size of the fragment amplified from the Rd 30S *lic2A*.



length within the lic2A gene of the 65 COPD samples to clinically relevant characteristics. (a) shows he putative repeat tract length within lic2A in relation to the severity of the individuals disease based on clinical diagnosis using the GOLD scale for COPD (the greater the number the greater the Figure 4.7. Putative lic2A SSR tract lengths within sputum supernatant samples from COPD patients. Figure shows the correlations of SSR tract severity). (b) shows the distribution of SSR unit number in relation to the clinical state of the host at time of sampling, such as during stable periods, or during exacerbations etc.

the *lic2A* amplicon contains an indel so that the gene is no longer in the ON state. Alternatively the fragments sizing is erroneous and these tracts should actually be the closest whole number SSR length. In the 104 sequences no additional or missing nucleotides were found in the region amplified by the SSR screening PCR. All differences in bp length (excluding SSR expansions) were either the presence or absence of 4 bp at the 3' end of the SSR tract, or the presence of an extra 2 bp within the SSR tract noted in one strain (which would produce an SSR number of X.5 in figure 4.7). If the X.25 and X.75 sizes are taken as erroneous and are actually reflective of the closest whole SSR tract length then this would mean that 44/65 (67.7 %) of COPD samples are potentially in an ON frame.

Secondly, as mentioned above and in section 4.3.1, there is a tetranucleotide (5'-CAAA-3') sequence present at the end of the SSR tract within a number of the *H. influenzae* strains analysed (including Rd 30S). As such some of the samples noted here to contain 22 repeats (and the other repeat tract sizes) may not have this 5'-CAAA-3' sequence. This will mean for instance that some of the samples listed here as 22x 5'-CAAT-3' repeats could in fact contain 23x 5'-CAAT-3' if they are missing the 5'-CAAA-3' motif at the terminal end of the SSR tract. However, this was the case for only 8.7 % (9/104) of the 104 GenBank sequences. If this observation applies to the COPD samples this would imply that miss classification of SSR tract only accounts for a small number of samples, and thus the overwhelming predominance of 22x CAAT remains a credible observation.

The last possibility for the majority of the samples being in the OFF state while one particular SSR length correlates to ON expression is a role of this ON state in a particular form of the disease. Association between repeat tract length and the severity of the individuals COPD and state at the time of sampling can be seen in figure 4.7 (a) and (b) respectively. In terms of the individuals disease severity, as measured according to GOLD scale (where number increases with severity and a rating of 5 is the most severe form of COPD), samples were from individuals with GOLD stages 1-4. Samples from individuals with GOLD stage 2 or GOLD stage 3 COPD made up the majority of the available samples, 35.8 % and 41.8 % respectively. Only a few samples from patients with GOLD stage 1 and 4 COPD were available for study, 2.99 % and 10.5 %, while a few samples did not have data available on their hosts GOLD stage (9 %).

No Gold stage 1 or 4 samples were associated with an SSR tract length of 22, likely due to the small sample numbers for these severities. Samples from individuals with Gold stage 2 COPD did not have 22x 5'-CAAT-3' as the most abundant SSR tract length. Only 3 of the 24 GOLD stage 2 samples contained 22 repeats within their SSR (12.5 %). Here, the most common SSR tracts were 15.75 and 20.75 SSR units, with these identified in only 5 (20.8 %) and 4 (16.7 %) samples respectively. The closest whole SSR unit number, 16 and 21, would encode for a full *lic2A* peptide and a non-functional *lic2A* gene in Rd 30S respectively. Thus, if these samples do contain a *lic2A* gene that has a 1 bp deletion then an SSR length of e.g. 16x 5'-CAAT-3' will no longer encode a functional product. As

such, with a lack of large discrepancies in SSR tract length frequencies and mixed putative ON/OFF states there is no obvious correlation between SSR length and this COPD severity.

For Gold stage 3 tract length measurements correlating to 22x 5'CAAT-3' were found in 12 of the 28 samples at this severity (42.9 %). This tract length was the most prominent in GOLD stage 3 samples, with 25 repeats and 12.75 repeats the next most prominent, each found in 10.7 % of samples (3/28 samples). An SSR length of 25 repeats would produce an ON state for *lic2A* in Rd 30S, while 12.75, which would correlate to an SSR length of 13 with a 1 bp deletion (or 12 with a 3 bp insertion), would represent an OFF state as the ON state in Rd 30S found with 13x 5'-CAAT-3' would be negated with such alterations. In this more severe state of COPD it appears that the ON state of *lic2A* may be favoured. It would be interesting to determine if there was any further correlation between *lic2A* SSR tract and the more severe GOLD stages 4 and 5. However, as only a limited numbers from the original 174 COPD samples available for study were from GOLD stage 4, and no GOLD stage 5 samples were available in this set for screening to further analyse this. Likewise, as only a small number of samples from the less severe GOLD stage 1 COPD samples was available it was not possible to see if further samples from here would emphasise a lack of 22x 5'-CAAT-3' *lic2A H. influenzae* samples to add to the notion of a positive correlation between disease severity and prevalence of potential *lic2A* ON *H. influenzae* strains.

Another association investigated was the association between lic2A SSR tract length and the disease state of the patient the sample originated from (i.e. stable COPD, during an exacerbation, or a point in time following an exacerbation). Only a small number of samples had known host patient states (stable = 17/65, 26.1 %; exacerbation = 13/65, 20.0 %; Post exacerbation = 9/65, 13.8 %). No real correlation could be observed, as there was an even distribution of SSR tracts across all of these states (Fig. 4.7 (b)). Samples with 22x 5'-CAAT-3' constituted 31.25 % of the total number of samples from patients undergoing exacerbations. Overall however, due to the low total number of samples where meta data on patient symptomatic state was available, and the wide range of SSRs noted in figure 4.7 there does not appear to be an association between SSR length and disease state.

4.8 Sequence analysis of the *lic2A* gene of *H. influenzae* Rd 30

Based on the observations for the association between repeat number and lic2A ON/OFF state for the 104 sequenced *H. influenzae* strains from the GenBank database, it appears that with an SSR tract length of 22x 5'-CAAT-3' within the *lic2A* ORF *lic2A* will be in the ON state. This is consistent with the observations of Zaleski et al. (2005) for the bacteriophage sensitive phenotype of *H. influenzae* Rd 30. To confirm this, the SSR number within the Rd 30 phase variants were then examined to determine if these were of a similar SSR length to that of Zaleski et al. (2005) and confirm the putative
ATGAGTGCTATTGAAAAATATTGTCATTAGTATGGAAAAATGCAACTGAACGTCGCAAACATATTACT (A) М S A I E N I V I S M E N A T E R R K H I Т AAGCAATTTGAATCAAAAAAACTTTCATTTAGTTTCTTTAATGCGTACACTTAT**CAATCAATCAAT** K Q F E S K K L S F S F F N A Y T Y Q S T N S NQSIN Q S I N Q S I N S Т Q Ι Ν Q S 0 S N S I L H N I E E S R I L T K G E K Ν 0 Ι G C L I S H F L L W N K C V N E N F E Y L K Ε D D V I L G E N A E V F L N Q N Т F E W T. AAAACACGTTTTGATTTTAATGATATTTTTTTTTTTTTCGTTTAGAAACTTTTTTACAGCCAGTTAAA T R F D F N D I F I I R L E T F L Q Κ P V K CTTGAGAAACAAACTAAAATTCCACCTTTTAATTCTAGGAACTTTGATATTTTAAAATCAACTCAC K Q T K I P P F N S R N F D I L Κ S T. E Т Н TGGGGGACGGCAGGTTATATTATTTCTCAAGGTGCGGCTAAATATGTAATTGAATATTTAAAGAAT GTAGYI I SQGAAKYVIE Y L K N ATTCCTTCTGATGAAATTGTTGCAGTTGATGAACTTATTTTTAATAAATTAGTTGATGTTGATAAT I P S D E I V A V D E L I F N K L V D V D N TATATTGTCTATCAACTGAATCCAGCAATTTGTATTCAAGAACTCCAAGCGAATCAAAGTAAAAGT I V Y Q L N P A I C I Q E L Q A N Q S K S GTATTAACAAGCGGATTGGAAAAAGAACGACAAAAACGTTCAAAAATCCGTAAGAAAAAGACATTG T S G L E K E R Q K R S K I R K K K V L т T. AAACAGCGTTTAACTCGGATAAAAGAAAATATTATACGAGCTTTAAATAGAAAAAAATGGAAAGAA K Q R L T R I K E N I I R A L N R K K W K E CAACAACGTATTAAAGAAATGCAAGGTAAAGAAATTGTTCGTTTTATGTAG Q R I K E M Q G K E I V R F M * 0

Figure 4.8. Nucleotide and amino acid sequences of the *lic2A* gene of the *H. influenzae* Rd 30 variants used in this study. (a) shows the sequence of the bacteriophage sensitive isolate Rd 30S, while (b) shows the sequence of the HP1c1 resistant variant Rd 30R. The 5'-CAAT-3' SSR tract is highlighted in bold and underlined. Asterisk (*) indicates stop codons within the amino acid sequences listed below the nucleotide sequences.

ATGAGTGCTATTGAAAATATTGTCATTAGTATGGAAAATGCAACTGAACGTCGCAAACATATTACT М S A IENIVISMENATERRKHIT AAGCAATTTGAATCAAAAAAACTTTCATTTAGTTTCTTTAATGCGTACACTTAT**CAATCAAT** K Q F E S K K L S F S F F N A Y T Y Q S I N S I N Q S I N Q S I N Q S I N Q S I N Q S 0 **ATCAATCAATCAATCAATCAAT**CAAACTCCATATTACATAATATAGAGGAATCTAGAATACT ІКЬНІТ* Ν Q S Y R G Ι Ν Т Y * R R K R L F N * P F L I M E * V C E * K TTTTGAATATCTCAAAATTTTTGAAGATGATGTAATTCTTGGCGAGAATGCGGAAGTATTTTTGAA F * I S Q N F * R * C N S W R E C G S I F Е CCAAAATGAATGGTTAAAAAACACGTTTTGATTTTAATGATATTTTTATTATTCGTTTAGAAACTTT K * M V K N T F * F * * Ρ Y F Y Υ S F R Ν F TTTACAGCCAGTTAAACTTGAGAAACAAACTAAAATTCCACCTTTTAATTCTAGGAACTTTGATAT F T A S * T * E T N * N S T F * F * E L * Y TTTAAAATCAACTCACTGGGGGACGGCAGGTTATATTATTATTTCTCAAGGTGCGGCTAAATATGTAAT F K I N S L G D G R L Y Y F S R C G * I C N TGAATATTTAAAGAATATTCCTTCTGATGAAATTGTTGCAGTTGATGAACTTATTTTTAATAAATT IFKEYSF**NCCS** т ү F Т AGTTGATGATGATAATTATATTGTCTATCAACTGAATCCAGCAATTTGTATTCAAGAACTCCAAGC S * C * * L Y C L S T E S S N L Y S R T P S GAATCAAAGTAAAAGTGTATTAACAAGCGGATTGGAAAAAGAACGACAAAAACGTTCAAAAATCCG E S K * K C I N K R I G K R T T K T F K N Ρ TAAGAAAAAGACATTGAAACAGCGTTTAACTCGGATAAAAGAAAATATTATACGAGCTTTAAATAG E K D I E T A F N S D K R K Y Y T S F Κ AAAAAATGGAAAGAACAACGTATTAAAGAAATGCAAGGTAAAGAAATTGTTCGTTTTATGTA Κ K M E R T T T Y * R N A R * R N C S F Y V G ?

(B)

Figure 4.8 (cont.). Nucleotide and amino acid sequences of the *lic2A* gene of the *H. influenzae* Rd 30 variants used in this study. (a) shows the sequence of the bacteriophage sensitive isolate Rd 30S, while (b) shows the sequence of the HP1c1 resistant variant Rd 30R. The 5'-CAAT-3' SSR tract is highlighted in bold and underlined. Asterisk (*) indicates stop codons within the amino acid sequences listed below the nucleotide sequences.

lic2A state that their SSR numbers encode.

Upon examining the putative ON/OFF states within the *H. influenzae* Rd30 derivatives it was observed that Rd30S was in fact phase ON, while Rd30R was in the OFF phase for *lic2A*. It can be seen from figure 4.8 (a) that the Rd 30S strain contained 22x 5'-CAAT-3' repeats, thus identical to the sensitive Rd 30 strain of Zaleski et al. (2005). This 22x 5'-CAAT-3' *lic2A* thus encodes a protein \sim 300 amino acids in length.

Figure 4.8 (b) shows the nucleotide sequence of the Rd 30R phase variant and its associated amino acid sequence. From the nucleotide sequence it was noted that the 5'-CAAT-3' repeat tract of Rd 30R consists of 26 iterations of the repeat sequence, not 21 repeats as used by Zaleski et al. (2005). This number of repeats will still encode a truncated peptide sequence of \sim 80 amino acids in length. However, although 26x 5'-CAAT-3' would code for a truncated amino acid sequence from the start codon displayed in figure 4.8 (b), it has been observed by Dixon et al. (2007) that *lic2A* activity can be detected (using a LacZ reporter) within this reading frame due to a start codon present 37 bp upstream of the displayed start site. However, when monoclonal antibodies targeting structures that require the Lic2A enzyme for synthesis were tested on phase variants within this frame, no reactivity was observed. Dixon et al. (2007) concluded that expression of *lic2A* from this frame was too weak for surface epitope production could be observed.

Thus from the sequence analysis of these Rd 30 lic2A phase variants it was concluded that the Rd 30S strain was in frame for lic2A, and will thus be sensitive to bacteriophage challenge, while the Rd 30R variant is out of frame for lic2A and as such should be completely resistant to bacteriophage infection. However, Rd 30R contains an SSR number that may result in weak lic2A expression, and although Dixon et al. (2007) were unable to observe a lic2A dependent phenotype using antibodies, bacteriophage infectivity has not previously been tested on strains with SSR numbers associated with this frame.

4.9 Determination of bacteriophage resistance phenotype of Rd 30 following bacteriophage HP1c1 challenge

As *lic2A* was in frame for Rd 30S and out of frame for Rd 30R it was likely that the HP1c1 sensitivity/resistance should also be observed. Therefore an infectivity assay was conducted to determine if the lack of an in frame *lic2A* in Rd 30R does confer the HP1c1 resistance phenotype observed by Zaleski et al. (2005).

The results of the testing of the ability of $\sim 1 \ge 10^9$ PFU ml⁻¹ of bacteriophage HP1c1 to form plaques on a lawn of *H. influenzae* Rd 30S and Rd 30R can be seen in figure 4.9. Here it was observed that Rd 30S



Figure 4.9. Infectivity assay of bacteriophage HP1c1 against the *lic2A* phase variant strains *H. in-fluenzae* Rd30S and Rd30R. Bar chart represents the estimation of number of infective particles of bacteriophage HP1c1 as observed through plaque formation. Rd30S (black bar) contains twenty-two 5'-CAAT-3' repeats resulting in a functional *lic2A* gene, while Rd30R (no bar) contains twenty-six 5'-CAAT-3' repeats within its SSR and does not produce a functional *lic2A* gene. N = 3, Error bars = \pm SEM.

was highly susceptible to bacteriophage HP1c1 challenge, meanwhile Rd 30R was completely resistant to the bacteriophage. This showed that Rd 30S was significantly more susceptible to bacteriophage HP1c1 infection compared to Rd 30R (two sample *t*-test: t = 4.142, P < 0.05; mean \pm SEM PFU ml⁻¹ values for bacteriophage HP1c1 infectivity on Rd 30S = $8.88 \pm 2.14 \times 10^8$; Rd 30R = 0; Figure 4.9), as spotting HP1c1 on a lawn of Rd 30R did not produced any plaques at all. This would mean that this *H. influenzae* isolate was completely resistant to bacteriophage HP1c1 challenge.

These results echo the observations of both Zaleski et al. (2005) and the bioinformatic analysis of these isolates in section 4.4. As such further work can be carried out on these *lic2A* phase variant isolates to determine the role that phase variation of these genes may have in the survival of *H. influenzae* to bacteriophage HP1c1 challenge.

4.10 Discussion

The first point of interaction between bacterial cells and their environmental is the external cell surface. These interactions have led to evolution in bacteria of mechanisms for generating adaptations of cell surface molecules to produce the necessary contingent responses enabling cells to survive and flourish in volatile environments. In *H. influenzae*, such adaptations have been hypothesised to have adaptive roles. For example, the altered expression of iron uptake channels may be beneficial in environments of differing haem availability (Moxon et al., 2006); adhesins benefit invasion of a new niche but will become targets for the host immune systems with prolonged expression (Davis et al., 2014); and altered decoration of LOS may maintain a degree of immune evasion (Clark et al., 2013). This latter point is pertinent as expression of one such evasion loci, *lic2A*, while facilitating resistance to host immune challenge causes the bacterial cell to now become highly susceptible to viral predators (Zaleski et al., 2005; Dixon et al., 2007). These pressures create a need for the bacterial population to be in a state of preparation for the arrival of either of these conflicting selection pressures; the 'preparation' in this case achieved through phase variation.

Phase variation of lic2A is caused by a tetranucleotide repeat within the ORF of lic2A that, through expansion or contraction of the SSR, can cause the production of a truncated polypeptide sequence or reinstate a disrupted sequence to produce a full polypeptide. Zaleski et al. (2005) noted that with 22x 5'-CAAT-3' repeats *H. influenzae* Rd 30S was susceptible to infection by bacteriophage HP1c1, while resistant cells were found to contain 21x 5'-CAAT-3' repeats. Thus these isolates would serve as an ideal system to study the dynamics of the effects of phase variable resistance to bacteriophage on bacteriophage-host dynamics. However, following difficulties reproducing the bacteriophage resistance observations of Zaleski et al. (2005) for the phase variable restriction modification system (discussed in section 3), the first aim was to confirm the association between lic2A SSR numbers and the putative expression of lic2A through a survey of the 125 *H. influenzae* genome sequences available in GenBank. The analysis of 125 genome sequences of *H. influenzae* in the GenBank database makes this current study the largest analysis of the distribution of SSRs within the lic2A gene to date.

The *lic2A* gene is well conserved across *H. influenzae* strains. For instance Garmendia et al. (2014) screened 18 *H. influenzae* isolates from sputum of patients with respiratory diseases and found that *lic2A* was present in all isolates tested. McCrea et al. (2008) identified a *lic2A* gene in 80 of 88 *H. influenzae* strains screened, and Martí-Lliteras et al. (2011) further identified *lic2A* in all 111 strains of the *H. influenzae* strains they tested. The present observations were consistent with the conserved nature of *lic2A*, as here a *lic2A* homologue was identified in 121 of the 125 *H. influenzae* strains in GenBank. Only 4 *H. influenzae* strains did not have an identifiable *lic2A* gene, thus 96.8 % of *H. influenzae* genomes from the GenBank database contained a *lic2A* homologue. However, in 17 genomes of the 121 where *lic2A* was found only in part within one contig, or distributed between multiple contigs, the 17 sequences were excluded from analysis leaving 104 *lic2A* sequences for the analysis of the presence and distribution of the 5'-CAAT-3' SSR sequence.

Previous surveys into the distributions of lic2A SSR tract lengths in *H. influenzae* have shown the large diversity in SSR tract lengths within the lic2A gene. High et al. (1996) investigated the distribution

of *lic2A* SSR units in a collection of 26 *H. influenzae* strains. They noted that the SSR tract lengths varied between the strains analysed, from 7-22x 5'-CAAT-3', with 16 iterations the most prominent. Power et al. (2009) conducted analysis of SSR tract lengths within 16 *H. influenzae* genomes that are available on GenBank (with these sequences also used in the present analysis). They observed that across these 16 genomes the *lic2A* SSR ranged from 5-25x 5'-CAAT-3' in length, with 14 and 25 5'-CAAT-3' units the most regularly observed in the sequences studied. Martí-Lliteras et al. (2011) noted a wide range of *lic2A* product sizes following PCR of the gene, attributed to the 5'-CAAT-3' repeat sequences with further analysis of 19 of their isolates showing the SSR length varied from 7-33x 5'-CAAT-3'. Fox et al. (2014) noted a similar distribution in *lic2A* SSR number in otitis media samples, with SSR numbers ranging from 7-35x 5'-CAAT-3'.

Similar to the above mentioned studies, a 5'-CAAT-3' motif was present at the SSR region in every strain studied, although with a more expansive range of tract lengths. It was observed that the number of 5'-CAAT-3' units within the SSR region of the *H. influenzae lic2A* gene can vary from 1-77 iterations in length. The isolate containing 77x 5'-CAAT-3' repeats, 552_HINF, may well contain the longest SSR tract to date should this not be a sequencing artifact, with the 57x 5'-CAAT-3' repeats found within the *H. influenzae lic1A* gene being the previous largest (High et al., 1996; Moxon et al., 2006). For the 57 repeat containing *lic1A* gene the SSR tract composes ~20 % of the genes total length, while for 552_HINF its *lic2A* SSR well encompass ~27.5% of the ORF. Despite these outliers, the majority of observed SSR tract lengths fall within the ranges identified in the aforementioned studies, with the most commonly identified SSR tract containing 13 repeats. Furthermore, 22 repeats of 5'-CAAT-3' was the second most prevalent SSR tract length, this being the repeat number observed by Zaleski et al. (2005) to produce a bacteriophage sensitive phenotype.

Zaleski et al. (2005) noted in their study that the *lic2A* ON state (bacteriophage sensitive) was associated with 22x 5'-CAAT-3' repeats, while the OFF state (bacteriophage resistant) was with 21x 5'-CAAT-3' repeats. Similar observations were made in this study, with all the *H. influenzae* genomes containing a *lic2A* with 22x 5'-CAAT-3' in the ON phase and those with 21x 5'-CAAT-3' repeats were in the OFF phase. Furthermore, the only isolate with 23x 5'-CAAT-3' (the +1 reading frame) was in the OFF phase. This correlates then with the observations on *lic2A* functionality by Zaleski et al. (2005). In fact the high level of homology found within the *H. influenzae lic2A* gene extends this association so that tracts of 1, 4, 7, 10, 13, 16, 19, 22, 25... etc, are also in the ON state a trend that can be readily observed from the GenBank sequences. There is however a small number (8/104 [7.7 %]) of cases that do not conform to this rule and this was due to the deletion of a 5'-CAAA-3' element located at the terminal end of the 5'-CAAT-3' SSR tract. Thus the majority of *H. influenzae* strains appear to adhere to the 22x 5'-CAAT-3' to ON pattern of SSR numbers noted by Zaleski et al. (2005). A key finding made with the 104 genome sequences from the H. influenzae was that lic2A was in the ON state in the majority of strains. A high level of ON expression implies that lic2A serves an important function during host colonisation by *H. influenzae*. A caveat is that this result may be due to the inherent sampling bias of sequences in the collective genome databases, as these isolates tend to be collected from individuals suffering from disease. For example, H. influenzae is usually found within the human host as a commensal organism and yet only 4 genomes were from isolates known to be isolated from healthy individuals. Thus it may be that due to the GenBank sampling collection containing a larger number of disease associated isolates than commensals the ON state prevails across the GenBank genomes, and should more commensals be analysed the prevalence of genomes containing *lic2A* in the ON state may reduce, as its expression has been shown to be less crucial at least during asymptomatic colonisation. In a human colonisation study it has been observed that in healthy individuals *lic2A* appears not to be necessary during initial colonisation of the host (Poole et al., 2013). In this study an average of 1.3 % of the initial inoculum were in the phase ON state, with the colonising populations remaining around this phase until reaching its highest occurrence at the end of day 6 where only 3 % of the colonising *H. influenzae* population where phase ON. Similarly, Weiser and Pan (1998) noted that in 5 isolates from asymptomatic colonisation the lic2A ON state was observed in only 1 isolate, as opposed to 5 isolates from individuals with *H. influenzae* associated pneumonia where 5/5 isolates had *lic2A* in the ON state. Thus it may be the case that prevalence of disease associated/lack of commensal genomes skew the *lic2A* state to the ON state in the current analysis.

With infection by *H. influenzae* associated with a wide range of clinical conditions, such as otitis media, conjunctivitis, and sinusitis; to more severe conditions such as bacteremia and meningitis (Ladhani et al., 2010; Laupland et al., 2011; Van Eldere et al., 2014); and its presence in certain individuals occurring as a complication of other clinical conditions, such as pneumonia, cystic fibrosis, and COPD (Burns et al., 1998; Cordero et al., 2000; Barker et al., 2015), here the putative ON/OFF state of *lic2A* was correlated to the disease state of the host from whom the bacteria was isolated, to determine if there was any association between the disease state of the host and the phase of *lic2A*. This analysis indicated that meningitis and bacteraemia were associated with ON expression of *lic2A*.

