

Northumbria Research Link

Citation: Duignan, Libby A. M. (2021) The temperate bacteriophages of *Pseudomonas aeruginosa* and their role in chronic lung infection. Doctoral thesis, Northumbria University.

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/48237/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

**The Temperate Bacteriophages of
Pseudomonas aeruginosa and Their
Role in Chronic Lung Infection**

Libby Duignan

PhD

2021

The Temperate Bacteriophages of
Pseudomonas aeruginosa and Their
Role in Chronic Lung Infection

Libby A M Duignan

A thesis submitted in partial fulfilment of
the requirements of the University of
Northumbria at Newcastle for the degree
of Doctor of Philosophy

Research undertaken in the Faculty of
Health and Life Sciences

January 2021

*Dedicated in loving memory of Evelyn McAllister, my Granny,
the most wonderful person I have ever known, without whose
love and support I would never be where I am today.*

1933 -2019

Abstract

Pseudomonas aeruginosa (Pa) is an opportunistic pathogen common in respiratory infections of both cystic fibrosis (CF) and bronchiectasis (BR) patients and is associated with reduced lung function and poor clinical outcome. Pa is proficient at establishing chronic infections of the lung due to its ability to adapt and evolve under environmental selection. It achieves this, in part, by being naturally competent, pliable and open to horizontal transfer of nucleic acids by conjugative plasmids and bacteriophages (phages) etc. The focus of this research is temperate phages, bacterial viruses that infect and integrate into the chromosome of their host bacteria either being replicated like any other genetic loci or induced to a replicative state that lyses the cell, allowing release for further bacterial infection. In their integrated, prophage state, their diversity and gene carriage offer altered cell function. This work illustrates that temperate phages are extremely common in Pa isolated from the lung, offering diversity that can aid bacterial fitness and virulence through the subversion and modification of bacterial metabolic function.

Using LC-MS and metabolomics analysis this research demonstrates that each of the Liverpool Epidemic Strain (LES) prophages subvert cell function and alter their Pa host's metabolism. This was then broadened to include studies focussing on the phages isolated from different aetiologies of chronic respiratory infection to see if these metabolic changes occur within a larger panel of bacteriophages. This determined that whilst there are limitations in aligning function in a high number of metabolites found using this approach, there are phage-mediated differences between lysogen and wild-type bacterial host, where a core shift in metabolism was observed in Pa converted with phages from different arms of these studies. Importantly, it shows that temperate phages play an intrinsic role in altering bacterial cell metabolism.

To complement this work there is focus on the wider genetic diversity of Pa prophages, which allowed a collaboration with Prof. Roger Levesque in Laval, Canada and access to the International *Pseudomonas* Consortium Database. From >1000 Pa genomes 8 groups of phages clustered by type were determined (e.g. F10-like phage). Lysogens were

created in PA14 with individual phages from each (4 out of 8 groups). Host virulence genes across the International Pseudomonas aeruginosa consortium database (IPCD) phages was also investigated illustrating carriage across the panel, something that is not this widespread when reported in other bacterial backgrounds.

This research identifies possible transfer between Pa morphology variants, also different MLST types, isolated from the same patient at the same timepoint, something that has been hypothesised, but not shown prior to this study. However, this only shows a single time-point and snapshot time and an important focus must be the longitudinal relationship between bacterium and phage, evoking the well-named 'evolutionary arms race' and promotion of selection in the chronic lung. This longitudinal relationship was studied by surveying the carriage of Pa prophage in the chronically infected lung of BR patients, which illustrates both genome expansion and reduction in these virus genomes over time. We focus further on this genetic shift/drift in these longitudinal samples. Cross infection of phages induced from these isolates did not offer the host restriction over time we envisaged, which further complexes the story of phage sensitivity over time.

As a whole, this research demonstrates the complexity temperate bacteriophages add to the diversity of Pa in the chronic lung, influencing adaptation and evolution that promotes continuing colonisation. Lysogens derived from a lab strain of Pa and BR/CF phages illustrated lowered pathogenicity in the *Galleria wax* moth larval model compared to the wild-type host. This is a trait that is seen in the CF/BR lung with lowered immunogenicity of Pa in later stages of CF/BR chronic respiratory infection. This work provides the foundations and hypotheses to further explore the phage-bacterial relationship in the lung and its role in persistence and disease.

Table of contents

1.	INTRODUCTION.....	1
1.1	NON-CYSTIC FIBROSIS BRONCHIECTASIS.....	1
1.1.1	Non-Cystic-Fibrosis-Bronchiectasis and its contributing factors.....	1
1.1.2	Microbial colonisation in Bronchiectasis.....	3
1.1.3	Bronchiectasis relationship to Cystic Fibrosis.....	3
1.2	CYSTIC FIBROSIS.....	4
1.3	COMMON CF AND BR PATHOGENS.....	8
1.3.1	Bacterial Pathogens.....	8
1.3.2	Fungal and viral pathogens.....	8
1.4	<i>PSEUDOMONAS AERUGINOSA</i>	9
1.4.1	Virulence in CF and BR.....	10
1.4.2	Surface structures.....	10
1.4.3	Biofilm production.....	13
1.4.4	<i>Pseudomonas aeruginosa</i> Genomics.....	14
1.4.5	Mobile genetic elements of <i>Pseudomonas aeruginosa</i>	17
1.5	BACTERIOPHAGES.....	17
1.5.1	Introduction to bacteriophages.....	17
1.5.2	Phage classification.....	18
1.5.3	Phage structure and morphology.....	21
1.5.3.1	Caudovirales.....	22
1.5.3.2	Inoviridae.....	22
1.5.3.3	Capsid and packaging of gDNA.....	22
1.5.3.4	Tail fibre.....	23
1.5.4	Mosaicism and phage heterogeneity.....	25
1.5.5	Phage conversion.....	25
1.5.6	Bacteriophages of <i>Pseudomonas aeruginosa</i>	26
1.5.7	The life-cycle of bacteriophages.....	27
1.5.7.1	Phage adsorption and infection.....	28
1.5.7.2	Lytic phages.....	29
1.5.7.3	Lysogenic phages.....	29
1.5.7.3.1	The lysogeny decision.....	34
1.5.7.3.2	Prophage.....	34

1.5.8	Mechanisms of phage resistance	35
1.5.8.1	Inhibition of adsorption	35
1.5.8.2	Superinfection exclusion (Sie).....	35
1.5.8.3	The abortive infection systems (Abis).....	36
1.5.8.4	CRISPR/cas	36
1.5.9	Lysogenic conversion	37
2	MATERIALS AND METHODS	40
2.1	MATERIALS AND GROWTH MEDIA CONSTITUENTS.....	40
2.1.1	Sterilisation	40
2.1.2	Media	40
2.1.3	<i>Pseudomonas aeruginosa</i> bacterial growth media.....	41
2.1.4	Storage and maintenance of bacterial cells.....	41
2.2	BACTERIAL STRAINS	41
2.2.1	Quantification of viable bacteria in a sample	42
2.3	<i>PSEUDOMONAS AERUGINOSA</i> CULTURE PARAMETERS	42
2.3.1	<i>Pseudomonas</i> culture in liquid broth	42
2.3.2	Use of artificial sputum medium (ASM) to culture bacteria	42
2.4	BACTERIAL MOTILITY ASSAY	43
2.4.1	Twitching motility	43
2.5	GENERAL BACTERIOPHAGE PROTOCOLS	43
2.5.1	Temperate Phage induction of the lytic life cycle.....	43
2.5.2	Separation of free infective bacteriophages from bacterial host.....	43
2.5.3	Standard plaque assay	43
2.5.4	Spot assay.....	44
2.5.5	Bacteriophage cross-infections and bacterial sensitivity.....	44
2.5.6	Lysogen preparation	44
2.6	DNA EXTRACTION AND PREPARATION	45
2.6.1	Bacterial Phenol chloroform DNA extraction method.....	45
2.6.2	Bacterial genomic miniprep kit method	45
2.6.3	Phage isolation and DNA extraction	46
2.6.4	QUBIT 2.0	46
2.7	GENERAL PCR METHODOLOGY	46
2.7.1	Crude DNA preparation using boiled bacterial colonies.....	46
2.7.2	Standard polymerase chain reaction (PCR).....	47
2.7.3	Visualisation of PCR products	49
2.8	GENOME SEQUENCING.....	49
2.9	<i>IN-VIVO</i> ANIMAL MODEL	49
2.9.1	<i>Galleria mellonella</i> larvae model	49

2.10	LCMS METABOLOMICS PREPARATION AND RUN	50
2.10.1	Metabolomic preparation	50
2.10.2	Ultra High Pressure Liquid chromatography (UHPLC) or LCMS	50
2.11	BIOINFOMATIC TOOLS FOR ANALYSIS	50
2.11.1	PHAge Search Tool Enhanced Release (PHASTER)	50
2.11.2	MUMmer	51
2.11.3	BLAST.....	51
2.11.4	Prokka	51
2.11.5	Roary	52
2.11.6	Saturn V.....	52
2.12	VISUALISATION TOOLS.....	52
2.12.1	Artemis and Artemis comparison tool (ACT)	52
2.12.2	Mauve	53
2.12.3	Circos.....	53
2.12.4	Fig tree	53
2.13	METABOLOMIC ANALYSIS.....	53
3.	EVALUATION AND OPTIMISATION OF PROPHAGE IDENTIFICATION FROM <i>PSEUDOMONAS AERUGINOSA</i> GENOMES.....	55
3.1	INTRODUCTION.....	55
3.1.1	Prophages	55
3.1.2	Identification of prophages from bacterial genomes	56
3.2	AIMS.....	59
3.3	OBJECTIVES	59
3.4	RESULTS.....	60
3.4.1	Evaluation of prophage identification tools.....	60
3.4.2	Accuracy and assurance.....	61
3.5	DISCUSSION.....	64
3.6	SUMMARY	65
4.	THE CARRIAGE, DIVERSITY AND TRANSFER OF PROPHAGES IDENTIFIED BETWEEN <i>PSEUDOMONAS AERUGINOSA</i> STRAINS WITHIN THE CHRONICALLY INFECTED LUNG	66
4.1	INTRODUCTION.....	66
4.1.1	Non-clonal bacterial populations within chronically infected lungs.....	67
4.1.2	<i>P. aeruginosa</i> genomes from BR patients.....	67
4.1.3	Carriage of prophages by <i>P. aeruginosa</i>	68
4.1.4	Temperate phage insertion as a source of genetic variation for evolution.....	69
4.1.5	Transduction of temperate phages between Pa and co-infection.....	70
4.2	AIMS.....	70
4.3	OBJECTIVES	71

4.4	RESULTS.....	71
4.4.1	Identification of prophages within Pa genomes	71
4.4.2	Integration points of prophage	74
4.4.3	Comparison of phage carriage across each isolate.....	79
4.4.4	Induction, DNA extraction and sequencing of phages from isolates to unequivocally identify the mobilisable phages.....	97
4.5	DISCUSSION.....	99
4.5.1	Phenotypic differences in Pa from the chronically infected lung	99
4.5.2	Bacteriophage identification and integration	100
4.5.3	Prophage carriage and comparison between co-colonising Pa isolates	101
4.6	SUMMARY	102
5.	WHAT ROLE DO TEMPERATE PHAGES PLAY IN SUBVERTING THE METABOLIC FUNCTION OF THEIR <i>P. AERUGINOSA</i> HOST?.....	103
5.1	INTRODUCTION.....	103
5.1.1	Effects temperate phages can have on their bacterial hosts.....	104
5.1.2	LES bacteriophages	104
5.1.3	Metabolomics to analyse prophage-host interactions	106
5.1.4	Changes in virulence caused by temperate phages in <i>P. aeruginosa</i> in chronic lung infections	108
5.2	AIMS.....	108
5.3	OBJECTIVES	109
5.4	RESULTS.....	109
5.4.1	Metabolomics results for PAO1 and LES phage lysogens	109
5.4.2	Metabolomics results for lysogens of PAO1 and clinical <i>P. aeruginosa</i> lysogenic phages from BR and CF patients	121
5.4.3	Virulence testing of PAO1 lysogens containing the clinical <i>P.aeruginosa</i> phages from BR and CF patients	133
5.5	DISCUSSION.....	138
5.5.1	Subversion of the PAO1 metabolism caused by LES prophages	138
5.5.2	Effects of temperate phage communities from Pa isolated from different aetiologies	142
5.5.3	Temperate phage effects of the virulence of Pa.....	145
5.6	SUMMARY	146
6	GENOME COMPARISON OF TEMPERATE PHAGES FROM THE INTERNATIONAL <i>PSEUDOMONAS</i> CONSORTIUM DATABASE.....	148
6.1	INTRODUCTION.....	148
6.1.1	The International <i>Pseudomonas</i> Consortium Database – IPCD	148
6.1.1.1	Why are pan-genome studies of <i>Pseudomonas aeruginosa</i> important?	149
6.1.2	The carriage of virulence factors by phages	150

6.2	AIMS.....	151
6.3	OBJECTIVES	151
6.4	RESULTS.....	151
6.4.1	Prophage identification from <i>P. aeruginosa</i> genomes in the IPCD.....	151
6.4.2	SaturnV diversity analysis of <i>Pseudomonas aeruginosa</i> prophage	164
6.5	DISCUSSION.....	168
6.5.1	Prophage Identification.....	168
6.5.2	Comparison of origin of Pa host and prophage carriage	169
6.5.3	Presence of virulence factors within Pa prophage genomes.....	170
6.5.4	Imaging of induced phages	172
6.5.5	Genetic diversity of Pa prophages identified from the IPCD	172
6.6	SUMMARY	175
7	METABOLOMIC AND FUNCTIONAL COMPARISON OF LYSOGENS CONVERTED WITH PHAGES FROM THE INTERNATIONAL <i>PSEUDOMONAS</i> CONSORTIUM DATABASE (IPCD).	177
7.1	INTRODUCTION.....	177
7.1.1	Importance of understanding the effects of temperate phages on their bacterial host on a large scale.....	177
7.2	AIMS.....	178
7.3	OBJECTIVES	178
7.4	RESULTS.....	179
7.4.1	Characterising bacteriophage gene frequency using the ROARY bioinformatic tool. Subsequent selection of target species from the IPCD panel that carry only one predicted inducible phage from each clade.	179
7.4.2	Induction and spot assay to test infectivity of phages.....	183
7.4.3	Formation and confirmation of PA14 lysogens containing phages from each clade	187
7.4.4	Metabolomic analysis and comparison of IPCD Phage-PA14 lysogens.....	189
7.4.5	Pathway analysis of metabolomics	197
7.5	DISCUSSION	200
7.5.1	Lysogen formation	200
7.5.2	Metabolic subversion of Pa by different types of prophage.....	202
7.6	SUMMARY.....	206
8	COMPARISON OF <i>PSEUDOMONAS AERUGINOSA</i> TEMPERATE BACTERIOPHAGES ISOLATED FROM CHRONICALLY INFECTED NON-CYSTIC FIBROSIS BRONCHIECTASIS PATIENT ISOLATES, OVER A PERIOD OF 10 YEARS.....	207
8.1	INTRODUCTION.....	207
8.1.1	Longitudinal studies of Pa from chronically infected BR patient lungs.....	207
8.1.2	Evolution of temperate phages in <i>Pseudomonas aeruginosa</i> over time	209
8.2	AIMS.....	209

8.3	OBJECTIVES	210
8.4	RESULTS.....	210
8.4.1	Selection of longitudinal Pa isolates to sequence.....	210
8.4.2	Genomic Diversity of longitudinal Pa isolates from BR patients.....	218
8.4.3	Prophage identification from the longitudinal Pa isolates from each BR patient	221
8.4.4	Phage Induction of longitudinal isolates and cross-infection and host range analysis	240
8.5	DISCUSSION.....	244
8.5.1	Changes to antibiotic resistance of Pa from longitudinally during chronic infection	244
8.5.2	Motility alteration to Pa from over the course of chronic infection.....	245
8.5.3	Analysis of prophage genomes from longitudinal Pa isolates to determine evolutionary adaptation over time	246
8.6	SUMMARY	249
9	GENERAL DISCUSSION.....	250
9.1	THE CONTEXT OF THIS THESIS.....	250
9.1.1	Temperate phage transfer in the chronic lung environment	250
9.1.2	Metabolomics approach to identify bacterial physiological changes caused by temperate phages.....	253
9.1.3.	Ability of prophages to change the virulence of their Pa host.....	257
9.1.4.	<i>Pseudomonas aeruginosa</i> prophage genetic diversity	259
9.1.5	Investigation of prophage diversity within Pa from the International <i>Pseudomonas</i> Consortium Database (IPCD).....	259
9.1.6	Evolution of prophages within Pa isolates over time in the chronically infected lung	262
9.2	FURTHER STUDY.....	264
10	REFERENCES	266
11	APPENDICES	289

List of Figures

Figure 1.1 - The effects of BR on the lungs	2
Figure 1.2 How mutated/ non-functioning CFTR channels cause dehydrated sticky mucus in patients with CF	5
Figure 1.3 - Age specific prevalence of respiratory organisms in cystic fibrosis patients	6
Figure 1.4 Circular map of the <i>P. aeruginosa</i> LES genome.	16
Figure 1.5 Diagram of the different types of bacteriophages.	19
Figure 1.6 Electron microscopy images of bacteriophage with different tail types.	19
Figure 1.7 Lambda phage tail assembly pathway.	24
Figure 1.8 Lytic and lysogenic life cycle. This schematic diagram illustrates the life cycle that temperate phage lambda can exhibit.	31
Figure 1.9. A short segment of the λ DNA molecule that controls the lysogeny decision.	32
Figure 4.1. The number of prophages identified using PHASTER in 13 Pa genomes	72
Figure 4.2a-f. MAUVE comparisons of prophage regions.	81
Figure 4.3 – An example of a Circos diagram	85
Figure 4.4a-c Circos comparison of phage genomes from Pa isolates C86 and C87 from patient 73	86
Figure 4.5a-c Circos comparison of phage genomes from Pa isolates C12 and C13 from patient 42.	87
Figure 4.6a-c Circos comparison of phage genomes from Pa isolates C107 and C108 from patient 84	88
Figure 4.7a-c Circos comparison of phage genomes from Pa isolates C109 and C110 from patient 85	89
Figure 4.8a-c Circos comparison of phage genomes from Pa isolates A100 and A106 from patient 148.	90
Figure 4.9a-c Circos comparison of phage genomes from Pa isolates C125, C127 and C129 from patient 92	91
Figure 5.1 – PCoA analysis of all LES lysogens's metabolomic profiles compared with one another and the control.	110
Figure 5.2 – Dendrogram and heatmap of all the LES lysogens' metabolomic profiles compared with one another and the control	111
Figure 5.3 – PCoA, dendrogram and heatmap of the LES2 lysogens metabolomic profiles compared to the control.	112
Figure 5.4 – PCoA, dendrogram and heatmap of the LES3 lysogens metabolomic profiles compared to the control	113
Figure 5.5– PCoA, dendrogram and heatmap of the LES4 lysogens metabolomic profiles compared to the control	114
Figure 5.6 – PCoA, dendrogram and heatmap of the LES2+3+4 lysogens metabolomic profiles compared to the control	115
Figure 5.7– PCoA, dendrogram and heatmap of the LES3+4 lysogens metabolomic profiles compared to the control	116