Meningitis and bacteraemia are often associated diseases, with meningitis occurring as a bi-product of the migration of an initial case of bacteraemia across the blood-brain barrier or the blood-cerebrospinal barrier (Nassif et al., 2002; Kim, 2003). For studies into *H. influenzae* associated bacteraemia and meningitis Hib strains are the most commonly utilised, although since the widespread utilisation of the Hib vaccine NTHi are responsible for the greatest incidence in human populations (Ladhani et al., 2010; Laupland et al., 2011). Hib meningitis is in the majority of cases the result of an initial bacteraemia, wherein the bacterial population undergoes a period of rapid expansion in bacterial numbers in the blood stream followed by migration into the CSF across the blood-brain barrier via the chorioid plexuses (Smith, 1987; Nassif et al., 2002). Between this time however, the bacterial cells have undergone a high degree of immune challenge within the blood stream (Smith, 1987). With lic2A expression contributing to the bacterial populations immune evasion it could be that in both bacteraemia and meningitis there is a predominance of the lic2A ON state due to the initial selection for the lic2A ON state for *H. influenzae* isolates in the blood. Selection for ON cells in the blood that will then migrate across the blood-brain barrier would result in these lic2A ON cells then being found in the CSF. This conclusion was supported by the analysis of the sampling location from which the GenBank strains were isolated, which showed that there was a high level of lic2A ON cells when this organism was isolated from blood and CSF.

It remains unclear why such invasive conditions occur. H. influenzae is often found as a commensal organism within the upper respiratory tract of the host, yet can cause conditions such as bacteraemia and meningitis following migration from the upper respiratory tract to the more invasive sites within the human host. The reason behind the migrations of the bacterium from the upper respiratory tract to these more invasive sites is unknown. However, possible reasons for the development of more invasive disease include susceptibility to disease, e.g. immune competency etc.; random chance; and/or the development of a phenotype within the host by the bacterium within the human host that allows the cell to go to more invasive locations (Margolis and Levin, 2007). The development of bacteraemia itself appears to be the product of a bottleneck event as the resulting bacterial population within the blood are believed to be the product of the progeny from one bacterial cell type (Moxon and Murphy, 1978; Margolis and Levin, 2007). That is to say that although in the upper respiratory tract microbial populations are highly diverse in population composition, cells isolated from blood are often noted to be monoclonal by comparison (Sandgren et al., 2004). One means that has been suggested to facilitate the adaptation that allows the development of these invasive clonal population is the selection of a putative transient phase variant (Margolis and Levin, 2007). *lic2A* itself has been signified as allowing the development of a more invasive phenotype (Griffin et al., 2005). Thus a monocolonal populations which express *lic2A* on their LOS may be better suited for survival under these conditions. This is not a definitive rule however, as there is OFF variants noted in each site also. These OFF variants may be expressing other LOS modifications/arrangements of LOS modifications that provide the required level of immunity for bacterial survival within these niches.

As discussed in section 4.7 the opportunity arose to study the distribution of lic2A SSR lengths in 174 sputum supernatant samples from individuals with COPD. From the genescan analysis of the lic2A fragment sizes produced following successful PCR amplification of 65 of these, it was observed that putative tract lengths ranged from 1-31 SSRs in length, consistent with the observations of SSR distribution for the GenBank isolates in section 4.3.1, as well as the previous studies into lic2A SSR distribution discussed earlier. Furthermore, similar to the analysis of the lic2A SSR lengths of the GenBank isolates, the most commonly observed fragment size correlated to a tract length of 22x 5'-

CAAT-3'. For the sequenced strains within the GenBank database from individuals with COPD, there was a greater number of *lic2A* ON isolates than *lic2A* OFF, although the sample size was too small to make any definitive conclusions. In the sputum samples 22x 5'-CAAT-3' was the most common SSR tract length (16/65 sputum supernatant samples had this tract length), which would encode an ON lic2A should these isolates contain the 5'-CAAA-3' sequence at the terminal end of the SSR as observed for the the majority of GenBank H. influenzae lic2A sequences. The remainder of the sputum samples however were predicted to encode OFF *lic2A* genes. The prediction of multiple OFF is largely due to the presence of samples with X.25 and X.75 repeats (33/65 COPD samples were identified with peaks correlating to these sizes) that may be errors and actually represent the closest integer SSR length to X. Should this be the case then the majority of the samples will be in the ON state (41/65 of)the sputum supernatant samples), making sputum sample analysis again consistent with the limited number of COPD samples from GenBank. The misattribution of SSR length in the X.25 and X.75 samples is the most likely scenario, as analysis of the 104 GenBank isolates did not detect any *lic2A* gene that would produce this tract length. Additionally if the majority of samples were in the OFF state it is unclear why the most frequent single SSR number would correlate with a ON *lic2A*. Thus it is likely that the majority of these COPD samples are in the ON state, with 22x 5'-CAAT-3' repeats being the most common tract length.

With the *lic2A* gene functioning as a means of survival to immune challenge it is likely that its ON expression would be as an adaptation to immune stress. COPD is characterised by a widespread inflammation within the airways, involving both the innate and adaptive immune response, to which *H. influenzae* colonisation is believed to be a contributor (Moghaddam et al., 2011). *H. influenzae* is found within the lower respiratory tract of ~ 30 % of individuals with COPD (Bandi et al., 2001). These individuals have a high level of immune activity, it is possible therefore that *lic2A* may be playing a role in immune evasion and hence may also be facilitating the survival of the bacterial populations survival under conditions of heavy immune challenge.

Furthermore, as *H. influenzae* with 22x 5'-CAAT-3' repeats was found to predominate in the COPD samples this may actually bode well for the future treatment of COPD. As should with 22x 5'-CAAT-3' repeats being associated with ON expression of *lic2A* this would also imply that the bacterial population will be at least sensitive to bacteriophage HP1c1, and as such this may mean that these ON cells could be targeted by other bacteriophage should bacteriophage therapy be applied in the future. That being said, other defence mechanisms may be active against bacteriophage, such as restriction modification systems (e.g. the Type I system, or the Type III system). Furthermore, LOS structures differ from strain to strain, and although *lic2A* may be in the ON state, the position of its LOS extension may differ between strains (e.g. which heptose residue the Gal*p*-(1-4)- β -D-Glc*p* is added), which in turn may effect the sensitivity of the bacterium to HP1c1-like bacteriophage predation, in a similar manner to where the position of phosphorylcholine is added on the *H. influenzae* LOS alters the extent of killing

by C-reactive protein (Weiser et al., 1998; Lysenko et al., 2000a). There is however no literature at present on the development of bacteriophage as therapeutics against H. influenzae. That is not to say that bacteriophage therapy against H. influenzae cannot be developed, and observations such as ours may be a worth while consideration in encouraging their development.

An interesting finding was that Rd30R contained 26x 5'-CAAT-3' within its SSR. This would mean that Rd 30R is in the +1 frame compared to the ON frame of *lic2A*. The +1 frame has previously been identified as producing partial ON expression of *lic2A* due to translation from an alternative initiation codon (Dixon et al., 2007). However, while that Rd 30S is highly sensitive to bacteriophage HP1c1 infection, Rd 30R displays absolute resistance to bacteriophage HP1c1 infection. This indicates that the +1 frame, while conferring weak *lic2A* expression, does not produce sufficient quantities of modified LOS to facilitate bacteriophage infection.

In summary, although the observations of Zaleski et al. (2005) for a bacteriophage resistance phenotype for the phase variable HindI restriction modification system could not be replicated, their observations for the phase variable glycosyltransferase, *lic2A* conferring resistance to bacteriophage HP1c1 was observed. Furthermore their observations further correlate with the putative ON/OFF states found in the vast majority of *H. influenzae* strains available for analysis within the GenBank database. It was also observed that, based on the *lic2A* sequence data from GenBank, *lic2A* expression may allow survival of bacteria during invasive conditions such as bacteraemia and meningitis. This may be due to the immune resistance associated with lic2A expression, whereby either the lic2A ON cells are those that are more able to migrate from the respiratory tract into the blood, or an event where the invasive H. influenzae population adapts within the blood to a *lic2A* ON state due to the immune selection. In support of this hypothesis, *lic2A* is not found in the ON state in *H. influenzae* strains isolated from other locations of the body, possibly due to lower levels, or different forms of immune selection in these other niches. Furthermore, following analysis of the *lic2A* tract lengths from sputum samples from individuals with COPD, 22x 5'-CAAT-3' repeats was the most commonly observed SSR number. This may prove a useful observation if bacteriophage therapy is to be applied as a antimicrobial therapy in the treatment of COPD. Finally, as it was possible to replicate the observations of a phase variable bacteriophage sensitive/resistant phenotype for lic2A it was now possible to investigate the effect of phase variation on bacteriophage-host dynamics.

Investigation of the effect of the Phase Variable
Bacteriophage Receptor, *lic2A*, on Population dynamics
of *H. influenzae*, and Bacteriophage HP1c1

5.1 Abstract

With the bacteriophage resistance phenotype associated with the phase variable lic2A gene of H. influenzae successfully observed, this chapter aimed to investigate the effect of phase variation on bacteriophage/host dynamics. Bacteriophage/host dynamics were investigated through a complementation of in vitro experiments and mathematical modelling of these experiments. For this a novel experiment, termed the 'oscillating prey assay' was designed. In the oscillating prey assay bacteriophage were serially passaged through cultures of either lic2A ON or lic2A OFF cells to simulate a population composed of varying ratios of ON:OFF cells. From this it was observed that mixed populations containing both lic2A ON and lic2A OFF cells were able to reduce bacteriophage spread through the population, even in the scenario tested where the bacterial population contained predominantly lic2A ON cells (66% ON). These observations were then further validated though generation of a mathematical model based on the results of the oscillating prey assay.

A modification of the oscillating prey assay was also used to observe the effect of phase variation generated heterogeneity on bacteriophage spread through a fixed area. It was observed in this assay that again the mixed phenotype populations could reduce bacteriophage spread, in the process creating discrete populations of bacteriophage densities that may play a key role in the eventual diversification of members within the larger bacterial macropopulation.

5.2 Introduction

Phase variable bacteriophage resistance mechanisms are found in a number of bacterial species, such as *V. cholerae*, *C. difficile*, and *S. enterica* (Cota et al., 2012; Seed et al., 2012; Sekulovic and Fortier, 2015). These can confer expression changes in the surface structure that act as the bacteriophage receptor, or in specific molecules involved in bacteriophage defence such as Sie systems, and restriction modification systems (Sørensen et al., 2011; Sekulovic and Fortier, 2015; Anjum et al., 2016).

One phase variable gene associated with a bacteriophage resistance in *H. influenzae*, *lic2A*, contains a 5'-CAAT-3' SSR tract that causes ON-OFF and OFF-ON switching of a Galp-(1-4)- β -D-Gal extension on the *H. influenzae* LOS. In the ON state lic2A is believed to be involved in immune evasion, increasing the survival of the bacterium under immune challenge (Clark et al., 2013). However, when the Galp-(1-4)- β -D-Gal extension is expressed the cell is now vulnerable to infection by bacteriophage HP1c1 (Zaleski et al., 2005). When *lic2A* switches to the OFF state the bacterial population shows absolute resistance to bacteriophage infection (Zaleski et al., 2005). As such the bacterial population will be exposed to two opposing selections, one requiring an active *lic2A*, while the other opposes its expression.

The involvement of bacteriophage as a selection pressure has been shown previously to facilitate the evolution of mutator phenotypes, populations with increased global mutation rates (Pal et al., 2007). In essence phase variable loci are simply regions within the chromosome which display increased plasticity, without the costs of global hypermutation. Furthermore, the stochastic nature of the expression plasticity they confer allows the bacterial population to alter its composition based on the strength of selection pressure with relative ease, due to the constant shedding of cells of the converse phenotype even in the face of strong opposing selection. Although the resistance conferred by *lic2A* phase variation will be binary at the cellular level, at the population level the heterogeneity generated will cause resistance to be modular. Wherein with increasing bacteriophage selection the bacterial population will be forced into shifting to a greater density of OFF variants than ON, and vice versa in scenarios of immune selection. Although these characteristics are known for bacterial adaptation following application of selection pressure, and the importance of phase variation has been suggested in the ecology and evolution of bacteriophage, to date no studies have examined what effect this mutational mechanism can have on bacteriophage propagation dynamics.

As such, this study aimed to investigate the effects of phase variation on bacteriophage propagation through heterogeneous populations with differential expression states of lic2A expression. This was to be achieved through the construction of mathematical models (constructed by Dr Andrey Morozov, University of Leicester) based on the experimental data obtained here. The model would aim to replicate a novel experiment, termed an oscillating prey assay, which allows transfer of bacteriophage to sequential cultures of host populations with different lic2A expression phenotypes that allow simple simulation of a fluctuating lic2A phenotype. By definition population heterogeneity consists of a mixture of phenotypes, in the case of a binary phenotype like the ON/OFF state of the lic2A gene, this will be dependent on the ratio of ON:OFF cells. To achieve this, at each transfer bacteriophage from a culture of one lic2A phenotype would be transferred to cultures of either the same, or opposing lic2A phenotype, in a sequential pattern depending on the level of phenotype mixing being simulated (e.g. ON-OFF-ON-OFF-ON... would be a simulation of a 50 % ON population [see figure 2.1 for more detail]). In order to achieve this the bacteriophage population and the host population would be separated after one viral replicative cycle before being transferred to a fresh bacterial population, the reasoning for this was four-fold: (1) to prevent the continued selection caused by bacteriophage lysis altering the bacterial population's lic2A phase composition, (2) preventing the development of alternative resistance mechanisms by the bacteria, (3) this allowed us to maintain the prey bacterial population at a constant density at the point of each bacteriophage transfer, and (4) through separation of bacteriophage and host the bacteriophage population could be reintroduced to an alternative lic2A phenotype in sequential patterns and thus simulate mixed populations.

Furthermore, the oscillating prey assay would also encumber the bacteriophage population with a cost following each transfer. This point is pertinent as with other resistance mechanisms such as restriction modification systems, Sie systems, and CRISPR-cas systems, bacteriophage will be consistently lost from the environment with every failed infection, due to the initial adsorption process requiring irreversible binding to the host cell. Thus, with every attempted infection of bacteria with these systems that successfully defends itself from bacteriophage infection the bacteriophage population will reduce in number by 1. Although this may not appear a significant contribution to diminishing the numbers of the bacteriophage population, its effects can be extensive when it is considered that multiple bacteriophage can attach to a single bacterium. Depending on the efficiency of the resistance mechanism this may mean that all bacteriophage infection attempts are repelled, and with it a large scale reduction in bacteriophage titre. Resistance through modification or loss of a bacteriophage receptor on the other hand, by definition will not. As with receptor expression changes the initial adsorption stage is prevented, and with it the irreversible binding that reduces bacteriophage numbers.

Instead the major means of reduction in bacteriophage numbers when the bacteriophage are unable to bind to a host are through environmental factors, or the kinetics applied to the population. This can be due to a number of factors, including spontaneous 'death' of the viral particle, inactivation by debris (including cell debris from previously colonising cells with a competent receptor), inactivation by environmental factors (e.g. temperature, pH, chelation, kinetic stress, etc. (Jończyk et al., 2011)), system clearing (e.g. wind/water movement in nature, bowel movements in the animal gut, cilial action in the ears and lungs), or on a larger scale the bottleneck of bacteriophage migration from one system to another (e.g. the successful fraction of bacteriophage that migrates from one individual to another via cough droplets). Thus it would be necessary to provide a means with which bacteriophage will be lost from the environment, as otherwise, in the current system when faced with a high number of *lic2A* OFF cells, and no significant change to the environmental conditions the bacteriophage population would likely stay constant, as in the current closed system no binding to the bacterial cell would occur and nor would environmental factors influence the bacteriophage densities.

As such through the above assay it would be possible to generate a model of the effects of population heterogeneity on bacteriophage survival, allowing expansion of the experimental results beyond what is feasible in the laboratory. Furthermore, this model would then be expanded upon, through construction of a methodology to determine how the effects of phase variation may alter the dispersal of bacteriophage through a spatially heterogeneous environment.

5.3 One-Step growth curve of bacteriophage HP1c1 on *H. influenzae* Rd 30S and Rd 30R

In constructing the method for measuring the effect of bacteriophage spread through heterogeneous populations, a number of characteristics of the bacteriophage must be determined. First, if bacteriophage are to be cycled through phase variant populations after a single bacteriophage infective cycle the length of time required for the completion of a single bacteriophage replicative cycle must be determined. To this end one-step growth curves of bacteriophage HP1c1 using each of the *H. influenzae* Rd 30 phasotypes were conducted.

The $\sim 1 \ge 10^8$ CFU ml⁻¹ bacterial population was initially inoculated with 10⁵ PFU ml⁻¹ bacteriophage HP1c1 with excess unbound bacteriophage removed following a period of binding by washing. This allows removal of background unbound HP1c1 (for methodology details see section 2.4.2). However, despite this, consistently, $\sim 1 \ge 10^3$ PFU ml⁻¹ remained present in the suspension for the Rd 30S phasotype (Fig. 5.1), while all detectable unbound bacteriophage were consistently removed from the cultures containing Rd 30R. Therefore, although this may appear as Rd 30S and Rd 30R have differing initial inocula, this is not the case, but this is unbound HP1c1 that could not be readily removed despite three washing steps. The reason for this observation remains unknown.

From the one-step growth curve (Fig. 5.1) initially it was expected to see only replication of bacteriophage HP1c1 in the *H. influenzae* Rd 30S phasotype, as this bacteriophage population is sensitive to bacteriophage predation. However, a small quantity of bacteriophage replication within the Rd 30R phase variant was also observed. As expected however, the level of replication observed for bacteriophage HP1c1 on the Rd 30R culture was significantly lower than that of the Rd 30S, even at the time point where the highest bacteriophage titre was observed for Rd 30R (two-sample *t*-test: t = 1.56, P < 0.001; mean \pm SEM PFU ml⁻¹ values for bacteriophage titre for the 55 minute time point for



Figure 5.1. One-step growth curve of bacteriophage HP1c1 on *H. influenzae lic2A* phase variants. Initial cultures of ~ 1 x 10⁸ CFU ml⁻¹ *H. influenzae* Rd 30S (\bullet) or Rd 30R (\blacktriangle) were incubated with bacteriophage HP1c1 at an MOI of 0.001 for 10 minutes to allow bacteriophage attachment. Suspensions were then centrifuged, the supernatant discarded, and the bacterial pellet then resuspended in fresh sBHI with this washing process repeated a further two times to remove as much unbound free HP1c1 as possible. After the final washing step 2 ml aliquots were then produced for each time point, and incubated at 37°C. Samples were taken every 5 minutes for 60 minutes to determine bacteriophage titre. N = 5, Error bars = mean \pm SEM.

replication of bacteriophage HP1c1 on Rd $30S = 1.90 \pm 1.21 \times 10^6$; 55 minute time point for Rd $30R = 2.11 \pm 0.87 \times 10^4$; Figure 5.1).

We observed that the HP1c1 latent phase, that is the length of time from the onset of infection (i.e. adsorption of bacteriophage to the host cell) until the first production of viral progeny, lasted for around 20 minutes based on the data for Rd 30S, and around 25 minutes based on the one-step growth curve of HP1c1 on Rd 30R. This time period will encompass the eclipse period, the time following adsorption before the production of viral particles within the cell; and at least part of the intracellular rise phase, the time of production of viral particles within the cell.

From the end of the latent period, the rise period, the time at which bacteriophage are actively

being released from the cells, can be noted to last from 20 minutes to around 45 minutes in the Rd 30S population, thus giving a rise period of around 25 minutes. In Rd 30R the rise phase lasts from 25 minutes to around 45 minutes, thus meaning the rise face in the Rd 30R cultures lasted 20 minutes. Burst was determined from the PFU ml⁻¹ values obtained for each of the growth curves and determined to be 18 ± 2.2 PFU ml⁻¹ for the Rd 30S populations, and 0.7 ± 0.1 PFU ml⁻¹ for Rd 30R. After these points, in both populations the bacteriophage titres plateaued, with a subsequent decrease in bacteriophage titre only observed in the Rd 30S population after 55 minute (likely due to the initiation of the period of adsorption for commencement of a second infective cycle).

From the one step growth curve the bacteriophage HP1c1 replication through both the Rd 30S and Rd 30R populations was determined to be completed after around 45 minutes. The decrease in HP1c1 titre within the Rd 30S population at 60 minutes is likely to be the onset of a secondary infective cycle, therefore the optimal window for separation of bacteriophage from bacteria after one replicative cycle would be between 45-55 minutes post inoculation. As such, 50 minutes was selected for the length of time of each incubation period in the oscillating prey assay.

5.4 The effect of Population Heterogeneity on Bacteriophage HP1c1 Spread

To simulate the effect of phase variation on the spread of bacteriophage through bacterial populations an assay termed the oscillating prey assay was developed. This assay was developed to generate experimental data that can then be applied to the construction of a mathematical model allowing expansion of the experimental results beyond what is capable in the laboratory setting. The oscillating prey assay involves the continual cycling of bacteriophage through host cultures of either *lic2A* ON or OFF phenotype following one bacteriophage replicative cycle. The cycling pattern of host *lic2A* expression state was altered to simulate different levels of phenotype mixing within the population due to phase variation. At each cycling step the bacteriophage population is diluted 10-fold to place a burden on the bacteriophage population at each transfer. A detailed experimental methodology can be found in section 2.4.6. The assay aimed to determine the effect of varying degrees of population phenotype mixing on bacteriophage propagation through six populations structures: (1) 100 % ON (HP1c1 cycled only through *lic2A* ON populations), (2) 66 % ON (2:1 *lic2a* ON:OFF), (3) 50 % ON (1:1 *lic2A* ON:OFF, beginning in an ON culture), (4) 50 % OFF (1:1 *lic2A* ON:OFF, beginning in an OFF culture), (5) 66 % OFF (1:2 *lic2A* ON:OFF), and (6) 100 % OFF (HP1c1 cycled only through *lic2A* OFF populations). The result of this assay can be observed in figure 5.2.

From the oscillating prey assay it was observed that the chance of bacteriophage survival was a function of the number of lic2A ON hosts the bacteriophage encounters (Figure 5.2). Extinction events were





present in all simulated populations except 66 % ON and the 100 % ON populations. The simulations of populations composed of entirely *lic2A* OFF phase variants resulted in extinction of HP1c1 from the environment before all other population structures, reaching below the minimum detectable threshold at cycle number 5. Extinction events occurred thereafter for the populations composed of 66 % *lic2A* OFF cells, and both variations of 50 % OFF/ON *lic2A*, with extinction events at cycles 8, 13, and 16 respectively.

It is however of note that although the 66 % and 100 % ON allowed bacteriophage survival, the survival was significantly greater at the final transfer in the 100 % ON population than the 66 % ON simulation (Fig. 1; two sample t-test; t = 4.298, P < 0.05; mean \pm SEM PFU ml⁻¹ values for bacteriophage at end of 20 cycles in simulations of 100 % lic2A ON population = $3.85 \pm 0.9 \times 10^9$; 66 % lic2A ON population = $2.82 \pm 0.88 \times 10^4$). Thus the 66 % ON population, despite being composed of a majority of *lic2A* ON cells was able to restrict bacteriophage propagation compared to that of a wholly sensitive population.

When analysed individually, the bacteriophage population significantly increased in density over time when propagated through a fully sensitive 100 % ON population, as would be expected (Fig. 1; paired t-test: t = 4.97, P < 0.05; mean \pm SEM PFU ml⁻¹ values for bacteriophage density at for propagation though the 100 % ON simulation at cycle 0 = 9.61 \pm 1.15 x 10⁵; at cycle 20 = 3.85 \pm 0.90 x 10⁹). In contrast the mean bacteriophage density decreased over time following bacteriophage HP1c1 propagation through the 66 % lic2A ON population simulation (Fig. 1; paired t-test: t = 4.97, P < 0.05; mean \pm SEM PFU ml⁻¹ values for bacteriophage density for propagation though the 66 % ON simulation at cycle 0 = 7.16 \pm 0.26 x 10⁵; at cycle 20 = 2.82 \pm 0.90 x 10⁴).

Thus should the experiment have continued the 66 % ON population would likewise have caused the bacteriophage densities to declined into extinction. While, as bacteriophage propagation progressed uninhibited in the 100 % *lic2A* ON population simulation, if this were a natural environment, without an adaptive response by the bacterial population, or an external intervention that caused a reduction in bacteriophage populations, the bacterial population would likely have gone extinct from the bacteriophage pressure. Thus from this experiment it appeared that bacterial population heterogeneity for *lic2A* expression allows a large degree of control over bacteriophage propagation, even when the majority of the population is still sensitive to bacteriophage predation.

5.5 Determination of adsorption rate of bacteriophage HP1c1 to Rd 30S

Bacteriophage infections are essentially a three step process. First the bacteriophage must adsorb to the host cell; this is followed by generation of new viral particles within the cell; and finally culminates with the release of the viral progeny following lysis of the host cell. While data had been obtained that can be utilised for the modelling of each of the latter two steps, in order to reflect the adsorption process within the model, the rate of bacteriophage adsorption to the cell must be determined. For this a bacteriophage adsorption assay was conducted to determine the adsorption kinetics of bacteriophage HP1c1 in the presence of *H. influenzae* Rd 30S.

This adsorption assay was to be conducted over 60 minutes (encompassing the length of each cycle of the oscillating prey assay), with the decline in numbers of unadsorbed bacteriophage measured at regular intervals. However, as the measurement of free bacteriophage would be effected by the release of bacteriophage from the cells following bacteriophage replication within cells a bacteriostatic antibiotic was applied to prevent bacteriophage replication. Therefore, the assay was conducted using 2 μ g ml⁻¹ chloramphenicol. The results of the adsorption experiment can be observed in figure 5.3.