Figure 5.8– PCoA, dendrogram and heatmap of the LES3 lysogens metabolomic profiles compared to the control	118
Figure 5.9 - Pathway analysis of LES phage lysogens	122
Figure 5.10 – Metabolomics results of all 546 metabolites for all lysogens from each of the four clinical groups compared with one another and the control	123
Figure 5.11 – Metabolomics results of 142 streamlined metabolites for all lysogens from each of the four clinical groups compared with one another and the control.	124
Figure 5.12 – PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from the adult_CF clinical group compared with the PAO1 control	124
Figure 5.13 – PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from the paed_CF clinical group compared with the PAO1 control.	125
Figure 5.14 – PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from >10 BR_CF clinical group compared with the PAO1 control.	126
Figure 5.15 PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from <10 BR_CF clinical group compared with the PAO1 control	127
Figure 5.16- Pathway analysis of lysogens containing clinical Pa phages from four clinical groups	129
Figure 5.17a-e Virulence results for galleria model tests against 5 PAO1 lysogens clinical group - early stages of CF (aged under 16 years).	131
Figure 5.18a-e Virulence results for galleria model tests against 5 PAO1 lysogens clinical group - later stages of CF (aged over 16 years).	132
Figure 5.19a-e Virulence results for galleria model tests against 5 PAO1 lysogens clinical group - early stages of non-CF-BR (diagnosed >10 years before)	133
Figure 5.20a-e Virulence results for galleria model tests against 5 PAO1 lysogens clinical group - late stages of non-CF-BR (diagnosed <10 years before)	134
Figure 6.1 Distribution of prophages in Pa from IPCD identified by PHASTER	150
Figure 6.2 -The average number of prophages from each class (intact, questionable, and incomplete) per isolate depending on their origin.	151
Figure 6.3a-o – Crude transmission electron microscopy (TEM) images of induced phages from Pa isolates that are from the IPCD.	159
Figure 6.4 Comparison of the intact temperate phages identified in all the IPCD isolates using SaturnV.	163
Figure 7.1 – Electrophoresis gel showing the results of the PCR indicating phage infection.	184
Figure 7.2 – PCoA of the metabolic profile comparison of lysogens from 4 different clade (phage types) grown in ASM	191
Figure 7.3 – Metabolic comparison of all lysogens metabolic profiles from all clades compared against one another and the PA14	192
Figure 7.4 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 1 compared to PA14 control	193

Figure 7.5 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 2 compared to PA14 control	194
Figure 7.6 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 5 compared to PA14 control.	195
Figure 7.7 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 8 compared to PA14 control.	196
Figure 7.8 - Pathway analysis showing the ‘impact’ of the phages in each clade on the metabolic pathways compared to the PA14 control without phages	199
Figure 8.1 – Disease progression and sample collection of 6 BR patients.	212
Figure 8.2a-f Antibiotic resistance results of the longitudinal Pa isolates from BR patients A-F that were taken forward for sequencing	214
Figure 8.3 – Bacterial diversity of longitudinal Pa isolates from BR patients performed using ROARY pangenome analysis	220
Figure 8.4 - Carriage of prophages by Pa isolates collected longitudinally from chronically infected BR patients.	222
Figure 8.5 – Diversity of prophages identified in longitudinal Pa isolates from BR patient	225
Figure 8.6a – Prophage comparison from the prophages identified in Pa isolates from patient A	228
Figure 8.6b – Prophage comparison from the prophages identified in Pa isolates from patient B	229
Figure 8.6c – Prophage comparison from the prophages identified in Pa isolates from patient E	230
Figure 8.7 – Prophage comparison from the prophages identified in Pa isolates from patient C	232
Figure 8.8 – Prophage comparison from the prophages identified in Pa isolates from patient D	234
Figure 8.9 – Prophage comparison from the prophages identified in Pa isolates from patient F	236

List of Tables

Table 2.1 List of Media and their constituents	40
Table 2.2 Oligonucleotide primers used in this study	47
Table 3.1 Comparison of prophages identifications tools	61
Table 3.2 Comparison of PHASTER and VirSorter of indentifying LES phages from LESB58	63
Table 4.1 The integration positions of prophages identified from Pa genomes	75
Table 4.2 The length of each phages genome, percentage identity and coverage of each phage to the most similar other phage from the comparison isolate from the same patient.	92
Table 4.3 Phage lysate sequencing results of phages induced from Pa strains isolated from BR patients	96
Table 4.4 BLASTn comparison of sequenced phages induced from Pa strains isolated from BR patients	96
Table 5.1 Prophages identified in <i>Pseudomonas aeruginosa</i> strain LESB58.	103
Table 6.1 – Prophage carriage of virulence genes identified with genome comparison to the VFDB	154
Table 7.1 Most prevalent genes in phages from each clade from prophages of the IPCD isolates	176
Table 7.2 - <i>P. aeruginosa</i> isolates selected from the IPCD for phage induction	179
Table 7.3 Spot assay results for the phages selected and induced from the IPCD isolates	181
Table 8.1 – Twitching motility of longitudinal Pa isolates from BR patients	216
Table 8.2 - Gene annotation results from the prophage expansion region in Pa isolate 3 from patient A.	239
Table 8.3a – Cross-infection of Pa isolates from patient A against phages induced from them	242
Table 8.3b - Cross-infection of Pa isolates from patient B against phages induced from them	242
Table 8.3c - Cross-infection of Pa isolates from patient C against phages induced from them	243
Table 8.3d - Cross-infection of Pa isolates from patient D against phages induced from them	243
Table 8.3e - Cross-infection of Pa isolates from patient E against phages induced from them	244
Table 8.3f - Cross-infection of Pa isolates from patient F against phages induced from them	244

Acknowledgements

Firstly, I'd like to thank Dr Darren Smith wholeheartedly for giving me this opportunity and supporting me throughout this work. I'd also like to thank him for the constant reassurance in the practical work and tolerance with my thesis writing, also for many conferences that I've had the pleasure of attending during my PhD, which has helped me develop as a researcher. I will be forever grateful for the support he has given me and for always believing in me, even when I did not, and for the navy and conference stories, which cheered me up when everything was getting a little too much. All in all, I could not have asked for a better supervisor and academic Dad. I need to thank Northumbria University for funding my studentship and giving me the opportunity to move to, and fall in love with, the North East, which may be 'miles from anywhere' but I've come to learn that's a good thing.

I would like to thank Prof Roger Levesque for the opportunity to travel to Quebec for a research visit and to learn from him. It was a wonderful experience even without knowing French. I'd like to thank Roger and the whole Levesque lab for the warmest welcome, especially to Jeremie, whose patience and teaching was super. Without this opportunity, much of this work could not have been possible. I would like to extend my thanks to Dr Anthony de Soyza and Audrey Perry, who provided the patient samples that much of this work was based on. I also need to give thanks to Prof Craig Winstanley and Dr Jo Fothergill for not only opening my eyes to phage research during my MRes, which set me on my path towards a PhD, but also for collaborating with us throughout my PhD and now even employing me as a Post-doc.

This thesis would not be possible without the huge amount of support from my fellow Smith lab buddies, Adnan Tariq and Giles Holt, who were great mentors and friends to me throughout. Also, to my fellow PhD students, I really felt like we were in this together and a special mention goes to Louise Francis and Laura Duffy, whose unconditional friendship has really helped me over the last four years. I also want to thank to Greg Young and Lewis Cuthbertson, my big brothers, who supplied great advice, laughs and hugs when I needed it.

Most importantly, I need to thank my wonderful boyfriend Callum. Not only has he listened to me moan and been there through every high and low over the last 4 years, but he also had to do a PhD alongside me. I cannot be more grateful for the support and encouragement he has given me, which extends to his family, especially his mum and dad who have adopted me and made me feel right at home in the North East. I think I even support Sunderland now! I would also like to give a massive thanks to my friends far and wide, from Aberdeen to Liverpool, especially to Pip Robinson, my oldest and dearest friend who has been there for 3 hour phone calls whenever I needed her and has kept me sane through these 4 years.

Finally, thank you to my Granny Mac, who sadly I started this journey with but had to finish without. I need to thank her for her love, care and her phone calls on a Sunday to see how I was doing. She was my biggest motivation for everything I've done, so this one is for her. We got there!

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Name: Libby Duignan

Signature:

Date: 31/01/2021

List of Abbreviations

LIST OF ABBREVIATIONS

Abis	Abortive infection systems
ACLAME	A CLAssification of Mobile Genetic Elements
ACT	Artemis comparison tool
AMR	Antimicrobial resistance
ASM	Artificial Sputum Medium
BAVS	Bacterial and Archaeal Viruses Subcommittee
BLAST	Basic Local Alignment Search Tool
BR	Non-cystic-fibrosis-bronchiectasis
BTS	British Thoracic Society
CDS	Coding sequences
CF	Cystic fibrosis
CFTR	Cystic Fibrosis Transmembrane conductance Regulator
COPD	Chronic Obstructive Pulmonary Disease
CoA	Coenzyme A
CRISPR	Clustered regularly interspersed palindromic repeat
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
GSM	Glycerol Skimmed Milk
HGT	Horizontal Gene Transfer
ICTV	International Committee on Taxonomy of Viruses
IGPS	Indole-3-glycerol phosphate synthase
IPCD	International Pseudomonas Consortium Database
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Luria-Bertani
LC-MS	Liquid chromatography–mass spectrometry
LES	Liverpool Epidemic Strains
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization-Time Of Flight

MDR	Multi-Drug Resistant
MGEs	Mobile Genetic Elements
MLST	Multi-Locus Sequence Typing
Mr	Molecular mass
NFLX	Norfloxacin
OR	Operator Region
Pa	<i>Pseudomonas aeruginosa</i>
PAMDB	<i>Pseudomonas aeruginosa</i> metabolite database
PCoA	Principal Components Analysis
PCR	Polymerase Chain Reaction
PES	Prairie Epidemic Strains
Pf	Filamentous phages
p.f.u.	plaque forming units
PHASTER	PHAge Search Tool Enhanced Release
pl	promoter integration
pL	promoter leftward
PLS-DA	Partial Least Squares - Discriminant Analysis
pR	promoter rightward
pR''	second rightward promoter
pRE	promoter repressor establishment
pRM	promoter repressor maintenance
Sie	Superinfection exclusion
SMP	Skimmed Milk Powder
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
TMP	Thiamine monophosphate
TPP	Thiamine pyrophosphate
UHPLC	Ultra High Pressure Liquid chromatography
vAMGs	virus Auxiliary Metabolic Genes

VFDB	Virulence factor database
VIP	Variable Importance in Projection

1. INTRODUCTION

1.1 NON-CYSTIC FIBROSIS BRONCHIECTASIS

1.1.1 Non-Cystic-Fibrosis-Bronchiectasis and its contributing factors

Non-cystic-fibrosis-bronchiectasis (BR) is a long-term condition caused by the abnormal widening of the airways in the lungs, first described in 1891 by René Laënnec (1819; Roguin, 2006). The pathophysiology of the disease leads to a build-up of excessive sticky mucus that can make the lungs more vulnerable to opportunistic bacterial infection (Figure 1.1) (Pasteur et al., 2010). BR is symptomatically similar to cystic fibrosis (CF) (described in 1.2). However, BR is caused by the bronchial tree undergoing events of localised dilation and inflammation that is not genetically inherited (ten Hacken et al., 2007), unlike CF, which is an autosomal recessive genetic disorder affecting multiple organs from birth. BR can be linked to lung trauma, or in some cases an allergic response that allows ensuing bacterial infection (Athanasio, 2012). The onset of BR is multifactorial, and the causation is sometimes unclear and termed idiopathic BR. The lack of mucociliary clearance has been linked to autoimmunity and epithelial damage (Morillas et al., 2007). The incidence rate of BR in the UK is 1 in 1000 adults and 1 in 10000 children (NHS.uk, 2016). It is predominately a disease of the elderly, with the majority of BR patients in the UK being over 60 years of age (Neves et al., 2011).

The prevalence of BR in the UK is frequently underestimated, and the disease is perceived by many to be rare and easily managed (De Soyza et al., 2014). However, the incidence is much higher than CF with reports in 2012 showing that ~210,000 people in the UK were living with bronchiectasis (Strachan et al., 2014), which is four times higher than the estimate published by the NHS for CF (NHS.uk, 2016), which reported only ~10,000 cases. Even though incidence is higher, it is still perceived as a rare disease and is commonly overlooked (De Soyza et al., 2014). Therefore, research into BR has been limited with no specific drug therapies currently available for BR patients (Pasteur et al., 2010). The first diagnostic and treatment guidelines for BR were only developed in the UK in 2010 by the British Thoracic Society (BTS) (Pasteur et al., 2010), further highlighting that research into BR has been limited. Patients were previously being misdiagnosed due to the severity of symptoms and

crossover of symptoms with other respiratory diseases, such as chronic obstructive pulmonary disease (COPD).

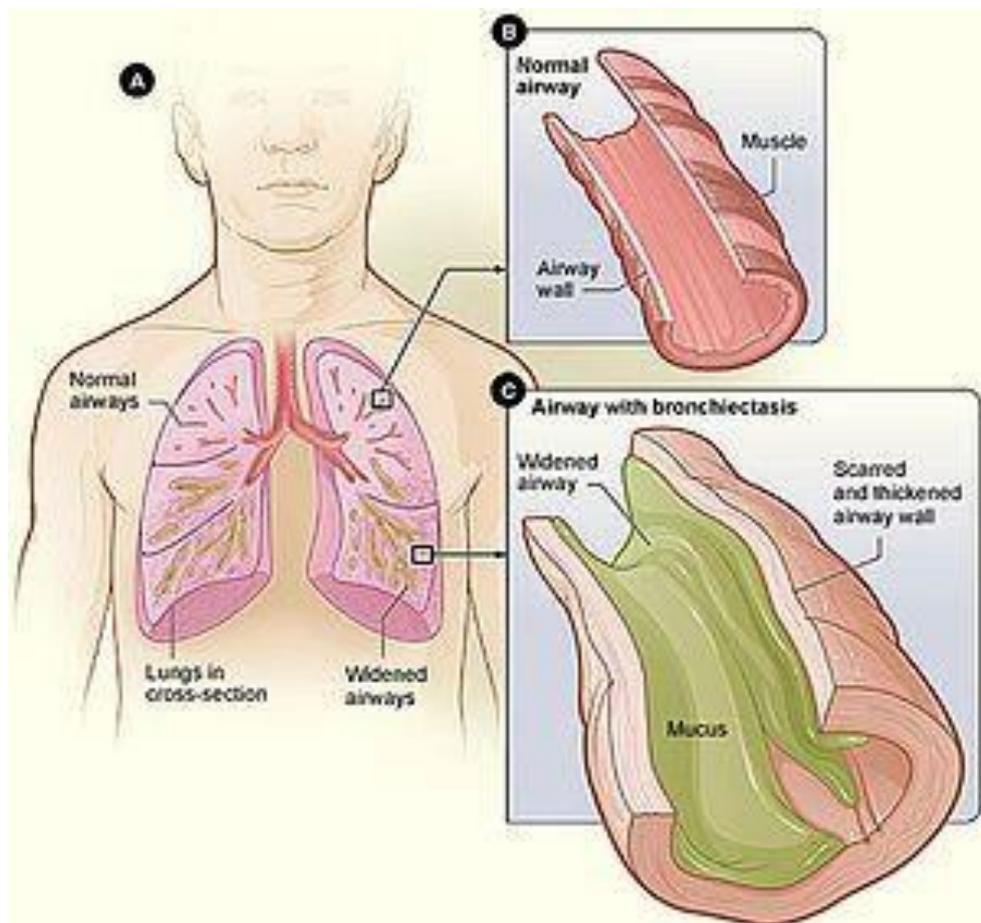


Figure 1.1 - The effects of BR on the lungs. A) A comparison between a cross section of healthy lung and the lung of a BR patient. B) Normal bronchiole walls in a healthy lung. C) The widened bronchiole, thickened walls and build-up of thick mucus in a BR lung. Illustration courtesy of NIH/NHLBI (NHLBI, 2018). Copyright approved.

1.1.2 Microbial colonisation in Bronchiectasis

The lungs of BR patients are commonly colonised by multiple bacterial species within a community, termed poly-microbial, shown by numerous studies using 16S rRNA gene amplification and bacterial community sequencing (Willner et al., 2009, Pragman et al., 2012, Guss et al., 2011). The key bacterial species involved in these infections usually include *Pseudomonas aeruginosa*, *Haemophilus influenza* and *Streptococcus pneumonia* (Angrill et al., 2002). However, other species are seen including *Klebsiella pneumonia* (Bopaka et al., 2015) and *Burkholderia cepacia* (Hauser and Orsini, 2015). All these microbes are opportunistic in nature and are able to colonise the BR lungs due to the nutrient rich environment. The BR lung environment is favourable to colonisation by these bacterial species due to a failure in host defences to maintain sterility of the respiratory tract, caused by the over production of mucus and impairment of mucociliary clearance. The colonisation then causes a host-mediated chronic inflammatory response to occur. (Wilson et al., 1985). This, in conjunction with bacterial factors, such as adherence (Plotkowski et al., 1994), biofilm formation (Ohgaki, 1994) and avoidance of phagocytosis (Skerrett et al., 2007), aids colonisation of the lower respiratory tract and impairs antibiotic efficacy (Ohgaki, 1994).

1.1.3 Bronchiectasis relationship to Cystic Fibrosis

Both BR and CF have been shown to be poly-microbial in nature, which means that these different bacterial species can interact and influence community structure. However, there is usually a dominant species present in the lungs (Cox et al., 2010, Purcell et al., 2014, Silby et al., 2011). A study by Duff et al. (2013) determined through culture and targeted 16S rRNA gene bacterial community sequencing that the BR lung is more complex in community structure as they have greater species diversity compared to the CF lung. However, the use of multiple antibiotics and a small sample size made it difficult to identify whether the observed levels of diversity were correlated with acute and chronic antibiotic use, patient age or exacerbation status (Duff et al., 2013). Much of the knowledge on the management and treatment of BR is based on CF, due to the similarity in pathophysiology

and symptoms. However, CF is recognised predominantly as a disease of children and young adults, whereas BR affects mainly the elderly, which adds complexity in their treatment compared to younger patients. This is mainly due to the addition of age-related conditions such as cognitive impairment or co-morbidities, which can independently affect adherence to poly-pharmacological treatments (Bellelli et al., 2016).

1.2 CYSTIC FIBROSIS

Cystic fibrosis (CF) is a genetic disorder affecting more than 10,400 people in the UK (CFTrust, 2018b). There is currently no cure for CF but a number of treatments are available to help control the symptoms and prevent complications of bacterial infections that relate to the pathophysiology of the disorder (NHS.uk, 2016). CF is caused by mutations in the gene encoding for the cystic fibrosis transmembrane conductance regulator (CFTR), a transmembrane ion channel protein required for chloride ion transport across epithelial cell membranes (Bear et al., 1992). CF is known as a monogenic disease as it is associated to mutation of a single gene. Over 1900 mutations have been reported within the CFTR gene, where each mutation can be assigned into six classes that determine disease severity (Bonadia et al., 2014). These mutations lead to the CFTR protein not functioning as efficiently, not being produced in sufficient quantities or not produced at all. The malfunctioning or missing CFTR proteins cause an ionic imbalance (Hopf et al., 1999), which leads to the production of abnormally thick, sticky mucus (Figure 1.2). This affects a number of systems within the body, notably the respiratory, pancreatic and the digestive systems (O'Sullivan and Freedman, 2009) but also the endocrine (Marshall et al., 2005) and the reproductive systems (Hull and Kass, 2000). Within these anatomical sites, especially the lungs, the ion imbalance leads to the production of abnormal thickening of the mucus through dehydration. This mucus builds up and provides a nutrient rich environment that opportunistic bacteria can colonise, resulting in frequent lung infections (Figure 1.3).

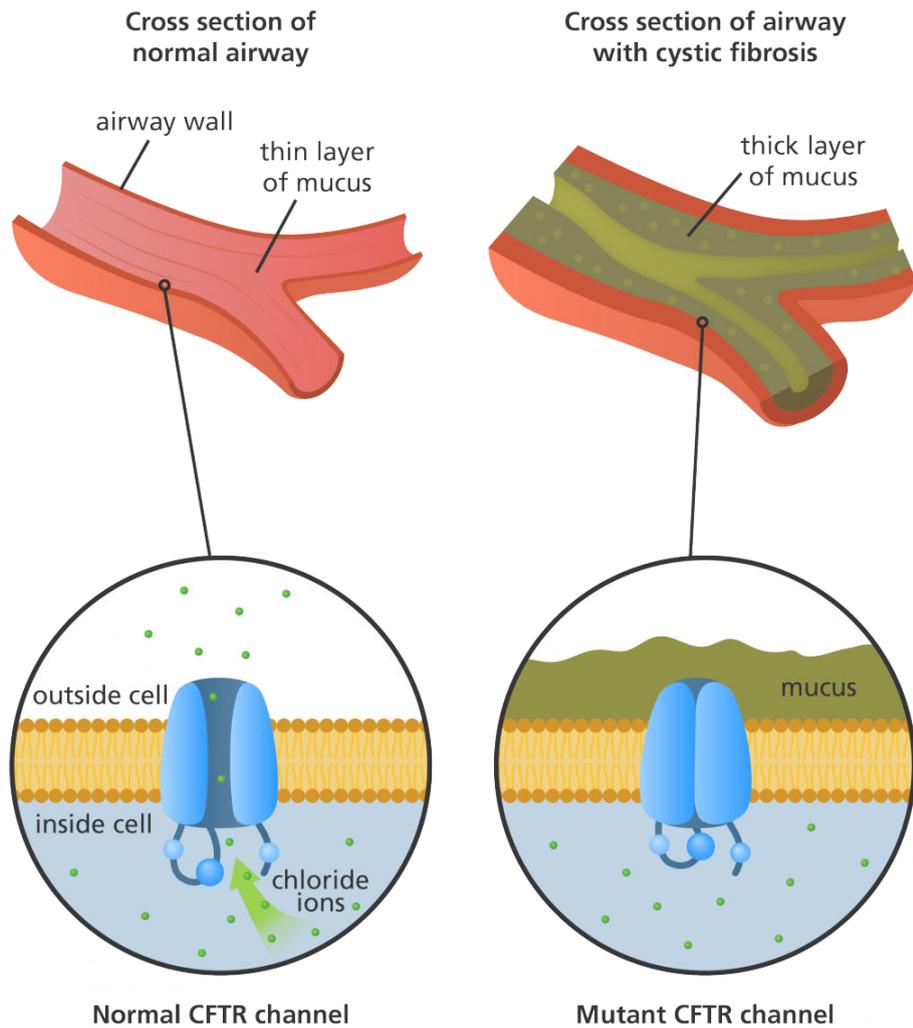


Figure 1.2 How mutated/ non-functioning CFTR channels cause dehydrated sticky mucus in patients with CF. A normal functioning CFTR (on left) allowing the movement of chloride ions to the outside of the lung epithelial cell, while the mutated/ non-functioning CFTR channel (on right) does not, causing mucus to become dehydrated and to build up on the outside of the cell. Reproduced from yourgenome (2020) copyright approved.