An adsorption rate coefficient (k) was calculated (see section 2.4.3) for bacteriophage HP1c1 to *H. in-fleunzae* Rd 30S. Over the entire length of the assay (i.e. from 0-60 minutes) the adsorption rate constant (k) was found to be 3.41×10^{-10} ml min⁻¹. This implies adsorption of bacteriophage HP1c1 to Rd 30S occurs at an intermidiate rate, as for instance, bacteriophage T4 is known to bind rapidly to its host and displays a k of 2.4×10^{-9} ml min⁻¹, while the M13 bacteriophage displays a slower binding rate of 3×10^{-11} ml min⁻¹ (Kropinski, 2009). It should be noted however, that a wide range of factors can alter the adsorption rate constant including characteristics of the bacterial cells physiology, and environmental factors (Kropinski, 2009).

Although $3.41 \ge 10^{-10}$ ml min⁻¹ was the overall rate of adsorption during the full 60 minutes, it can be seen in figure 5.3 that the HP1c1 population adsorbs to Rd 30S in two phases. The first phase, lasting from 0 minutes until around 40 minutes involves a faster adsorbing fraction with a constant of $4.80 \ge 10^{-10}$ ml min⁻¹. The second phase involves a slower adsorbing subpopulation, occurring at a rate of $6.14 \ge 10^{-11}$ ml min⁻¹. This second phase of adsorption is thus considerably slower than the first, occurring at a rate $\sim 7.8 \ge$ slower than the initial adsorption phase. Although it may first be suggested that this reduction is due to the reduction in the numbers of cell surface receptors that are free to recieve bacteriophage, it should be noted that the adsorption experiments were conducted at an MOI of 0.1, and thus the bacterial cells (which are decorated with more than one bacteriophage receptor themselves) outnumber the bacteriophage particles, thus there would still be a high number



Figure 5.3. Adsorption assay of bacteriophage HP1c1 to *H. influenzae* Rd 30S. Points indicate the bacteriophage titres remaining in free suspension (i.e. unbound bacteriophage) at each time point as a fraction of the initial inocula following incubation of bacteriophage HP1c1 with *H. influenzae* Rd 30S (\bullet) and in the absence of cells (\blacktriangle). N = 5.

of free receptors for the bacteriophage particles to bind to. This second fraction is therefore most likely to be due to the behaviour of the bacteriophage population. It is not unusual for bacteriophage to display reduced adsorption rates over time (e.g. Moldovan et al. 2007), as bacteriophage populations often can be seen to contain a slower adsorbing subfraction of particles (Kropinski, 2009; Gallet et al., 2012).

As each of the cycles of the oscillating prey assay lasted for 50 minutes the adsorption rate from 0-50 minutes is of key interest for use in the model. Over this time period, bacteriophage adsorption was observed to occur at a rate of 3.93×10^{-10} ml min⁻¹.

For the cell free control only a single 1adsorption' rate was observed (although being discussed as 'adsorption' for cell free control for simplicity this would be inhibition rather than adsorption) during the 60 minute incubation. Bacteriophage were inactivated in this suspension at a rate of 1.98×10^{-10} ml min⁻¹. This is significantly lower than the rate observed for bacteriophage incuba-

tion with the Rd 30S cells over the same period (two sample t-test: t = 5.563, P < 0.01; mean \pm SEM adsorption rate of bacteriophage (ml min⁻¹) from 0-60 minutes in the cell free suspension = $1.98 \pm 0.06 \times 10^{-10}$; following incubation with Rd 30S = $3.41 \pm 0.26 \times 10^{-10}$; Figure 3.4). Although slower than the rate following incubation with cells this implies the rate of inactivation of bacterio-phage in the absence of cells is relatively high (e.g. bacteriophage is lost at a faster rate than the rate that the M13 bacteriophage binds its cognate receptor).

Through determination of the bacteriophage HP1c1 adsorption rate it was possible to add this information to the known parameters and proceed in generating the model of the effects of phase variation on the propagation dynamics of bacteriophage through phase variable populations.

5.6 Description of the mathematical model facilitating simulation of the oscillating prey assay

To expand the oscillating prey assay a mathematical model, able to replicate the observations was constructed by Dr Andrey Morozov (University of Leicester), using the empirical data on bacteriophagehost interactions. To begin, a system of ordinary differential equations were constructed to describe the interaction of bacteriophage and their hosts during each cycle number n of incubation time T. The differential equations are as follows:

$$\begin{split} \frac{dB_0}{dt} &= -KPB_0 \\ \frac{dB_i}{dt} &= KPB_{i-1} - KPB_i \\ \frac{dB_N}{dt} &= KPB_{N-1} \\ \frac{dP}{dt} &= -PK\sum_{i=0}^{N-1} B_i - mP + bH(t - T_0)KB_0(t - T_0)P(t - T_0) \qquad nT < t < T(n+1) \end{split}$$

Here, B_0 describes the number of bacterial cells unassociated with bacteriophage (i.e. each unadsorbed cell) while B_i is the number of bacterial cells with *i* bacteriophage adsorbed to the cell; *N* represents the maximum number of bacteriophage attachments; and *P* is the number of bacteriophage free in solution. Furthermore, *K* is the rate at which bacteriophage adsorb to the cell surface; *m* is the rate of bacteriophage mortality; *b* is the burst size of the bacteriophage; and *H* is the Heaviside function, given by H(x) = 0, x < 0; H(x) = 1, $x \ge 0$. The Heaviside function is required to show bacteriophage replication occurs after the replication time.

As, in the experimental study, bacterial densities were kept constant at the initiation of the experiment

and were not measured or carried over at the end of each cycle, in the mathematical model, for simplicity, the bacterial density is assumed constant throughout the experiment. The assumption is that the total bacterial densities are maintained at that of the initial B_0 value throughout the experiment, thus bacteriophage release does not alter bacterial densities and nor does the dilution effect at each step, such that:

$$\sum_{i=0}^{N} B_i = B_0(t^+ = Tn) \qquad nT < t < T(n+1)$$

where t^+ is the time just after the dilution step (i.e. the beginning of the next cycle).

An additional assumption for simplicity, is that the adsorption rate of bacteriophage to bacterial cells (K) is independent of the maximum number of attachments (N). In terms of adsorption, sensitive bacteria, i.e. those expressing *lic2A* (Rd 30S), show adsorption of K > 0, while the resistant cells, those with *lic2A* in the OFF state (Rd 30R), show no adsorption, K = 0. Thus for the Rd 30R cells, it was assumed that the replication noted in section 5.3 was due to the presence of a small number of Rd 30S cells, and that the Rd 30R cells show absolute resistance. Furthermore, it was also assumed that K is constant across each cycle of T length. That is to say that rather than the two phase adsorption noted in section 5.5, one blanket adsorption rate was applied across the length of each cycle.

At any time during the cycle, three components can be described within the culture solution, bacterial cells which are not bound by bacteriophage (B_0) , bacterial cells that are bound by *i* number of bacteriophage (B_i) , and the total number bacteriophage that are free in suspension (will contain both the initial bacteriophage from a previous cycle [i.e. those that did not produce a viable infection in a previous cycle, but nor were lost from the system at the dilution step] and the new progeny that were produced from infection of the bacterial host). Thus at the beginning of each infective cycle, in the model these components were set at the following densities:

$$B_0(t^+ = Tn) = B_0; \quad B_i(t^+ = Tn) = 0, \quad i = 1, ...N; \quad P(t^+ = Tn) = P(t^- = Tn)C$$

where C is the dilution enforced on the bacteriophage population at each transfer step. In the oscillating prey assay of section 5.4 this was the 10-fold dilution applied at each step (thus C = 0.1). Also here, t⁻ represents the time just before the dilution (i.e. at the immediate end of the previous cycle).

Bacteriophage densities at the end of each cycle were calculated using the following equations:



Figure 5.4. Examples of the outcome from the model of the three components added at the beginning of each cycle. P (blue line) represents the density of the initial bacteriophage population added at T =0 across the 50 minutes (in this case $P = 4.28 \times 10^7$). This can be seen to decrease at a constant rate in line with the rate of adsorption ($K = 5 \times 10^{10}$ for sensitive bacteria and K = 0 for resistant bacteria) and additionally the rate of bacteriophage mortality (0.005 ± 0.003 bacteriophage min⁻¹). B_0 (orange line) are the number of bacterial cells unbound by bacteriophage, this was set to 1.75×10^8 . Note that bacteriophage produced following burst from the bacterial cells are not shown in this diagram, as this diagram represents only the fate of the initial 'inoculum' used in the modelling simulation. B_1 (green line) and B_2 (purple line) represent the number of bacterial cells with 1 and 2 bacteriophage adsorbed to the cell respectively. This would continue until B_{10} in the simulations, however the densities of $B_{> 2}$ were very low and thus are not represented in the diagram.

$$P(t^{-} = Tn) = \sum_{i=1}^{N} B_{i}(Tn - T_{0})b + P_{s}(t^{-} = Tn)$$

or
$$P(t^{-} = T(n+1)) = \sum_{i=1}^{N} B_{i}(T(n+1) - T_{0})b + P_{s}(t^{-} = T(n+1))$$

Thus, as mentioned earlier the bacteriophage density at the end of each cycle is the sum of the densities of both the bacteriophage numbers produced following infection of a bacterial cell (i.e. the product of the bacteriophage progeny bursts from the bacterial cells adsorbed with at least one bacteriophage), and the number surviving bacteriophage from the previous infection that are still free in suspension (denoted here as P_s).

Through application of these equations it was possible to replicate the data of the oscillating prey

assay. With the experimental data providing validation for the model, it was then possible to attempt to expand on the results observed from the empirical oscillating prey assay.

5.7 Modelling of the effect of phase variation on bacteriophage survival

With a model developed that replicates the experimental data, the mathematical model was then applied to expand the experimental data. In the oscillating prey assay it was possible to extend the assay as far as 20 bacteriophage replicative cycles, through 5 iterations of 6 fixed pattern cycling regimes. Through application of the mathematical model it was possible to simulate bacteriophage infection over a far longer time, a total of 105 cycles. Furthermore, though modelling the pattern of encounter for each simulated population level of mixed *lic2A* phenotypes could be randomised. That is to say that in a 50 % ON population, rather than cycling ON-OFF-ON-OFF... etc the bacteriophage were propagated for 105 cycles through populations that involved a total of 50 cycles through *lic2A* OFF hosts, with the order in which these cycles occur randomised. Furthermore, as this randomised pattern may significantly influence the observations, each condition tested was conducted in 200 iterations, with cycling pattern randomised in each iteration. Through the use of the model it was possible to simulate all population structures from 0-100 % *lic2A* ON.

The model was run with the standard parameters b = 40; T = 50 minutes; and $B_0 = 1.75 \times 10^8$. A maximum number of bacteriophage able to adsorb to each bacterial cell (N) was set at a value of N = 10. A rate of bacteriophage mortality (m), was applied to account for bacteriophage that died by means other than adsorption to the host cell, such as through damage to the particle, inactivation by free LPS etc. This rate was $m = 0.005 \pm 0.003$ bacteriophage minute⁻¹ based on the cell free control adsorption assay.

In the experimental assay the application of a 10-fold reduction in bacteriophage titre at each transfer was conducted. This was conducted to simulate loss of bacteriophage from the environment that would be seen in nature over time due to abiotic (pH change, salinity, etc) and biotic (inhibition due to noncell adhered receptors in the environment, cilial clearance etc) factors. As such, a dilution step was also applied at each bacteriophage transfer in the model simulations. This rate of dilution (C) was applied with some variance, as would be seen in the experimental process. This variance was applied as $C = C(1 + \epsilon)$ where ϵ is a normally distributed random variable with a mean of 0.3 and a variance of 0.3^2 . In the simulations here C = 0.1, thus reflecting the 10-fold dilution of the oscillating prey assay.

As the oscillating prey assay relies on the switching of the bacterial populations phenotype at each



Figure 5.5. Examples of bacteriophage densities during simulations from the mathematical model of the oscillating prey assay. Figure shows each cycle of 6 example iterations of the modelling of bacteriophage replication through bacterial populations with differing degrees of phenotype mixing for *lic2A* expression for 105 viral replicative cycles. Each point represents the bacteriophage densities $(\log(P))$ at the end of each incubation cycle. Individual iterations are represented by each coloured line. (a) shows the simulation of bacteriophage propagation through a 75 % *lic2A* ON population (r = 0.75), while (b) shows simulation of bacteriophage spread through populations that are 85 % ON for *lic2A* expression (r = 0.85). These simulations were conducted using the parameters outlined in the text. The sharp drop in bacteriophage titre at low concentrations in (a) is due to the extinction threshold of $P_0 = 100$

bacteriophage infective cycle this was also implemented in the model. Here, phenotype switching was facilitated through altering the adsorption rate, K, from $K = 5 \ge 10^{-10}$ for sensitive bacteria to K = 0 for replications through the populations of resistant cells. The number of cycles through bacterial populations with either of these adsorption rates was dependent on the probability of encountering a sensitive bacterial population, p. Thus simulation of a bacterial population composed of 66 % sensitive cells (66 % *lic2A* ON) would be set at p = 0.66.

Finally, as the simulation does not aim to describe an infinite system in terms of bacteriophage densities, that is to say that bacteriophage can undergo extinction events. A minimum threshold was applied which, once reached, does not allow further cycling of the lineage. If the density of the bacteriophage population falls below this threshold then the bacteriophage population cannot recover and the population is deemed to be extinct with P = 0 for the remainder of the simulation. This threshold value, P_0 , was set to a consistent level across all cycles ($P_0 = 100$).

All the above were applied to simulate the effects of bacterial population mixing for bacteriophage resistance phenotypes on bacteriophage survival following migration through these populations over the course of 105 viral replications. For each cycle T = 50 minutes would correlate to a total time course of 5250 minutes, or ~3.65 days of bacteriophage HP1c1 replication. Examples of the bacteriophage population densities during the 105 cycles can be seen in figure 5.5.

From the example simulations the randomisation of the host state can be clearly seen, with examples of tumultuous periods of increases and decreases in bacteriophage densities observable in both figure 5.5 (a) and (b). Furthermore, the potential for the pattern of the bacterial populations *lic2A* expression can be seen to have a key role in determining the bacteriophage populations chances for survival. In the set of example cycling patterns illustrated in figure 5.5 (a), in the 6 bacteriophage propagation simulations only 1 simulation allows the bacteriophage populations survival by the end of the 105 cycles. Although 5 of the populations in the simulation of figure 5.5 (a) undergo extinction (i.e. their populations densities fell below the $P_0 = 100$ threshold) the times at which extinction arises is varied. Extinction events in these populations can be seen to have occurred at cycles 16, 41, 61, 89, and 97. Furthermore, these extinction events can still occur as early as cycle 16 (and potentially earlier) despite figure 5.5 (a) simulating a bacterial population structure that is composed of 75 % *lic2A* ON cells.

In figure 5.5 (b) the bacteriophage densities at each cycle during replication in an 85 % ON population can be seen. Here no extinction events can be seen to have occurred during the 6 simulations. That is not to say that extinction events cannot occur in this population, as an 85 % OFF population would allow a total of 16 cycles through resistant bacterial populations. Even with bacteriophage densities of $P = 1 \ge 10^{10}$ the threshold of $P_0 = 100$ will be reached should 8 cycles with resistant populations occur in a row. Although this will be a rare event, with lower densities the number of required consecutive cycles to produce an extinction event will be lower.



Figure 5.6. Modelled probability of bacteriophage survival following replication through heterogeneous bacterial populations for *lic2A* expression after 105 cycles. (a) shows the mean proportion of bacteriophage lineages that survived the length of the 105 cycles. Coloured areas represent the proportions of bacteriophage populations that survived till the end of the experiment. Colour correlations are shown on the bar to the right of the graph with 1 representing 100 % of lineages surviving, and 0 representing 100 % of the bacteriophage lineages had gone extinct. The proportion of surviving lineages was calculated based on the number of lineages where bacteriophage densities had not fallen below the extinction threshold of $P_0 = 100$ after 200 iteration of random pattern cycling. The parameters used are those outlined previously in the text. (b) shows a graphic representation of the role of population heterogeneity in bacteriophage survival chances. As the number of sensitive cells increases (green circles) from the initial 100 % resistant (pink circles) so to does the proportion of bacteriophage lineages that survive until the end of the 105 cycles.

Using the model, how phase variation effected the chances of survival of bacteriophage populations at the end of the 105 cycles was determined over 200 repetitions. Survival was measured by the ratio of populations that had maintained survival of the bacteriophage through the entire length of the 105 cycles against the number of iterations where extinction events had occurred. The chances of bacteriophage survival will be strongly influenced by two of the variables, the rate of dilution after each cycle (C), and the degree of phenotype mixing within the host population for *lic2A* expression that the bacteriophage is replicating through. As such, simulations using a range of bacteriophage populations structures for *lic2A* expression were conducted, with each population structure also modelled over a range of dilution levels. The results of which can be observed in figure 5.6 (a).

These variables can be seen from figure 5.6 (a) to be key factors in the survival chances of bacteriophage. In terms of bacterial resistance phenotype mixing, with increasing proportions of resistant bacterial cells came a decrease in the proportion of bacteriophage able to survive the length of the 105 cycles (5.6 (b)). The degree to which heterogeneity was able to alter the levels of bacteriophage survival however, was strongly influenced by the rate at which bacteriophage were being lost from the system. It can be seen that with a higher level of bacteriophage being carried over between each cycle (e.g. C = 0.6, meaning 60 % of bacteriophage is carried over) bacteriophage were able to survive over a larger range of bacterial *lic2A* populations structures. For C = 0.6 it was found that 100 % of lineages could survive even when ~60 % of the bacterial populations were composed of resistant *lic2A* OFF cells. Meanwhile, when C = 0.02 bacteriophage were only able to survive at a high proportion when the bacterial populations were close to 100 % susceptible.

From this it could be said that the ability of phase variation to generate bacterial populations structures where the proportions of *lic2A* OFF resistant phenotype and the *lic2A* ON sensitive phenotype are adaptable dependant on the rate of loss of bacteriophage from the environment could allow control of the bacteriophage population density, preventing unconstrained increases in bacteriophage density and ultimately aiding in bacterial survival. For instance, if C = 0.6 then bacteriophage extinction events occur with $p = \sim 0.2$ (as the probability of encountering resistance in figure 5.6 (a) is 0.8), and as such 20 % of the bacterial population is sensitive, should any of this 20 % have gone uninfected by the bacteriophage population, these can now flourish under bacteriophage free conditions should a selection be applied that benefits their replication. Furthermore, even if all 20 % of total bacterial population that is sensitive to bacteriophage had been killed, following the extinction of the bacteriophage population, through phase variation the remaining resistant isolates can now once again stochastically generate bacteriophage sensitive phase variants that can survive if an event arises that selects for death of the bacteriophage resistant isolates, such as a large immunological response.

Although in the mathematical model it was determined that these heterogeneous bacterial populations may play a key role in reducing the bacteriophage populations expansion and causing the occurrence of extinction events, to further study how these inhibition and extinction events would effect bacteriophage expansion, an these dynamics within a fixed area were next studied.

5.8 Multi-directional oscillating prey assay

From the empirical oscillating prey assay, and extended simulations through mathematical modelling it was observed that population heterogeneity can play a key role in reducing the densities of bacteriophage compared to those populations composed of only sensitive bacterial cells. These assays were linear in nature, while the biological space has multiple dimensions. From the example cycles shown in section 5.7 it can be seen that the sequence of hosts faced by the bacteriophage populations could result in the bacteriophage population undergoing extinction shortly after introduction, or later in the cycles, or even facilitating the bacteriophage populations survival until completion of the cycling simulations (i.e. figure 5.5 (a)).

As an improved representation of biological scenarios the oscillating prey assay was extended into two dimensions. This assay, termed the multi-directional oscillating prey assay, is conducted in a way that allows for the propagation of bacteriophage from one central point of invasion outward in multiple directions. This assay, allow observation of the implications of a variable spatial distribution on the forward expansion of bacteriophage over a fixed area.

In this assay bacteriophage cycling was conducted in a manner similar to that of the oscillating prey assay of section 5.4. That is to say that bacteriophage populations were incubated with hosts for one bacteriophage replicative cycle, after which point they were then passaged into the next well, in the process undergoing a 10-fold dilution. However, in this current assay the cycling experiment was conducted across a number of 96-well plates. Each well of the 96-well plates was pre-designated as either receiving H. influenzae Rd 30S or Rd 30R depending on the population structure being simulated. With the assay conducted across a total of 631 wells, population phenotype mixing was simulated by pre-assigning a proportion of these wells to receive either *lic2A* ON cells or *lic2A* OFF cells (i.e. a 66 % ON simulation would have 416 wells containing Rd 30S). The position of these well designations was randomised (for the randomly generated well designations see section 2.4.7), with the wells only receiving their bacterial inoculum at the point of bacteriophage transfer.

The bacteriophage population was propagated initially from a single the central region of inoculation in eight directions, with further lineages of bacteriophage then arising in the vertical and horizontal directions from the lineages propagating diagonally (as can be seen in fig. 5.7). Thus through monitoring of the bacteriophage densities it was possible to observe how the population heterogeneity generated through phase variation influences bacteriophage survival following spread in a fixed environment.

From this assay it was again observed that non-heterogeneous environments (i.e. propagation of HP1c1 through the 100 % ON and 100 % OFF populations [Fig. 5.7 (a) and (f) respectively]) resulted in far less disconsonant bacteriophage densities. For the 100 % ON population bacteriophage densities



Figure 5.7. Multi-directional oscillating prey assay of bacteriophage HP1c1 through *H. influenzae* populations of varying degrees of phenotype mixing for *lic2A* expression. The simulated populations are as follows: (a) 100 % *lic2A* ON, (b) 66 % *lic2A* ON, (c) 50 % *lic2A* ON, (d) 50 % *lic2A* ON, (e) 33 % *lic2A* ON, and (f) 100 % *lic2A* OFF. Each node represents a well in which bacteriophage density was measured. The colour of each node indicates the concentration of bacteriophage noted at each sampled well. Lines indicate the route taken from the initially inoculated location to the observed position, with the length of each line is proportional to the number of cycles that occurred between each node.

were found to be continuously high in all directions, with values of 10^9 and 10^{10} PFU ml⁻¹ achieved beyond the 3rd cycle. In contrast, for the 100 % OFF population, bacteriophage densities rapidly decreased with each cycle. This resulted in bacteriophage densities reaching $<10^2$ PFU ml⁻¹ beyond cycle 5. Thus in this population structure the bacteriophage populations had undergone extinction in all directions.

The heterogeneous populations on the other hand displayed diverse bacteriophage densities dependant on the direction of propagation. For instance, although no densities of $\leq 10^2$ PFU ml⁻¹ were observed for the 66 % ON population (Fig 5.7 (b)) densities reached as low as 10³, and as high as 10¹⁰ PFU ml⁻¹. This contrast in densities can be most readily noted in this population structure when comparing the bottom left and bottom right diagonal lineages. The bottom left lineage is consistently maintained at 10⁹ PFU ml⁻¹ on the main diagonal lineage. Meanwhile bacteriophage that migrated along the bottom right diagonal lineage faced more harsh conditions, with resulting densities mainly found to be below that of the initial inoculum (i.e. the initial inoculum of the experiment was 10⁶ PFU ml⁻¹). Although on the main diagonal lineage of the bottom left diagonal showed consistently high densities two of the divergent lineages from this main line resulted in differential effects. In the first (both directions) the bacteriophage densities fall below that of the population which initially invaded the system. That being said, in the second and third branches the bacteriophage densities never fell below 10⁷ PFU ml⁻¹.

For the 50 % ON population of figure 5.7 (c) vastly diverse bacteriophage population densities can be observed. Although this population was found to undergo extinction by cycle 16 using the fixed cycling pattern of the first oscillating prey assay of section 5.4, here only one lineage appeared to undergo extinction, that being the main vertical lineage. However, although only one lineage appeared to produce an extinction event, a larger number of below inoculum, and densities similar to the densities of the initial inoculum, were observed. Furthermore, some lineages can be seen to have thrived in these conditions. For instance the top right diagonal, and a number of its descendant lineages can be seen to have achieved densities as high as $10^8/10^9$ over their 20 cycles. This is due to this direction having a high number of sensitive populations for replication in, and thus its success is through chance.

The 50 % OFF population structure (Fig. 5.7 (d)) is the reverse of the 50 % ON population. Thus each well that received an inoculum of Rd 30S in the 50 % ON assay now received Rd 30R. For instance the top right diagonal now contained a large number of Rd 30R wells and hence extinction events occurred in high numbers along the lineages from both top diagonal progressions. For all other lineages bacteriophage were consistently present, all be it at lower densities than those seen in the 50 % ON population. The bacteriophage densities in for instance the central vertical lineages were similar to the initial inoculum in a number of nodes by the end. This is likely due to the effect of the central well. In the 50 % OFF population (and the other majority Rd 30R populations) the initial well contains Rd 30R, and as such this population will not gain an initial boost to bacteriophage population

densities that occurred in the more Rd 30S population structures.

For the 66 % OFF population there was an even more restricted pattern of bacteriophage population densities (Fig. 5.7 (e)). Here the top diagonal lineages have dropped consistently in density to $\leq 10^2$ PFU ml⁻¹ for a number of transfers, and thus are assumed to have undergone extinction. The main lower right diagonal lineage, one of its offspring lineages, and the right horizontal lineages, are the only lineages that maintained bacteriophage densities above 10^2 PFU ml throughout the length of their stems. All other lineages showed $\leq 10^2$ PFU ml⁻¹ densities at the end of their lines of propagation.