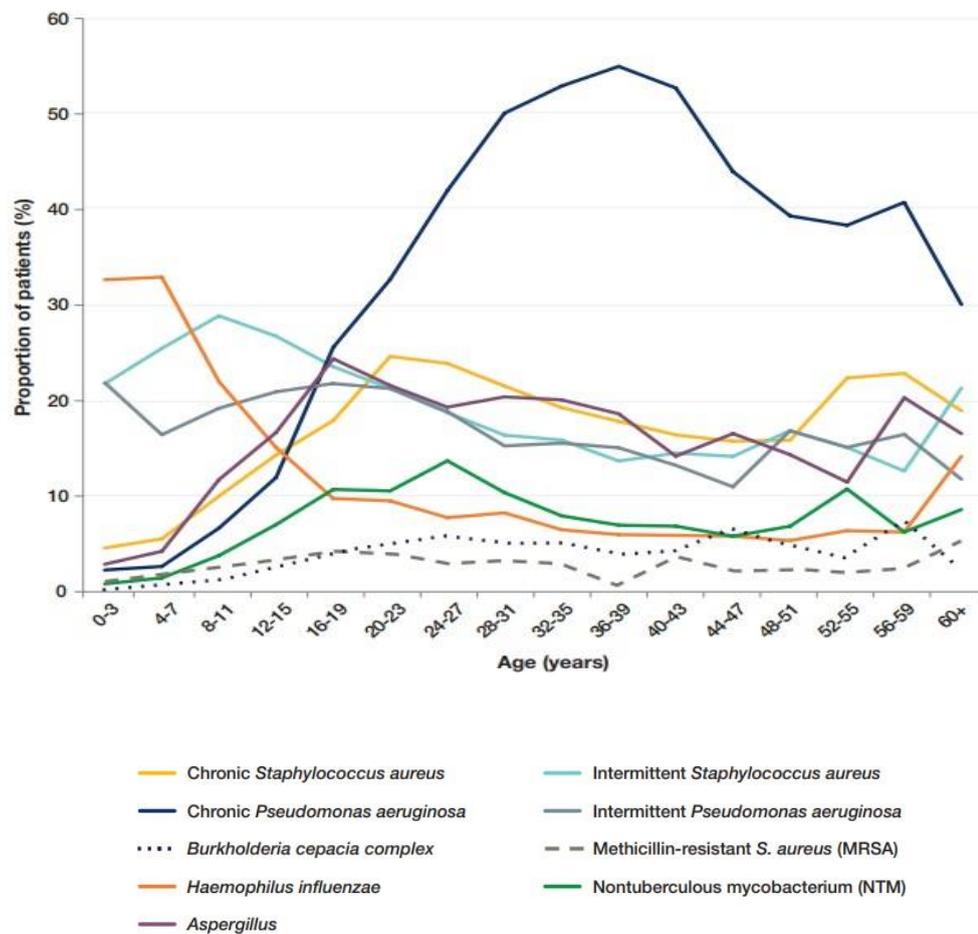


Figure 1.3 - Age specific prevalence of respiratory organisms in cystic fibrosis patients. Data from 2018 showing the species of bacterium isolated from CF patients with lung infections at different age ranges, whilst also stating whether the infection is chronic or intermittent. The key bacteria associated with greater patient colonisation from the age of 15 is *Pseudomonas aeruginosa*. Reproduced from CF Trust (2018a) Copyright approved.

In CF patients, an innate inflammatory immune response is initiated during bacterial infection, but it does not result in total clearance of the infection (Hart and Winstanley, 2002), which in turn leads to the viscous cycle of chronic inflammation and lung damage. This inflammatory response stimulates neutrophils to migrate into the airways, which produce proteases and reactive oxygen species (ROS) (Sagel et al., 2002), under normal physiological conditions, cells control ROS levels by balancing the generation of ROS with their elimination by scavenging systems. But under oxidative stress conditions, excessive ROS can damage cellular proteins, lipids and DNA, leading to fatal lesions in the cell that

contribute to carcinogenesis. The protease/antiprotease imbalance and oxidative stress have multiple downstream effects, including impaired mucus clearance (Boucher, 2007), increased and prolonged inflammation, and impaired immune responses (Armstrong et al., 1997), thus facilitating and promoting bacterial infections. Although neutrophils help to control infections, when present in great excess and for extended periods of time they cause harm to the lung, by producing many proteases and radical oxygen species. These cause an imbalance and oxidative stress has multiple downstream effects, including, increased and self-perpetuating inflammation, eventually leading to tissue damage and reduced lung function (Massimo Conese, 2017). Respiratory infections play a significant role in the mortality of CF patients, with 80-95% of deaths associated with respiratory failure caused by chronic bacterial infection (Lyczak et al., 2002).

It is still not fully understood why CF patients are initially so susceptible to bacterial infections, but numerous factors contribute to the abnormal composition of airway secretions. Water from CF sputa is absorbed at a high rate which causes dehydration and therefore depletion of the lubricating periciliary liquid layer (a component of airway surface liquid), preventing the transport of mucus and subsequent clearance of bacteria (Matsui et al., 1998). Also, the increased viscosity of the mucus prevents neutrophil chemotaxis (Matsui et al., 2005) and defective ion transport causes a pH imbalance, which inhibits the natural bactericidal activity of airway surface liquid, which increases susceptibility to infection (Pezzulo et al., 2012)

The prevalence of the microbes that cause respiratory infections is reported to be age dependent (Figure 1.3). In children and adolescents *Staphylococcus aureus* is the most prevalent, changing to *P. aeruginosa* in adulthood (CF Trust, 2018a). Chronic *P. aeruginosa* infection is the main bacterial agent in the loss of respiratory function and results in high rates of mortality due to epithelial damage, scarring and the plugging of airways (Lyczak et al., 2002).

1.3 COMMON CF AND BR PATHOGENS

1.3.1 Bacterial Pathogens

There is a commonality in the opportunistic bacterial pathogens colonising the lung in CF and BR, which include but are not limited to; *Staphylococcus aureus*, *Pseudomonas aeruginosa*, (Lyczak et al., 2002, Angrill et al., 2002, Al-Aloul et al., 2004) and the *Burkholderia cepacia* complex (Hauser and Orsini, 2015). Chronic *P. aeruginosa* and *S. aureus* colonisation are the most common airway infections of the lungs of adult CF patients. Interestingly, there is a switch from chronic *S. aureus* being the most prevalent pathogen in children and teenagers to chronic *P. aeruginosa* being the most prevalent infection in adults (Figure 1.3). Previously, culture-based microbiological methods were the most frequently used methods to identify bacteria from the CF airway, but now other molecular and sequencing-based methods, including PCR and microbiome analyses are used and have added to the number of microbial taxa recognised to be associated with the CF airways (Pattison et al., 2013). Increased research into the bacterial pathogens of the CF and BR lungs over the last 30 years has led to improved identification methods and changes in infection control and treatment regimens in both CF and BR (Taccetti et al., 2012). Evidence indicates that the incidence of *H. influenzae* and *S. aureus* have decreased with age from child (under 18) to adult (Millar et al., 2009). However, incidences of MRSA have increased, along with the rare and previously undetected species *Stenotrophomonas maltophilia* (Millar et al., 2009). In the UK Cystic Fibrosis Registry's Annual Data Report for 2018, a fall in the incidence of chronic *P. aeruginosa* compared to previous annual reports was noted.

1.3.2 Fungal and viral pathogens

There are fungi and viruses that are commonly detected in the lung sputa from CF and BR patients. The predominant fungal species is *Aspergillus fumigatus*, colonising over fifth of adults chronically, with occurrence increasing with age (Mortensen et al., 2011). Recent studies of fungal communities in the CF airway have used new molecular techniques, such as mycological culture with CF-derived fungal specific culture media and direct DNA extraction from patient sputa. Both approaches have been shown to increase the sensitivity of fungal detection, making possible the identification of fungal, it has been shown they were

not found using conventional culture methods (Nagano et al., 2010). Through the use of direct DNA extraction from sputum and fungal DNA PCR sequencing and DNA sequencing, it was confirmed that an increase in diversity of fungal species could be detected, compared to using conventional culturing methods.

The acute viral respiratory infections that CF children suffer from are caused by the same aetiological agents and have a similar frequency and duration to healthy children. However, children with CF have longer periods of lower respiratory tract symptoms compared to healthy children (van Ewijk et al., 2008). It is thought that exacerbations can be triggered by respiratory viral infections, such as Influenza A, Respiratory Syncytial Virus and Rhinovirus, which have been detected in the throat swabs of patients (Wark et al., 2012). In addition, viruses are linked to an increase in *P. aeruginosa* numbers in the lungs (Wark et al., 2012). A pilot study into viruses of BR patients demonstrated that respiratory viruses (mainly influenza A and B) are frequently detected in patients with stable BR and are also detected in asymptomatic periods with multiple viruses presenting simultaneously (Mitchell et al., 2018), which is similar to results shown in CF.

1.4 PSEUDOMONAS AERUGINOSA

Pseudomonas aeruginosa (*Pa*) is a Gram-negative bacterium, which is rod shaped in appearance and is commonly found in soil but can also be found in numerous clinical and environmental settings (Wiehlmann et al., 2007). *Pa* was first discovered in Paris by Carle Gessard in 1882. He described how it auto-fluoresced under UV light and later described it as an opportunistic pathogen of humans and plants (Gessard, 1984). As an opportunistic pathogen, *Pa* causes disease where normal host defences have been compromised, such as in immunocompromised individuals (Al-Hasan et al., 2008), people with severe burns (Estahbanati et al., 2002), patients on ventilators and more recently discovered in contact-lens users (Lyczak et al., 2000). Despite being an opportunistic pathogen, *Pa* is able to express a large number of virulence factors, which are specific to the type of infection, whether it be chronic or acute (Rumbaugh et al., 1999). It has been shown that *Pa* needs to express multiple virulence factors in order to colonise and cause infection of individuals (Manos et al., 2013).

1.4.1 Virulence in CF and BR

The lung epithelial inflammatory responses that are associated with a chronic *Pa* infection leads to progressive loss of lung function through scarring of the lung tissue and consequently respiratory failure (Callaghan and McClean, 2012, Costello et al., 2011). Once a *Pa* infection is established in the lungs it is difficult to clear using antibiotics (Hart and Winstanley, 2002). This is due to many strains of *Pa* encode multiple genes conferring resistance to antibiotics or the ability to adapt and evolve under selective pressure (Sanz-García et al., 2018). They are also able to form biofilms, which are known to protect the bacteria from phagocytosis, with slower growth rate and lower metabolic activity, biofilms are associated with a tolerance to antibiotics (Hoiby et al., 2010a). CF related *Pa* strains have been shown to adapt and evolve quickly. This adaptation can encompass alterations alter many characteristics of the bacteria, such as adhesion, alginate and antibiotic susceptibility with biofilm acting as diffusion barrier (Silbert et al., 2001). For these reasons, antibiotic therapy is only seen to be effective in clearing *Pa* infections fully in the early stages. Due to this, there is an eradication programme for CF patients when they are first diagnosed with a *Pa* lung infection (Taccetti et al., 2012). This regimen has now also been tested on BR patients with similar positive results (Orriols et al., 2015). There are numerous virulence factors that lead to the overall virulence of *Pa* in the lung environment, such as surface adhesion, motility and biofilm formation.

1.4.2 Surface structures

Cell surface proteins are expressed by *Pa* similar to other bacteria, such as pili and flagella and have multiple functions, including; motility (Bradley, 1980), adhesion (Doig et al., 1988) and biofilm formation (Deziel et al., 2001). These are virulence factors that play an important pathological role in the colonisation, the invasion of tissues and the survival of the bacteria.

Pili are hair-like projections on the surface of a bacterial cell and are on average 6-7 nm in length (Bucior et al., 2012). The names pili or pilus and fimbria are used interchangeably. However, fimbria is also known as one of the three types of pili, all of which are found on the surface of *P. aeruginosa* (Mattick, 2002) and are described in the following sections,

they are used for transfer of genetic information and as a surface structure their hydrophobic nature promotes attachment to cell surfaces (Irvin, 1993).

The conjugative pili, also known as "sex pili", facilitates the transfer of DNA between bacteria in a process called bacterial conjugation. The DNA that is transferred comprises genes necessary to express and transfer pili, commonly encoded on a plasmid. However, co-transference of other bacterial genes occurs frequently and this can lead to a spread of genes that may have advantageous traits, such as antibiotic resistance via horizontal gene transfer (Bradley, 1980).

Type IV pili are made of pilin protein subunits and are built with a leader peptide that is removed before pilus construction (Giltner et al., 2006). There are two main types of type IV pili, namely type IVa and type IVb. They differ in the length of the leader peptide sequence and have independent assembly systems (Mattick, 2002). These pili are involved in producing motile force to move the bacterium (Mattick, 2002). However, in *Pa* both types have been described and the type IVb is not associated with motility and is less characterised than type IVa (Bernard et al., 2009). The type IV pili are associated with multiple types of motility, including twitching motility. The distal ends of the pili bind to a surface within their surrounding area, including other bacteria, and movement occurs as the pili contracts and pulls the bacteria towards its binding surface, which produces a twitching movement of the cell (Bradley, 1980). Type IV pili are also associated with swarming motility and, more recently discovered, walking motility (Kazmierczak et al., 2015) and slingshot motility (Jin et al., 2011). Mutations within pili genes that construct the type IVa pilus can result in different phenotypes, ranging from complete absence of pili to hyperpilated, which affects the virulence of the bacteria (Chiang and Burrows, 2003).

Fimbriae are short pili and often referred to as "attachment pili". They are often found at the poles of the bacteria or evenly spread over the whole surface. Fimbriae mediate the attachment of the bacterium to the host surfaces, which can initiate the formation of a biofilm (Deziel et al., 2001). Bacteria that lack fimbriae cannot adhere to target surfaces. For example, in the case of *Pa*, this means biofilms cannot develop, which makes them much less virulent and incapable of causing disease (Vallet et al., 2004). Pili are also a common

attachment site for bacterial viruses, called bacteriophages or more simply as phages. Some phages attach or adsorb to specific receptors on the pili (Ceyssens et al., 2009, James et al., 2012). (discussed further in section 1.5).

Flagella are whip-like structures that project from the bacteria and their main function is propulsion of the cell. However, they also have sensory functions and are responsive to environmental changes, such as chemotaxis (Dasgupta et al., 2003) and changes outside the bacterial cell (Haiko and Westerlund-Wikstrom, 2013). Pa has a single polar flagellum, comprising the flagellin protein that is operated by a motor (Terashima et al., 2008). A flow of protons across a cell membrane powers the motor (proton motive force), which comprises two units, the stator (a static structure) and a rotor, which fit together and cause the flagellum to rotate (Blair, 2003). This enables motility, especially swimming and swarming motility in less solid environments (Murray and Kazmierczak, 2008). Flagella are considered important in virulence because of their role in motility and adhesion as they are necessary for colonisation of their host epithelial surface (Montie et al., 1982).

The flagella help Pa to colonise the epithelia through adherence of flagellin protein to epithelial cells due to them being hydrophobic (Irvin, 1993). This adherence causes hypersecretion of mucins, which is likely to have negative clinical effects (Ben Mohamed et al., 2012). The flagellin protein has been shown to be highly immunogenic and to stimulate an inflammatory response and has been detected in sputum from CF patients (Feldman et al., 1998). This is extremely dangerous for CF patients as this inflammatory response causes impairment of mucociliary clearance and host defences (Wilson et al., 1985), which is a cause of exacerbation and subsequent reductions in lung function (Hayashi et al., 2001a). Subsequently, a loss of the flagella is an effective tactic of evading the immune response (Patankar et al., 2013) and is recognised as an adaptive and evolutionary trait in CF Pa isolates (Workentine et al., 2013).

1.4.3 Biofilm production

Bacterial biofilms rarely comprise a single bacterial species, such as *Pa*, with biofilms with multiple species being recovered from patients' lungs with CF and BR. The bacteria produce a matrix of polysaccharides (Sutherland, 2001), proteins and DNA (Whitchurch et al., 2002) (Hoiby et al., 2010a) and are typically adhered to a surface but can also be a free floating mass. The matrix that is produced offers protection from antibiotic attack by acting as a diffusion barrier, protecting biofilm-grown cells from antibiotic penetration (Hill et al., 2005, Costerton et al., 1987). The hosts immune response including antibodies and inflammatory cells are also not able to penetrate to biofilms matrix. Biofilms can provide protection against phage predation in two ways, (i) inhibition of phage transport into the biofilm and (ii) the amyloid fibre network of CsgA (curli polymer) fibres protect cells by individually coating their surface and binding phage particles, so to prevent their attachment to the cell exterior (Vidakovic et al., 2018). *Pa* produces three extracellular polysaccharides: alginate, Psl and Pel (Franklin et al., 2011). When a *Pa* biofilm is formed, alginate plays a vital role in entrapping the bacteria and forming the primary structures of the biofilm, which forms a mushroom-like shape when attached to the surface. The alginate also helps to conserve the viability of *Pa* by increasing its resistance to phagocytosis and reducing susceptibility to antibody-dependent bactericidal mechanisms (Baltimore and Mitchell, 1980, Pier et al., 2001). The Pel polysaccharide is capable of playing both a structural and a protective role in *Pa* biofilms with studies showing it is crucial in maintaining cell-to-cell interactions and necessary to hold an aggregate together (Colvin et al., 2011). Psl is important for initial surface attachment (Ghafoor et al., 2011), creating a positive feedback loop regulating the production of cyclic di-GMP, which results in increased Psl levels and biofilm structure (Irie et al., 2012). If any of these genes are absent, changes to biofilm formation can occur, as shown in PA14 deletion of *pelB*, which results in severe biofilm deficiency (Colvin et al., 2011). However, this effect is strain-specific and loss of Pel production in PAO1 results in no difference to biofilm development. Instead, Psl proved to be the primary structural polysaccharide for biofilm maturity (Colvin et al., 2011). Pel and Psl serve redundant functions as structural scaffolds in mature biofilms but without both polysaccharides, poor biofilms are produced (Colvin et al., 2012).

Type IV pili on the cell surface of Pa also play a role in biofilm formation, specifically surface-attached biofilm formation, as they help with the surface adhesion of the biofilm. However, they are not essential as studies have indicated that pili-deficient mutants still adhere to surfaces and establish biofilms (Klausen et al., 2003). Established biofilms have the capability to spread by shedding, where cells are released from the biofilm in a process called dispersal. However, there is a study that shows the shedded cells is a mixture of biofilm with senescent cells, and some that are planktonic but scavenge from the biofilm cells (Berlanga and Guerrero, 2016). Interestingly, there is a filamentous bacteriophage called Pf4 that is integrated into the chromosome of PAO1, which has been shown to be vital for biofilm dispersal (Rice et al., 2009, Webb et al., 2004). Rice et al (2009) also showed that mice infected with the Pf4 mutant (Pf4 removed) survived significantly longer than those infected with its isogenic wild-type strain (functioning Pf4), demonstrating that Pf4 contributes to the virulence of Pa (Rice et al., 2009). The specific mechanism by which the Pf4 mediates these effects is currently unknown.