Overall from this assay it can be observed that heterogeneous populations of bacteriophage resistance phenotypes generated as a consequence of phase variation may play a role in generation of diverse bacteriophage population densities within an area being invaded by bacteriophage. Furthermore, if each well in this assay is taken as a either an individual microcolony of bacterial cells as one of a series of colonies covering the mucosal epithelia of an infected individual, with lineages indicating forward expansion of bacteriophage in these populations; or, if each well is imagined to be a single person colonised with a particular dominant phenotype of H. influenzae within a larger group of individuals, and the lineages represent person-person transmission of the bacteriophage outward from the central point, then it could be argued that phase variation can generate distinct macropopulations for bacteriophage densities within these populations.

The assays so far have focused on the effect of phase variation on the densities and propagation of bacteriophage. It was next aimed to determine what effect population phenotype mixing can have on bacterial survival to varying degrees of selection by these viral predators.

5.9 Determination of whether *lic2A* phase variation generated population heterogeneity aids in host survival against bacteriophage HP1c1

As the mathematical and empirical simulations of bacteriophage spread through heterogeneous H. infleunzae populations showed that bacteriophage spread was readily reduced in the heterogenous environments compared to replication through sensitive populations it was then decided to investigate what effect bacterial *lic2A* expression population heterogeneity would have on the survival of H. influenzae, using single bacterial populations.

To test this growth curves were conducted for discreet population structures for *lic2A* expression. The population structures were created by mixing exponentially growing cultures of *H. influenzae* Rd30S and Rd30R at differing ratios in fresh sBHI to give an initial total concentration of $\sim 1 \times 10^5$ CFU ml⁻¹.



Figure 5.8. Bacterial growth curves in the presence of various bacteriophage HP1c1 multiplicities for H. influenzae populations composed of distinct degrees of lic2A phenotypes. Each graph shows bacterial densities (OD₆₀₀) for one set of heterogeneous populations for lic2A expression when exposed to bacteriophage at differing MOIs. Individual lines represent each of the population structures tested. Graphs are as follows (a) MOI 0, (b) MOI 0.01, (c) MOI 0.1 (d) MOI 1, (e) MOI 10, and (f) MOI 100. N = 5.

Each bacterial population structure was tested using six replicate populations that each received a different initial MOI of bacteriophage HP1c1 to establish if increasing bacteriophage selective pressure on these *H. influenzae* population structures altered the survival of the *H. influenzae* populations.

The effects of bacteriophage predation were only identifiable following incubation of *H. influenzae* with bacteriophage at MOIs of 0.1 and above (fig. 5.8). A reduction in growth was observed in the 100 % *lic2A* ON population as compared to the other population structures exposed to bacteriophage at an MOI of 0.1 but was not statistically significant due to a high level of variance between replicates with this population structure (two-way ANOVA with Tukey's multiple comparison test: p > 0.05; at all time points, between 100 % ON populations and all other population structures; Figure 5.8 (a)).

All other population structures did show statistically significant differences in growth when exposed to MOI 1 or greater. For growth in the presence of bacteriophage at an MOI of 1, the 100 % sensitive (100 % *lic2A* ON) population showed significant reduction in growth in comparison to all other population structures from ~7 hours until ~11.5 hours (e.g. at 10 hours; two-way ANOVA with Tukey's multiple comparison test: p < 0.05; mean OD₆₀₀ values for 100 % ON *lic2A* population = 0.087; OD₆₀₀ for 100 % OFF *lic2A* population at 10 hours = 0.20; Figure 5.8 (b)). Thus all population structures were significantly less sensitive to bacteriophage when exposed to bacteriophage at an MOI of 1 than the completely sensitive population. No significant differences were observed between the heterogeneous populations and the 100 % resistant population.

Significant differences in growth was also observed between the 100 % *lic2A* ON population and all other *H. influenzae* populations following exposure to bacteriophage HP1c1 at an MOI of 10. Highly significant levels of growth supression in the 100 % *lic2A* ON culture compared to all other population *lic2A* structures was observed between ~6 and ~16 hours (e.g. at 10 hours; two-way ANOVA with Tukey's multiple comparison test: p < 0.001; mean OD₆₀₀ values for 100 % ON *lic2A* population = 0.05; OD₆₀₀ for 100 % OFF *lic2A* population = 0.21; Figure 5.8 (e)). Growth reductions were observed in the 75 % ON population compared to the 100 % OFF *lic2A* population (e.g. at 8 hours; two-way ANOVA with Tukey's multiple comparison test: p < 0.05; mean OD₆₀₀ values for 75 % ON *lic2A* population = 0.17; Figure 5.8 (e)). Additionally during this time the 75 % *lic2A* ON population displayed a significantly reduced density in comparison to other mixed *lic2A* phenotype population structures (i.e. the 25 % *lic2A* ON population and the 33 % ON population). This signifies that although the 75 % ON population is less sensitive than the 100 % ON population it is not as resistant as the more Rd 30R populations.

When the *H. influenzae* populations were exposed to an MOI of 100, highly significant suppression of growth in the 100% *lic2A* ON population compared to the resistant all population structures (e.g. at 9 hours; two-way ANOVA with Tukey's multiple comparison test: p < 0.0001; mean OD₆₀₀ values for 100 % ON *lic2A* population = 0.03; OD₆₀₀ for 100 % OFF *lic2A* population = 0.21; Figure 5.8 (f)).

In a number of the other populations (75 %, 66 %, 50 % and 25 % ON populations) significant growth reductions were noted also in comparison to the 100 % OFF population, in a *lic2A* ON proportion dependant manner. The 33 % ON population, although being lower in mean OD at the same time did not show a statistically significant difference compared to the 100 % OFF population. The 75 %, 66 %, and 50 % ON populations also showed a significant suppression in growth compared to the other heterogeneous populations, although no significant differences at any time were observed between the 66 % population when compared to either the 75 % ON population of the 50 % ON population.

From these population structures it can be noted that across the MOIs where bacteriophage predation caused a notable suppression in bacterial population densities, the heterogeneous populations showed a far lower degree of inhibition of growth. Furthermore, in some of the mixed phenotype populations no difference in population densities were observed in comparison to the densities of the 100 % *lic2A* OFF (and thus absolutely resistant to bacteriophage predation) population structure. This thus implies that phase variation can play a key role in maintaining bacterial population densities even when exposed to exceedingly high bacteriophage densities.

5.10 Discussion

Phase variation of the *H. influenzae lic2A* gene is associated with alterations in the sensitivity/resistance of the bacterium to bacteriophage HP1c1 infection and hence is likely to have strong ramifications for the bacteriophage population. The effect of phase variation on the propagation dynamics of bacteriophage HP1c1 was investigated using both empirical data, and data obtained following the generation of a mathematical model of bacteriophage propagation through phase variable populations.

First, the dynamics of bacteriophage replication in each of the *lic2A* phasotype populations was studied using the *lic2A* ON phasotype, *H. influenzae* Rd 30S, and the *lic2A* OFF phaseotype, *H. influenzae* Rd 30R. As the Rd 30R population is unable to produce plaques upon infection with HP1c1 on solid media (see section 4.9) one would expect a similar scenario in liquid media. However, replication of bacteriophage HP1c1 in both the Rd 30S and Rd 30R populations in liquid culture bacteriophage infection assay was observed. This is obviously counter intuitive as the Rd 30R population does not express the Gal*p*-(1-4)- β -D-Gal extension, which is encoded by *lic2A*, that facilitates bacteriophage HP1c1 adsorption. However Rd 30R (and Rd 30S) are natural phase variants, thus the 5'-CAAT-3' repeat tracts within *lic2A* will still enable this gene to undergo phase variation. This will mean that an Rd 30R population will contain a small number of *lic2A* ON cells, i.e. phenotypically Rd 30S. The observation of bacteriophage replication occurring within the Rd 30R in identifiable numbers in the homogeneous environment of a liquid suspension will therefore likely be due to this small number of *lic2A* ON cells. Assuming that the number of bacteriophage particles produced following replication in each of the populations is proportional to the number of cells initially bound by the bacteriophage population, and that the number of bound particles is reflective of the overall density of sensitive cells within the bacterial population, as the Rd 30S population produced $\sim 10^6$ PFU ml⁻¹ bacteriophage and the Rd 30R population produced $\sim 10^4$ PFU ml⁻¹ this would imply that there is 100-fold lower number of sensitive cells in the Rd 30R population than the Rd 30S. As both bacterial populations contained the same density of cells ($\sim 1 \times 10^8$ CFU ml⁻¹), it can therefore be assumed that for the Rd 30R population there is $\sim 1 \times 10^6$ CFU ml⁻¹ sensitive cells (i.e. $\sim 1 \%$ of the total density). Although this small fraction is accessible in liquid culture, in contrast the structured environment of a petri dish will restrict bacteriophage spread to any sub populations of Rd 30S cells, meaning production of no visible plaques. Both of these experimental environmental structures have comparable biologically relevant environments, e.g. the spatially structured environments of soil, biofilms, or in environments of secreted mucus; or the more homogeneous environments of lakes, saliva etc.

From the oscillating prey assay it was possible to conduct preliminary simulations of the bacteriophage propagation through populations heterogeneous for lic2A expression. From this it was observed a high proportion of lic2A OFF populations removes bacteriophage rapidly from the environment due to the inhibition of spread. Furthermore, with bacteriophage propagation though a majority lic2A ON population (66 % lic2A ON) bacteriophage propagation could be limited, ending the 20 cycle experiment at a significantly lower density than that of the initial inoculum. The reductions in bacteriophage density will then have ramifications on the onward spread of bacteriophage as, with high numbers, bacteriophage will have an increased chance of survival if faced by any reduction in numbers that occur in the migration process, or due to any antagonistic factors in the new niche. In contrast lower density populations will have lower chances of survival as large reductions in their numbers during migration etc will more regularly lead to extinction events.

These observations were then expanded by generating a mathematical model from the experimental data on bacteriophage replication through heterogeneous populations. This model was able to replicate the empirical observations. Through modelling it was possible to determine the effects of a number of variables that could not be tested with such breadth in the laboratory. It was observed that relatively low rates of loss of bacteriophage from host populations were able to reduce the chances of bacteriophage populations survival even when subsets of *lic2A* ON cells were present within the population. With higher rates of bacteriophage loss from the population, this same level of reduction in bacteriophage survival ratio was achieved even when the bacterial population constituted of large numbers of *lic2A* ON cells with only few resistant cells.

Phase variation of bacteriophage resistance in the form of altering lic2A expression thus may allow the bacterial population to control bacteriophage densities through altering the composition of their population in response to bacteriophage densities. An interesting example of this has been shown in
S. enterica.

In S. enterica the LPS biosynthesis associated inner membrane proteins OpvA and OpvB are subject to phase variation controlled by both Dam methylation and the transcriptional regulator OxyR (Cota et al., 2012). Four 5'-GATC-3' sites are located upstream of the opvAB operon (Cota et al., 2012). In the OFF state sites 2 and 4 are methylated, while in the ON state sites 1 and 3 are methylated (Cota et al., 2016). Differential methylation is believed to occur due to, OxyR binding sites which block 5'-GATC-5' sites 1 and 3, while for the ON state OxyR binds sites that block 5'-GATC-5' sites 2 and 4 (Cota et al., 2016). The latter binding pattern allows RNA polymerase transcription of the opvAB operon leading to expression of OpvA and OpvB. The sites bound by OxyR to produce the ON methylation pattern do not perfectly match the sequence for OxyR binding, while the site bound to produce the OFF phase methylation pattern do. This is hypothesised then to contribute to the observed ~600 fold lower rate of OFF to ON switching rate compared to the ON-OFF rate (Cota et al., 2016).

In S. enterica populations < 1 % of cells are found with methylation of the *opvA* operon in the correct pattern to allow expression of OpvA and OpvB (Cota et al., 2016). When this operon is in the ON state the O-antigen chains of the bacterial LPS are truncated compared to that of the OFF cells (Cota et al., 2012). When these shorter chain lengths are expressed the bacterial population is now resistant to challenge by bacteriophage from the families Podoviridae (P22), Myoviridae (Det7), and Siphoviridae (9NA) (Cota et al., 2015). Cota et al. (2015) observed that the phase variable WT population of S. enterica was able to survive challenge by each of these bacteriophage families to a greater extent than $\Delta opvAB$ strains. This wild-type population however, had greatly reduced growth while under predation in comparison to a fixed ON opvAB strain. This is significant as the low rate of OFF to ON and high ON-OFF switching would mean that the initial bacterial inoculum would contain only a minute number of opvAB ON strains in comparison to opvAB OFF.

Based on the oscillating prey assay and mathematical modelling of bacteriophage propagation, should the *H. influenzae* population adjust its population structure to a more *lic2A* OFF population, which, depending on the rate of loss of bacteriophage from the environment would prevent bacteriophage propagation and ultimately reduce bacteriophage densities to a level at which a lower selection pressure is imposed on the bacterial population for resistance. This could then lead to a re-expansion in numbers of the *lic2A* ON population. This was observed in the study by Cota et al. (2015), as following removal of members of the population that survived bacteriophage challenge to a new environment free of bacteriophage an increase in the prevalence of the *opvAB* OFF population was then observed.

In the case of *lic2A*, simply removing the bacteriophage pressure will not drive switching of bacterial populations into the ON state of *lic2A*, as phase variation from OFF to ON occurs at a slower rate than the switching from ON to OFF (Dixon et al., 2007). For phasotypes to revert in significant numbers

a counter selective pressure is required. In the case of lic2A this selection pressure would likely be in response to the human hosts immune system. lic2A is known to provide the bacterial population with increased resistance to the bactericidal action of anti-LPS antibodies, increasing host survival from ~25 % to ~50 % when only the Galp-(1-4)- β -D-Glcp extension is expressed, while if lgtC is also expressed the α -Gal-(1-4)- β -Gal extension, which requires *lic2A*, can increase survival to ~80 % (Clark et al., 2013). Thus, in the human host, it could be postulated that a bacteriophage population could select for an overly resistant population that would prevent bacteriophage propagation and result in a reduction in their density. The reduction in bacteriophage densities will then reduce the selection pressure on the bacterial population to be composed of mainly *lic2A* OFF cells, and should this pressure fall below that of the host immune response then a resurgence in the number of lic2AON cells may then be seen. In support of this point, Clark et al. (2013) observed that from an initial inoculum of *H. influenzae* Rd with *lic2A* in the OFF state, serial passage (they conducted 4 passages) of the bacterial population through human sera could resulted in low levels of bacterial survival until a gradual increase in survival occurred after the second round of passage, with the highest level of survival observed in the final (fourth) passage. Screening of the phase of 10 phase variable LOS genes showed that the only gene which had altered its phase was *lic2A*, switching from OFF to ON, indicating that the subpopulation of *lic2A* ON cells present in the original majority *lic2A* OFF inocula had facilitated the populations survival. This same switching of lic2A to the ON state was observed in three other H. influenzae strain backgrounds (Clark et al., 2013).

This coincides well with observations of other bacteriophage resistance mechanisms, where development of resistance to viral predation comes at a cost to the resistant bacterial population compared to their sensitive counterparts. This can be in factors such as their virulence (León and Bastías, 2015), or in overal growth deficits in comparison to the sensitive populations (Koskella et al., 2012). For the current work this cost would likely be a reduced ability to colonise/flourish in a niche where an immune selection is applied. In the case of the example of *S. enterica* and *opvAB* expression, fixed ON *opvAB* strains have been shown to have reduced survival following exposure to guinea pig sera, and a reduced ability to replicate within macrophage (Cota et al., 2012).

However in the current study it was observed that if phase variation were to allow the bacterial population a means of adjusting their population structure to control the level of selection pressure with which they are faced, the outcome population structure would depend heavily on what rate the bacteriophage were being remove from the system. In natural environments bacteriophage can be lost easily through factors such as exposure to ultra-violet (UV) light (Iriarte et al., 2007), clearance and expulsion from the macrohosts (Golshahi et al., 2008), or the response of the macrohosts immune system (Lusiak-Szelachowska et al., 2014). Bacteriophage themselves show differential stability, some bacteriophage showing stable levels of survival upon storage for months, while others show instability on a daily basis (Clokie et al., 2011; Jończyk et al., 2011).

In terms of bacteriophage loss in biological systems, bacteriophage of *Bacillus megaterium*, when introduced into the blood stream of dogs, was found to be initially rapidly eliminated over the first 10 minutes (bacteriophage densities fell from $\sim 3 \ge 10^7$ PFU ml⁻¹ to $\sim 8 \ge 10^5$ PFU ml⁻¹) following injection. After which point the bacteriophage rate of loss decreased to a more stable level, resulting in a 3 fold reduction in bacteriophage over the remaining 80 minutes (Keller and Zaltzman, 1959). Movement between systems will generate greater losses of bacteriophage densities. For instance in airborne transmission bacteriophage must contend with environmental conditions such as humidity, salinity, and temperature, all of which can have serious impact on the viability rates of bacteriophage(Ehrlich et al., 1964; Dubovi and Akers, 1970; Trouwborst and de Jong, 1973). Even if the bacteriophage population is able to survive the transmission process it will also have needed to reach a niche which is inhabited by a susceptible bacterial population.

If the bacterial population being discussed is that of microcolonies within a biological system then, the rates of bacteriophage loss will likely be relatively low. Based on the observations from the oscillating prey assay and mathematical modelling a lower rate of bacteriophage loss required a large number of cells resistant to viral attack to reduce bacteriophage densities. If the the potential effects are considered on a wider scale, say transmission from person to person with each individual, then the ON state could flourish in each individual due to a higher rate of loss of bacteriophage. It should however be remembered that in *H. influenzae* ON-OFF switching occurs at a higher rate than OFF-ON, and as such OFF populations would eventually dominate without exposure to a selective pressure requiring the *lic2A* ON state. Colonisation studies have shown *lic2A* expression is not readily selected for (Poole et al., 2013), and factors such as the prevalence of the OFF state in healthy individuals may contribute to the transient nature of identifying bacteriophage species, as bacteriophage are not ubiquitously identified in individual when screened for bacteriophage, and nor at high densities.

To extend the observations of this current study on bacteriophage propagation through heterogeneous bacterial populations a multi-directional oscillating prey assay, across a fixed area from a central point, was conducted. From this, heterogeneity for bacterial *lic2A* expression also caused heterogeneity in the bacteriophage densities measured in each of the sampled nodes. Thus the resistant wells could effectively act as barriers to bacteriophage spread. For instance exposure of HP1c1 to a series of Rd 30R containing wells will reduce bacteriophage densities, and if this decrease lead to an extinction event then the bacterial populations beyond this point would now have been saved from invasion by bacteriophage was rare in majority sensitive heterogenous populations, instances of vastly reduced densities compared to the 100 % *lic2A* population structure could be observed. These reductions in bacteriophage densities would prevent the bacteriophage from causing suppression of bacterial populations, as their densities will fall below that of the bacteriophage inundation threshold. The inundation threshold is the density of bacteriophage required before a notable effect can be seen on the bacterial populations density

(Cairns et al., 2009). Examples of the occurrence of bacteriophage achieving the inundation threshold can be seen in the host survival assays.

With the 100 % ON simulation the vast, uncontrolled propagation of bacteriophage HP1c1 was noted. Thus without bacterial adaptations such events would ultimately result in the bacterial populations extinction. Likewise in the 100 % OFF simulation bacteriophage underwent rapid extinction, which, without either a mutation within the bacteriophage population, or alleviation of the bacterial populations resistance then the bacteriophage population would rapidly undergo extinction. However, with a fixed OFF *lic2A* population this would mean that the bacterial population would never be able to benefit from any transduction events mediated by the bacteriophage nor utilise its receptor for immune evasion. In other bacteriophage-host systems, spatial structures have been shown to create more stable environments for co-existence. This has been shown in *P. aeruqinosa* for interaction with its bacteriophage PP7 in unmixed cultures (Brockhurst et al., 2006). While in eukaryotic systems, such as the parasitic wasp Nasonia vitripennis and its prey, the housefly Musca domestica and the common green bottle fly *Phaenicia sericata*, periods of stable coexistence of fly and wasp over time could be seen when physical barriers were implemented into the experimental setting. When wasp and fly were introduced into a system of interconnected cells, extinction of the parasite population was observed, yet introduction of a series of baffles to the system resulted in reduced dispersal of the parasite progeny and a more stable coexistance of the parasite and prev population (Pimentel et al., 1963). Thus the action of phase variation within bacterial populations may be providing a similar, adjustable, barrier which will ultimately have allowed the long term interaction of the bacterial and bacteriophage populations seen in nature.

The barricade created by the resistant populations generated distinct macropopulations of bacteriophage densities. This generation of varied bacteriophage densities across the bacterial macropopulation may have significant ramifications on the evolution of the *H. influenzae* population. Bacteriophage are strong drivers of bacterial evolution (Scanlan and Buckling, 2012). Thus through exposure to different selection pressures the need for bacterial populations to adapt will also be different. This has been shown in *Caulobacter crescentus* in response to varying degrees of selection by its bacteriophage ϕ CbK. When populations of transposon mutants of *C. crescentus* were exposed to varying MOIs of ϕ CbK the resulting mutations able to give resistance were likewise varied. When bacteriophage exposure was at an MOI of 0.1, 208 bacteriophage susceptibility genes were identified in the host, while when bacteriophage are added at more extreme densities (MOI 1000) only 58 genes were identified to provide resistance upon interruption by the transposon. Thus the potential diversity in bacterial phenotypes in response to bacteriophage challenge can be readily modulated by the bacteriophage densities faced (Christen et al., 2016).

In summary, bacterial populations must regularly contend with volatile environmental conditions to

survive. To combat these pressures genetic mechanisms have evolved in bacteria, such as phase variation, that provide a means of 'molecular bet-hedging', genetically pre-empting the population with contingencies for times of adversity. Phase variation is believed to have arisen due to unpredictable oscillatory environmental selection pressures, such as the presence of the viral predator of prokaryotes, the bacteriophage. Here it was shown that phase variation of one loci, lic2A, may have arisen as a means of controlling the densities and dissemination of bacteriophage through bacterial populations. It was found experimentally and mathematically that the population heterogeneity generated by phase variation can alter the densities, and expanse reached by bacteriophage populations. It was noted that, through phase variation induced population heterogeneity, bacterial populations can generate distinct macropopulations of bacteriophage densities, which may play a role in the long-term co-existence and evolutionary relationship between both bacteria and bacteriophage. The results presented herein indicate that fluctuating viral predation pressures could be a significant driver of evolution of phase variable contingency loci in *H. influenzae* and other bacterial species.

6. Concluding discussion

The respiratory tract associated bacterium H. influenzae contains a number of genes that undergo stochastic, reversible, expression changes at high frequency, a process known as phase variation (Bayliss et al., 2001). This allows a single H. influenzae cell the potential, over time, to generate lineages that lead to a population composed of a vast degree of diversity in gene expression. This diversity, although generated through the stochastic switching in gene expression, can be moulded by the demands of various selection pressures, both environmental and biotic (van der Woude and Bäumler, 2004). One such biotic pressure is predation by the viral predators of prokaryotes, the bacteriophage.

Bacteriophage predation will impose pressure on the bacterial populations to adapt to resist the viral attack (Buckling and Rainey, 2002a). These subversion mechanisms can be in the form of a number of mechanisms, such as internal mechanisms like restriction modification systems, CRISPR-cas systems, or abortive infection systems; or through changes to external surface, including the production of exopolysaccharide to block bacteriophage adsorption, or through the alteration or removal of bacterio-phage receptors to completely prevent the adsorption process (Labrie et al., 2010). In this current work two such mechanisms that have been observed in *H. influenzae* were investigated, the gene encoding the methyltransferase, *hsdM*, of a phase variable Type I restriction modification system, HindI, and a gene encoding a UDP-galactose-LOS-galactosyltransferases involved in LOS biosynthesis, *lic2A*. Both of these genes are known to undergo phase variation (Glover and Piekarowicz, 1972; High et al., 1993), and phase variable switching between resistance and sensitivity to bacteriophage has been observed (Zaleski et al., 2005).

In the present study, although the resistance phenotype was observed in phase variants of lic2A, no resistance was observed in *H. influenzae* RM118 isolates used here. The RM118 phasotypes used in this study, RM118 $_{r^+HinDI}$ m⁺HinDI and RM118 $_{r^-HinDI}$ contained 4x and 5x 5'-GAGAC-3' repeats respectively. These are the same numbers of repeats found in the *hsdM* phase variants of Zaleski et al. (2005) and thus would be expected to show a resistant and sensitive phenotype respectively upon bacteriophage HP1c1 challenge. However, it was found upon analysis of the reading frames associated with the *hsdM* gene of each of these repeat tracts that, with 4x and 5x 5'-GAGAC-3' repeats, *hsdM* would produce a truncated product. Furthermore, neither of these phasotypes were able to cause a

reduction in the efficiency of plating of bacteriophage HP1c1 when tested. Further analysis revealed that a 3x 5'-GAGAC-3' would produce a full length polypeptide product, and while work is still ongoing to generate a strains with this SSR number, it still remains uncertain why the 4x 5'-GAGAC-3' repeat conferred resistance in the work of Zaleski et al. (2005) and not here.