Not all biofilms are attached to the epithelial surface and some/a proportion have been shown to be free floating in CF sputa, suggesting that Pa may also grow as unattached biofilms in the CF lungs (Bjarnsholt et al., 2009). Most experimental work carried out thus far has looked at surface attached biofilms. Innovative *in vitro* approaches are being used to model Pa growth to replicate the unattached biofilms in the CF lung, most notably artificial sputum media (ASM). Here, Pa produces biofilm structures using the components of the media, such as mucin and DNA. The concentrations of free DNA and mucin both influence the viscosity of the sputum in the lung *in vivo* (Haley et al., 2012).

1.4.4 Genomics of *Pseudomonas aeruginosa*

The ability of *Pseudomonas aeruginosa* to adapt and evolve in microenvironments means they must be able to adapt rapidly to outcompete other pathogens or to gain longevity in an environment. Pa has a comparatively high GC content of 65-67% and the ability to carry multiple plasmids. Since 1999, when the first complete Pa genome was sequenced (PAO1) (Stover et al., 2000), many Pa genomes have been deposited to GenBank (5134 at the time of writing June 2020). This allows analysis of the Pa species at the genetic level and allows gene

functions to be determined. (Mathee et al., 2008, Wolfgang et al., 2003) PAO1 is the predominant lab strain of Pa used. The genome is 6.3 Mbp and when it was first sequenced in 1999, 5,570 open reading frames (ORFs) were predicted, with a high percentage being regulatory genes (8.4%) (Stover et al., 2000). The Pa core genome is interspersed with an accessory genome, representing between 7-18% of the whole genome, which explains the vast inter-strain genetic diversity seen (Klockgether et al., 2011). The majority of genes in the Pa accessory genome are hypothetical and may represent horizontally acquired elements, such as phages (Ozer et al., 2014). A study by Wolfgang et al. (2003) illustrated that the core genome is highly conserved in Pa. It also showed that strains such as PA7 have a significantly lowered core genome compared to other Pa reference strains, such as LESB58, PAO1 and PA14 (Roy et al., 2010, Ozer et al., 2014). A panel of Pa genomes from environmental and clinical strains were also compared, showing highly conserved virulence genes between strains from differing environments, suggesting that environmental strains conserve the same virulence genes and therefore also have the capability to colonise humans (Wolfgang et al., 2003).

The accessory genomes of clinical Pa isolates are extremely diverse with some genes being strain-specific (Pohl et al., 2014). Many accessory genes from clinical isolates are not found in other lab strains of Pa (Shen et al., 2006). Therefore, the accessory genomes of clinical isolates are an important area to examine and determine the function of genes that relate to colonisation and disease.

There are several reported CF epidemic strains, including Liverpool epidemic strains (LES) and Prairie epidemic strains (PES) accessory genomes, which enhance the bacteria and are thought to carry genes that facilitate the survival of the bacteria in the CF lung (Dettman et al., 2013). LESB58 was the first historical isolate of LES and, when sequenced, it was shown to have a large number of horizontally acquired elements within its accessory genome, with five genomic islands and six prophages (Figure 1.4). In addition, it was shown that one of these novel genomic islands is crucial for LES competitiveness in a rat model for chronic lung infection (Winstanley et al., 2009).

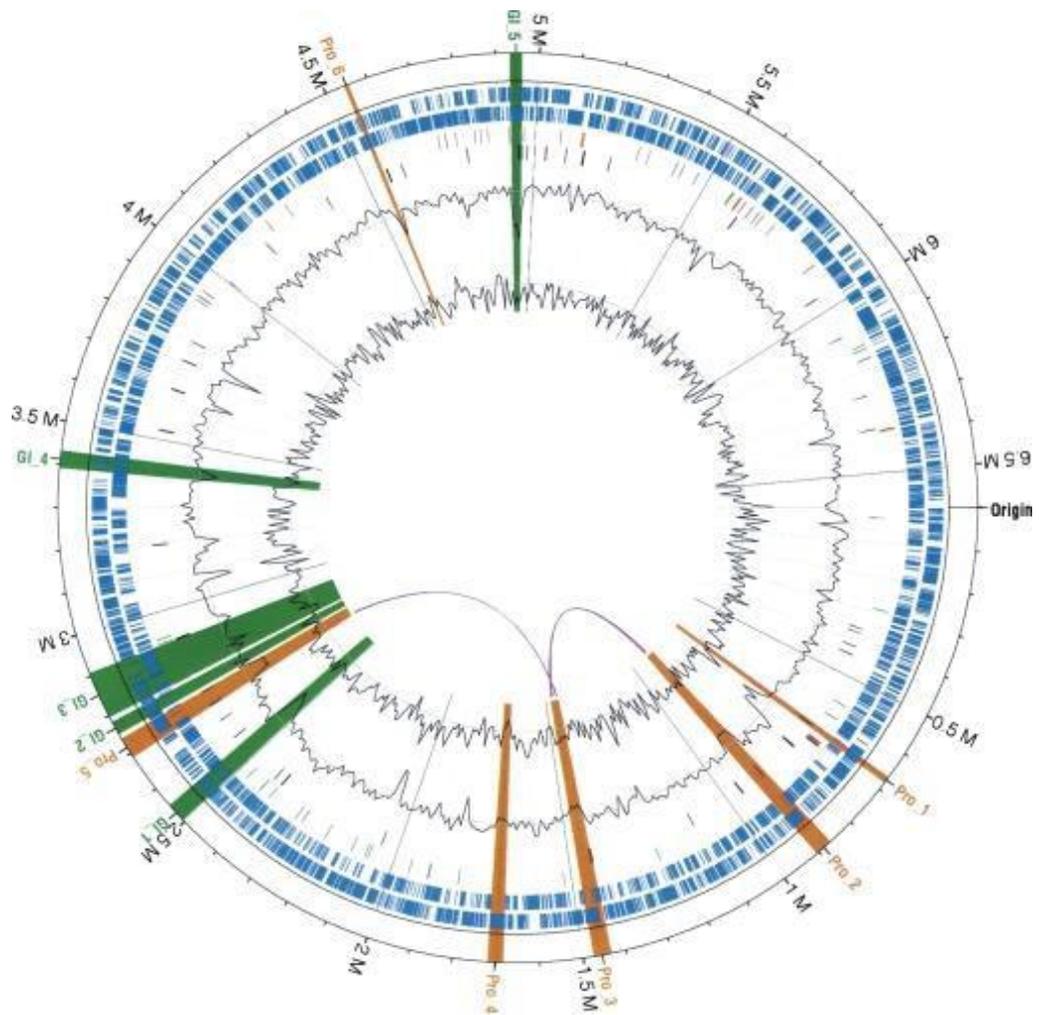


Figure 1.4 Circular map of the *P. aeruginosa* LES genome. Starting from outermost circle going inwards: major (500 kb) and minor tick (100 kb) measurements of the genome with estimated location of the origin of replication; prophage (orange) and genomic islands (GI) (green) are highlighted across all tracks; protein coding genes (blue) on plus (outer) and minus strand (inner); tRNAs (green), rRNAs (orange), and all other noncoding RNA genes (purple); signature tagged mutants (black); GC content (outer black line plot) with GC content average (gray line). Note that one region of low GC, upstream of the first noted prophage, plus additional smaller regions of low GC, contain ribosomal genes that are commonly known to have a lower GC in genomes. The location of two highly similar genomic regions within the prophages are marked with looping purple lines, between their locations on the innermost circle. The identified prophage and GIs are distributed around the genome, but there is one notable cluster of LESGI-1, LESGI-2, and LESGI-3, reflecting the non-random nature of GI insertion in Pa. Reproduced from Winstanley et al., (2009). Copyright approved.

1.4.5 Mobile genetic elements of *Pseudomonas aeruginosa*

Mobile genetic elements (MGEs) are genetic material that can move around within a genome, or that can be horizontally transferred from one bacterial species to another. MGEs are found in all bacteria, including Pa. There are several different types of MGEs. Transposons are DNA sequences that can move locations within a genome, which includes DNA transposons, replicons, plasmids and bacteriophages. These MGEs all play a distinct role in bacterial evolution by gene duplication and additive function (Wiedenbeck and Cohan, 2011), mutations in gene coding regions that alter protein function and gene rearrangements in the host genome. Pa plasmids have been associated with antibiotic-resistance, with some Pa strains acquiring plasmids encoding new β -lactamases that confer resistance to penicillin and cephalosporins (Sacha et al., 2008). Studies have also shown plasmid mediated resistance to a range of anti-pseudomonal antibiotics, including ciprofloxacin and amikacin (Shahid et al., 2003). Transposons have also been seen to mediate antibiotic resistance in Pa, such as transposon Tn1403 from a clinical *Pseudomonas*, which carries the tet(C) tetracycline resistance determinant (Stokes et al., 2007). Newly acquired genes from MGEs can not only facilitate antibiotic resistance but can increase fitness by gaining new or additional functions. Importantly, temperate bacteriophages have also been shown to increase the fitness of Pa (Davies et al., 2016a) and represents the focus of this thesis.

1.5 BACTERIOPHAGES

1.5.1 Introduction to bacteriophages

Bacteriophages or phages are the largest number of biological entities on the planet, with different genomes and life cycles (McGrath and van Sinderen, 2007). Phages are viruses that infect bacteria and populate all environments inhabited by bacteria. They were originally observed by Frederick Twort in 1915, which was later confirmed by Felix d'Herelle in 1917 in experiments associated with phage lysis of dysentery bacillus (Twort, 1915, d'Herelle, 1917). They were only proven to be viruses after they were viewed by electron microscopy.

Phages exhibit two life cycles; lytic (virulent) or lysogenic (temperate). The lytic life cycle

results in the death of the bacterial host and propagation of new phage particles, whereas in the lysogenic cycle the phage integrates its DNA into the bacterial genome and is partitioned into daughter cells with cell division. Temperate phages have the ability to switch between the lytic and lysogenic life cycles whereas obligatory lytic phages can only perform the lytic lifecycle. Temperate phages that integrate can lose genes over time which may include ability to switch from the lytic to the lysogenic cycle and are continued to be carried like any other genetic loci and are subject to evolutionary selection for continual carriage.

1.5.2 Phage classification

Phages, unlike bacteria, do not carry a universal gene such as 16S rRNA, although they do have genes that are common in all phages, due to their low level of similarity they cannot be used to determine difference in phylogenies like 16S can, however the ssDNA and RNA phages the capsid proteins are significantly different and can be differentiated. Historically, morphology has been used as the gold standard to order and classify phages (Ackermann, 2009) (Figure 1.5). They are different in shape, size and have distinct appendages which are linked to their infection strategies. Morphological characteristics include: long and short tails, contractile tails, non-contractile tails and whether they have an icosahedral capsids or elongated capsids (prolate heads). These features are most commonly visualised using electron microscopy (Figure 1.6). Phages can also be classified using other characteristics, such as host range, genome size, life cycle, sequence similarity and genome organisation (Ackermann, 2009).

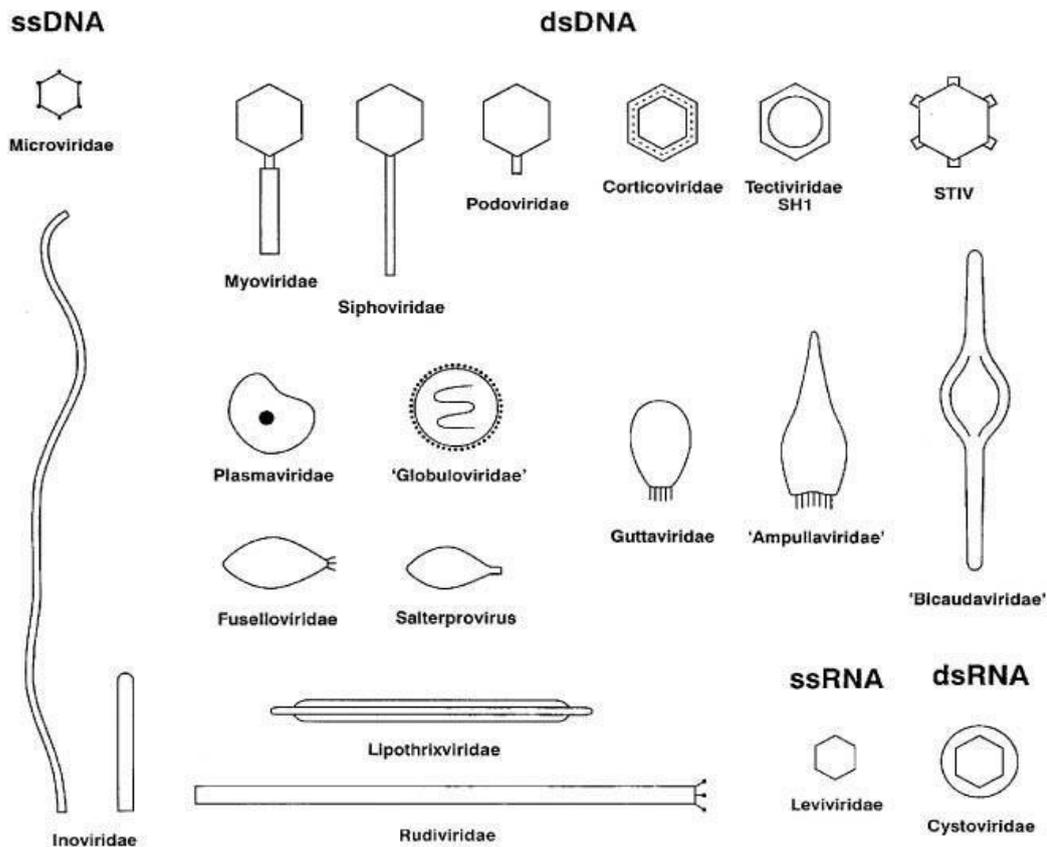


Figure 1.5 Diagram of the different types of bacteriophages. Diagrammatic representation of the families of viruses that can infect bacteria. Taxa are grouped according to whether the viral genome is DNA or RNA. If the genome is double stranded, it is labelled dsDNA or dsRNA. If it is single stranded, it is labelled ssDNA or ssRNA. Reproduced from Ackermann (2009). Copyright approved.

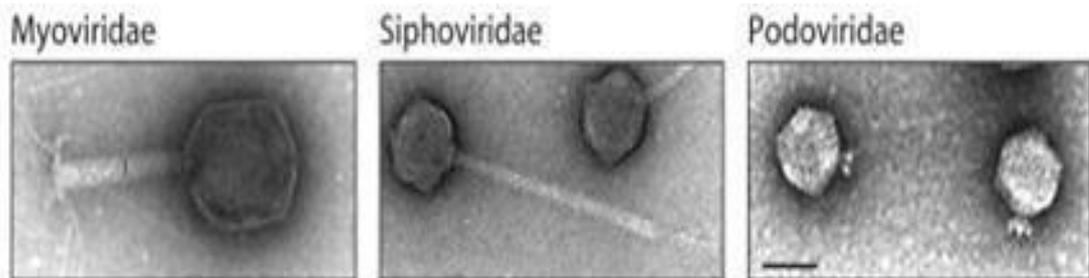


Figure 1.6 Electron microscopy images of bacteriophage with different tail types. The *Myoviridae* image shows a T4-like bacteriophage provided by Damien Maura and Laurent Debarbieux (Institute Pasteur). The *Siphoviridae* is a HK97 phage and *Podoviridae* is a P22 phage from Laurence et al. (2002). Scale bar represents 50nm. The images were taken from Krupovic et al. (2011). Copyright approved.

Phages can carry genetic information in different forms including viruses with double-stranded DNA (dsDNA), which are the vast majority, single-stranded DNA (ssDNA), single-stranded RNA (ssRNA), and double-stranded RNA (dsRNA), which are very rare. They can also be grouped in this, which is known as the Baltimore classification method of viruses. They can then be grouped into orders based on morphology, which are further subdivided into families, sub-families genera and species (King, 2011).

The increase in high throughput sequencing means more recently phages have been classified using both morphology, nucleic acid sequence and proteome comparisons, by the International Committee on Taxonomy of Viruses (ICTV) and subcommittee the Bacterial and Archaeal Viruses Subcommittee (BAVS). In the most recent report by the ICTV's BAVS, there were over 6 orders, 10 families, 35 subfamilies, 672 genera and 1976 species (including other viruses that infect other microbes) (Adriaenssens et al., 2020). The aim of the subcommittee is to provide a rational approach to phage classification using genome-based metrics.

1.5.3 Phage structure and morphology

The structure and morphology of phages are diverse. However, most isolated phages (96%) are tailed; other types such as filamentous or pleomorphic represent less than 4% of phages discovered (Ackermann, 2009), perhaps suggesting a bias in enrichment/isolation methods. The structure of virions varies from the simple, such as phiX174, which has a capsid made from a single monomer, to complex dsDNA phages with capsids that are constructed from multiple proteins. The phage's genetic material is encapsulated in a protein (or more rarely, lipid) capsid head which offers protection and enables survival in the environment. The basic structures of the three most frequently seen morphotypes are: an icosahedral head (20-sided) with a tail, an icosahedral without a tail and filamentous phage, although filamentous phage do not have a capsid. The tails of the phages can also be contractile or non-contractile and facilitate bacterial infection. These tails also denote the morphological family they are placed in. The most common type of phage to infect bacteria are in the order *Caudovirales*, which encompasses the tailed dsDNA phage families, the different morphological family are detailed in the following section.

1.5.3.1 Caudovirales

The order *Caudovirales* consist of the three families: *Siphoviridae*, *Myoviridae* and *Podoviridae* (King, 2011). Tail morphology determines which family a phage belongs to. *Siphoviridae* have long, non-contractile tails, *Myoviridae* have long, contractile tails, and *Podoviridae* have short tails, as shown by electron microscopy images in figure 1.6. Most *Pa* phages have long, non-contractile tails and are classed as *Siphoviridae* (Ackermann, 2007; Knezevic et al., 2009). *Pa* phages isolated from CF sputum, have been identified all belonging to the family of *Myoviridae*, *Siphoviridae* and *Podoviridae*, and notably majority of *Pa* temperate phages characterised thus far also have tails (Ojeniyi et al, 1991). Recent taxonomic changes have changed some of these groups with new families being created within this order based on comparative genomic approaches (e.g. *Herelleviridae*).

1.5.3.2 Inoviridae

The *Inoviridae* are a family of ssDNA filamentous phages. Pf1 was the first identified filamentous phage infecting *Pa*. It infects the cell via the type IV pilus and, like other filamentous phages, is unusual in that it does not cause cell lysis but instead extrudes virions through the cell poles (Bradley, 1973; Hill et al., 1991). Pf1-like phages have since been discovered to be widespread and are detectable in over 50% of *Pa* strains as prophages (Knezevic et al., 2015).

The *Inoviridae* are of interest because of their role in bacterial virulence, biofilm formation and have been readily found in chronically infected lungs of CF patients (Burgener et al., 2019). They are shown to promote the formation of biofilms in *Pa* by spontaneously assembling into a liquid crystalline structure when alongside polymers found in CF sputum. These liquid crystals also insulate against antibiotic entry by stopping diffusion through biofilms (Secor et al., 2015). They have also been shown to increase the virulence in animal models (Rice et al., 2009, Secor et al., 2017).

1.5.3.3 Capsid and packaging of DNA

Caudovirales phages mostly have a symmetrical icosahedral capsid, made from proteins, which forms an enclosed head unit in which the phage DNA is actively packaged. These capsid proteins are positioned on scaffold proteins, known as procapsids or proheads, prior to being packaged with DNA, after which a mature capsid is formed (Caspar and Klug, 1962). The proteins that connect the head to the tail are encoded by minor capsid genes

and are known as neck proteins (Casjens and Hendrix, 1974; Hendrix and Duda, 1998). The assembly pathway of *E.coli* phage HK97 is well studied; it begins with self-assembly of gp5 into pentamers and hexamers. A protease, called gp4, cleaves gp5 at its N-terminus. Attachment of a portal protein, gp3, coupled with conformational changes leads to the formation of a prohead, or procapsid, which is the precursor to the mature capsid (Duda et al., 1995). It has also been noted that while DNA is being packaged into the capsid, the capsid expands by approximately 5 nm and changes to an icosahedral shape.

The DNA is wound into the capsid when packaged and therefore supercoiled, which offers the energy potential required for gDNA injection during infection. This energy is partly created by the 12° rotation per 2 base pairs, as well as a turn produced by the packaging motor (Spakowitz and Wang, 2005). The scaffold proteins are removed from the capsid during maturation and are recycled to make new capsids or are cleaved to make smaller polypeptides and amino acids (Aksyuk and Rossmann, 2011; Casjens et al., 1985). There are two main methods for DNA packaging, the headful method where the replication of the phage DNA creates concatemers arrayed in head-tail axial and terminase resolves the issue by cleaving the DNA once the head is full and is not sequence specific, seen in the T4 phage. The other method seen in λ starts with the phage capsid entering the maturation step by the addition of linearised dsDNA that is cut at cos sites, which are sequence specific, by the terminase. The phage DNA then enters the capsid through the portal protein and in phages that do not have concatemeric DNA this event causes changes in the portal protein which triggers the terminase to cut the DNA (Black, 1989, Valpuesta and Carrascosa, 1994).

1.5.3.4 Tail fibre structure and formation

The tail fibre is the part of the phage that determines specificity of infection, more specifically the host adsorption apparatus at the distal tail end. They vary in size, composition, and morphology and their function includes specific host detection. Variations can be as simple, such as the tail tip protein or a complex like the phage baseplate. For myo- and siphoviruses the tail tape measure protein encoded determines the length of the tail (Xu et al., 2004).. The interaction between the phage adsorption apparatus (tail fibres/tail spikes) with its concomitant receptor often leads to irreversible binding to the host cell. Phage tails are usually encoded by an operon of genes that express the succession of proteins used to construct the tail, making the numbers of molecules that allow these structures to assemble. In the model *E.coli* temperate phage lambda the tail is expressed from an operon of 11 genes for gene products (gp) gpZ, gpU, gpV, gpG, gpGT, gpH, gpM, gpL, gpK, gpl and gpJ (Casjens and Hendrix, 1974, Xu et al., 2004). The tail assembly pathway can be divided roughly into three stages: formation of the initiator, tail polymerisation, and termination (Katsura and Kuhl, 1975; Figure 1.7).

Phages from Pa fall into the Caudovirales order and therefore have tail fibres. Lytic phages in the Siphoviridae family from Pa have been shown to have tails that vary in length from 110-170 nm (Sepulveda-Robles et al., 2012), whereas lytic Pa phages in the Myoviridae family are shown to have tails ranging from 10-302 nm (Sepulveda-Robles et al., 2012).

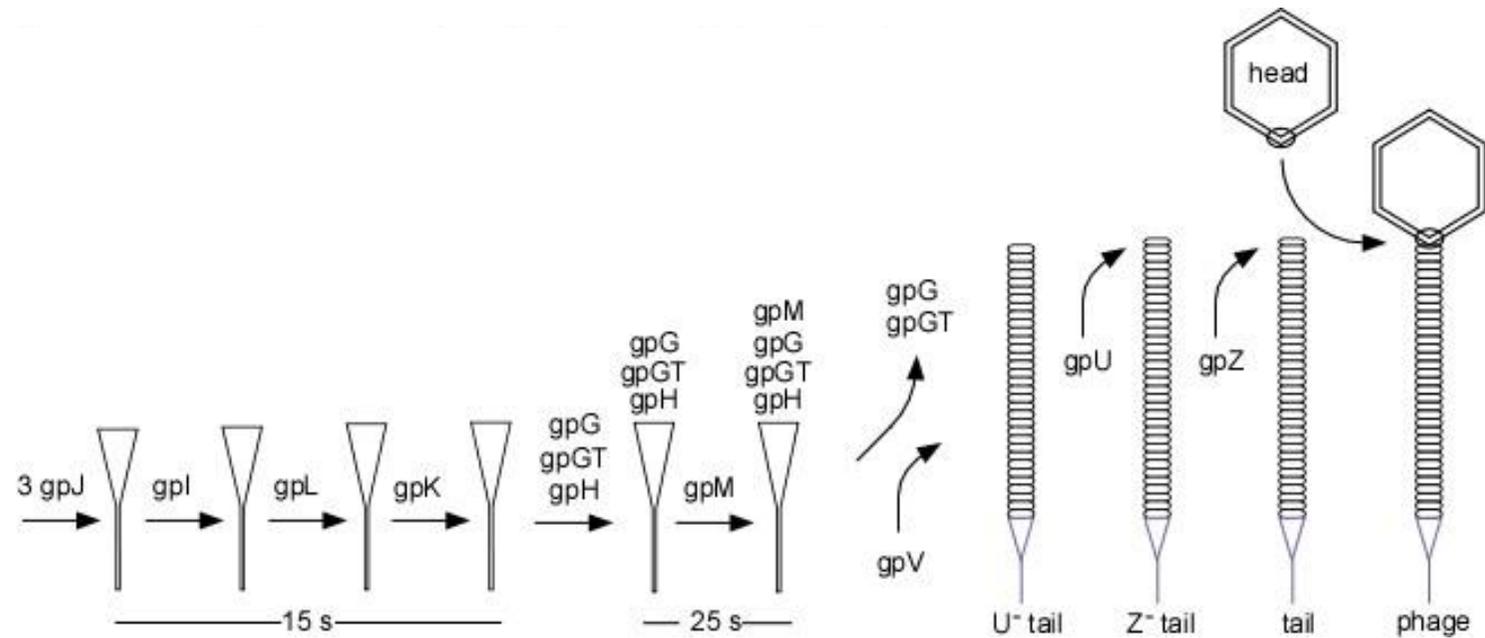


Figure 1.7 Lambda phage tail assembly pathway. The 11 proteins associated with phage tail assembly in lambda phage, showing the λ tail assembly pathway. The assembly pathway can be divided roughly into three stages: formation of the initiator, tail polymerisation, and termination. Image taken from (Xu et al., 2013). Copyright approved.

1.5.4 Mosaicism and phage heterogeneity

Temperate phage genomes are highly mosaic and are considered as a unique combination of regions that are interchangeable with other phages by recombination. Although, a key feature is; even though there is genetic exchange, there is conserved order of genes linked to the life cycle of the bacteriophage (Hendrix et al., 1999). The size of each region open to recombination, the rate at which they exchange and the phages that carry them all vary. This heterogeneity is also prevalent in bacteria, where genes are transferred by horizontal genetic exchange. However, the extent of intergenic heterogeneity in phage genomes is much higher (Bobay et al., 2013), which has become more apparent with the rise in phage genome sequences and the ability to compare them. This heterogeneity however, makes classification complicated as phages can obtain genes from distinctly different phages, sometimes from those that infect other species of bacteria via HGT. For example, phages may have tail genes with 90% similarity but very low sequence similarity across the rest of the genome which has been shown in phage P2 tail in genes G and H which have are similar those found in to Mu (Haggård-Ljungquist et al., 1992). In particular, the dsDNA phages are difficult to classify taxonomically as extensive horizontal gene transfer of multiple phages from differing bacterial species have regions of sequence similarity (mainly in temperate phages), suggestions of a shared ancestry have been reported between mycobacteriophages D29 and coliphages HK97 and HK022 (Hendrix et al., 1999). A recent paper shows that gene flux is much higher in temperate phages. Moreover, the work on lytic phage classification suggests that evolution is predominantly vertical in phages of this lifestyle.

1.5.5 Phage conversion

Phage conversion is infection of a cell and integration into the bacterial genome as a prophage. There is site specific integration where the phages use their integrases to carry out recombination between attachment sites on the phage and bacterial genomes, known as attP and attB. However, some phages carry out random recombination and can insert anywhere in the bacterial genomes, such as Mu like phage infection. Both methods of recombination are pertinent. Conversion with the phage can cause subversion of the bacterial metabolism outside of those linked directly to the life cycle of the phage and may

play a role in a change of cellular function. The integration of a phage into the genome and creation of a lysogen can impact the evolutionary selection of the bacterium by offering a selective advantage (Sun et al., 2013). For example, the *stx* gene, which encodes for the Stx toxin, is carried by a phage in *E.coli* (Plunkett et al., 1999). *E.coli* may benefit in an altruistic manner from Stx toxin production as a result of toxin-dependent killing of eukaryotic cells, such as unicellular predators or human leukocytes. Hence, the non-induced lysogens have an advantage. The model of STEC altruism described by Łoś et al. (2013) supports this. It has been proposed that Stx has evolved as a mechanism of defense against protozoa that confers a selective advantage for bacteria harboring *stx* phages (Stolfa and Koudelka, 2012).

Bacterial fitness can also be affected by phage conversion and this can increase disease severity, as well as genes that can help the bacteria evade the human immune system by changing the antigens on the surface of the bacteria shown in *Salmonella enterica serovar choleraesuis* by phage P14 (Nnalue et al., 1990). In lambda phage, the *lom* and *bor* genes are associated with increased virulence of adhesion of *E.coli* by increasing adhesion and serum resistance, respectively (Barondess and Beckwith, 1990; Vica Pacheco et al., 1997). Put into context, bacteriophages add further complexity as bacteria are usually infected with multiple bacteriophages throughout the bacterial chromosome. Some may not be inducible, yet their genome remains as a prophage known as cryptic/degenerate phages in comparison to viable phages that can be induced. An example of polylysogeny in Pa is the Liverpool Epidemic Strain (LES) where the strain LESB58 harbours five active prophages and is highly virulent.

1.5.6 Bacteriophages of *Pseudomonas aeruginosa*

There are 128 *Pseudomonas* phage genomes in the EMBL-EBI database (at the time of writing June 2020) and many more as complete or cryptic prophages in bacterial genomes (McWilliam et al., 2013). Pa phages isolated from geographically diverse locations can be closely related and classified into a limited number of subgroups (Ceyssens et al., 2009), although a more recent study of phages isolated in Mexico suggests that the diversity of Pa phages is larger than previously thought, as 12 different “species” were identified, six of which were novel (Sepulveda-Robles et al., 2012).

Bacteriophages infecting Pa mainly belong to the order of Caudovirales. Phages of the

Caudovirales can be found in PAO1 as a prophage, such as MD8 a 35kbp phage that belongs to the Siphoviridae family, PAO1 also carries a Pf as a prophage, which is a filamentous phage with a 15.7Kbp genome. MD8 cannot induce from the chromosome and produce an active viral progeny in any of the studies previously performed. Although studies looking at infection of PAO1 illustrate Pf induction (Rice et al., 2009). Other examples of well-known phages of Pa are phage phiCTX, which is an example of a Myoviridae phage and has a 35.5kbp genome, which encodes the CTX gene and CTX toxin in the PA158 strain (Hayashi et al., 1990). Phage D3112 is an example of Siphoviridae phages that use the type IV pili as a way to infect the bacterium via adsorption (Roncero et al., 1990). Phage D3112 is a particular class of temperate phage that is able to integrate anywhere within the bacterial chromosome to disturb the genes. During excision, they package part of the host chromosome, which facilitates horizontal gene transfer by carrying that region of bacterial DNA to the next bacteria they infect, integrating it into the new bacteria genome with the phage DNA (Morgan et al., 2002). Examples of Podoviridae phage that infect Pa are F116 and H66, which have genomes of 65.1 Kbp and 65.2 Kbp in length, respectively (Lammens and Lavigne, 2013). The Pa bacterial genome can contain a significant percentage (>20%) of functional and non-functional phage genes (Casjens, 2003). Subsequently, prophage genes can account for a large proportion of the accessory genome variation within Pa as a species. Whether they are functional or remnant phage genes, they all have potential to alter the bacterial genome and drive evolution, making phages particularly important in microbial evolutionary research.

1.5.7 The life-cycle of bacteriophages

Bacteriophages typically fall into two groups: lytic and lysogenic (Figure 1.8). Some phages only follow one lifecycle e.g. obligate lytic phages that are only capable of a lytic lifecycle, but others can follow either and are known as temperate phages. The first step of infection that must take place in both cycles is adsorption of the virion onto a site-specific point on the cell surface of the host/target bacteria that allows entry of phage DNA into the cell. One other stage that can happen after the phage DNA has entered the cell is called pseudolysogeny, which is a stage of stalled development of a phage when there is no multiplication of the phage genome (as in lytic development) or replication with the bacterial

chromosome (lysogenisation). During pseudolysogeny, no viral genome degradation occurs, which allows the subsequent restart of virus development when conditions are more favorable. Pseudolysogeny is usually caused by unfavourable growth conditions for the host (such as starvation), which initiates either true lysogeny or lytic growth when conditions improve for the host (Łoś and Węgrzyn, 2012).

1.5.7.1 Phage adsorption and infection

Adsorption of the phage to the host bacterial cell is the first stage in phage infection. There is an epitope-specific protein-protein interaction between the phage's tail spike protein and a receptor on the cell surface of the host, however it is not always a protein receptor – e.g. P22 binds to LPS. These interactions are highly specific and make phage infection a highly specific process. There are a range of target proteins and cell surface components that phages can use to infect bacteria (Shao and Wang, 2008). These include but are not limited to: cell surface receptors such as the TolC protein and the LamB protein on the surface of *E.coli* (German and Misra, 2001, Roa, 1979), structural proteins that span the peptidoglycan layer (Koebnik, 1999), porins (Nakae, 1976) and substrate receptors (Filippov et al., 2011). The bacterial flagella (Schade et al., 1967) and pili (Guerrero-Ferreira et al., 2011) may also be targeted. Phage-Chi is an example where the cellular target is the *E.coli* flagella (Schade et al., 1967). It interacts with a flagellum and winds to the base where the flagellum is adjacent to the surface of the bacteria. Once attached to the host cell, the entry of the viral DNA into the cell must take place. This is a two-step process, including (i) the degradation of the host cell envelope, which allows the virion to reach the plasma membrane, followed by (ii), the ejection of the genomic DNA into the cytoplasm of the infected cell. Phages carry enzymatic equipment (often as domains on tail fibres), which are used to penetrate through host cell envelopes and other protective structures, including capsules and proteinaceous layers (Fernandes and São-José, 2018). For example, ϕ 29 that infects *Bacillus subtilis* carries a cell wall degrading enzyme (Xiang et al., 2008). The DNA delivery mechanism depends on phage morphology and the type of phage tail. If a contractile phage, such as a myovirus, injects its DNA through the contractile tail, across the outer and inner membrane and into the cell itself, it may propagate using the cells transcription machinery or go into the lysogenic cycle.

1.5.7.2 Lytic phages

In the lytic life cycle, phage replication takes place using the host cell's transcription machinery to produce new phage virions, which are then released by cell lysis and cause the death of the host bacteria. An outline of the lytic and lysogenic lifecycles is shown in figure 1.8, using lambda as a model temperate phage that infects *E.coli*, as it is well characterised (Johnson et al., 1978, Ptashne, 1992). The lambda phage has immediate early, delayed early and late genes and the expression of these are regulated to ensure that the lytic life cycle proceeds correctly. The immediate early genes are required for the default lytic life cycle to proceed (and prevent the lysogenic life cycle), whilst delayed early genes enable phage DNA replication and late genes enable phage assembly (Oppenheim et al., 2005). The mechanism is described further in section 1.5.7.3.1. This mechanism of transcriptional control means that the genes required for DNA replication and phage particle formation are expressed at temporally discrete points in the phage life cycle, so the viral particles can be correctly assembled and are mostly infective upon lysis. Using RNA sequencing, Chevallereau et al. (2016) investigated the patterns of both host and phage transcription in *Pa* after infection by lytic phages and showed that infection alters the transcription of over a thousand host genes and temporally regulates the expression of its own genes, with most of the transcription targeted at phage replication. (Chevallereau et al., 2016)

1.5.7.3 Lysogenic phages

Lysogenic or temperate phages can undergo the lytic or lysogenic cycle. However, they have a tendency to enter the lysogenic cycle and integrate their viral genome into the bacterial genome, which can remain and replicate like any other genetic loci through cell division. The phage lysogeny decision is dictated by gene expression and regulation of a simple genetic switch region (described further in section 1.5.7.3.1). If the lysogenic life cycle takes place it leads to the transcription of the *int* region, which transcribes the integrase needed for the temperate phage to specifically integrate its DNA in the bacterial genome at location attB. The integrase is essential for phage insertion into the bacterial chromosome, confirmed by a *int* negative mutant (Zissler, 1967). These phages that are integrated into the bacterial host are termed 'prophages' and the bacteria that harbours the

prophage is termed a 'lysogen'. The prophages can evoke many different effects on their host, such as increased fitness and virulence (Davies et al., 2016a) by not only using viral genes but altering host gene regulation. Bacteria are also capable of carrying multiple prophages in their genome, termed 'polylysogens'. Multiple prophages have been shown to cause effects on the host bacteria when integrated together, which they do not cause as single lysogens (Burns et al., 2015).

Temperate phages are capable of switching back to the lytic cycle, in a process called induction where the prophage is excised from the bacterial genome (Figure 1.8). this process is commonly due to a change in host conditions, such as damage to the host cell DNA or host stress responses, as induction has evolved to be intrinsically linked to the health and environmental selective pressure on the host bacterial cell. Phages use bacterial stress signals to trigger the induction (described further in section 1.5.7.3.2). Induction can occasionally occur in the absence of any known environmental triggers and is known as spontaneous induction. This often occurs at low but detectable levels in bacterial cultures and varies between strains and has previously been reported to occur in Pa (Fothergill et al., 2011, James et al., 2012).

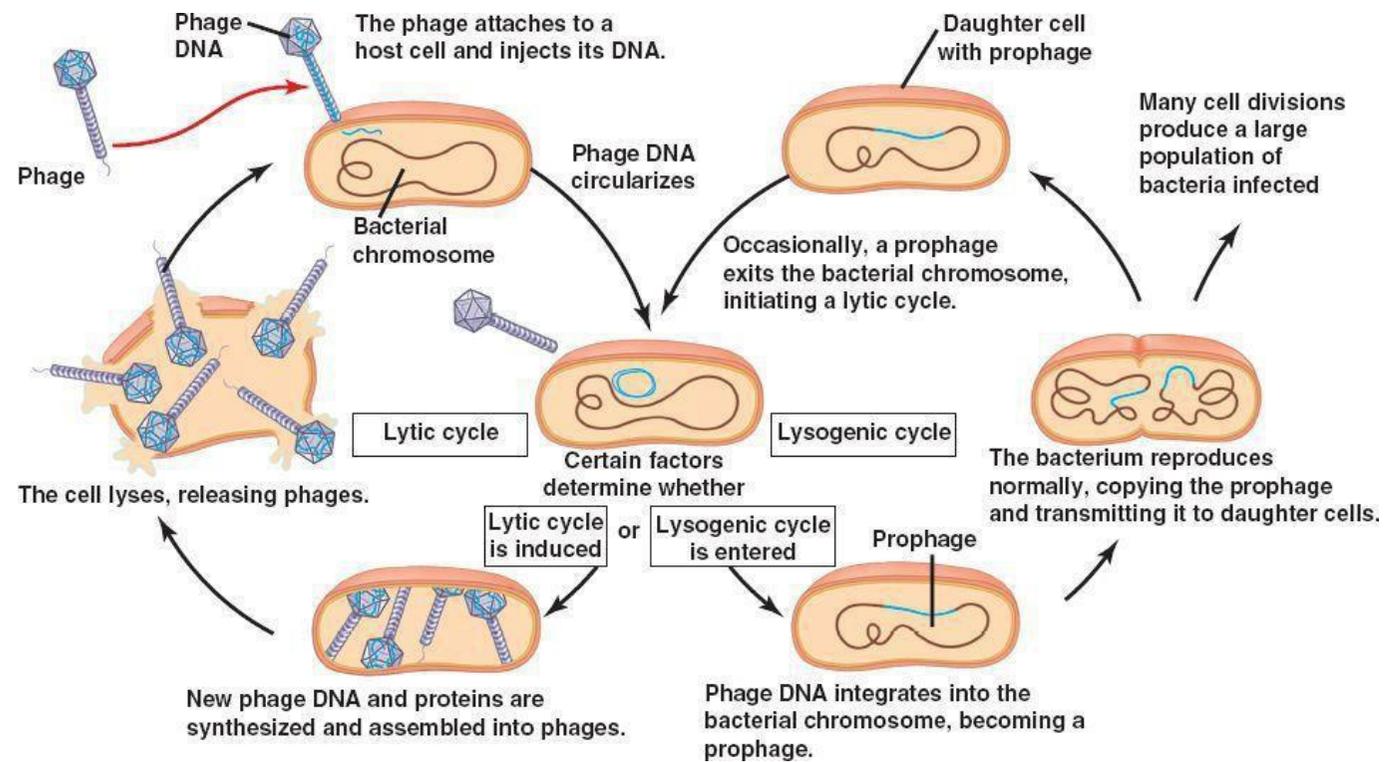


Figure 1.8 Lytic and lysogenic life cycle. This schematic diagram illustrates the life cycle that temperate phage lambda can exhibit. The life cycle on the left shows the lytic cycle, showing how the phage injects its DNA into the bacteria. New virions are produced and the cell lyses, releasing the phages. The lifecycle on the right shows the lysogenic cycle. After injection of the phage DNA it is integrated into the bacteria genome, replicating with the bacterial DNA into subsequent daughter cells. Image taken from (Reese J et al., 2011). Copyright approved.