Although no resistance phenotype was observed for hsdM phase variants, the confirmation of lic2A attributing bacteriophage resistance adds to a list of phase variable bacteriophage resistance mechanisms that have been identified across a range of bacterial species (see table 1.2). Phase variation of bacteriophage resistance mechanisms allows the bacterial population to consistently generate a small number of resistant variants even when the population is free of bacteriophage predation (e.g. < 1 % of the population expressing *opvAB* in *S. enterica* (Cota et al., 2016)). Thus when the bacterial population is then invaded by bacteriophage this resistant population will increase in proportion compared to the sensitive members as the bacteriophage population select against the sensitive cells, allowing cells with the resistant phenotype to flourish. Due to the reversible nature of phase variation, cells with this resistant phenotype will be continually generating sensitive cells. This can be useful when the selection pressure by the bacteriophage is removed, as bacteriophage resistance mechanisms are often associated with a cost to host fitness in comparison to cells with the sensitive phenotype (Lennon et al., 2007; Koskella et al., 2012; Scanlan et al., 2015).

For hsdM and lic2A, it was noted that, after screening of all H. influenzae genomes within the GenBank database, the majority of strains were expressing lic2A while hsdM tended to contain a number of repeats that would put it in the OFF frame. Both of these phenotypes would confer a more sensitive phenotype than if expression of each was in the opposing state. Thus in the range of strains available in the GenBank database a bacteriophage sensitive phenotype may be preferred. This would correlate well with there being a cost to these genes being in the phase which would provide bacteriophage resistance. For lic2A the cost may be in line with the host immune system, as lic2A is believed to play a role in immune resistance (Weiser and Pan, 1998; Clark et al., 2013). Thus, as the only known niche for H. influenzae is the human host, anti-immune responses will be readily required. For hsdM no benefits (except potential bacteriophage resistance) or costs to the fitness of the host are known, however expression of other Type I restriction systems have been known to reduce the ability of bacterial populations to colonise their host environment (Gauntlett et al., 2014).

Within *H. influenzae* expression of the phase variable modA gene of its Type III restriction modification system has been shown to come at a cost. Atack et al. (2015) observed increased sensitivity to opsonophagocytic killing than when the cells are in the ON phase for cells with the modA4 allele. Conversely, benefits to expression of this system have been observed. Atack et al. (2015) also observed differential changes to the susceptibility of cells to certain antibiotics dependent on which modA allele was being expressed (i.e. which specificity domain the modA gene had). The above is believed to be due to changes in the 'phasevarion' (Srikhanta et al., 2005), the phase variable regulation of gene expression through changes in chromosomal methylation patterns, altering gene expression across the genome of a cell following phase switching of the modA gene.

However, there is not a general conclusion that expression of restriction modification systems in the absence of bacteriophage constitutively produce negative effects on host fitness. For instance when modA2 allele was in the ON state cells were ~2-fold more sensitive to ampicillin, while when cells with the modA5 allele had their restriction system in the ON state cells were 2-fold less sensitive to erythromycin compared to those with modA5 in the OFF phase. Similarly a 2-fold decrease in gentamicin sensitivity was also observed when a strain with the modA10 allele was in the ON phase. Thus there are variable benefits and costs associated with phase variation of restriction modification systems. With the GenBank genomes analysed in the current study showing the hsdM in the OFF state it may therefore be that there is a cost to its expression, and that its benefit to the cell is transient.

Which loci are affected by epigenetic regulation from the restriction modification systems are determined by the target site of that cognate restriction modification system. For the modA gene, 21 modAalleles have been observed (Atack et al., 2015). For the H. influenzae phase variable HindI restriction modification in the current study 40 hsdS alleles were identified. Although it is likely a number of these alleles overlap in target sequence (those with identical variable domains with the only notable differences occurring within the central conserved domain) the presence of multiple hsdS alleles across H. influenzae species could mean that there is a large diversity in the putative phasevarions of H. influenzae from the HindI restriction modification system. One allele that may be of interest for further study is hsdS32, an allele identified in ~ 11 % of the *H. influenzae* strains investigated. Each strain containing a HindI Type I restriction modification system with this allele was observed to have hsdMin the ON state (as was the hsdR gene, implying a full restriction modification system). Thus with this hsdS allele found persistently ON, it may be playing a key role in the fitness of the strains containing the hsdS32 allele, especially considering that the *H. influenzae* strains within GenBank tend to have this system in the OFF state. Strains with the hsdS32 allele were isolated from individuals with a range of disease conditions (where host condition was known), namely bacteraemia, meningitis, and otitis media.

Although no link was found between ON/OFF state for hsdM expression and disease, for lic2A there did appear to be a greater number of lic2A ON isolates compared to OFF isolates in individuals with bacteraemia and meningitis. These states are overlapping as meningitis often occurs as a result of bacteraemia (Nassif et al., 2002; Kim, 2003), a similar lic2A phase for each of these conditions would be expected. As bacteraemia is by definition the presence of bacteria within blood, these cells will thus have to face a strong immune challenge (Smith, 1987). The cells that are able to invade this niche

will therefore likely only be those with a high level of resistance to the immune challenge. As cells expressing *lic2A* have increased resistance to human sera (Weiser and Pan, 1998; Clark et al., 2013) these cells may be more readily able to migrate into the blood, and then on to cause meningitis.

In this study, while the ON state predominated across the GenBank isolates for lic_{2A} , and the OFF state predominated for the hsdM gene, when sputum supernatants from individuals with the respiratory disease COPD were analysed it was observed that both of these genes may be predominantly in the ON state. However, these observations did come with a caveat in that, due to sequence variations either side of the hsdM gene, and sequence disparities at the 3' end of the lic2A SSR in a small number of strains, discrepancies may exist in some of the samples between SSR lengths in the assay and their actual sizes. A more focused study should therefore be applied in order to determine if this is in fact a true trend, similar to the cohort studies that have allowed the study of phase variable loci in N. meningitidis from individuals (Alamro et al., 2014). However, if the current observations were then reflected in such a study this would be of major interest, as *H. influenzae* is a significant aetiological agent associated with this respiratory condition (Finney and Ritchie, 2014). In terms of speculation towards the reason behind each of these genes being in the ON state in COPD, multiple reasons could be suggested. For lic2A its expression may be due to the presence of selection for the ON state in the form of immune attack. In COPD a immune responses to bacteria are readily observable, both in sera and the mucosal layer (Bakri et al., 2002). Furthermore, it is known that in individuals with COPD a large proportion of immune responses are directed towards *H. influenzae* (Groeneveld et al., 1990). Thus with the Galp-(1-4)- β -D-Glcp extension of the *H. influenzae* LOS, produced following *lic2A* expression, being associated with increased immunity to host defences (Clark et al., 2013), expression of *lic2A* may be required to increase the bacterial populations survival in the respiratory tract in the face of the ongoing immune challenge. Additionally the α -Gal-(1-4)- β -Gal structure that is associated with immune resistance (Weiser and Pan, 1998), produced if lqtC is also phase ON, can increase the resistance of the cells to immune challenge to a higher degree than solely *lic2A* expression (Clark et al., 2013). The α -Gal-(1-4)- β -Gal again requires *lic2A* expression, however investigation of ON/OFF state of *lqtC* could be investigated in future to determine if it is the single, or di-galactoside epitope expressed in these cells. Likewise, with hsdM in the ON state the expression of hsdM may be playing a role in increasing the bacterial populations survival in this niche through alterations in the phasevarion that in-turn alters the expression of unknown gene(s) that increase bacterial fitness in this niche.

Alternatively bacteriophage predation could be playing a role in the switching of *hsdM*. With the respiratory tract of individuals with COPD characterised by a high level of mucus hypersecretion, and bacteriophage hypothesised to become embedded in mucosal layers (Barr et al., 2013) (which also can promote bacteria/bacteriophage interactions (Barr et al., 2015)) this may mean that bacteriophage resistance mechanisms are required in response. However as the expression of the majority of samples

tested were ON for lic2A (thus bacteriophage would be able to bind) this selection pressure may not be particularly high in comparison to the immune system selection. Lower immune selection pressures would likely increase the proportion of *lic2A* based resistance phenotypes. In terms of crossover, i.e. the number of samples where lic_{2A} was in a putative ON state, as was hsdM, this only occurred in 14/44 sputum supernatant samples (31.8 % of samples). This does however rely on the observations from the genecan assay for both genes, and as the exact figure must be viewed with caution. It could realistically be the case that bacteriophage invasion in COPD is transient, and as such the expression of hsdM is likewise transient. In the similar respiratory condition of cystic fibrosis researchers have detected a large diversity of bacteriophage within sputum samples (compared to non-cystic fibrosis individuals) with these bacteriophage targeting a number of bacterial species, larger than that seen in healthy individuals citepWillner2009. Thus as the presence of bacteriophage within the respiratory tract of individuals with cystic fibrosis differ between individuals this may also be the case for the COPD sample of this study, leading to only certain samples being ON for hsdM. Alternatively, other mechanisms may be providing bacteriophage resistance in the hsdM OFF samples that were lic2AON, such as the Type III restriction modification system. As such the hsdM gene was not selected for, with the alternative resistance mechanism providing a sufficient level of resistance.

If the observations from this study were confirmed, this would be promising from a therapeutic standpoint. If a bacteriophage therapeutic was given, this hypothetical cocktail could target *lic2A*, then even if resistant isolates do arise their ability to colonise the respiratory tract may be severely attenuated due to the loss of *lic2A* expression in response as a means of resistance. Loss of *lic2A* will increase the cells sensitivity to the hosts immune system, which based on the large number of *lic2A* ON samples from the COPD samples may be under selection to be maintained. Furthermore, if administered at high MOI then replication of the bacteriophage would not needed for lysis to occur, in a phenomenon termed lysis from without (Abedon, 2011).

The successful application of a bacteriophage based therapeutic would rely on their being a cost in COPD associated with the loss of lic2A expression, such as the lic2A OFF cells being readily cleared by the host immune response. Successful therapy would likely required this as, in the current work, when *H. influenzae* populations were exposed to bacteriophage HP1c1, the populations heterogeneous for *lic2A* expression were significantly less suppressed than those populations only containing *lic2A* ON cells. From the empirical and mathematical simulations of bacteriophage propagation through heterogeneous populations in this current study it was observed that heterogeneous populations for *lic2A* expression could reduce bacteriophage propagation through these populations. The ratio of ON:OFF cells at which this was possible was dependent on the rate of loss of the bacteriophage from the population. In scenarios where the rate of bacteriophage loss from the environment is low then a greater level of switching from a *lic2A* ON population to a *lic2A* OFF population will be required to reduce bacteriophage less even majority ON populations

can survive an invasion by bacteriophage. Thus if a switch to a more lic2A OFF population can remove bacteriophage, a selection pressure targeting the lic2A OFF state will be required to effectively eliminate the bacterial population, which in COPD may be the immune response in the respiratory tract since the majority of samples tested were lic2A ON. The complementation of the applied bacteriophage and the patients own immune response may thus facilitate the reduction or elimination in bacterial densities.

From the observations in this study it appears phase variation of bacterial populations can not only provide bacteriophage resistance to the bacterium at the cellular level, but also at the population level. With the presence of resistant cells benefiting the survival chances of the sensitive cells. When a multi-directional propagation assay was conducted it was observed that this may be in part due to the resistant bacterial cells acting as a barrier between the bacteriophage and sensitive population that allow the bacteriophage numbers to expand. Through the resistant populations providing a period of bacteriophage density reduction, the bacteriophage load then faced by the sensitive population will be reduced, increasing their survival chances by facing a lower bacteriophage load that otherwise would have been higher. The introductions of barriers that can constrain predator migration through an environment have been shown in other predator-prey systems such as the introduction of baffles to the environment that reduces the dispersal of *Nasonia vitripennis*, allowing more stable co-existence with this parasitic wasp with its prey populations, the housefly *Musca domestica*, and the common green bottle fly *Phaenicia sericata* (Pimentel et al., 1963). Thus phase variation may serve the greater bacterial population in a similar way, effectively acting as a buffer between the sensitive cells and the propagating bacteriophage that allows a period in which bacteriophage densities are reduced.

The diversity in bacteriophage densities faced by the bacterial populations may result in diversity in the evolutionary trajectory of the bacterial populations. When introduced in similar densities to parallel bacterial cultures bacteriophage cause diverse and distinct adaptations in their bacterial prey (Marston et al., 2012). However, differing levels of bacteriophage selection have been shown to orchestrate differing evolutionary responses in the bacterial population under predation (Christen et al., 2016). As such, within a macropopulation, there will be a number of parallel discrete populations under bacteriophage predation, and if a number of these populations face different levels of bacteriophage predation, an expansive array of adapted phenotypes may be displayed within the macropopulation. Thus phase variation may also then be playing a role in the generation of greater diversity within bacterial populations beyond only the loci under phase variable control.

The evolution of phase variable loci is presumed to have arisen due to the requirement for bacterial populations to cope with uncertain environments where selection fluctuates between opposing selection pressures (Moxon et al., 1994). The genomes from GenBank indicate that lic2A is regularly expressed within *H. influenzae*, indicating its expression may be beneficial to survival within the human host.

Bacteriophage HP1c1 predation on the other hand targets cells expressing lic2A meaning loss of expression is key to permit survival. Thus both of these pressures could present two fluctuating selective pressures that may have contributed to the development of the SSR tract within the lic2A gene. This has provided the bacterial population with a means of adjusting the number of lic2A ON/OFF populations in line with which of these selection pressures is providing the greatest degree of pressure at the local level (i.e. only altering the expression of one gene). Alternative adaptations to selection pressures include global hypermutation phenotypes, and mutator phenotypes have been shown to occur as an adaptive measures to bacteriophage predation in other bacterial systems (Pal et al., 2007). In comparison to global hypermutation, development of local hypermutation will negate some of the negative effects associated with long term global hypermutation that occur due to accumulation of effectively neutral mutations when under selection, that then become costly when pressure is removed (Giraud et al., 2001). Thus, through the development of phase variation as a mutational mechanism to resist bacteriophage and at the same time allow populations to cope with opposing pressures, such as immune selection, *H. influenzae* appears to have developed a targeted adaptive response rather than a general response that can have a number of detrimental side-effects.

Furthermore this work may potentially provide a new perspective on phase variation of bacteriophage resistance mechanisms. Phase variation is usually discussed at the level of a single populations survival to a selection pressure. However, here evidence is shown of the effect that interconnecting single populations can have on bacteriophage expansion. That is to say that through the generation of diversity for lic2A expression within a larger macropopulation by phase variation, bacterial populations can generate the larger structured populations that can ultimately create refuges for other individual populations.

In summary, this study observed the diversity in SSR tract lengths within the hsdM and lic2A genes of H. influenzae. When these tract lengths were correlated to a putative expression state for these genes it was revealed that while lic2A was ON in the majority of H. influenzae strains analysed, potentially indicating a benefit to its expression in H. influenzae cells, the inverse observation was made for hsdM, and as such its expression may be more transient. Contrastingly a putative link was detected between the respiratory condition COPD and SSRs that correlate to the ON state for hsdM, a link that may warrant further investigation. Furthermore, one target recognition allele hsdS32 may be of particular interest as in the genomes containing this allele the hsdM gene was consistently in the ON state, in comparison to the otherwise majority OFF background of the other H. influenzae genomes. Another major finding came from the dynamics of bacteriophage spread through populations heterogeneous for lic2A expression it was observed that phase variation may be able to reduce bacteriophage dispersal through bacterial populations. Additionally phase variation may further provide a means by which bacterial populations can adapt their densities to a level which can facilitate the control of the bacteriophage densities. Through generation of heterogeneous populations through phase variation

bacterial populations may be able to further create diverse bacteriophage densities across the bacterial macropopulation, potentially eliciting differential adaptive responses in bacteria across the larger area, ultimately altering the evolutionary trajectory of populations sets within the greater macropopulation. In conclusion this current work has brought evidence that through generation of adjustable heterogeneous populations phase variation may provide a means for the bacterial population to control bacteriophage densities that facilitates the bacterial populations long-term survival.

6.1 Potential further work

The present work has raised a number of observations that warrant further investigation. These questions relate both to the role of phase variation of *hsdM* and *lic2A*. For instance, does the 3x 5'-GAGAC-3' provide a bacteriophage resistance phenotype? A question that can be answered upon successful counter-selection the cassette containing RM118 strain with the 3x 5'-GAGAC-3' containing pCT2.

Additionally, as screening of the hsdM gene in sputum supernatants from individuals with COPD showed that hsdM may contain an SSR tract of a length that would encode a functional hsdM gene, does hsdM play a role in the colonisation of individuals with COPD by H. influenzae? This question could be answered by analysis of H. influenzae sequence data either following a metagenomic study, or through isolation of bacteria colonising the respiratory tract of afflicted individuals. The latter approach would be preferred, as any strains isolated could then be phenotypically tested similar in manner to previous studies conducted for the elucidation of the phenotypic effects of expression of the Type III restriction modification system (Srikhanta et al., 2005; Atack et al., 2015; VanWagoner et al., 2016). This same question may also be asked of the lic2A gene, with analysis revealing that this gene similarly was also found to be in the ON state in the majority of COPD samples tested here. An analysis of bacteriophage in the respiratory tract of individuals with COPD could aid in elucidating if their involvement has any effect on the alteration in the hsdM gene of H. influenzae colonising such individuals.

With the development of phase variable loci within H influenzae potentially the result of the counter selection between immune pressure and bacteriophage pressure, an assay similar to the experimental evolution assays of Brockhurst et al. (2007) whereby bacteriophage and host are continually passaged through environments where immune selection fluctuates (allowing oscillations in selection for maintenance of *lic2A*) may be of interest to investigate if the counter selection of the immune system aids in a more stable co-existence of *H. influenzae* and bacteriophage as hypothesised from the multi-directional assay of the current study. This however would rely on either the isolation of an obligately lytic bacteriophage or mutation of bacteriophage HP1c1 to loose the ability to integrate into the bacterial chromosome, as lysogeny within the population would likely develop, resulting in a lic2A ON population that can be completely bacteriophage resistant through super-infection immunity (the *H. influenzae* RM118_{r⁻HinDI} m⁻HinDI lysogen containing HP1c1 as a prophage shows complete resistance to HP1c1 following exposure).

7. Supplementary Information

Media/Buffer/Reagent	Composition
name	-
Hemin	40 mg Hemin (Sigma)
	40 mg L-Histadine (Sigma)
	160 μ l Triethanolamine (Sigma)
	Nanopure water to 40 ml
	Dissolve at 70 °C for 20 mins
LB 1% Agar	10 g LB broth, Miller powder (Fisher)
	4 g Agar bacteriological (agar No. 1) (Oxoid)
	Distilled water to 400 ml
	Adjust pH to 7.5 with NaOH
	Sterilise by autoclave
LB Broth	10 g LB broth, Miller powder (Fisher)
	Distilled water to 400 ml
	Adjust pH to 7.5 with NaOH
	Sterilise by autoclave
Leventhals	400 ml BHI broth
	200 ml defibrinated horse blood (Sigma)
	$2 \ \mu g \ ml^{-1} \ NAD$
	heat at 95°C for 40 mins, centrifuge at 4696 x g for 20 mins, and
	retain supernatant

Table S1. Table of Media/Buffer/Reagents used in this study

Media/Buffer/Reagent name	Composition
NAD	10 mg Nicotinamide adenine dinucleotide (Roche)
	Nanopure water to 10 ml
	Sterilise by filtration
sBHI 1% agar	14.8 g Brain-heart infusion (BHI) (Oxoid)
	$4.4~{\rm g}$ Agar bacteriological (agar No. 1) (Oxoid)
	Distilled water to 400 ml
	$40 \text{ ml Leventhals}^*$
	800 μ l NAD*
	*Added after autoclaving
sBHI Broth	14.8 g BHI (Oxoid)
	Distilled water to 400 ml
	$2 \ \mu \text{g ml}^{-1} \ \text{NAD}^*$
	$10 \ \mu \text{g ml}^{-1} \ \text{Hemin}^*$
	\ast Added after autoclaving to aliquots of BHI broth only before use
SM Buffer	5.8 g NaCl (Fisher)
	2 g MgSO4·7H2O (Fisher)
	50ml 1M Tris HCl (pH7.5)
	Distilled water to 1 L
	Sterilise by autoclave
SOB	5 g Yeast Extract (Oxoid)
	20 g tryptone (Oxoid)
	0.584 g NaCl (Fisher)
	0.186 g KCl (Fisher)
	$1 \text{ ml sterile } 1 \text{ M MgCl}_2 \text{ (Fisher)}$
	$1 \text{ ml sterile } 1 \text{ M MgSO}_4 \text{ (Fisher)}$
	Distilled water to 1 L
	Adjusted to pH 7.5 with NaOH (Fisher)

 ${\bf Table \ S1.}\ - {\rm Continued \ from \ overleaf}$

Media/Buffer/Reagent name	Composition
	Sterilise by autoclave
SOC media	990 ml SOB 10 ml sterile 2 M Glucose (Fisher)
50 X TAE buffer	484 g Tris-base (Fisher)
	200 ml 0.5 M Ethylenediaminetetraacetic acid (EDTA) (Sigma)
	114.2 ml Glacial Acetic Acid (Fisher)
	Distilled water to 2 L
	Dilute 50 fold in distilled water before use

Table S1. – Continued from overleaf

Primer name	Sequence	Source
adk-up	5'-GGTGCACCGGGTGCAGGTAA-3'	Meats et al. 2003
adk-dn	5'-CCTAAGATTTATCTAACTC-3'	Meats et al. 2003
atpG-up	5'-ATGGCAGGTGCAAAAGAGAT-3'	Meats et al. 2003
atpG-dn	5'-TTGTACAACAGGCTTTTGCG-3'	Meats et al. 2003
frdB-up	5'-CTTATCGTTGGTCTTGCCGT-3'	Meats et al. 2003
frdB-dn	5'-TTGGCACTTTCCACTTTCC-3'	Meats et al. 2003
fucK-up	5'-ACCACTTTCGGCGTGGATGG-3'	Meats et al. 2003
fucK-dn	5'-AAGATTTCCCAGGTGCCAGA-3'	Meats et al. 2003
dn-ubm	5'-TCATTGTATGATATTGCCCC-3'	Meats et al. 2003
mdh-dn	5'-ACTTCTGTACCTGCATTTTG-3'	Meats et al. 2003

 Table S2. Other primers Used in this Study

Source	Meats et al. 2003	Meats et al. 2003	Meats et al. 2003	Meats et al. 2003	This study	This study	This study	This study	Turner et al. 1999	10099 1999	This study
Sequence	5'-GGTGAAAAATCAATCGTAC-3'	5'-ATTGAAAGACCAATAGCTGA-3'	5'-ATGGCAACTCAAGAAGAAA-3'	5'-TTACCAAACATCACGCCTAT-3'	5'-CGGCATTCTACTGGAAATCATTG-3'	5'-ATAGCGGCATCAAGTTGTT-3'	5'-CGGGCAACTGAACGTCGCAA-3'	5'-CGCTCTCCTTAGTAAGTATTC-3'	5'-AGAGTTTGATCCTGGCTCAG-3'	5'-GACGGGCGGTGTGTRCA-3'	5'-CAATCTTGGTCGTCTCGTCTCGCTTGCATAAATTATGC-3'
Primer Name	pgi-up	pgi-dn	recA-up	recA-dn	HsdM-CT-F2	HsdM-GS-R2	Lic2a-GS-F	Lic2a-GS-R	8F	1319R	HsdM-3xGAGAC-2-R-1

Table S2. - Continued from overleaf

Table S2. – Continued from overleaf		
Primer Name	Sequence	Source
HsdM-3xGAGAC-3-F-1	5'-TAATTTATGCAAGCGAGACGAGACGAGACGACGACCAAGATTGTCG-3'	This study
HsdM-3xGAGAC-3-F-2	5'-TTTATGCAAGCGAGACGAGACGAGACGACGACCAAGATTGTCG-3'	This study
pCT1_screen_F	5'-TTCCACATCAGAAACCG-3'	This study
pCT_screen_R	5'-TAGGTTGCGTATTGGAAG-3'	This study

Bibliography

- Abedon, S. T. (2011). Lysis from without. Bacteriophage, 1(1):46-49.
- Abedon, S. T. (2014). Phage Therapy : Eco-Physiological Pharmacology. Scientifica, 2014:1–30.
- Adam, H. J., Richardson, S. E., Jamieson, F. B., Rawte, P., Low, D. E., and Fisman, D. N. (2010). Changing epidemiology of invasive Haemophilus influenzae in Ontario, Canada: evidence for herd effects and strain replacement due to Hib vaccination. *Vaccine*, 28(24):4073–8.
- Adamczyk-Poplawska, M., Lower, M., and Piekarowicz, A. (2011). Deletion of one nucleotide within the homonucleotide tract present in the hsdS gene alters the DNA sequence specificity of type I restriction-modification system NgoAV. *Journal of bacteriology*, 193(23):6750–9.
- Agrawal, A. and Murphy, T. F. (2011). Haemophilus influenzae infections in the H. influenzae type b conjugate vaccine era. *Journal of clinical microbiology*, 49(11):3728–32.
- Alamro, M., Bidmos, F. A., Chan, H., Oldfield, N. J., Newton, E., Bai, X., Aidley, J., Care, R., Mattick, C., Turner, D. P. J., Neal, K. R., Ala'aldeen, D. A. A., Feavers, I., Borrow, R., and Bayliss, C. D. (2014). Phase variation mediates reductions in expression of surface proteins during persistent meningococcal carriage. *Infection and immunity*, 82(6):2472–84.
- Alikhan, M. M. and Lee, F. E.-H. (2014). Understanding nontypeable Haemophilus influenzae and chronic obstructive pulmonary disease. *Current opinion in pulmonary medicine*, 20(2):159–64.
- Allen, B., Willner, D., Oechel, W. C., and Lipson, D. (2010). Top-down control of microbial activity and biomass in an Arctic soil ecosystem. *Environmental microbiology*, 12(3):642–8.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–10.
- Anderson, C. W. and Eigner, J. (1971). Breakdown and exclusion of superinfecting T-even bacteriophage in Escherichia coli. *Journal of virology*, 8(6):869–86.
- Anderson, R., Wang, X., Briere, E. C., Katz, L. S., Cohn, A. C., Clark, T. A., Messonnier, N. E., and

Mayer, L. W. (2012). Haemophilus haemolyticus isolates causing clinical disease. *Journal of clinical microbiology*, 50(7):2462–5.

- Andreatti Filho, R. L., Higgins, J. P., Higgins, S. E., Gaona, G., Wolfenden, a. D., Tellez, G., and Hargis, B. M. (2007). Ability of bacteriophages isolated from different sources to reduce Salmonella enterica serovar enteritidis in vitro and in vivo. *Poultry science*, 86(9):1904–9.
- Anjum, A., Brathwaite, K. J., Aidley, J., Connerton, P. L., Cummings, N. J., Parkhill, J., Connerton, I., and Bayliss, C. D. (2016). Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in Campylobacter jejuni strain NCTC11168. *Nucleic acids research*, page gkw019.
- ApE (2016). ApE: a plasmid editor.
- Aro, E. M., Virgin, I., and Andersson, B. (1993). Photoinhibition of Photosystem II. Inactivation, protein damage and turnover. *Biochimica et biophysica acta*, 1143(2):113–34.
- Atack, J. M., Srikhanta, Y. N., Fox, K. L., Jurcisek, J. a., Brockman, K. L., Clark, T. a., Boitano, M., Power, P. M., Jen, F. E.-C., McEwan, A. G., Grimmond, S. M., Smith, A. L., Barenkamp, S. J., Korlach, J., Bakaletz, L. O., and Jennings, M. P. (2015). A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable Haemophilus influenzae. *Nature* communications, 6:7828.
- Atanasiu, C., Su, T.-J., Sturrock, S. S., and Dryden, D. T. F. (2002). Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic acids research*, 30(18):3936–44.
- Aubrey, R. and Tang, C. (2003). The Pathogenesis of Disease Due to Type b Haemophilus influenzae. In *Hemophilus influenzae Protocols*, volume 71, pages 29–50. Humana Press, New Jersey.
- Bächi, B., Reiser, J., and Pirrotta, V. (1979). Methylation and cleavage sequences of the EcoP1 restriction-modification enzyme. *Journal of Molecular Biology*, 128(2):143–163.
- Baena-González, E. and Aro, E.-M. (2002). Biogenesis, assembly and turnover of photosystem II units. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 357(1426):1451–9; discussion 1459–60.
- Bakri, F., Brauer, A. L., Sethi, S., and Murphy, T. F. (2002). Systemic and mucosal antibody response to Moraxella catarrhalis after exacerbations of chronic obstructive pulmonary disease. *The Journal* of infectious diseases, 185(5):632–40.
- Bandi, V., Apicella, M. A., Mason, E., Murphy, T. F., Siddiqi, A., Atmar, R. L., and Greenberg, S. B. (2001). Nontypeable Haemophilus influenzae in the lower respiratory tract of patients with chronic bronchitis. *American journal of respiratory and critical care medicine*, 164(11):2114–2119.

- Bandi, V., Jakubowycz, M., Kinyon, C., Mason, E. O., Atmar, R. L., Greenberg, S. B., and Murphy, T. F. (2003). Infectious exacerbations of chronic obstructive pulmonary disease associated with respiratory viruses and non-typeable Haemophilus influenzae. *FEMS Immunology and Medical Microbiology*, 37(1):69–75.
- Barker, B. L., Haldar, K., Patel, H., Pavord, I. D., Barer, M. R., Brightling, C. E., and Bafadhel, M. (2015). Association between pathogens detected using quantitative polymerase chain reaction with airway inflammation in COPD at stable state and exacerbations. *Chest*, 147(1):46–55.
- Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A. S., Doran, K. S., Salamon, P., Youle, M., and Rohwer, F. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences*, 110(26):10771–10776.
- Barr, J. J., Auro, R., Sam-Soon, N., Kassegne, S., Peters, G., Bonilla, N., Hatay, M., Mourtada, S., Bailey, B., Youle, M., Felts, B., Baljon, A., Nulton, J., Salamon, P., and Rohwer, F. (2015). Subdiffusive motion of bacteriophage in mucosal surfaces increases the frequency of bacterial encounters. *Proceedings of the National Academy of Sciences*, 112(44):201508355.
- Bartual, S. G., Otero, J. M., Garcia-Doval, C., Llamas-Saiz, A. L., Kahn, R., Fox, G. C., and van Raaij,
 M. J. (2010). Structure of the bacteriophage T4 long tail fiber receptor-binding tip. Proceedings of the National Academy of Sciences of the United States of America, 107(47):20287–92.
- Bayliss, C. (2009). Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. *FEMS microbiology reviews*, 33(3):504–20.
- Bayliss, C., Sweetman, W. a., and Moxon, E. R. (2004). Mutations in Haemophilus influenzae mismatch repair genes increase mutation rates of dinucleotide repeat tracts but not dinucleotide repeat-driven pilin phase variation rates. *Journal of bacteriology*.
- Bayliss, C., Sweetman, W. a., and Moxon, E. R. (2005). Destabilization of tetranucleotide repeats in Haemophilus influenzae mutants lacking RnaseHI or the Klenow domain of PolI. *Nucleic acids research*, 33(1):400–8.
- Bayliss, C. D., Callaghan, M. J., and Moxon, E. R. (2006). High allelic diversity in the methyltransferase gene of a phase variable type III restriction-modification system has implications for the fitness of Haemophilus influenzae. *Nucleic acids research*, 34(14):4046–59.
- Bayliss, C. D., Field, D., and Moxon, E. R. (2001). The simple sequence contingency loci of Haemophilus influenzae and Neisseria meningitidis. *The Journal of clinical investigation*, 107(6):657– 62.
- Bebeacua, C., Lorenzo Fajardo, J. C., Blangy, S., Spinelli, S., Bollmann, S., Neve, H., Cambillau, C.,

and Heller, K. J. (2013). X-ray structure of a superinfection exclusion lipoprotein from phage TP-J34 and identification of the tape measure protein as its target. *Molecular microbiology*, 89(1):152–65.

- Beck, N. A., Krukonis, E. S., and DiRita, V. J. (2004). TcpH influences virulence gene expression in Vibrio cholerae by inhibiting degradation of the transcription activator TcpP. *Journal of bacteriology*, 186(24):8309–16.
- Bendler, J. W. and Goodgal, S. H. (1968). Prophage S2 mutants in Haemophilus influenzae: a technique for their production and isolation. *Science (New York, N.Y.)*, 162(3852):464–5.
- Berkner, K. L. and Folk, W. R. (1979). The effects of substituted pyrimidines in DNAs on cleavage by sequence-specific endonucleases. *The Journal of biological chemistry*, 254(7):2551–60.
- Bertani, G. and Weigle, J. J. (1953). Host controlled variation in bacterial viruses. *Journal of bacte*riology, 65(2):113–121.
- Bertozzi Silva, J., Storms, Z., and Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. FEMS microbiology letters, 363(4):1–11.
- Bickle, T. A. and Krüger, D. H. (1993). Biology of DNA restriction. *Microbiological reviews*, 57:434–450.
- Bikard, D. and Marraffini, L. a. (2012). Innate and adaptive immunity in bacteria: mechanisms of programmed genetic variation to fight bacteriophages. *Current opinion in immunology*, 24(1):15–20.
- Bitzan, M., Richardson, S., Huang, C., Boyd, B., Petric, M., and Karmali, M. A. (1994). Evidence that verotoxins (Shiga-like toxins) from Escherichia coli bind to P blood group antigens of human erythrocytes in vitro. *Infection and immunity*, 62(8):3337–47.
- Blomfield, I. C., Vaughn, V., Rest, R. F., and Eisenstein, B. I. (1991). Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Molecular microbiology*, 5(6):1447–57.
- Boling, M. E., Allison, D. P., and Setlow, J. K. (1973). Bacteriophage of Haemophilus influenzae. 3. Morphology, DNA homology, and immunity properties of HPlcl, S2, and the defective bacteriophage from strain Rd. *Journal of virology*, 11(4):585–91.
- Boling, M. E., Setlow, J. K., and Allison, D. P. (1972). Bacteriophage of Haemophilus influenzae. I. Differences between infection by whole phage, extracted phage DNA and prophage DNA extracted from lysogenic cells. *Journal of molecular biology*, 63(3):335–48.
- Bonnain, C., Breitbart, M., and Buck, K. N. (2016). The Ferrojan Horse Hypothesis: Iron-Virus Interactions in the Ocean. *Frontiers in Marine Science*, 3(June):1–11.
- Brabetz, W., Müller-Loennies, S., and Brade, H. (2000). 3-Deoxy-D-manno-oct-2-ulosonic acid (Kdo)

transferase (WaaA) and kdo kinase (KdkA) of Haemophilus influenzae are both required to complement a waaA knockout mutation of Escherichia coli. *The Journal of biological chemistry*, 275(45):34954–62.

- Braun, V., Schaller, K., and Wolff, H. (1973). A common receptor protein for phage T5 and colicin M in the outer membrane of Escherichia coli B. *Biochimica et biophysica acta*, 323(1):87–97.
- Brockhurst, M., Buckling, A., and Rainey, P. B. (2006). Spatial heterogeneity and the stability of host-parasite coexistence. *Journal of evolutionary biology*, 19(2):374–9.
- Brockhurst, M., Morgan, A. D., Fenton, A., and Buckling, A. (2007). Experimental coevolution with bacteria and phage. The Pseudomonas fluorescens–Phi2 model system. *Infection, genetics* and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 7(4):547–52.
- Brüssow, H. (2001). Phages of dairy bacteria. Annual review of microbiology, 55:283–303.
- Brüssow, H., Canchaya, C., Hardt, W.-d., and Bru, H. (2004). Phages and the Evolution of Bacterial Pathogens : from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and molecular biology reviews*, 68(3):560–602.
- Buckling, A. and Rainey, P. B. (2002a). Antagonistic coevolution between a bacterium and a bacteriophage. Proceedings. Biological sciences / The Royal Society, 269(1494):931–6.
- Buckling, A. and Rainey, P. B. (2002b). The role of parasites in sympatric and allopatric host diversification. *Nature*, 420(6915):496–9.
- Burns, J. L., Emerson, J., Stapp, J. R., Yim, D. L., Krzewinski, J., Louden, L., Ramsey, B. W., and Clausen, C. R. (1998). Microbiology of sputum from patients at cystic fibrosis centers in the United States. *Clinical infectious diseases : an official publication of the Infectious Diseases Society* of America, 27(1):158–63.
- Cairns, B. J., Timms, A. R., Jansen, V. a. a., Connerton, I. F., and Payne, R. J. H. (2009). Quantitative models of in vitro bacteriophage-host dynamics and their application to phage therapy. *PLoS Pathogens*, 5(1):1–10.
- Campbell, A. (2003). The future of bacteriophage biology. Nature reviews. Genetics, 4(6):471–7.
- Canchaya, C., Desiere, F., McShan, W. M., Ferretti, J. J., Parkhill, J., and Brüssow, H. (2002). Genome analysis of an inducible prophage and prophage remnants integrated in the Streptococcus pyogenes strain SF370. Virology, 302(2):245–58.
- Carroll, P. a., Tashima, K. T., Rogers, M. B., DiRita, V. J., and Calderwood, S. B. (1997). Phase vari-

ation in tcpH modulates expression of the ToxR regulon in Vibrio cholerae. *Molecular microbiology*, 25(6):1099–111.

- Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? Molecular Microbiology, 49(2):277–300.
- Cerdeño-Tárraga, A. M., Efstratiou, A., Dover, L. G., Holden, M. T. G., Pallen, M., Bentley, S. D., Besra, G. S., Churcher, C., James, K. D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M. A., Rabbinowitsch, E., Rutherford, K. M., Thomson, N. R., Unwin, L., Whitehead, S., Barrell, B. G., and Parkhill, J. (2003). The complete genome sequence and analysis of Corynebacterium diphtheriae NCTC13129. Nucleic Acids Research, 31(22):6516–6523.
- Cerdeño-Tárraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. a., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M. T. G., Larke, N., Line, A., Lord, A., Norbertczak, H., Ormond, D., Price, C., Rabbinowitsch, E., Woodward, J., Barrell, B., and Parkhill, J. (2005). Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science (New York, N.Y.)*, 307(5714):1463–5.
- Chakraborty, S., Mukhopadhyay, A. K., Bhadra, R. K., Ghosh, A. N., Mitra, R., Shimada, T., Yamasaki, S., Faruque, S. M., Takeda, Y., Colwell, R. R., and Nair, G. B. (2000). Virulence genes in environmental strains of Vibrio cholerae. *Applied and environmental microbiology*, 66(9):4022–8.
- Chibeu, A., Ceyssens, P.-J., Hertveldt, K., Volckaert, G., Cornelis, P., Matthijs, S., and Lavigne, R. (2009). The adsorption of Pseudomonas aeruginosa bacteriophage phiKMV is dependent on expression regulation of type IV pili genes. *FEMS microbiology letters*, 296(2):210–8.
- Christen, M., Beusch, C., Bösch, Y., Cerletti, D., Flores-Tinoco, C. E., Del Medico, L., Tschan, F., and Christen, B. (2016). Quantitative Selection Analysis of Bacteriophage \u03c6CbK Susceptibility in Caulobacter crescentus. Journal of Molecular Biology, 428(2):419–430.
- Cianciotto, N. P., Long, R., Eisenstein, B. I., and Cary Engleberg, N. (1988). Site-specific mutagenesis in Legionella pneumophila by allelic exchange using counterselectable ColE1 vectors. *FEMS Microbiology Letters*, 56(2):203–207.
- Clark, S. E., Eichelberger, K. R., and Weiser, J. N. (2013). Evasion of killing by human antibody and complement through multiple variations in the surface oligosaccharide of Haemophilus influenzae. *Molecular microbiology*, 88(3):603–18.
- Clark, S. E., Snow, J., Li, J., Zola, T. A., and Weiser, J. N. (2012a). Phosphorylcholine allows for evasion of bactericidal antibody by Haemophilus influenzae. *PLoS pathogens*, 8(3):e1002521.
- Clark, T. A., Murray, I. A., Morgan, R. D., Kislyuk, A. O., Spittle, K. E., Boitano, M., Fomenkov,

A., Roberts, R. J., and Korlach, J. (2012b). Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Research*, 40(4).

- Clokie, M. R. J., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011). Phages in nature. Bacteriophage, 1(1):31–45.
- Clokie, M. R. J., Shan, J., Bailey, S., Jia, Y., Krisch, H. M., West, S., and Mann, N. H. (2006). Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environmental microbiology*, 8(5):827–35.
- Cope, L. D., Hrkal, Z., and Hansen, E. J. (2000). Detection of phase variation in expression of proteins involved in hemoglobin and hemoglobin-haptoglobin binding by nontypeable Haemophilus influenzae. *Infection and immunity*, 68(7):4092–101.
- Cordero, E., Pachón, J., Rivero, A., Girón, J. A., Gómez-Mateos, J., Merino, M. D., Torres-Tortosa, M., González-Serrano, M., Aliaga, L., Collado, A., Hernández-Quero, J., Barrera, A., and Nuño, E. (2000). Haemophilus influenzae pneumonia in human immunodeficiency virus-infected patients. The Grupo Andaluz para el Estudio de las Enfermedades Infecciosas. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 30(3):461–5.
- Cota, I., Blanc-Potard, A. B., and Casadesús, J. (2012). STM2209-STM2208 (opvAB): a phase variation locus of Salmonella enterica involved in control of O-antigen chain length. *PloS one*, 7(5):e36863.
- Cota, I., Bunk, B., Spröer, C., Overmann, J., König, C., and Casadesús, J. (2016). OxyR-dependent formation of DNA methylation patterns in OpvABOFF and OpvABON cell lineages of Salmonella enterica. *Nucleic acids research*, 44(8):3595–609.
- Cota, I., Sánchez-Romero, M. A., Hernández, S. B., Pucciarelli, M. G., García-del Portillo, F., and Casadesús, J. (2015). Epigenetic Control of Salmonella enterica O-Antigen Chain Length: A Tradeoff between Virulence and Bacteriophage Resistance. *PLOS Genetics*, 11(11):e1005667.
- Cowan, G. M., Gann, A. A., and Murray, N. E. (1989). Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell*, 56(1):103–9.
- Datta, D. B., Arden, B., and Henning, U. (1977). Major proteins of the Escherichia coli outer cell envelope membrane as bacteriophage receptors. *Journal of bacteriology*, 131(3):821–9.
- Davis, G. S., Marino, S., Marrs, C. F., Gilsdorf, J. R., Dawid, S., and Kirschner, D. E. (2014). Phase variation and host immunity against high molecular weight (HMW) adhesins shape population dynamics of nontypeable Haemophilus influenzae within human hosts. *Journal of theoretical biology*, 355:208–18.
- Davison, S., Couture-Tosi, E., Candela, T., Mock, M., and Fouet, A. (2005). Identification of the Bacillus anthracis (gamma) phage receptor. *Journal of bacteriology*, 187(19):6742–9.

- Dawid, S., Barenkamp, S. J., and St Geme, J. W. (1999). Variation in expression of the Haemophilus influenzae HMW adhesins: a prokaryotic system reminiscent of eukaryotes. Proceedings of the National Academy of Sciences of the United States of America, 96(3):1077–82.
- De Bolle, X., Bayliss, C., Field, D., Van De Ven, T., Saunders, N. J., Hood, D. W., and Moxon, E. R. (2000). The length of a tetranucleotide repeat tract in Haemophilus influenzae determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Molecular Microbiology*, 35(1):211–222.
- D'Herelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. CR Acad. Sci. Paris, 165:373–375.
- D'Herelle, F. (2007). On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr.F. D'Herelle, presented by Mr. Roux. 1917. *Research in microbiology*, 158(7):553–4.
- Dixon, K., Bayliss, C. D., Makepeace, K., Moxon, E. R., and Hood, D. W. (2007). Identification of the functional initiation codons of a phase-variable gene of Haemophilus influenzae, lic2A, with the potential for differential expression. *Journal of bacteriology*, 189(2):511–21.
- Dryden, D. T., Murray, N. E., and Rao, D. N. (2001). Nucleoside triphosphate-dependent restriction enzymes. *Nucleic acids research*, 29(18):3728–41.
- Dubovi, E. J. and Akers, T. G. (1970). Airborne stability of tailless bacterial viruses S-13 and MS-2. Appl Microbiol, 19(4):624–628.
- Duplessis, M. and Moineau, S. (2001). Identification of a genetic determinant responsible for host specificity in Streptococcus thermophilus bacteriophages. *Molecular microbiology*, 41(2):325–36.
- Dworkin, M. S., Borchardt, L. P., and M, S. (2007). The changing epidemiology of invasive Haemophilus influenzae disease, especially in persons (greater-than or equal to)65 years old. *Clinical Infectious Diseases*, 44(6):810–816.
- Dy, R. L., Richter, C., Salmond, G. P. C., and Fineran, P. C. (2014). Remarkable Mechanisms in Microbes to Resist Phage Infections. Annual review of virology, 1(1):307–31.
- Dybvig, K., Sitaraman, R., and French, C. T. (1998). A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23):13923–8.
- Ehrlich, R., Miller, S., and Idoine, L. S. (1964). Effects of environmental factors on the survival of airborn T-3 coliphage. Applied microbiology, 12(6):479–82.
- Eklund, M. W., Poysky, F. T., and Reed, S. M. (1972). Bacteriophage and the toxygenicity of Clostridium botulinum type d. *Nature new biology*, 235:16–17.

- Eklund, M. W., Poysky, F. T., Reed, S. M., and Smith, C. A. (1971). Bacteriophage and the toxigenicity of Clostridium botulinum type C. *Science (New York, N.Y.)*, 172(3982):480–2.
- Emerson, J. E., Reynolds, C. B., Fagan, R. P., Shaw, H. a., Goulding, D., and Fairweather, N. F. (2009). A novel genetic switch controls phase variable expression of CwpV, a Clostridium difficile cell wall protein. *Molecular Microbiology*, 74(3):541–556.
- Erwin, A. L., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., McCrea, K. W., Gilsdorf, J. R., and Smith, A. L. (2006). Heterogeneity in tandem octanucleotides within Haemophilus influenzae lipopolysaccharide biosynthetic gene losA affects serum resistance. *Infection and immunity*, 74(6):3408–14.
- Esposito, D., Fitzmaurice, W. P., Benjamin, R. C., Goodman, S. D., Waldman, A. S., and Scocca, J. J. (1996). The complete nucleotide sequence of bacteriophage HP1 DNA. *Nucleic acids research*, 24(12):2360–8.
- Fahy, J. V. and Dickey, B. F. (2010). Airway mucus function and dysfunction. The New England journal of medicine, 363(23):2233–47.
- 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. *Journal of clinical microbiology*, 32(10):2382–6.
- Fan, X., Pericone, C. D., Lysenko, E., Goldfine, H., and Weiser, J. N. (2003). Multiple mechanisms for choline transport and utilization in Haemophilus influenzae. *Molecular Microbiology*, 50(2):537–548.
- Faruque, S. M., Islam, M. J., Ahmad, Q. S., Faruque, A. S. G., Sack, D. A., Nair, G. B., and Mekalanos, J. J. (2005a). Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proceedings of the National Academy of Sciences of the United States of America*, 102(17):6119–24.
- Faruque, S. M., Naser, I. B., Islam, M. J., Faruque, a. S. G., Ghosh, a. N., Nair, G. B., Sack, D. a., and Mekalanos, J. J. (2005b). Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proceedings of the National Academy of Sciences of the United States of America*, 102(5):1702–7.
- Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., and Herskovits, A. A. (2015). A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nature reviews. Microbiology*, 13(10):641–50.
- Fernaays, M. M., Lesse, A. J., Cai, X., and Murphy, T. F. (2006). Characterization of igaB, a second immunoglobulin A1 protease gene in nontypeable Haemophilus influenzae. *Infection and immunity*, 74(10):5860–70.
- Finney, L. and Ritchie, A. (2014). Lower airway colonization and inflammatory response in COPD: a focus on Haemophilus influenzae. *International journal of ...*, pages 1119–1132.

- 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., and Merrick, J. M. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science (New York, N.Y.)*, 269(5223):496–512.
- Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *The British Journal of Experimental Pathology*, 10:226–236.
- Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., and Weitz, J. S. (2011). Statistical structure of host-phage interactions. *Proc Natl Acad Sci U S A*, 108(28):E288–97.
- Fokine, A. and Rossmann, M. G. (2014). Molecular architecture of tailed double-stranded DNA phages. Bacteriophage, 4(1):e28281.
- Fortier, L.-C. and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, 4(5):354–365.
- Fox, K. L., Atack, J. M., Srikhanta, Y. N., Eckert, A., Novotny, L. a., Bakaletz, L. O., and Jennings, M. P. (2014). Selection for phase variation of LOS biosynthetic genes frequently occurs in progression of non-typeable Haemophilus influenzae infection from the nasopharynx to the middle ear of human patients. *PloS one*, 9(2):e90505.
- Fox, K. L., Dowideit, S. J., Erwin, A. L., Srikhanta, Y. N., Smith, A. L., and Jennings, M. P. (2007). Haemophilus influenzae phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. *Nucleic acids research*, 35(15):5242–52.
- Fox, K. L., Yildirim, H. H., Deadman, M. E., Schweda, E. K. H., Moxon, E. R., and Hood, D. W. (2005). Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in Haemophilus influenzae, identification of a gene for adding O-acetyl groups. *Molecular microbiology*, 58(1):207–16.
- Fralick, J. A. (1996). Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of Escherichia coli. *Journal of bacteriology*, 178(19):5803–5.
- Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature, 399(6736):541–8.
- Gallet, R., Lenormand, T., and Wang, I. N. (2012). Phenotypic Stochasticity Protects Lytic Bacteriophage Populations From Extinction During The Bacterial Stationary Phase. *Evolution*, 66(11):3485– 3494.
- Garbe, J., Bunk, B., Rohde, M., and Schobert, M. (2011). Sequencing and characterization of Pseudomonas aeruginosa phage JG004. BMC microbiology, 11(1):102.
- García-Rodríguez, J. Á. and Fresnadillo Martínez, M. J. (2002). Dynamics of nasopharyngeal col-

onization by potential respiratory pathogens. J.Antimicrob.Chemother., 50 Suppl S(0305-7453 (Print)):59–73.