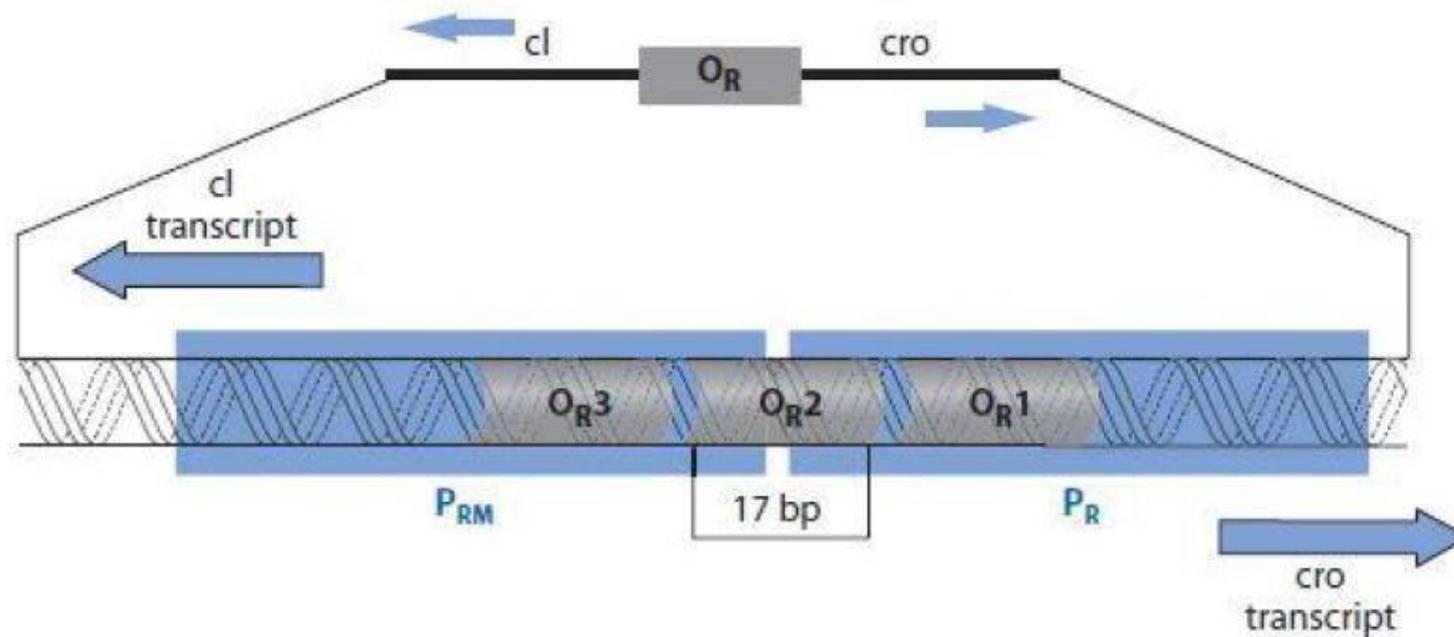


Figure 1.9. A short segment of the λ operator module that controls the lysogeny decision. Promoters (P_{RM} and P_R) can send polymerase travelling in opposite directions depending on what cycle has been initiated: left to transcribe the repressor gene (*cl*) and rightward to transcribe the *cro* gene, with the three-part right operator (O_R) overlapping the two promoters. Image taken from (Ptashne, 2004). Copyright approved.

1.5.7.3.1 The lysogeny decision

The beginning of both lytic and lysogenic phage life cycle is infection, initiated by adsorption. Once the phage has adsorbed, infection continues with the injection of the phage nucleic acids from the phage capsid, into the bacterial cell. In most dsDNA phages once the phage genome is in the host cell it circularises at the two cohesive ends or *cos* sites using DNA ligase and DNA gyrase (De Wyngaert and Hinkle, 1979). At this point, the lytic and lysogenic life cycles diverge. Temperate phages are capable of both lytic and lysogenic infection, and the complex molecular process that defines which path is taken is known as the lytic-lysogenic decision (Ptashne, 2004). Lysogeny is thought to be favoured in environments with low nutrients in which the bacteria are starving, because phages require cells that are metabolically active to grow in the lytic cycle. The ability to sense this change in cellular status is important for the temperate phage and is driven in lambdoid phages by an articulate, yet simple genetic switch. (Figure 1.9). The lytic-lysogenic decision in phage lambda is well described and relies on the by presence or absence of a protein known as CII, which is one of the first proteins to be synthesised by the incoming phage (Ptashne, 2004). The process that leads to the lytic cycle starts with transcription from *p_R* that encodes *cro*, which is a repressor regulator of the lytic cycle. It does this by binding to domains of an operator region (*OR*), which prevents premature synthesis of *C_I* (the regulatory protein of phage lysogeny and repressor maintenance) (Johnson et al., 1978). The expression of phage DNA replication proteins *O* and *P* stimulates circular phage DNA replication at location *ori* in the phage DNA (Ptashne, 2004). This then leads to the production of the *Q* anti-termination protein, which is a regulator of late genes, such as capsid and tail genes, through modification of the host RNA polymerase.

The CII activator protein is rapidly degraded by host proteases in metabolically-active bacterial cells, which allows the lytic cycle to progress. Therefore, high levels of CII protein accumulates in inactive cells that have low protease levels, driving the synthesis of the phage *C_I* repressor and integrase proteins, which blocks the lytic replication of the phage and allow stable integration into the host chromosome. Gene expression from *p_L* and *p_{RE}*

regulate *CII* and *CIII* which are necessary for producing *CI* from promoter pRE and then switches to promoter *Pint*. This transcribes integrase to specifically integrate phage DNA into the bacterial genome at location attB, as described previously (section 1.5.2.3). *CIII* is transcribed from the pL promoter and regulates the lysogenic pathway by stabilising *CII* (Altuvia et al., 1987) and stabilised *CII* activates *CI* gene expression (Kornitzer et al., 1991). *CI* has two promoters: pRE, for establishing lysogeny and pRM, for maintaining lysogeny (Dodd et al., 2001). Due to this lysogeny, the *CI* repressor is constitutively expressed, repressing the production of all genes required for phage replication (Ptashne, 2004).

The prophage replicates alongside with the host genome unless an induction signal is detected, such as SOS response of the host cell. The bacterial SOS response is an ancient gene expression cascade which is activated by DNA-damage (Baharoglu and Mazel, 2014, Michel, 2005). Upon interaction with single-stranded DNA, the bacterial RecA protein becomes 'activated' (Sassanfar and Roberts, 1990), though the exact molecular basis of RecA protein 'activation' is not fully understood. Consequently, the phage lytic replication pathway proceeds. Specifically, the prophage excises from the host chromosome, replicates its DNA, synthesises and packages itself inside a protein capsid and tail, and induces cell lysis by the synthesis of proteins which hydrolyse the bacterial cell wall.

1.5.7.3.2 Prophage Induction

Even if lysogeny is established, the prophage can revert to the lytic cycle at any time by a process called induction. The induction of a phage from a bacterium is often stimulated by damage to bacterial DNA and involves excision of the prophage from the bacterial chromosome. There are a number of ways bacterium can be artificially forced to induce phages, as well as phages going through spontaneous induction. Experimental induction can be done either chemically by using antibiotics, such as mitomycin C (Humphrey et al., 1995) and norfloxacin (Matsushiro et al., 1999), or with UV light (Takebe et al., 1967), which forces the prophage into the lytic cycle as it makes the bacterial genome unstable by stimulating the same bacterial response mechanism towards DNA damage. This happens via the SOS response that activates RecA, which additionally causes auto-cleavage of the *cI*/repressor, essential for maintenance of lysogeny. This triggers a cascade of genes involved in the λ

lytic cycle, including excisionase, which allows the phage DNA to excise from the bacterial genome (Ptashne, 2004). Excisionases are needed for excision from the genome, including some other genetic factors that help with DNA bending.

1.5.8 Mechanisms of phage resistance

Many bacteria evolve to carry phage resistance mechanisms within their genome to prevent or terminate phage infection as a survival tactic, as the majority of phage infections result in lysis of the cell. The phage-bacteria co-evolutionary arms race, proposed as the Red Queen hypothesis, suggests that phages are constantly adapting and evolving in a host/prey relationship. Other mechanisms by which bacteria attempt to resist the infection of phages include superinfection exclusion and the abortive infection system. Bacteria have numerous mechanisms to prevent or terminate phage infection. The simplest mechanism involves changing the sites where phages infect, through masking or mutation. This inhibits the absorption of the phage and can limit or stop phage infection (detailed below in section 1.5.8.1). Loss of a target epitope for the phage can make the bacterium immune to specific phages. However, phages can also evolve to enable them to infect by using other epitopes.

1.5.8.1 Inhibition of adsorption

The prevention of adsorption is the first and simplest way a bacteria can evade infection. As the phages need the specific receptor to enable adsorption, the absence or alteration of the receptor prevents this interaction. Type IV pili are a common receptor of Pa phages (Ceyssens, Noben et al. 2009). Pili mutations within the Pa *piliA* gene results in a loss of motility, as well as removing the phage target receptor, conferring bacterial immunity. Immunity is dependent upon which *pil* gene is mutated as mutations in *pilU* do not lead to phage resistance (Whitchurch and Mattick 1994).

1.5.8.2 Superinfection exclusion (Sie)

By inhibiting bacteriophages from injecting their DNA into the cell, this mechanism can support immunity to the targeting bacteriophage. Sie proteins encoded by prophages already present in the bacterial cell fulfill this role against related phages. If the phage DNA manages to enter the bacterial cytoplasm it can be cleaved or methylated, which makes it resistant to restriction and inhibits the phage lytic cycle (Gingeras and Brooks, 1983).

1.5.8.3 The abortive infection systems (Abis)

Abis are another strategy encoded by the bacterium to negate incoming genetic material. This is an altruistic system to protect surrounding bacteria, where cell death or senescence occurs in the cell to inhibit phage infection and replication (Chopin et al., 2005, Fineran et al., 2009) by termination of cellular metabolism (Dy et al., 2014). AbiE and AbiF are gene products identified to cause an abortive infection (Garvey et al., 1995). The AbiE system has been shown to halt phage proliferation by having a bacteriostatic effect (Dy et al., 2014). Abis are used as last resort, where the bacterial cell sacrifices itself to protect other bacteria from phage infection. There have been no *Pa* specific Abi systems yet described in the literature. Such systems are common in the *Lactococcus lactis* (used in the cheese making industry), where 20 Abi systems have been identified (Chopin et al, 2005).

1.5.8.4 CRISPR/cas

The Clustered Regularly Interspersed Palindromic Repeat or CRISPR-Cas system is another bacterial response to phage infection and can offer immunity to phage infection based on the incoming phage gene sequence. It was first reported in 1987 by Ishino et al. (1987) who found a 29 nucleotide repeat with 32 nucleotide spacer sequences. However, the function was unknown at this point. The CRISPR locus was further characterised by Francisco Mojica and Ruud Jansen and termed CRISPR (Jansen et al., 2002) but it was not until 2005 that the spacer sequences were shown to be complementary to bacteriophages (Mojica et al., 2005). The carriage of these recognition regions illustrates adaptive bacterial phage immunity to phage infection by genetically similar bacteriophages. The CRISPR system has been identified in many bacterial species (Barrangou et al., 2007), with one study showing it was present in 36% of clinical *Pa* strains tested (Cady et al., 2011). Two types of CRISPR- Cas have been uncovered in *Pa*, type I-E (similar to *E.coli*) and the more common type I-F (*Yersinia*-like) (Cady et al., 2011). This has been shown experimentally in PA14, a *Pa* strain with a known functional type I-F system, and a panel of phages. It demonstrated that the bacteria acquired CRISPR-mediated resistance in response to phage challenge of the infection (Cady et al., 2011).

1.5.9 Lysogenic conversion

Lysogenic conversion occurs when new genetic information (phage genome) is added to a bacterial genome by a lysogenic phage. This new genetic information then can change the properties of a bacterium through infection with a temperate phage, therefore lysogenic conversion constitutes the phenotypic effects of prophage carriage to its host cell. Due to this they not only alter the biology of their bacterial hosts but in turn can influence the surrounding community of host and non-host cells. Three such effects of lysogenic conversion, that have been previously reviewed are mentioned here. First, integrating prophages can engineer a host's genome (Menouni et al., 2015) and help to regulate gene expression and function such that integration and excision alters cellular processes (Feiner et al., 2015). Second, prophages can change cell physiology by introducing novel functions or altering pre-existing ones, such as virulence factor production, metabolism, cell development and immunity to phages (Hargreaves et al., 2014). Third, through transduction, temperate phages can facilitate the transfer of bacterial DNA that potentially confer new phenotypes such as antibiotic resistance (Davies et al., 2016). *E. coli* phage lambda is regularly used as a model of lysogenic conversion (Oppenheim and Slonim, 1971). One study has shown when *E. coli* K-12 is infected with lambda phage and lysogenically converted cells produced temperature-sensitive repressors, and were significantly more resistant to normal serum than the uninfected organisms (Muschel et al., 1968). Infection of *E. coli* K-12 with a lambdoid phage, phi80, whose prophage attachment site is different from that of lambda, did not result in a detectable change in serum resistance, showing how conversion can benefit the host, with these benefits being phage specific. Another classic example of lysogenic conversion benefiting the host is the Shiga-toxin encoding phage ϕ 24B, which as the name suggest carries a gene that encodes for the Shiga-toxin in *E. coli* drastically changing the virulence of the host (Friedman and Court, 2001). This shows the importance in studying lysogenic conversion and the effects of prophages on their hosts.

1.6 AIMS

- Evaluate prophage identification tools and compare their features, and compare the best prophage identification tools to identify prophage regions from the well-defined Pa genome LESB58.
- Compare prophage genomes carried in colony variants of Pa isolated from the same patient's lung, and ascertain whether phages transferred in the lung environment.
- Investigate how the temperate LES phages and temperate phages from clinical Pa isolates from CF and BR patients subvert the host bacterial cell's metabolism, using individual and polylysogens of common Pa lab strains.
- Determine if the metabolic changes caused by the addition of Pa temperate phages from CF and BR patient have the ability to change the virulence of Pa.
- Identify temperate phages from the sequenced Pa genomes in the IPCD, and compare the prophage genomes to identify any prevalent phage types in a global Pa dataset.
- Determine if the phages from each clade in the IPCD subvert the metabolic physiology of their Pa host. To compare between the phage groupings and to compare/contrast which metabolic pathways are subverted by phage conversion.
- Use longitudinally collected Pa samples from BR patients to determine if prophage carriage changes over time and if they play a role in evolution of the bacteria in the lung environment in chronic infection.

1.7 Objectives

- Compare the features of commonly used prophage identification tools to evaluate which would be the best and most suitable for our needs in each of our projects.
- To compare the putative prophages in these co-colonising isolates to ascertain the number of prophage regions per bacterial genome, identify the attB site of prophage integration assess whether this would impact on cellular function and lead to phenotypic alterations and determine whether phages are shared between isolates taken from the same patient indicating transfer.

- Using LC-MS and metabolomics analysis, investigate how phages change the bacterial metabolism of PAO1 in Artificial Sputum Media (ASM) to better represent the environment of lung sputa. Firstly, LES phage lysogens of PAO1 will be investigated and then PAO1 that have been infected with clinical phages induced from CF and BR patient Pa samples. Using a Galleria virulence model, compare the infection of PAO1 with PAO1 that have been infected with clinical phages induced from CF and BR patient Pa samples.
- Use PHASTER to identify each phage class (intact, questionable and incomplete) prophages within the IPCD Pa genomes, then compare the intact phage genomes using SaturnV (which uses protein-protein comparison in an 'all versus all' manner) to group similar prophages carried by Pa isolates from the IPCD and determine the diversity of temperate phages in Pa.
- Induce phages from Pa isolates from each clade identified that only carry one phage and infect a suitable lab Pa strain to make a lysogen and check integration using PCR and primers for each clade, then grow each lysogen in ASM and prepare for metabolomics analysis to determine how phages from each clade effect the Pa host.
- Use pairwise comparison to compare the prophage genomes within Pa samples collected longitudinally from BR patients between isolates from the same patient to identify whether they differ and how this may play a part in evolution of Pa in the chronically infected lungs of BR patients.

2 MATERIALS AND METHODS

2.1 MATERIALS AND GROWTH MEDIA CONSTITUENTS

2.1.1 Sterilisation

Sterilisation was performed by autoclaving at 121 °C for a 20 minute cycle at a pressure of 15 psi before use.

2.1.2 Media

Table 2.1 List of Media and their constituents

Media	Ingredients
Bottom LB agar for plates;	12.5 g (2.5% w/v) of LB broth (Sigma Aldrich, Gillingham, UK) and 7.5 g (1.5% w/v) (Lab M, Heywood, UK) of phage agar in 500 ml of distilled water.
1M CaCl ₂	147.02 g dissolved in 800 ml , make up to 1 L
Soft (top) agar	5 g (2.5% w/v) of LB broth and 0.8 g (0.4% w/v) of phage agar, 0.01 M calcium chloride before having 200 ml of distilled water added.
Norfloxacin (NFLX) (Sigma Aldrich, Gillingham, UK)	1 mg/ml stocks, in water, few drops of 1 M NaOH were added to alter pH so drug dissolves into solution
Phage buffer/Growth media LB Broth, plus 0.01 M CaCl ₂	12.5 g (2.5% w/v) of LB broth + 5 ml of 1 M CaCl ₂ stock in 495 ml of distilled water

2.1.3 *Pseudomonas aeruginosa* bacterial growth media

All Pa strains were grown at 37°C, unless stated otherwise, either in liquid or on solid Luria Bertani media (LB) (Sigma Aldrich, Gillingham, UK) + 0.1 M CaCl₂ (Sigma Aldrich, Gillingham, UK) and high clarity agar (Lab M, Heywood, UK) used for soft agar overlay (0.4%) or bottom agar (1.5%).

2.1.4 Storage and maintenance of bacterial cells

Strains of Pa were stored in Glycerol Skimmed Milk (GSM) at -80°C. GSM was prepared using 20 ml Glycerol (ReAgents, Runcorn, UK) 180 ml dH₂O (pH 5.6) 6 g TSB (Lab M, Heywood, UK) 4 g skimmed milk powder (SMP) (Tesco, Newcastle upon Tyne, UK) 1 g D-Glucose (Fisher Scientific, Loughborough, UK).

2.2 BACTERIAL STRAINS

Longitudinal bacterial strains were obtained from the Freeman Hospital at Newcastle upon Tyne, NHS trust. The patients were selected, their sample taken, typed using Matrix-Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) and stored by members of the trust. The patient and isolate numbers for BR patients are described in appendix A Table 1. Pa isolates were supplied by Craig Winstanley's lab at the University of Liverpool, from BR patients with Pa samples with multiple strains (described in appendix A, Table 2). They also supplied the LES lysogens of PAO1 (described in appendix A Table 3).

Pa strains were also obtained from the IPCD via Roger Levesque, with the owner's permission. These are Pa samples from all over the world collated in a database at the university of Laval Quebec, Canada and details for these strains are described in appendix A, Table 4, There were also lysogens that were made using phages infected into a lab strain of PAO1 from clinical isolates and stored (described in appendix B, Table 5).

Ethical approval was granted (REC reference: 11082).