- Garmendia, J., Viadas, C., Calatayud, L., Mell, J. C., Mart??-Lliteras, P., Euba, B., Llobet, E., Gil,
 C., Bengoechea, J. A., Redfield, R. J., and Li??ares, J. (2014). Characterization of nontypable
 Haemophilus influenzae isolates recovered from adult patients with underlying chronic lung disease
 reveals genotypic and phenotypic traits associated with persistent infection. *PLoS ONE*, 9(5).
- Gauntlett, J. C., Nilsson, H.-O., Fulurija, A., Marshall, B. J., and Benghezal, M. (2014). Phase-variable restriction/modification systems are required for Helicobacter pylori colonization. *Gut pathogens*, 6(1):35.
- Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E., and Hoch, J. a. (1983). Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. *Journal of bacteriology*, 153(3):1424–31.
- Geller, B. L., Ivey, R. G., Trempy, J. E., and Hettinger-Smith, B. (1993). Cloning of a chromosomal gene required for phage infection of Lactococcus lactis subsp. lactis C2. *Journal of Bacteriology*, 175(17):5510–5519.
- German, G. J. and Misra, R. (2001). The TolC protein of Escherichia coli serves as a cell-surface receptor for the newly characterized TLS bacteriophage. *Journal of molecular biology*, 308(4):579– 85.
- Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., and Taddei, F. (2001). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science (New York, N.Y.)*, 291(5513):2606–8.
- Glover, S. W. and Piekarowicz, A. (1972). Host specificity of DNA in Haemophilus influenzae: restriction and modification in strain Rd. *Biochemical and biophysical research communications*, 46(4):1610–7.
- Golomidova, A., Kulikov, E., Isaeva, A., Manykin, A., and Letarov, A. (2007). The diversity of coliphages and coliforms in horse feces reveals a complex pattern of ecological interactions. *Applied* and environmental microbiology, 73(19):5975–81.
- Golshahi, L., Seed, K. D., Dennis, J. J., and Finlay, W. H. (2008). Toward modern inhalational bacteriophage therapy: nebulization of bacteriophages of Burkholderia cepacia complex. *Journal of* aerosol medicine and pulmonary drug delivery, 21(4):351–60.
- Gómez, P. and Buckling, A. (2011). Bacteria-phage antagonistic coevolution in soil. Science (New York, N.Y.), 332(6025):106–9.

- González, E., Padilla, C., Saavedra, C., and Vásquez, C. (1994). The expression of the bstVIM gene from Bacillus stearothermophilus V is restricted to vegetative cell growth. *Microbiology (Reading, England)*, 140 (Pt 6(1 994):1337–40.
- Griffin, R., Bayliss, C. D., Herbert, M. A., Cox, A. D., Makepeace, K., Richards, J. C., Hood, D. W., and Moxon, E. R. (2005). Digalactoside expression in the lipopolysaccharide of Haemophilus influenzae and its role in intravascular survival. *Infection and immunity*, 73(10):7022–6.
- Groeneveld, K., Eijk, P. P., van Alphen, L., Jansen, H. M., and Zanen, H. C. (1990). Haemophilus influenzae infections in patients with chronic obstructive pulmonary disease despite specific antibodies in serum and sputum. *The American review of respiratory disease*, 141(5 Pt 1):1316–21.
- Groenewegen, K. H. and Wouters, E. F. M. (2003). Bacterial infections in patients requiring admission for an acute exacerbation of COPD; a 1-year prospective study. *Respiratory Medicine*, 97(7):770–777.
- Grumann, D., Nübel, U., and Bröker, B. M. (2014). Staphylococcus aureus toxins-their functions and genetics. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 21:583–92.
- Gunnarsson, R. K., Holm, S. E., and Söderström, M. (1998). The prevalence of potential pathogenic bacteria in nasopharyngeal samples from healthy children and adults. *Scandinavian journal of* primary health care, 16(1):13–17.
- Hammerschmidt, S., Müller, A., Sillmann, H., Mühlenhoff, M., Borrow, R., Fox, A., van Putten, J., Zollinger, W. D., Gerardy-Schahn, R., and Frosch, M. (1996). Capsule phase variation in Neisseria meningitidis serogroup B by slipped-strand mispairing in the polysialyltransferase gene (siaD): correlation with bacterial invasion and the outbreak of meningococcal disease. *Molecular microbiology*, 20(6):1211–20.
- Hanahan, D. (1985). Techniques for transformation of E. coli. In In: Glover D M, editor. DNA cloning: a practical approach. Vol. 1. Oxford, United Kingdom: IRL Press, pages 109–135.
- Hardy, G. G., Tudor, S. M., and St Geme, J. W. (2003). The pathogenesis of disease due to nontypeable Haemophilus influenzae. *Methods in molecular medicine*, 71:1–28.
- Harm, W. and Rupert, C. (1963). Infection of transformable cells of Haemophilus influenzae by bacteriophage and bacteriophage DNA. Zeitschrift für Vererbungslehre, 348(1963):336–348.
- Harvey, H. a., Swords, W. E., and Apicella, M. a. (2001). The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic neisseria and haemophilus. *Journal of autoimmunity*, 16(3):257–62.
- Hatfull, G. and Hendrix, R. (2011). Bacteriophages and their genomes. *Current opinion in virology*, 1(4):298–303.

- Heidelberg, J. F., Eisen, J. a., Nelson, W. C., Clayton, R. a., Gwinn, M. L., Dodson, R. J., Haft, D. H., Hickey, E. K., Peterson, J. D., Umayam, L., Gill, S. R., Nelson, K. E., Read, T. D., Tettelin, H., Richardson, D., Ermolaeva, M. D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleishmann, R. D., Nierman, W. C., White, O., Salzberg, S. L., Smith, H. O., Colwell, R. R., Mekalanos, J. J., Venter, J. C., and Fraser, C. M. (2000). DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature*, 406(6795):477–83.
- Henning, U. and Jann, K. (1979). Two-component nature of bacteriophage T4 receptor activity in Escherichia coli K-12. Journal of bacteriology, 137(1):664–6.
- Henson, B. J., Hartman, L., Watson, L. E., and Barnum, S. R. (2011). Evolution and variation of the nifD and hupL elements in the heterocystous cyanobacteria. *International journal of systematic and* evolutionary microbiology, 61(Pt 12):2938–49.
- Heo, Y.-J., Lee, Y.-R., Jung, H.-H., Lee, J., Ko, G., and Cho, Y.-H. (2009). Antibacterial efficacy of phages against Pseudomonas aeruginosa infections in mice and Drosophila melanogaster. *Antimi*crobial agents and chemotherapy, 53(6):2469–74.
- Herbert, M. a., Hayes, S., Deadman, M. E., Tang, C. M., Hood, D. W., and Moxon, E. R. (2002). Signature Tagged Mutagenesis of Haemophilus influenzae identifies genes required for in vivo survival. *Microbial pathogenesis*, 33(5):211–223.
- High, N. J., Deadman, M. E., and Moxon, E. R. (1993). The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the Haemophilus influenzae lipopolysaccharide epitope alpha Gal(1-4)beta Gal. *Molecular microbiology*, 9(6):1275–82.
- High, N. J., Jennings, M. P., and Moxon, E. R. (1996). Tandem repeats of the tetramer 5'-CAAT-3' present in lic2A are required for phase variation but not lipopolysaccharide biosynthesis in Haemophilus influenzae. *Molecular microbiology*, 20(1):165–74.
- Hinz, R., Zautner, A. E., Hagen, R. M., and Frickmann, H. (2015). Difficult identification of Haemophilus influenzae, a typical cause of upper respiratory tract infections, in the microbiological diagnostic routine. *European journal of microbiology & immunology*, 5(1):62–7.
- Ho, T. D. and Slauch, J. M. (2001). OmpC is the receptor for Gifsy-1 and Gifsy-2 bacteriophages of Salmonella. *Journal of bacteriology*, 183(4):1495–8.
- Honarvar, S., Choi, B.-K., and Schifferli, D. M. (2003). Phase variation of the 987P-like CS18 fimbriae of human enterotoxigenic Escherichia coli is regulated by site-specific recombinases. *Molecular Microbiology*, 48(1):157–171.
- Hood, D. W., Cox, a. D., Wakarchuk, W. W., Schur, M., Schweda, E. K., Walsh, S. L., Deadman, M. E., Martin, a., Moxon, E. R., and Richards, J. C. (2001). Genetic basis for expression of the major

globotetraose-containing lipopolysaccharide from H. influenzae strain Rd (RM118). *Glycobiology*, 11(11):957–67.

- Hood, D. W., Deadman, M. E., Jennings, M. P., Bisercic, M., Fleischmann, R. D., Venter, J. C., and Moxon, E. R. (1996). DNA repeats identify novel virulence genes in Haemophilus influenzae. *Proceedings of the National Academy of Sciences of the United States of America*, 93(20):11121–5.
- Hosking, S. L., Craig, J. E., and High, N. J. (1999). Phase variation of lic1A, lic2A and lic3A in colonization of the nasopharynx, bloodstream and cerebrospinal fluid by Haemophilus influenzae type b. *Microbiology (Reading, England)*, 145 (Pt 1(11):3005–11.
- Hoskisson, P. a. and Smith, M. C. M. (2007). Hypervariation and phase variation in the bacteriophage 'resistome'. *Current opinion in microbiology*, 10(4):396–400.
- Hyman, P. and Abedon, S. T. (2010). Bacteriophage host range and bacterial resistance., volume 70. Elsevier Inc., 1 edition.
- Iida, S., Streiff, M. B., Bickle, T. A., and Arber, W. (1987). Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology*, 157(1):156–66.
- Imamovic, L. and Muniesa, M. (2012). Characterizing RecA-independent induction of Shiga toxin2encoding phages by EDTA treatment. *PloS one*, 7(2):e32393.
- Iriarte, F. B., Balogh, B., Momol, M. T., Smith, L. M., Wilson, M., and Jones, J. B. (2007). Factors affecting survival of bacteriophage on tomato leaf surfaces. *Applied and Environmental Microbiology*, 73(6):1704–1711.
- Jarosik, G. P. and Hansen, E. J. (1994). Identification of a new locus involved in expression of Haemophilus influenzae type b lipooligosaccharide. *Infection and immunity*, 62(11):4861–7.
- Johnston, J. W. (2012). An improved counterselection cassette for use in Haemophilus influenzae. Gene, 492(1):325–8.
- Jończyk, E., Kłak, M., Międzybrodzki, R., and Górski, A. (2011). The influence of external factors on bacteriophages-review. *Folia microbiologica*, 56(3):191–200.
- Kan, N. C., Lautenberger, J. A., Edgell, M. H., and Hutchison, C. A. (1979). The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes. *Journal of Molecular Biology*, 130(2):191–209.
- Kannan, P., Cowan, G. M., Daniel, A. S., Gann, A. A. F., and Murray, N. E. (1989). Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *Journal of Molecular Biology*, 209(3):335–344.
- Karch, H., Schmidt, H., Janetzki-Mittmann, C., Scheef, J., and Kröger, M. (1999). Shiga toxins even

when different are encoded at identical positions in the genomes of related temperate bacteriophages. *Molecular* \mathcal{E} general genetics : MGG, 262(4-5):600–7.

- Karlyshev, A. V., Linton, D., Gregson, N. A., and Wren, B. W. (2002). A novel paralogous gene family involved in phase-variable flagella-mediated motility in Campylobacter jejuni. *Microbiology* (*Reading, England*), 148(Pt 2):473–80.
- Keen, E. C. (2015). A century of phage research: bacteriophages and the shaping of modern biology. BioEssays : news and reviews in molecular, cellular and developmental biology, 37(1):6–9.
- Keller, R. and Zaltzman, M. L. (1959). Studies on the Factors Concerned in the Disappearance of Bacteriophage Particles from the Animal Body.
- Kelly, D., McAuliffe, O., Ross, R. P., O'Mahony, J., and Coffey, A. (2011). Development of a broadhost-range phage cocktail for biocontrol. *Bioengineered bugs*, 2(1):31–7.
- Kelly, T. J. and Smith, H. O. (1970). A restriction enzyme from Hemophilus influenzae. II. Journal of molecular biology, 51(2):393–409.
- Killmann, H., Braun, M., Herrmann, C., and Braun, V. (2001). FhuA barrel-cork hybrids are active transporters and receptors. *Journal of bacteriology*, 183(11):3476–87.
- Kim, K. S. (2003). Pathogenesis of bacterial meningitis: from bacteraemia to neuronal injury. Nature reviews. Neuroscience, 4(5):376–85.
- Kim, M. and Ryu, S. (2012). Spontaneous and transient defence against bacteriophage by phasevariable glucosylation of O-antigen in Salmonella enterica serovar Typhimurium. *Molecular microbiology*, 86(2):411–25.
- Koskella, B. (2013). Phage-mediated selection on microbiota of a long-lived host. Current biology : CB, 23(13):1256–60.
- Koskella, B. and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS microbiology reviews*, 38(5):916–31.
- Koskella, B., Lin, D. M., Buckling, A., and Thompson, J. N. (2012). The costs of evolving resistance in heterogeneous parasite environments. *Proceedings of the Royal Society B: Biological Sciences*, 279(1735):1896–1903.
- Koskella, B. and Meaden, S. (2013). Understanding bacteriophage specificity in natural microbial communities. Viruses, 5(3):806–23.
- Koskella, B., Thompson, J. N., Preston, G. M., and Buckling, A. (2011). Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. *The American naturalist*, 177(4):440–51.

- Kroll, J. S. and Moxon, E. R. (1988). Capsulation and gene copy number at the cap locus of Haemophilus influenzae type b. *Journal of bacteriology*, 170(2):859–64.
- Kropinski, A. M. (2009). Measurement of the Rate of Attachment of Bacteriophage to Cells. In Methods in molecular biology (Clifton, N.J.), volume 501, pages 151–155.
- Krüger, D. H., Barcak, G. J., Reuter, M., and Smith, H. O. (1988). EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic acids research*, 16(9):3997–4008.
- Krüger, D. H. and Bickle, T. a. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiological reviews*, 47(3):345–60.
- Krüger, D. H. and Schroeder, C. (1981). Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. *Microbiological reviews*, 45(1):9–51.
- Kuhnlein, U. and Arber, W. (1972). Host specificity of DNA produced by Escherichia coli. XV. The role of nucleotide methylation in in vitro B-specific modification. J Mol Biol, 63(1):9–19.
- Kulik, E. M. and Bickle, T. A. (1996). Regulation of the activity of the type IC EcoR124I restriction enzyme. Journal of molecular biology, 264(5):891–906.
- Kutter, E. and Sulakvelidze, A. (2004). Bacteriophages: biology and applications. CRC Press.
- Labrie, S. J., Samson, J. E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nature reviews. Microbiology, 8(5):317–27.
- Ladhani, S., Slack, M. P. E., Heath, P. T., von Gottberg, A., Chandra, M., and Ramsay, M. E. (2010). Invasive Haemophilus influenzae Disease, Europe, 1996-2006. *Emerging infectious diseases*, 16(3):455–63.
- Lafontaine, E. R., Cope, L. D., Aebi, C., Latimer, J. L., McCracken, G. H., and Hansen, E. J. (2000). The UspA1 protein and a second type of UspA2 protein mediate adherence of Moraxella catarrhalis to human epithelial cells in vitro. *Journal of bacteriology*, 182(5):1364–73.
- Latino, L., Midoux, C., Hauck, Y., Vergnaud, G., and Pourcel, C. (2016). Pseudolysogeny and sequential mutations build multiresistance to virulent bacteriophages in Pseudomonas aeruginosa. *Microbiology (Reading, England)*, 162(5):748–63.
- Laupland, K. B., Schønheyder, H. C., Østergaard, C., Knudsen, J. D., Valiquette, L., Galbraith, J., Kennedy, K. J., Gradel, K. O., and International Bacteremia Surveillance Collaborative (2011). Epidemiology of Haemophilus influenzae bacteremia: a multi-national population-based assessment. *The Journal of infection*, 62(2):142–8.
- Le, S., Yao, X., Lu, S., Tan, Y., Rao, X., Li, M., Jin, X., Wang, J., Zhao, Y., Wu, N. C., Lux,
R., He, X., Shi, W., and Hu, F. (2014). Chromosomal DNA deletion confers phage resistance to Pseudomonas aeruginosa. *Scientific Reports*, 4:1–8.

- Leanord, A. and Williams, C. (2002). Haemophilus influenzae in acute exacerbations of chronic obstructive pulmonary disease. *International journal of antimicrobial agents*, 19:371–375.
- Lennon, J. T., Khatana, S. A. M., Marston, M. F., and Martiny, J. B. H. (2007). Is there a cost of virus resistance in marine cyanobacteria? *The ISME journal*, 1(4):300–12.
- Lenski, R. E. and Levin, B. R. (1985). Constraints on the Coevolution of Bacteria and Virulent Phage: A Model, Some Experiments, and Predictions for Natural Communities.
- León, M. and Bastías, R. (2015). Virulence reduction in bacteriophage resistant bacteria. Frontiers in microbiology, 6(April):343.
- Li, X., Lockatell, C. V., Johnson, D. E., and Mobley, H. L. T. (2002). Identification of MrpI as the sole recombinase that regulates the phase variation of MR/P fimbria, a bladder colonization factor of uropathogenic Proteus mirabilis. *Molecular microbiology*, 45(3):865–74.
- Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., Kettler, G., Sullivan, M. B., Steen, R., Hess, W. R., Church, G. M., and Chisholm, S. W. (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, 449(7158):83– 6.
- Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M., and Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*, 438(7064):86–9.
- Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):11013–8.
- Linn, S. and Arber, W. (1968). Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proceedings of the National Academy of Sciences of the United States of America, 59(4):1300–6.
- Lobocka, M., Hejnowicz, M. S., Dąbrowski, K., Gozdek, A., Kosakowski, J., Witkowska, M., Ulatowska, M. I., Weber-Dąbrowska, B., Kwiatek, M., Parasion, S., Gawor, J., Kosowska, H., and Głowacka, A. (2012). Genomics of Staphylococcal Twort-like Phages - Potential Therapeutics of the Post-Antibiotic Era, volume 83.
- Loc-Carrillo, C. and Abedon, S. T. (2011). Pros and cons of phage therapy. *Bacteriophage*, 1(2):111–114.

- Loenen, W. a. M., Dryden, D. T. F., Raleigh, E. a., and Wilson, G. G. (2014a). Type I restriction enzymes and their relatives. *Nucleic acids research*, 42(1):20–44.
- Loenen, W. a. M., Dryden, D. T. F., Raleigh, E. a., Wilson, G. G., and Murray, N. E. (2014b). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic acids research*, 42(1):3–19.
- Loenen, W. a. M. and Raleigh, E. a. (2014). The other face of restriction: modification-dependent enzymes. *Nucleic acids research*, 42(1):56–69.
- Lusiak-Szelachowska, M., Zaczek, M., Weber-Dąbrowska, B., Międzybrodzki, R., Kłak, M., Fortuna, W., Letkiewicz, S., Rogóż, P., Szufnarowski, K., Jończyk-Matysiak, E., Owczarek, B., and Górski, A. (2014). Phage neutralization by sera of patients receiving phage therapy. Viral immunology, 27(6):295–304.
- Lysenko, E., Richards, J. C., Cox, A. D., Stewart, A., Martin, A., Kapoor, M., and Weiser, J. N. (2000a). The position of phosphorylcholine on the lipopolysaccharide of Haemophilus influenzae affects binding and sensitivity to C-reactive protein-mediated killing. *Molecular microbiology*, 35(1):234–45.
- Lysenko, E. S., Gould, J., Bals, R., Wilson, J. M., and Weiser, J. N. (2000b). Bacterial phosphorylcholine decreases susceptibility to the antimicrobial peptide LL-37/hCAP18 expressed in the upper respiratory tract. *Infection and immunity*, 68(3):1664–71.
- Mann, N. H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003). Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature*, 424(6950):741.
- Manso, A. S., Chai, M. H., Atack, J. M., Furi, L., De Ste Croix, M., Haigh, R., Trappetti, C., Ogunniyi, A. D., Shewell, L. K., Boitano, M., Clark, T. a., Korlach, J., Blades, M., Mirkes, E., Gorban, A. N., Paton, J. C., Jennings, M. P., and Oggioni, M. R. (2014). A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nature communications*, 5:5055.
- Margolis, E. and Levin, B. R. (2007). Within-Host Evolution for the Invasiveness of Commensal Bacteria: an Experimental Study of Bacteremias Resulting from Haemophilus influenzae Nasal Carriage. The Journal of Infectious Diseases, 196(7):1068–1075.
- Marston, M. F., Pierciey, F. J., Shepard, A., Gearin, G., Qi, J., Yandava, C., Schuster, S. C., Henn, M. R., and Martiny, J. B. H. (2012). Rapid diversification of coevolving marine Synechococcus and a virus. *Proceedings of the National Academy of Sciences of the United States of America*, 109(12):4544–9.
- Martí-Lliteras, P., López-Gómez, A., Mauro, S., Hood, D. W., Viadas, C., Calatayud, L., Morey, P., Servin, A., Liñares, J., Oliver, A., Bengoechea, J. A., and Garmendia, J. (2011). Nontypable

Haemophilus influenzae displays a prevalent surface structure molecular pattern in clinical isolates. *PloS one*, 6(6):e21133.

- Martínez-García, E., Jatsenko, T., Kivisaar, M., and de Lorenzo, V. (2015). Freeing Pseudomonas putida KT2440 of its proviral load strengthens endurance to environmental stresses. *Environmental* microbiology, 17(1):76–90.
- Maskell, D. J., Szabo, M. J., Butler, P. D., Williams, A. E., and Moxon, E. R. (1991). Molecular analysis of a complex locus from Haemophilus influenzae involved in phase-variable lipopolysaccharide biosynthesis. *Molecular microbiology*, 5(5):1013–22.
- Mason, E. O., Kaplan, S. L., Wiedermann, B. L., Norrod, E. P., and Stenback, W. A. (1985). Frequency and properties of naturally occurring adherent piliated strains of Haemophilus influenzae type b. *Infection and immunity*, 49(1):98–103.
- May, T. B., Shinabarger, D., Maharaj, R., Kato, J., Chu, L., DeVault, J. D., Roychoudhury, S., Zielinski, N. A., Berry, A., and Rothmel, R. K. (1991). Alginate synthesis by Pseudomonas aeruginosa: a key pathogenic factor in chronic pulmonary infections of cystic fibrosis patients. *Clinical microbiology reviews*, 4(2):191–206.
- Mazzocco, A., Waddell, T. E., Lingohr, E., and Johnson, R. P. (2009a). *Enumeration of bacteriophages* by the direct plating plaque assay., volume 501.
- Mazzocco, A., Waddell, T. E., Lingohr, E., and Johnson, R. P. (2009b). Enumeration of bacteriophages using the small drop plaque assay system. *Methods in molecular biology (Clifton, N.J.)*, 501:81–85.
- McCrea, K. W., Xie, J., LaCross, N. C., Patel, M., Mukundan, D., Murphy, T. F., Marrs, C. F., and Gilsdorf, J. R. (2008). Relationships of nontypeable Haemophilus influenzae strains to hemolytic and nonhemolytic Haemophilus haemolyticus strains. *Journal of clinical microbiology*, 46(2):406–16.
- McCusker, M. P., Turner, E. C., and Dorman, C. J. (2008). DNA sequence heterogeneity in Fim tyrosine-integrase recombinase-binding elements and functional motif asymmetries determine the directionality of the fim genetic switch in Escherichia coli K-12. *Molecular Microbiology*, 67(1):171–187.
- 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. *Journal of clinical microbiology*, 41(4):1623–36.
- Meisel, A., Bickle, T. a., Krüger, D. H., and Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, 355(6359):467–9.

- Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., and Lenski, R. E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science (New York, N.Y.)*, 335(6067):428–32.
- Millard, A., Clokie, M. R. J., Shub, D. a., and Mann, N. H. (2004). Genetic organization of the psbAD region in phages infecting marine Synechococcus strains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):11007–12.
- Mills, S., Shanahan, F., Stanton, C., Hill, C., Coffey, A., and Ross, R. P. (2013). Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. *Gut microbes*, 4(1):4–16.
- Mira, A., Ochman, H., and Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. Trends in Genetics, 17(10):589–596.
- Moghaddam, S. J., Ochoa, C. E., Sethi, S., and Dickey, B. F. (2011). Nontypeable Haemophilus influenzae in chronic obstructive pulmonary disease and lung cancer. *International journal of chronic obstructive pulmonary disease*, 6:113–23.
- Moldovan, R., Chapman-McQuiston, E., and Wu, X. L. (2007). On kinetics of phage adsorption. Biophysical journal, 93(1):303–15.
- Morgan, G. J., Hatfull, G. F., Casjens, S., and Hendrix, R. W. (2002). Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. *Journal of molecular biology*, 317(3):337–59.
- Morton, D. J., Madore, L. L., Smith, A., Vanwagoner, T. M., Seale, T. W., Whitby, P. W., and Stull, T. L. (2005). The heme-binding lipoprotein (HbpA) of Haemophilus influenzae: role in heme utilization. *FEMS microbiology letters*, 253(2):193–9.
- Moxon, E. R., Bayliss, C., and Hood, D. W. (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annual review of genetics*, 40:307–33.
- Moxon, E. R. and Murphy, P. A. (1978). Haemophilus influenzae bacteremia and meningitis resulting from survival of a single organism. Proceedings of the National Academy of Sciences of the United States of America, 75(3):1534–1536.
- Moxon, E. R., Rainey, P. B., Nowak, M. a., and Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current biology* : CB, 4(1):24–33.
- Muniesa, M., Hammerl, J. A., Hertwig, S., Appel, B., and Brüssow, H. (2012). Shiga toxin-producing Escherichia coli O104:H4: a new challenge for microbiology. *Applied and environmental microbiology*, 78(12):4065–73.
- Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., and van Sinderen, D. (2013). Bacteriophage orphan

DNA methyltransferases: insights from their bacterial origin, function, and occurrence. Applied and environmental microbiology, 79(24):7547–55.