2.2.1 Storage of viable bacteria in a sample

All strains came from previously prepared glycerol stocks stored at -80°C. These were streaked out for single colonies and incubated at 37°C overnight to see if the sample was viable. A single colony was taken and re-streaked and incubated at 37°C overnight again. For the strains collected from the Freeman hospital, the colonies were also tested using a MALDI-TOF to confirm that the colony is Pa. New glycerol stocks were made from the second overnight if it was confirmed to be Pa (for the samples from the Freeman) and stored at -80°C.

2.3 PSEUDOMONAS AERUGINOSA CULTURE PARAMETERS

2.3.1 *Pseudomonas* culture in liquid broth

Unless stated in the text the standard culture conditions for *Pseudomonas aeruginosa* samples are as follows. All stocks of bacterial isolates were held at Ultra Low Temperature (-80°C) in 50% glycerol (Sigma Aldrich), stocks were raised as streak plates on Luria Bertani agar. From the streak plate, single colonies were inoculated into 10 ml of phage buffer and grown ~18 h at 37 ° C (overnight), shaking at 200 rpm.

2.3.2 Use of artificial sputum medium (ASM) to culture bacteria

2.3.2.1 Preparation of ASM

ASM was prepared as described in (Kirchner et al., 2012) (Appendix B) and stored at 4°C for a maximum of eight weeks.

2.3.2.2 *Pseudomonas* culture in ASM

Bacterial cultures were set up using standard culture conditions from a single colony (37°C, shaking at 200 rpm) of *P. aeruginosa* for ~18 h. The culture was diluted in LB to an OD600 of 0.05 (± 0.01), then further diluted 1:100 in fresh ASM (total volume 1.8 ml). This ASM bacteria mixture was then added to each well of a 24-well tissue culture sterile plate. Three wells contained ASM only for use as a blank. The plates were then secured with parafilm and incubated for 3 days (72 hours) at 37°C, with shaking at 75 rpm.

2.4 BACTERIAL MOTILITY ASSAY

2.4.1 Twitching motility

A colony was stabbed through to the bottom of a LB agar plate and incubated for 24 hours. The agar was removed with tweezers and 5 ml 0.25 % (w/v) crystal violet (TCS Biosciences, Botolph Claydon, UK) added to the plate and left at room temperature for 30 minutes to stain the area of bacterial growth. The crystal violet was poured off and plate rinsed with water. The diameter of bacterial growth was measured at the widest point. An isolate with a diameter <10 mm was considered to have impaired twitching motility. LESB58 was used as a negative control during these assays as it is known to be non-motile (Winstanley et al., 2009). PAO1 was used as a positive control.

2.5 GENERAL BACTERIOPHAGE PROTOCOLS

2.5.1 Temperate Phage induction of the lytic life cycle

Temperate bacteriophages were chemically induced using norfloxacin from bacterial isolates (Matsushiro et al., 1999). Overnight cultures (37°C, shaking at 200 rpm) were sub cultured 2 % (v/v) into 10 ml Luria Bertani (LB) Broth (Sigma Aldrich, Gillingham, UK) supplemented with 100 mM CaCl₂. At mid exponential growth phase 0.5-0.6 (OD_{600nm}) the bacteria were exposed to norfloxacin (1 µg mL⁻¹) (Sigma Aldrich) for 1 hour (37°C, 200 rpm). Cultures were diluted in LB (1:10) to reduce the effects of norfloxacin and to begin the recovery for 2 hours at 37°C. A negative control was always used to validate results by negating the addition of the bacteria.

2.5.2 Separation of free infective bacteriophages from bacterial host

Bacterial cultures were filtered through a 0.2 µm filter (Millipore) to remove cells and debris, unless otherwise stated. The phage-containing filtrate was stored at 4°C for < 1 week.

2.5.3 Standard plaque assay

In order to determine whether the lysates contained viable temperate phages, plaque assays were performed. A dilution series of the lysate was made and 100 µl of each dilution was added to 100 µl of a suitable indicator strain, usually PAO1 grown to mid-exponential growth phase 0.5-0.6 (OD_{600nm}), vortexed and left at 37°C for 20 minutes. Then 5 ml of molten soft agar was added and the tube slowly inverted three times to minimise bubble formation. The mixture was overlaid onto a standard LB agar (1% w/v) plate and allowed to

set, before incubation overnight at 37°C. Plaques were counted and plaque forming units (p.f.u.) per ml calculated as an estimate of the number of infective phages present. Three technical replicates were performed per bacteriophage dilution.

2.5.4 Spot assay

Where a more high-throughput method was required, a modification of the traditional plaque assay was used. The suitable indicator strain was mixed with soft agar and overlaid onto LB agar. Spots of 20 µl of lysate dilutions were spotted on top of the soft agar and allowed to dry, before incubation overnight at 37°C. If a zone of clearing was seen, the lysates were diluted to define individual plaques to negate clearing due to pyocins and demonstrate productive infection.

2.5.5 Bacteriophage cross-infections and bacterial sensitivity

In order to assess the infection profiles of the temperate phages in the lysates, spot assays were firstly performed against a range of Pa bacterial hosts, each as a lawn on a separate plate, to assess which hosts the phages could infect to form an infection profile of each lysate. These plates were grown overnight at 37°C and observed for zones of clearance. Plaque assays (section 2.3.3) were then performed with the corresponding lysates that showed clearance against the same host lawn, to define individual plaques to negate clearing due to pyocins.

2.5.6 Lysogen preparation

Firstly, phages were checked to determine whether they infected the specific bacteria needed to make a lysogeny, by performing a plaque assay (section 2.3.3). A 100 µl volume of purified or mixed phage lysate (either having one type of phage or a mixed population of phages) was added to 100 µl of mid-exponential bacterial host and incubated at 37°C for 20 minutes. A serial dilution of the mix was made and 100 µl of each dilution was plated onto a solid LB plate and incubated overnight at 37°C. Five to ten colonies were selected and PCR was used to confirm lysogeny of the phage. A crude DNA extraction (section 2.6.1.) was used on each of the isolates to make a template DNA for the PCR (section 2.5). If lysogeny had occurred, this was shown by a positive band. DNA from the original isolate was used as the control. For unknown phages, or to confirm the insertion sites, DNA extraction (section 2.5.1.2) was used and the DNA was sequenced (section 2.7) to confirm

the phage insertion. The positive lysogens were stored at -80°C.

2.6 DNA EXTRACTION AND PREPARATION

2.6.1 Bacterial Phenol chloroform DNA extraction method

A culture of each bacteria was grown overnight in 10 ml of LB. A 1.5 ml aliquot of overnight culture was transferred to a microcentrifuge tube, the cells were pelleted through centrifugation at 12,000 x g for 2 minutes and the supernatant was discarded. Tubes were then transferred to a gas hood and one volume of phenol:chloroform:isoamyl alcohol (25:24:1) (Thermo Fisher, Loughborough, UK) was added to each sample and vortexed thoroughly for approximately 20 seconds. Samples were then centrifuged (within the hood) at room temperature for 5 minutes at 16,000 x g. The upper aqueous phase was then removed and transferred to a fresh tube, taking care not to carry over any phenol from below. An ethanol precipitation step was carried out next. 1 µl of glycogen (20 µg/µl) was added to each sample, followed by 0.5 volumes of 7.5 M NH₄OAc (ammonium acetate). To precipitate the DNA from the sample 2.5 volumes of 100% v/v ethanol was then added, vortexed and placed at -20°C overnight. The samples were then centrifuged at 4°C for 30 minutes at 16,000 x g to pellet the cDNA. The supernatant was then removed without disturbing the DNA pellet and 150 µl of 70% v/v ethanol was added and centrifuged again at 4°C for 2 minutes at 16,000 x g and the supernatant removed. The pellet was then dried in a concentrator (SpeedVac, Thermo Scientific) at room temperature for 5-10 minutes. The DNA pellets were then resuspended in 150 µl of TE (Tris- EDTA) buffer by pipetting up and down 30-40 times and then stored at -80°C until ready to be sequenced.

2.6.2 Bacterial genomic miniprep kit method

The GenElute bacterial genomic DNA kit (Sigma Aldrich, Gillingham, UK) was used to extract DNA from all bacterial isolates, except the longitudinal isolates. Samples were processed using the manufacturer's protocol and stored at -80°C until further processing.

2.6.3 Phage isolation and DNA extraction

After confirmation of phage presence in the lysate (method 2.4.1), any remaining bacterial DNA and RNA within the lysate was degraded using 1 µl of TURBO DNase and 1 µl of RNase Cocktail (Life Technologies Limited). The lysate was incubated at 37 °C for 30 minutes. The DNase and RNase were inactivated at 65°C in 15 mM EDTA (final concentration) for 10 minutes. NORGEN Phage DNA Isolation Kits (Geneflow Limited, Lichfield, UK) were used to extract viral DNA. The manufacturer's protocol was followed apart from the final DNA elution step where 50 µl of elution buffer was used and passed through the elution column twice to increase the total yield. The DNA samples were stored at -80°C until they were sequenced.

2.6.4 QUBIT 2.0

In order to quantify the DNA concentration prior to sequencing, the QUBIT dsDNA high sensitivity assay kit was used in conjunction with a QUBIT 2.0 fluorometer (Life Technologies Ltd, Paisley, UK) was used. Briefly, 197 µl of HS buffer was combined with 1 µl of fluorescence dye and 2 µl of DNA. The tubes were vortexed for 5 seconds prior to incubation in the dark for 2 minutes, in order to let any bubbles generated by vortexing settle. The sample was then read on the fluorometer, selecting the dsDNA option. Care was taken to observe for bubbles and scratches on the tube, which could influence the output reading.

2.7 GENERAL PCR METHODOLOGY

2.7.1 Crude DNA preparation using boiled bacterial colonies

A small portion of a single colony was transferred into 50 µl SDW in a 0.2 ml PCR tube. The suspension was heated to 100°C for 5 minutes and then used immediately in the standard PCR.

2.7.2 Standard polymerase chain reaction (PCR)

Standard PCR was carried out in 96 well PCR plates or 0.2 ml PCR tubes, depending on the number of samples. Each reaction contained: 2.5 µl of 10x taq buffer (New England Biolabs), 0.5 µl of 10 µM dNTPs (New England Biolabs), 0.5 µl each of 10 µM forward and reverse primers (Eurofins), (Table 2.2) and 1 µl of crude DNA template. SDW was added to a final volume of 25 µl. All primer sequences and PCR cycling conditions are described in Table 2.2. PCR cycling was carried out in an Eppendorf MasterCycler Gradient.

Table 2.2 Oligonucleotide primers used in this study

Name	5→3 sequence	Product size (bp)	Target gene/ region	Cycling conditions
CL1F	GAGGACTTCGAYGCGGTGG	206	Clade 1 phage region	95°C, 4 min then 30 cycles: 95°C, 30 s;
CL1R	TGCTTCTTCGCGCTGTTAC			
CL2F	GCTCTGCGCGTACATCCTC	155	Clade 2 phage region	58°C, 30 s; 72°C, 30 s; final extension step, 72°C, 7 min
CL2R	TGTTGCCCTTGTTGCGGTA			
CL3F	GTTGCTGCCTGGAATTCCG	157	Clade 3 phage region	
CL3R	CAACGCAACCGGACACGC			
CL4F	CTCGATCGGYAACAGCGCG	315	Clade 4 phage region	
CL4R	GCAGTGTCGATCTGCGCC			
CL5F	GTTCTACTGAAAGGCGAGTT	155	Clade 5 phage region	95°C, 4 min then 30 cycles: 95°C, 30 s; 53°C, 30 s; 72°C, 30 s; final extension step, 72°C, 7 min
CL5R	GACTCGGTGCGATAGGCC			
CL6F	TATAACGTCCACGACACCG C	137		

CL6R	TTTTCGCCCTTGATCTCGCT		Clade 6 phage region	95°C, 4 min then 30 cycles: 95°C, 30 s;
CL7F	GCACCGCGCCGTTCTGCTC	164	Clade 7 phage region	58°C, 30 s; 72°C, 30 s; final extension
CL7R	GTCGAGGTCACGGCGGATG C			
CL8F	CGGCGGGCGGTTTCAT	207	Clade 8 phage region	step, 72°C, 7 min
CL8R	GTTTCGTAGGTCAGGCTTTC C			

2.7.3 Visualisation of PCR products

PCR products were visualised by agarose gel electrophoresis. Molecular grade agarose (Sigma Aldrich, Gillingham, UK) was dissolved in 0.5X TBE buffer. A 3% (w/v) agarose gel was used for PCR products below 650 bp and a 1% (w/v) gel used for larger DNA fragments, or visualisation of intact genomic DNA. The agarose was melted by heating and 1 µl of SYBR safe (Thermo Fisher) was added. PCR products were mixed with 1 part loading dye (New England Biolabs), and 5 µl was added to each well. A 1 KB+ or 100bp marker (New England Biolabs) was run alongside for size determination of products. Electrophoresis was performed for 25-30 minutes at 120V or 10V/cm and DNA was visualised under an UV transilluminator (Syngene InGenius).

2.8 GENOME SEQUENCING

Sequencing was partly conducted at the NU-OMICS DNA sequencing facility (Northumbria University at Newcastle, UK) on an Illumina MiSeq, and partly at Laval University, Quebec (IBIS) on an Illumina MiSeq platform. The phage genomes and the lysogens were sequenced in house by NU-OMICS and the longitudinal bacterial Pa isolates (chapter 4) were sequenced at Laval University. Libraries were prepared using Illumina NexteraXT and run on the V2 500bp chemistry. The samples were demultiplexed on the instrument and >Q30 paired end sequencing reads were provided as FASTQ files (NU-OMICS, Northumbria University at Newcastle, UK).

2.9 IN-VIVO ANIMAL MODEL

2.9.1 *Galleria mellonella* larvae model

The *in vivo* assay was conducted on a wax moth larvae model (*G. mellonella*, Livefoods Direct, Sheffield, UK) by Zuzanna Drulis-Kawa's group at the University of Wroclaw using

the following protocol. The *in vivo* assay was conducted on a wax moth larvae model (*G. mellonella*, Livefoods Direct, Sheffield, UK). Prior to each experiment, the larvae were subjected to a 7-day acclimatization period in the dark at 15 °C. Experiments were performed in triplicate (ten larvae per trial). For survival control, we observed both untouched larvae and larvae injected with 20 µl of sterile saline buffer. Depending on the growth rate of the tested bacterial strains, experiments were conducted up to 96 h. Freshly plated bacterial cultures were inoculated into Luria Broth (LB, Sigma-Aldrich) tubes and incubated at 37 °C overnight under agitation. After 18 h, the bacterial suspension was diluted with physiological saline to OD₆₀₀ = 0.1 and this was used as a starting point for serial dilutions. For the measurement of *P. aeruginosa* virulence, 10 µl of diluted bacterial suspension (serially diluted to 10⁻⁸ CFU) were used to infect larvae. Bacteria were administered to larvae by injection into the ventral side of the last pair of pseudopods. After injection, the larvae were incubated for 72 h at 37 °C. The effects of infection were checked at 8, 24, 48, 72, and 96 h after injection by assessment of survival and macroscopic appearance. The results were expressed as the percentage survival rates. The experiments were performed at least three separate occasions. The experiments were controlled by observation of uninfected larvae, sham- infected larvae, larvae receiving phage lysate only, and infected but phage untreated larvae.

2.10 LCMS METABOLOMICS PREPARATION AND RUN

2.10.1 Metabolomic preparation

Experiments were replicated in triplicate (n=9 for each strain). Bacterial cultures were grown in ASM (1.8 ml) as previously described (section 2.1.3.2). After 72 hours of growth the culture was transferred to a microcentrifuge tube.

The samples were then placed at -80C for at least an hour prior to lyophilisation. Lyophilisation was performed by using a multi-pipe lab freeze dryer with manifold vacuum lyophiliser, frozen samples had perforated holes placed in the tops of the tubes and were placed inside were setting to freeze dry for at least 18hrs. Lyophilised samples were then weighed and normalised for weight/vol (2.5 mg/ml) with methanol, which were then vortexed for 30 seconds. The samples were then placed in liquid nitrogen for 1 minute and then allowed to thaw on ice. This was repeated twice more, then vortexed for 30 seconds. The supernatant was filtered

through 0.22 µm filter (Millipore). This was then concentrated in a vacuum concentrator (SpeedVac Thermo Scientific) and stored at -80°C until ready to be run on the LCMS.

2.10.2 Ultra High Pressure Liquid chromatography (UHPLC) or LCMS

Metabolite profiling of the ASM bacterial samples were acquired on a Dionex 3000 Ultra High Pressure Liquid chromatography (UHPLC) system connected to the Q-Exactive classic high resolution mass spectrometer system (ThermoScientific, Bremen, Germany). All solvents and ionisation agents used were of analytical grade. The UHPLC or LCMS run was carried out in the metabolomics lab by Dr William Cheung, using materials and methods detailed in appendix D.

2.11 BIOINFOMATIC ANALYSIS

2.11.1 PHAge Search Tool Enhanced Release (PHASTER)

PHASTER (Zhou et al., 2011, Arndt et al., 2016) version 2, was used to identify prophage sequences within the assembled contigs of each genome. All contigs <2 Kb were discarded before submitting the assembly to PHASTER, as PHASTER processes only contigs of length ≥ 2 kb. PHASTER identifies phages within bacterial genomes and labels them as incomplete, questionable and intact phages by rapidly performing a number of database comparisons as well as phage “cornerstone” feature identification steps. This then forms a score and, if the score is over a minimum threshold, it is identified as a prophage prediction these predicted prophage regions are then annotated (Arndt et al., 2016). The default settings were used. This gave putative values for how many phages were in each bacterial genome. These were manually curated to check that they were phage-like.

2.1.1 MUMmer

MUMmer is a system for rapidly aligning entire genomes, whether incomplete or draft form. The NUCmer command in MUMmer was used to generate nucleotide alignments between reference and query genomes (both bacterial and phage genomes). Default parameters were used to align the two inputs. However, if the alignment was too sensitive or not sensitive enough, the minimum match length and cluster sizes could be adjusted accordingly. Circos (section 2.11.4) was then used to visualise the comparisons.

2.1.2 BLAST

BLAST (basic local alignment search tool) (Altschul et al., 1990) is an algorithm for comparing genome sequences, such as the amino-acid sequences of proteins or the nucleotides of DNA. A BLAST search was used to compare a subject protein or nucleotide sequences (called a query) by using either BLASTn, comparing nucleotide to nucleotide, or blastx, which translated nucleotides into proteins. This could then be compared with either the online database of sequences, or when using command line, be compared against our own databases. The following default parameters were used in BLASTn: word_size = 7, gapopen = 5, gapextend = 2, reward = 2, penalty = -3 when comparing bacterial sequences. When phage genome sequences were used as the query, BLAST was used to identify the most similar phage type it resembled. Command line BLAST was used as a tool to locate similar phages or phage regions in bacterial genomes with an output that showed percentage similarity.

2.1.3 Prokka

Prokka is a genome annotation tool that finds and annotates features (both protein coding regions and RNA genes) present in a sequence (Seemann, 2014). Prokka v1.11 was used through the Galaxy platform (Giardine et al., 2005) using default parameters of an e-value threshold of 10^{-6} . Prokka uses a two-step process for the annotation of protein coding regions: firstly the coding regions are identified using Prodigal (Hyatt et al., 2010), secondly, the function of the identified proteins are predicted by similarity to proteins on databases. Prokka is a software tool that can be used to annotate bacterial, archaeal and viral genomes quickly, generating standard output files in GenBank, EMBL and GFF formats. This was used to annotate bacterial and phage genomes.

2.1.4 Roary

Roary is a high-speed standalone pan genome pipeline, which takes annotated assemblies in GFF format (produced by Prokka (section 2.11.4) and calculates the pangenome (Page et al., 2015). The default parameters were used to create a pangenome with the threshold default being 99% and the minimum percentage identity for sequence comparisons performed by BlastP default being 95%. The gene presence and absence spreadsheet lists each gene and which samples they are present in and how many. Therefore, the genes that

are highly prevalent in a cohort of bacteria or phage are highlighted, allowing identification of core, accessory and singleton genes.

2.1.5 Saturn V

Saturn V is software used to compare two or more genomes in order to determine the core and accessory proteins (Jeukens et al., 2017). Saturn V is a protein-wise comparison, which compares the presence and absence of proteins between genomes. The core proteins are the set of proteins that are present in all genomes of a given dataset in bacteria. However, due to the rate of diversity, when comparing the phage genomes the percentage identity was lowered to $\geq 50\%$ between proteins and $\geq 85\%$ of alignment coverage for the proteins to be considered orthologues. This was carried out by the bioinformatics team at the University of Laval, Quebec, in collaboration with Roger Levesque.