- Murray, N. E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiology and molecular biology reviews : MMBR*, 64(2):412–34.
- Naito, T., Kusano, K., and Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. Science (New York, N.Y.), 267(5199):897–9.
- Nassif, X., Bourdoulous, S., Eugène, E., and Couraud, P.-O. (2002). How do extracellular pathogens cross the blood-brain barrier? *Trends in microbiology*, 10(5):227–32.
- Nobelprize.org (2014). The Nobel Prize in Physiology or Medicine 1978.
- Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., and Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*, 21(6):528–34.
- Nyengaard, N. R., Falkenberg-Klok, J., and Josephsen, J. (1996). Cloning and analysis of the restriction-modification system LlaBI, a bacteriophage resistance system from Lactococcus lactis subsp. cremoris W56. Applied and environmental microbiology, 62(9):3494–8.
- Otczyk, D. C., Clancy, R. L., and Cripps, A. W. (2011). Haemophilus influenzae and smoking-related obstructive airways disease. *International journal of chronic obstructive pulmonary disease*, 6:345–351.
- Pal, C., Maciá, M. D., Oliver, A., Schachar, I., and Buckling, A. (2007). Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, 450(7172):1079–1081.
- Palmer, M. E. and Lipsitch, M. (2006). The Influence of Hitchhiking and Deleterious Mutation Upon Asexual Mutation Rates. *Genetics*, 173(1):461–472.
- Panek, H. and O'Brian, M. R. (2002). A whole genome view of prokaryotic haem biosynthesis. *Micro-biology (Reading, England)*, 148(Pt 8):2273–82.
- Pang, B., Winn, D., Johnson, R., Hong, W., West-Barnette, S., Kock, N., and Swords, W. E. (2008). Lipooligosaccharides containing phosphorylcholine delay pulmonary clearance of nontypeable Haemophilus influenzae. *Infection and immunity*, 76(5):2037–43.
- Papi, A., Bellettato, C. M., Braccioni, F., Romagnoli, M., Casolari, P., Caramori, G., Fabbri, L. M., and Johnston, S. L. (2006). Infections and airway inflammation in chronic obstructive pulmonary disease severe exacerbations. *American journal of respiratory and critical care medicine*, 173(10):1114– 21.
- Patel, I. S., Seemungal, T. A. R., Wilks, M., Lloyd-Owen, S. J., Donaldson, G. C., and Wedzicha,

J. A. (2002). Relationship between bacterial colonisation and the frequency, character, and severity of COPD exacerbations. *Thorax*, 57(9):759–64.

- Pelicic, V., Reyrat, J. M., and Gicquel, B. (1996). Expression of the Bacillus subtilis sacB gene confers sucrose sensitivity on mycobacteria. *Journal of bacteriology*, 178(4):1197–9.
- 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. *Clinical microbiology reviews*, 13(2):302–317.
- Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. a., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. a., and Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature*, 409(6819):529–33.
- Pimentel, D., Nagel, W., and Madden, J. (1963). Space-time structure of the environment and the survival of parasite-host systems. *American Naturalist*, 97(894):141–167.
- Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: Structure and mechanism. *Cellular and Molecular Life Sciences*, 62(6):685–707.
- Pingoud, A. and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. Nucleic acids research, 29(18):3705–27.
- Pingoud, A., Wilson, G. G., and Wende, W. (2014). Type II restriction endonucleases-a historical perspective and more. *Nucleic acids research*, 42(12):7489–527.
- Pirone-Davies, C., Hoffmann, M., Roberts, R. J., Muruvanda, T., Timme, R. E., Strain, E., Luo, Y., Payne, J., Luong, K., Song, Y., Tsai, Y.-C., Boitano, M., Clark, T. A., Korlach, J., Evans, P. S., and Allard, M. W. (2015). Genome-wide methylation patterns in Salmonella enterica Subsp. enterica Serovars. *PloS one*, 10(4):e0123639.
- Pittman, M. (1931). Variation and type specificity in the bacterial species Hemophilus influenzae. The Journal of experimental medicine, 53(4):471–92.
- Plunkett, G., Rose, D. J., Durfee, T. J., and Blattner, F. R. (1999). Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product. *Journal of bacteriology*, 181(6):1767–78.
- Poje, G. and Redfield, R. (2003). Transformation of Haemophilus influenzae, volume 71. Humana Press.

- Pommier, T., Douzery, E. J. P., and Mouillot, D. (2012). Environment drives high phylogenetic turnover among oceanic bacterial communities. *Biology Letters*, 8(4):562–566.
- Poole, J., Foster, E., Chaloner, K., Hunt, J., Jennings, M. P., Bair, T., Knudtson, K., Christensen, E., Munson, R. S., Winokur, P. L., and Apicella, M. A. (2013). Analysis of nontypeable haemophilus influenzae phase-variable genes during experimental human nasopharyngeal colonization. *The Journal* of infectious diseases, 208(5):720–7.
- Power, P. M., Sweetman, W. a., Gallacher, N. J., Woodhall, M. R., Kumar, G. a., Moxon, E. R., and Hood, D. W. (2009). Simple sequence repeats in Haemophilus influenzae. *Infection, genetics* and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 9(2):216–28.
- Public Health England (2015a). Laboratory reports of Haemophilus influenzae infection by serotype and year: England, 1990 to 2014.
- Public Health England (2015b). Laboratory reports of Haemophilus influenzae type b infection by age group and quarter: England, 1990 2014.
- Public Health England (2016). Laboratory reports of Haemophilus influenzae by age group and serotype, England and Wales, October to December 2015. *Infection Reports*, 10(8).
- Quesada, J. M., Soriano, M. I., and Espinosa-Urgel, M. (2012). Stability of a pseudomonas putida KT2440 bacteriophage-carried genomic island and its impact on rhizosphere fitness. *Applied and Environmental Microbiology*, 78(19):6963–6974.
- Ramsey, D. M. and Wozniak, D. J. (2005). Understanding the control of Pseudomonas aeruginosa alginate synthesis and the prospects for management of chronic infections in cystic fibrosis. *Molecular microbiology*, 56(2):309–22.
- Randall-Hazelbauer, L. and Schwartz, M. (1973). Isolation of the bacteriophage lambda receptor from Escherichia coli. *Journal of bacteriology*, 116(3):1436–46.
- Rao, D. N., Dryden, D. T. F., and Bheemanaik, S. (2014). Type III restriction-modification enzymes: a historical perspective. *Nucleic acids research*, 42(1):45–55.
- Ren, Z., Jin, H., Whitby, P. W., Morton, D. J., and Stull, T. L. (1999). Role of CCAA nucleotide repeats in regulation of hemoglobin and hemoglobin-haptoglobin binding protein genes of Haemophilus influenzae. *Journal of bacteriology*, 181(18):5865–70.
- Reynolds, C. B., Emerson, J. E., de la Riva, L., Fagan, R. P., and Fairweather, N. F. (2011). The Clostridium difficile cell wall protein CwpV is antigenically variable between strains, but exhibits conserved aggregation-promoting function. *PLoS pathogens*, 7(4):e1002024.

- Ricci, V. and Piddock, L. J. V. (2010). Exploiting the role of TolC in pathogenicity: identification of a bacteriophage for eradication of Salmonella serovars from poultry. *Applied and environmental microbiology*, 76(5):1704–6.
- Richter, C., Chang, J. T., and Fineran, P. C. (2012). Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. Viruses, 4(10):2291–311.
- Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K., Dryden, D. T. F., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Krüger, D. H., Lacks, S., Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P.-C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M. B., Wilson, G. G., and Xu, S.-y. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic acids research*, 31(7):1805–12.
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, 43(D1):D298–D299.
- Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasić, L., Thingstad, T. F., Rohwer, F., and Mira, A. (2009). Explaining microbial population genomics through phage predation. *Nature reviews. Microbiology*, 7(11):828–36.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics (Oxford, England)*, 16(10):944–5.
- Salaün, L., Linz, B., Suerbaum, S., and Saunders, N. J. (2004). The diversity within an expanded and redefined repertoire of phase-variable genes in Helicobacter pylori. *Microbiology (Reading, England)*, 150(Pt 4):817–30.
- Samson, J. E., Magadán, A. H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature reviews. Microbiology*, 11(10):675–87.
- Samuels, J. and Clarke, J. K. (1969). New bacteriophage of Haemophilus influenzae. Journal of virology, 4(5):797–8.
- Sanabria-Valentín, E., Colbert, M.-T. C., and Blaser, M. J. (2007). Role of futC slipped strand mispairing in Helicobacter pylori Lewisy phase variation. *Microbes and infection / Institut Pasteur*, 9(14-15):1553–60.

- Sandgren, a., Sjostrom, K., Olsson-Liljequist, B., Christensson, B., Samuelsson, A., Kronvall, G., and Henriques Normark, B. (2004). Effect of clonal and serotype-specific properties on the invasive capacity of Streptococcus pneumoniae. *The Journal of infectious diseases*, 189(5):785–96.
- São-José, C., Baptista, C., and Santos, M. A. (2004). Bacillus subtilis operon encoding a membrane receptor for bacteriophage SPP1. *Journal of bacteriology*, 186(24):8337–46.
- Saunders, N. J., Jeffries, A. C., Peden, J. F., Hood, D. W., Tettelin, H., Rappuoli, R., and Moxon, E. R. (2000). Repeat-associated phase variable genes in the complete genome sequence of Neisseria meningitidis strain MC58. *Molecular microbiology*, 37(1):207–15.
- Scanlan, P. D. and Buckling, A. (2012). Co-evolution with lytic phage selects for the mucoid phenotype of Pseudomonas fluorescens SBW25. *The ISME journal*, 6(6):1148–58.
- Scanlan, P. D., Hall, A. R., Blackshields, G., Friman, V.-P., Davis, M. R., Goldberg, J. B., and Buckling, A. (2015). Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations. *Molecular biology and evolution*, 32(6):1425–35.
- Schultz, I. J., Chen, C., Paw, B. H., and Hamza, I. (2010). Iron and porphyrin trafficking in heme biogenesis. *The Journal of biological chemistry*, 285(35):26753–9.
- Schweda, E. K. H., Richards, J. C., Hood, D. W., and Moxon, E. R. (2007). Expression and structural diversity of the lipopolysaccharide of Haemophilus influenzae: implication in virulence. *International journal of medical microbiology : IJMM*, 297(5):297–306.
- Scott, A. E., Timms, A. R., Connerton, P. L., Loc Carrillo, C., Adzfa Radzum, K., and Connerton, I. F. (2007). Genome dynamics of Campylobacter jejuni in response to bacteriophage predation. *PLoS pathogens*, 3(8):e119.
- Scott, J., Thompson-Mayberry, P., Lahmamsi, S., King, C. J., and McShan, W. M. (2008). Phageassociated mutator phenotype in group A streptococcus. *Journal of bacteriology*, 190(19):6290–301.
- Seed, K. D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLOS Pathogens*, 11(6):e1004847.
- Seed, K. D., Faruque, S. M., Mekalanos, J. J., Calderwood, S. B., Qadri, F., and Camilli, A. (2012). Phase variable O antigen biosynthetic genes control expression of the major protective antigen and bacteriophage receptor in Vibrio cholerae O1. *PLoS pathogens*, 8(9):e1002917.
- Seib, K. L., Peak, I. R. A., and Jennings, M. P. (2002). Phase variable restriction-modification systems in Moraxella catarrhalis. *FEMS immunology and medical microbiology*, 32(2):159–65.
- Sekulovic, O. and Fortier, L.-C. (2015). Global transcriptional response of Clostridium difficile carrying the CD38 prophage. Applied and environmental microbiology, 81(4):1364–74.

- Sekulovic, O., Ospina Bedoya, M., Fivian-Hughes, A. S., Fairweather, N. F., and Fortier, L.-C. (2015). The Clostridium difficile cell wall protein CwpV confers phase-variable phage resistance. *Molecular Microbiology*, 98(2):329–342.
- Shapiro, O. H., Kushmaro, A., and Brenner, A. (2010). Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *The ISME journal*, 4(3):327–36.
- Shin, H., Lee, J.-H., Kim, H., Choi, Y., Heu, S., and Ryu, S. (2012). Receptor diversity and host interaction of bacteriophages infecting Salmonella enterica serovar Typhimurium. *PloS one*, 7(8):e43392.
- Smith, A. L. (1987). Pathogenesis of Haemophilus influenzae meningitis. The Pediatric infectious disease journal, 6(8):783–6.
- Smith, H. O. and Wilcox, K. W. (1970). A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. *Journal of molecular biology*, 51(2):379–91.
- Sorek, R., Lawrence, C. M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem, 82:237–266.
- Sørensen, M. C. H., van Alphen, L. B., Harboe, A., Li, J., Christensen, B. B., Szymanski, C. M., and Brøndsted, L. (2011). Bacteriophage F336 recognizes the capsular phosphoramidate modification of Campylobacter jejuni NCTC11168. *Journal of Bacteriology*, 193(23):6742–6749.
- Srikhanta, Y. N., Fox, K. L., and Jennings, M. P. (2010). The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nature reviews. Microbiology*, 8(3):196–206.
- Srikhanta, Y. N., Gorrell, R. J., Steen, J. a., Gawthorne, J. a., Kwok, T., Grimmond, S. M., Robins-Browne, R. M., and Jennings, M. P. (2011). Phasevarion mediated epigenetic gene regulation in Helicobacter pylori. *PloS one*, 6(12):e27569.
- Srikhanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M., and Jennings, M. P. (2005). The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15):5547–51.
- Stachura, I., Mckinley, F. W., Leidy, G., and Alexander, H. E. (1969). Incomplete bacteriophage-like particles in ultraviolet-irradiated haemophilus. *Journal of bacteriology*, 98(2):818–20.
- Stern, A. and Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 33(1):43–51.
- Stojiljkovic, I. and Perkins-Balding, D. (2002). Processing of heme and heme-containing proteins by bacteria. DNA and cell biology, 21(4):281–95.

- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift Mutations and the Genetic Code.
- Stroeher, U. H., Karageorgos, L. E., Brown, M. H., Morona, R., and Manning, P. A. (1995). A putative pathway for perosamine biosynthesis is the first function encoded within the rfb region of Vibrio cholerae O1. *Gene*, 166(1):33–42.
- Sumby, P. and Smith, M. C. M. (2002). Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2). *Molecular microbiology*, 44(2):489–500.
- Sumby, P. and Smith, M. C. M. (2003). Phase variation in the phage growth limitation system of Streptomyces coelicolor A3 (2). *Journal of bacteriology*, 185(15):4558–4563.
- Summers, W. C. (2011). In the beginning..2. Bacteriophage, 1(1):50-51.
- Sun, X., Göhler, A., Heller, K. J., and Neve, H. (2006). The ltp gene of temperate Streptococcus thermophilus phage TP-J34 confers superinfection exclusion to Streptococcus thermophilus and Lactococcus lactis. *Virology*, 350(1):146–157.
- Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. Microbial ecology, 28(2):237–43.
- Suttle, C. A. (2005). Viruses in the sea. *Nature*, 437(7057):356–61.
- Suttle, C. A. (2007). Marine viruses-major players in the global ecosystem. Nature reviews. Microbiology, 5(10):801–12.
- Swords, W. E., Buscher, B. A., Ver Steeg Ii, K., Preston, A., Nichols, W. A., Weiser, J. N., Gibson, B. W., and Apicella, M. A. (2000). Non-typeable Haemophilus influenzae adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Molecular microbiology*, 37(1):13–27.
- Szczepankowska, A. (2012). Role of CRISPR/cas system in the development of bacteriophage resistance. Advances in virus research, 82:289–338.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution*, 30(12):2725–9.
- Tarr, P. I., Gordon, C. A., and Chandler, W. L. (2005). Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *Lancet (London, England)*, 365(9464):1073–86.
- Taubenberger, J. K., Hultin, J. V., and Morens, D. M. (2007). Discovery and characterization of the 1918 pandemic influenza virus in historical context. Antiviral Therapy, 12(4 B):581–591.
- Tauseef, I., Ali, Y. M., and Bayliss, C. (2013). Phase variation of PorA, a major outer membrane pro-

tein, mediates escape of bactericidal antibodies by Neisseria meningitidis. *Infection and immunity*, 81(4):1374–80.

Team, R. C. (2016). R: A Language and Environment for Statistical Computing.

- Terns, M. P. and Terns, R. M. (2011). CRISPR-based adaptive immune systems. Current opinion in microbiology, 14(3):321–7.
- Thiel, K. (2004). Old dogma, new tricks–21st Century phage therapy. *Nature biotechnology*, 22(1):31–36.
- Tock, M. R. and Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. *Current opinion in microbiology*, 8(4):466–72.
- Tong, H. H., Blue, L. E., James, M. A., Chen, Y. P., and DeMaria, T. F. (2000). Evaluation of phase variation of nontypeable Haemophilus influenzae lipooligosaccharide during nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infection and immunity*, 68(8):4593–7.
- Trouwborst, T. and de Jong, J. C. (1973). Interaction of some factors in the mechanism of inactivation of bacteriophage MS2 in aerosols. *Applied microbiology*, 26(3):252–7.
- Turner, S., Pryer, K. M., Miao, V. P., and Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *The Journal of eukaryotic microbiology*, 46(4):327–338.
- van der Ende, a., Hopman, C. T., and Dankert, J. (2000). Multiple mechanisms of phase variation of PorA in Neisseria meningitidis. *Infection and immunity*, 68(12):6685–90.
- van der Woude, M. W. (2011). Phase variation: how to create and coordinate population diversity. Current opinion in microbiology, 14(2):205–11.
- van der Woude, M. W. and Bäumler, A. J. (2004). Phase and antigenic variation in bacteria. Clinical microbiology reviews, 17(3):581–611, table of contents.
- Van Eldere, J., Slack, M. P. E., Ladhani, S., and Cripps, A. W. (2014). Non-typeable Haemophilus influenzae, an under-recognised pathogen. *The Lancet. Infectious diseases*, 14(12):1281–92.
- van Ham, S., van Alphen, L., Mooi, F. R., and van Putten, J. P. (1993). Phase variation of H. influenzae fimbriae: Transcriptional control of two divergent genes through a variable combined promoter region. *Cell*, 73(6):1187–1196.
- VanWagoner, T. M., Atack, J. M., Nelson, K. L., Smith, H. K., Fox, K. L., Jennings, M. P., Stull, T. L., and Smith, A. L. (2016). The modA10 phasevarion of nontypeable Haemophilus influenzae R2866 regulates multiple virulence-associated traits. *Microbial pathogenesis*, 92:60–7.

- Vasu, K. and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiology and molecular biology reviews : MMBR*, 77(1):53–72.
- Virji, M., Weiser, J. N., Lindberg, A. A., and Moxon, E. R. (1990). Antigenic similarities in lipopolysaccharides of Haemophilus and Neisseria and expression of a digalactoside structure also present on human cells. *Microbial pathogenesis*, 9(6):441–50.
- Vogel, U., Weinberger, A., Frank, R., Müller, A., Köhl, J., Atkinson, J. P., and Frosch, M. (1997). Complement factor C3 deposition and serum resistance in isogenic capsule and lipooligosaccharide sialic acid mutants of serogroup B Neisseria meningitidis. *Infection and immunity*, 65(10):4022–9.
- Wagner, P. L., Acheson, D. W., and Waldor, M. K. (2001). Human neutrophils and their products induce Shiga toxin production by enterohemorrhagic Escherichia coli. *Infection and immunity*, 69(3):1934–7.
- Waldron, D. E. and Lindsay, J. a. (2006). Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into Staphylococcus aureus and between S. aureus isolates of different lineages. *Journal of bacteriology*, 188(15):5578–85.
- Walkinshaw, M. D., Taylor, P., Sturrock, S. S., Atanasiu, C., Berge, T., Henderson, R. M., Edwardson, J. M., and Dryden, D. T. F. (2002). Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Molecular cell*, 9(1):187–94.
- Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., and Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nature Communications*, 1(9):147.
- Wang, Z., Bafadhel, M., Haldar, K., Spivak, A., Mayhew, D., Miller, B. E., Tal-Singer, R., Johnston, S. L., Ramsheh, M. Y., Barer, M. R., Brightling, C. E., and Brown, J. R. (2016). Lung microbiome dynamics in COPD exacerbations. *European Respiratory Journal*, 47(4):1082–1092.
- Weinbauer, M. G. and Rassoulzadegan, F. (2004). Are viruses driving microbial diversification and diversity? *Environmental Microbiology*, 6(1):1–11.
- Weiser, J. N., Love, J. M., and Moxon, E. R. (1989). The molecular mechanism of phase variation of H. influenzae lipopolysaccharide. *Cell*, 59(4):657–65.
- Weiser, J. N. and Pan, N. (1998). Adaptation of Haemophilus influenzae to acquired and innate humoral immunity based on phase variation of lipopolysaccharide. *Molecular microbiology*, 30(4):767–75.
- Weiser, J. N., Pan, N., McGowan, K. L., Musher, D., Martin, A., and Richards, J. (1998). Phosphorylcholine on the lipopolysaccharide of Haemophilus influenzae contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. *The Journal of experimental medicine*, 187(4):631–40.

- Weiser, J. N., Shchepetov, M., Chong, S. T., and Chong, S. T. H. (1997). Decoration of lipopolysaccharide with phosphorylcholine : a phase-variable characteristic of Haemophilus influenzae . Decoration of Lipopolysaccharide with Phosphorylcholine : a Phase-Variable Characteristic of Haemophilus influenzae. 65(3):943–950.
- White, D. C. and Granick, S. (1963). Hemin Biosynthesis in Haemophilus. *Journal of bacteriology*, 85(November):842–50.
- Wilhelm, S. W. and Suttle, C. a. (1999). Viruses and Nutrient Cycles in the Sea. BioScience, 49(10):781.
- Williams, B. J., Golomb, M., Phillips, T., Brownlee, J., Olson, M. V., and Smith, A. L. (2002). Bacteriophage HP2 of Haemophilus influenzae. *Journal of bacteriology*, 184(24):6893–905.
- Williams, H. T. P. (2013). Phage-induced diversification improves host evolvability. BMC evolutionary biology, 13:17.
- Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F. E., Silva, J., Tammadoni, S., Nosrat,
 B., Conrad, D., and Rohwer, F. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PloS one*, 4(10):e7370.
- Wilson, J. H., Luftig, R. B., and Wood, W. B. (1970). Interaction of bacteriophage T4 tail fiber components with a lipopolysaccharide fraction from Escherichia coli. *Journal of molecular biology*, 51(2):423–34.
- Xia, G., Corrigan, R. M., Winstel, V., Goerke, C., Gründling, A., and Peschel, A. (2011). Wall teichoic Acid-dependent adsorption of staphylococcal siphovirus and myovirus. *Journal of bacteriology*, 193(15):4006–9.
- Xu, S.-Y., Corvaglia, A. R., Chan, S.-H., Zheng, Y., and Linder, P. (2011). A type IV modificationdependent restriction enzyme SauUSI from Staphylococcus aureus subsp. aureus USA300. Nucleic acids research, 39(13):5597–610.
- Xu, S.-y., Nugent, R. L., Kasamkattil, J., Fomenkov, A., Gupta, Y., Aggarwal, A., Wang, X., Li, Z., Zheng, Y., and Morgan, R. (2012). Characterization of type II and III restriction-modification systems from Bacillus cereus strains ATCC 10987 and ATCC 14579. *Journal of bacteriology*, 194(1):49– 60.
- Yi, K., Rasmussen, A. W., Gudlavalleti, S. K., Stephens, D. S., and Stojiljkovic, I. (2004). Biofilm formation by Neisseria meningitidis. *Infection and immunity*, 72(10):6132–8.
- Young, R. E. B. and Hood, D. W. (2013). Haemophilus parainfluenzae has a limited core lipopolysaccharide repertoire with no phase variation. *Glycoconjugate journal*, 30(6):561–76.
- Zaleski, P., Wojciechowski, M., and Piekarowicz, A. (2005). The role of Dam methylation in phase

variation of Haemophilus influenzae genes involved in defence against phage infection. *Microbiology* (*Reading, England*), 151(Pt 10):3361–9.

- Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H., and O'Toole, G. A. (2009). Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of Pseudomonas aeruginosa. *Journal of bacteriology*, 191(1):210–9.
- Zhang, M., Li, Q., Zhang, X.-Y., Ding, X., Zhu, D., and Zhou, X. (2010). Relevance of lower airway bacterial colonization, airway inflammation, and pulmonary function in the stable stage of chronic obstructive pulmonary disease. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 29(12):1487–93.
- Zhang, X., McDaniel, A. D., Wolf, L. E., Keusch, G. T., Waldor, M. K., and Acheson, D. W. (2000). Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *The Journal of infectious diseases*, 181(2):664–70.
- Zhou, K., Aertsen, A., and Michiels, C. W. (2014). The role of variable DNA tandem repeats in bacterial adaptation. *FEMS microbiology reviews*, 38(1):119–41.