2.2 VISUALISATION TOOLS

2.2.1 Artemis and Artemis comparison tool (ACT)

Artemis is a genome browser and annotation tool that enables visualisation of bacterial and phage sequences (Rutherford et al., 2000), and allowed visualisation of the insertion sites of the prophages that were predicted using PHASTER. ACT was used to visualise the BLASTn or MUMmer comparisons between genomes and similarities/differences could be seen.

2.2.2 Mauve

Once the prophage regions had been identified and extracted using PHASTER (section 2.11.1), regions were then compared at the nucleotide level and visualised, which was achieved using MAUVE (Darling et al., 2004). MAUVE visually highlighted regions of similarity between the phages showing regions and level of heterogeneity.

2.2.3 Circos

Circos is a software package for visualising a range of data and information. Circa (a GUI for making circos plots) (Krzywinski et al., 2009) was used as it is a fast, user friendly way to make complex images. It visualises data in a circular layout, making it ideal for exploring relationships between objects or positions on a genome. Nucmer was used to compare various genomes either one on one, or concatenated genomes so that one genome could be compared to multiple others in one comparison and any regions of similarity or difference could be identified. The insertion sites of the phages made it possible to also show them in a circos plot.

2.2.4 Fig tree

FigTree is a program used to graphically view phylogenetic trees and was used to map the comparisons between prophage genomes from the International Pseudomonas Consortium database (IPCD) using the output from Saturn V (section 2.11.6). This allowed the colouring of the nodes to depict the origins of the phages.

2.3 METABOLOMIC ANALYSIS

Firstly, to streamline the LCMS features identified down to only the metabolites that were significant, analysis of the dataset was performed using the Statistical analysis mode of MetaboAnalyst 4.0 (<http://www.metaboanalyst.ca>) (Xia et al., 2009), from which the peak table is uploaded, and the parameters set to Peak Integrity Table with the format (unpaired). Data filtering was performed using non-parametric relative standard deviation (MAD/median), with normalisation achieved by log transformation and pareto scaling. The resulting analysis path facilitated both Univariate and Chemometric analysis as well as clustering analysis. By using the Partial Least Squares - Discriminant Analysis (PLS-DA)

(data not shown), it gave the Variable Importance in Projection (VIP) score and any metabolite with a VIP score under 1 was discarded. The VIP is a weighted sum of squares of the PLS loadings considering the amount of explained Y-variation in each dimension, therefore the higher the VIP score, the more significant the metabolite change in concentration.

The *Pseudomonas aeruginosa* metabolite database (PAMDB) was used to generate a compound list prior to analysis using MetaboAnalyst 4.0, by searching the given molecular mass of each of the significant metabolites against the database and selecting the top result each time. This provided a list of compounds to take forward. Firstly, the compound list was ranked lowest to highest based upon column retention time with each metabolite assigned a metabolite number (M). Next, each condition/group of lysogens or control was assigned a class value and samples were renamed S1, S2 and so on. For software compatibility, the molecular weight and retention time columns were removed and the spreadsheet saved as a .csv file. Analysis of the dataset was performed using the Statistical analysis node of MetaboAnalyst 4.0 as previously described, this time with fewer metabolites and with compound names. The resulting analysis path facilitated both Univariate and Chemometric analysis as well as clustering analysis and heatmaps.

3. EVALUATION AND OPTIMISATION OF PROPHAGE IDENTIFICATION FROM *PSEUDOMONAS AERUGINOSA* GENOMES

3.1 INTRODUCTION

There are two commonly used methods to identify prophages; experimentally and computationally. The experimental approach involves stimulating stress in the bacterium, which elicits the induction of the phage from the host bacteria. This method determines the existence of viable phages if there is a sensitive bacterial host it can infect. *Pseudomonas aeruginosa* (Pa) are known to carry high numbers of prophages within their genomes, shown in a number of studies by the induction of the phages and sequencing of the phage genomes (Silby et al., 2011). However, induction, phage purification and sequencing are time consuming and costly. Not all prophages within the Pa genomes can be induced and to confirm induction of the phages it is necessary for the phage to infect a sensitive host and plaque purification to be performed. For some species of bacteria plaque assays cannot be performed as the bacteria cannot be grown as a lawn. Therefore, the second method by identifying prophages in the bacterial genome computationally is a cost effective and rapid way of looking at prophage content without experimentation. This can also relate to an environment, thus by examining the genomes of bacteria isolated from an infected lung would mean prophages within them are likely to have also evolved in the lung environment. With the increase of genome sequencing, the computational prediction of prophages from bacterial genomes has become the more commonly used method as many temperate phages are not able to be induced. These non-inducible prophage regions that remain dormant in the bacterial genome, and are not lost over time, may be beneficial to the bacterium and therefore must not be seen as junk DNA inserts. There are numerous bioinformatics programs that can be used for both the identification of prophage regions within bacterial genomes, enabling the analysis and comparison of phage genomes.

3.1.1 Prophages

Prophages make up a significant proportion (> 20%) of the Pa genome but the number and composition can differ significantly between bacterial species with more restricted genomes, such as *Campylobacter*. Prophages are major contributors of diversification in microbes

(Hendrix et al., 1999) as they are one of the key drivers of inter-strain variability in *Pa* (Casjens, 2003) with many remnant prophages (some within the genome) in a state of mutational decay (Srividhya et al., 2007). Collectively genes of remnant and viable prophages may offer positive environmental and evolutionary selection for the bacterium. The impact of prophages on bacterial gene expression has not been significantly researched and reviewed compared to their lytic counterparts (Canchaya et al., 2004; Brussow et al., 2004; Tinsley et al., 2006; Rice et al., 2009). Even more so, the impact of phages on their bacterial host's physiology is a relatively new field of research and often overlooked. One question that is a driving point of this research and for others who work in this area is: if a prophage slightly changes the metabolism and phenotype of a bacterium, is it the same bacterium? Within the lung different phenotypes of *Pa* have been isolated, which could be due to prophages. These differing phenotypes could be due to prophage infections that give a selective advantage, such as the mucoid phenotype in *Pa* (Webb et al., 2004) that is prevalent in the lung infection of CF patients (Valenza et al., 2008). It is known that prophages mediate and drive adaptive evolution of *Pa* (Davies et al., 2016b). Therefore, it is important to identify them within *Pa* genomes, especially from chronically infected patients, to help investigate the role that prophages may have in the chronicity of infection.

3.1.2 Identification of prophages from bacterial genomes

As experimental identification targets only mobilisable phage fractions, it does not reveal incomplete/remnant prophages (Casjens, 2003). Therefore, computational identification of prophages from bacterial genomes has become a useful tool to study these prophage regions in clinical isolates, to compare inducible to remnant prophage regions. Prophages can be considered as a region of an ordered cluster of phage-like genes that compare to previously identified bacteriophages (Zhou et al., 2011). Bioinformatics tools use clustering algorithms to determine if these genes are sequential and related locally enough to each other to constitute a prophage region (Lima-Mendez et al., 2008, Zhou et al., 2011). Identifying prophages within bacterial genome sequences is not simple, as prophage sequences are heterogeneous and are often poorly annotated with many orphan and hypothetical proteins, making unequivocal identification difficult (Casjens, 2003). Prior to these automated bioinformatics tools, identification was completed manually by inspection of

gene regions of altered G + C content and disruption of genes usually found in the backbone of the bacterial host, to allow ways to determine definition of the boundaries of prophage regions. These simple approaches proved to be too unreliable (Nelson et al., 2002), therefore, further research was directed towards detecting 'cornerstone genes' for the purpose of prophage identification (Fouts et al., 2004). Cornerstone genes are those that are essential to the temperate phages biology and encoding its life cycle. An example of this is the integrase gene, which was one of the first proteins used as they are usually identifiable. However, although all temperate phages have an integrase gene, this is not a good indicator of the presence of a prophages (Casjens, 2003). The function of integrase is generated from specific motifs linked to catalysing the strand exchange of attP and attB. These motifs can be located at different locations on the gene or genome and therefore are subject to high levels of diversity outside of these regions. This is apparent in this study by Balding et al. (2005), which used a strategy targeting multiple sites across the gene to determine presence or absence (Balding et al., 2005). This illustrates the high level of intergenic heterogeneity in phage genes but it does make it extremely difficult to use single target genes as a robust method of identifying bacteriophage conversion and lysogeny. Recent genomics methods have looked at more rounded and integrated approaches to identify prophages. These methods bring together pairwise sequence comparisons and alignment to known prophage genes, and also comparisons to known bacterial genes (Altschul et al., 1990, Fouts, 2006). This combination of methods is found in a number of bioinformatics tools, making phage identification more accessible and faster. These tools are reviewed in this chapter.

3.1.2.1 Bioinformatics tools used to identify prophages

Some examples of the bioinformatics tools that were first developed to identify prophage regions in bacterial genome include; Phage_Finder (Fouts, 2006), Prophinder (Lima-Mendez et al., 2008) Prophage Finder (Bose and Barber, 2006), PHAST (Zhou et al., 2011) and the updated version of PHAST, PHASTER (Arndt et al., 2016) (described in section 2.11.1). Phage_Finder searches against a collection of bacteriophage sequences and results from HMMSEARCH (analysis of 441 phage-specific hidden Markov models [HMMs]) locate prophage regions. By using MUMMER (Marcais et al., 2018)

or BLASTN (Altschul et al., 1990), potential attachment (att) sites of the phage region(s) can be found. Prophinder was one of the first web tools for prophage detection. It searches coding sequences (CDS) that are similar to those found in the ACLAME (A CLAssification of Mobile Genetic Elements) database (Leplae et al., 2010) using BLAST. Based on the annotation of the ACLAME database, Prophinder selects the genes with the best correspondence to a potential prophage. Prophage Finder was developed to predict prophage loci based upon clusters of phage-related gene products encoded within DNA sequences. This application provides results detailing several features of these clusters to enable rapid prediction and analysis of prophage sequences without the need for annotated genomes. PHAST was one of the later tools to be developed (2011) and is an integrated search and annotation tool that combines gene prediction and translation (via GLIMMER), protein identification (via BLAST matching and annotation by homology), phage sequence identification (via BLAST matching to a phage-specific sequence database), tRNA identification, attachment site recognition and sequence annotation text mining. PHAST also evaluates the completeness of the putative prophage, allowing the user to focus on phages that are more likely to be inducible, if necessary. PHAST was then upgraded to PHASTER in 2016.

These tools have evolved to offer much faster and more accurate identification of prophages from bacterial genomes. Most phage identification programs mentioned previously were developed between 2005 and 2011. At this time the methods they used involved identifying the 'corner stone genes' (Zhou et al., 2011). The downfall of this approach is the high level of heterogeneity in bacteriophage genomes and low sequence similarity. This extends to the inability to annotate the genes needed for identification of the phage, which is further hindered by the low numbers of phage genomes published to compare against. Therefore, new algorithms for phage identification have been developed recently to improve and update the original programs (Arndt et al., 2016), which aid identification through increasing sensitivity and accuracy.

Important developments in the field since 2011 include the increased number of sequenced and published phage genomes, with the size of the NCBI phage database increasing 15-

fold since 2011 from 600 to 9,000 sequenced genomes. These greatly increased datasets help overcome the high levels of heterogeneity in phage genes and has enabled the development of higher resolution search tools, including PhiSpy (Akhter et al., 2012), VirSorter (Roux et al., 2015a) and ProphET (Reis-Cunha et al., 2019), in accordance with new and increased demands of phage research, such as working with metagenomic data. A final tool that has been developed very recently (January 2021) named PhageBoost uses a generalising machine learning method based on feature space to facilitate novel prophage discovery (Sirén et al., 2021). The approach calculates biological features for each gene from both nucleotide and amino acid sequences, and then uses machine learning to predict which genes belong to bacteria or phages. This is an exciting new development in prophage identification that could be utilised in future analyses (developed following completion of data analysis of this thesis).

3.2 AIMS

- Evaluate prophage identification tools and compare their features.
- Compare the best prophage identification tools to identify prophage regions from the well-defined Pa genome LESB58.
- Determine how accurate these phage predictions are between the chosen prophage identification tools compared to the published phages.

3.3 OBJECTIVES

- Compare the features of commonly used prophage identification tools to evaluate which would be the best and most suitable for our needs in each of our projects.

3.4 RESULTS

3.4.1 Evaluation of prophage identification tools

The prophage identification tools were evaluated for accuracy, assurance, ease of use and ability to input the data that was available for the genomes above. Therefore, a bacterial genome LESB58 with known prophages was analysed with multiple tools first to select the best tool.

The first stage of this approach was to look at the most cited and recently published prophage identification tools from the literature to date (October 2016). The programs chosen for evaluation were; Phage_Finder (Fouts, 2006), ProPhinder (2008), PHAST (Zhou et al., 2011) PhiSpy (Akhter et al., 2012), VirSorter (Roux et al., 2015a) and PHASTER (Arndt et al., 2016). Each tool was assessed based on the following criteria: (i) input data format required, as for parts of this study the genomes available were unannotated fasta files. Therefore, a tool was needed that could accept this format; (ii) how they match to known phage sequences e.g. which database was being used to assess if the tool had any limitations; (iii) the detection method used to identify the prophages, to assess which was the most accurate; (iv) the platform the programs were accessed on, to see how accessible each programme was (Windows/Unix/Linux).

One of the first criteria that needed to be met was that some of the bacterial genomes had inaccurate annotations. We therefore looked to use tools where the input data was in an unannotated format, such as an assembled genome, possibly in multiple contiguous sequences. Of these, PHASTER and VirSorter would accept this type of data. These were then compared whether one of the other programs offered greater function and subsequent bacterial genome annotation and curation, which would offer added value to the study. After comparing all tools, PHASTER and VirSorter were the preferred options to take forward (Table 3.1).

Table 3.1 Comparison of prophages identifications tools

Program (year published)	Input data required	Matching to known phage sequences	Prophage detection method	Database searched	Platform
Phage_Finder (2006) (Fouts, 2006)	Pre-annotated contigs accepted	HMMER (Eddy, 1998), BLASTP (Altschul et al., 1990)	Knowledge based rules/ metrics, gene function	NCBI	Unix-based OSs (e.g. Linux, MacOS) (download)
Prophinder (2008) (Lima-Mendez et al., 2008)	Pre-annotated	BLASTP	Statistical metrics	NCBI	Web service via Perl script run on Unix-based Oss
PHAST (2011) (Zhou et al., 2011)	Unannotated FASTA or pre-annotated	BLASTP	Knowledge based rules/ metrics, gene function	NCBI	Web-based
PhiSpy (2012) (Akhter et al., 2012)	Pre-annotated	None	Similarity agnostic statistical metrics, phage insertion point, gene function	ACLAME phage protein database	Unix-based OSs (e.g. Linux, MacOS) (download)
VirSorter (2015) (Roux et al., 2015a)	Unannotated FASTA contigs accepted	HMMER, BLASTP	Statistical metrics, gene function	Custom Refseq database	CyVerse discovery environment (Merchant et al., 2016)
PHASTER (2016) (Arndt et al., 2016)	Unannotated FASTA or pre-annotated contigs accepted	BLASTP	Knowledge-based rules/ metrics, gene function	NCBI	Web-based

3.4.2 Accuracy and assurance

PHASTER (Arndt et al., 2016) and VirSorter (Roux et al., 2015a) were tested for accuracy in identifying prophages within the well-curated genome annotation of Pa LES B58 (Winstanley et al., 2009), which has well-studied prophages and genomic islands. The

results of the comparison between PHASTER and VirSorter demonstrated that both VirSorter and PHASTER accurately identified the 6 prophage regions in LES B58 (Table 3.2). However, VirSorter falsely identified two of the genomic islands as prophages and PHASTER falsely identified one of these genomic islands as a prophage. PHASTER identified one of the genomic islands as an incomplete prophage and the other two genomic islands identified by VirSorter were category 3 (possible phage). VirSorter only classified one of the prophages (Prophage 5) as category 1 (most confident predictions) and the other 5 as category 2 (likely predictions), whereas PHASTER classified 4 out of 6 prophages as intact and classified the other 2 as questionable and incomplete, respectively. It was determined that PHASTER, regarding the identification of prophages and inducible prophages that can be mobilised, had the greater utility. Both tools also gave the coordinates of the predicted prophage region to allow comparison to the original data published by Winstanley et al (2009), where the co-ordinates for all prophages and genomic islands were added. In LES B58, four of the prophages are known to be inducible and therefore definite genome ends are known for LES phages 2-5. The accuracy of the tool to identify the ends of the prophages could be seen best by comparing against these phages. PHASTER predicted the LES phage 2 genome to be 9.5kb larger compared to data produced by Winstanley et al. (2009), whereas VirSorter only predicted it to be 820bp larger. However, overall, PHASTER predicted two out of four phages accurately enough that no extra genes were included in the genome, whereas VirSorter only predicted one to have no change in gene number. Therefore it was realised that there may be some need for manual curation downstream if the phages could be sequenced and genomes and ends determined.

Table 3.2 Comparison of PHASTER and VirSorter of indentifying LES phages from LESB58

Feature in LES B58	Coordinates of LES phages from Winstanley 2009	VirSorter results			PHASTER results		
		VirSorter result	Comparison with prophage region in paper		PHASTER result	Comparison with prophage region in paper	
Prophage1	665561-680385	Prophage -2	1328 bp larger	1 more gene	Questionable	3646 bp larger	2 more genes
Prophage 2	863875–906018	Prophage -2	820 bp larger	1 more gene	Intact	9499 bp larger	8 more genes
Prophage 3	1433756–1476547	Prophage -2	1679 bp larger	3 more genes	Intact	1 bp larger	No change
Prophage 4	1684045–1720850	Prophage -2	1835 bp larger	1 more gene	Intact	107 bp larger	No change
Genomic Island 1	2504700–2551100	Prophage -3	149622 bp larger	88 more genes	Incomplete	34299 bp shorter	18 fewer genes
Prophage 5	2690450–2740350	Prophage -1	14 bp larger	No change	Intact	2735bp larger	1 more genes
Genomic Island 2	2751800–2783500	None	-	-	None	-	-
Genomic Island 3	2796836–2907406	None	-	-	None	-	-
Genomic Island 4	3392800–3432228	None	-	-	None	-	-
Prophage 6	4545190–4552788	Prophage -2	654 bp larger	No change	Incomplete	10524bp larger	9 more genes
Genomic Island 5	4931528–4960941	Prophage -3	9589 bp larger	8 more genes	None	-	-

3.5 DISCUSSION

There are very few bioinformatics programs developed specifically for the analysis of temperate bacteriophages. Hence, those that are available for phage or bacterial analysis were tested to see if they were compatible with our study. The first aim was to evaluate which phage identification tool was the most suitable for our needs, namely, the rapid identification of phages with a user-friendly GUI and the ability to work online with low computing power. The tool also needed to have the ability to be used with command line and to have offline capabilities to be used on a server with a higher computing power when searching for large numbers of bacterial genomes. Six commonly used phage identification tools were compared to look at a number of important factors, which included the input data required for the tool, the prophage detection method and the method used to match to known phage sequences (Table 3.1). The majority of the projects in this PhD required the bacterial genomes to be sequenced from clinical Pa samples that were not pre-annotated. Therefore, the tool must be able to work directly from the unannotated genome fasta files. This narrowed the tools available for use to two: PHASTER (Arndt et al., 2016) and VirSorter (Roux et al., 2015a). They were the most up to date tools, being released in 2016 and 2015, respectively, and also both tools use the Blast database to match to known phages, which is constantly being updated. The second part of this optimisation was to establish which tool had the best accuracy rate at predicting prophages in a bacterial genome. PHASTER and VirSorter were tested against each other using the well annotated Pa genome LESB58, which has six well defined prophages, four of which have been shown to be viable, and five genomic islands (Winstanley et al. 2009). As Pa from patients with chronically infected lungs were the main source of samples analysed in subsequent studies, LESB58 was a suitable strain to test the tools in this chapter, as it was first isolated from a chronically infected CF patient lung. Comparing the output of these two tools to the prophages that are experimentally defined and validated previously by Winstanley et al. (2009), allowed an assessment of how accurately these tools predict the prophages, by what classification the phage is predicted to be (i.e. intact phage or incomplete) and how close the insertion sites are to the previously published data. As there are also known genomic islands in this

genome, these tools were assessed to ascertain whether they could distinguish between a prophage and a genomic island and that would constitute a false positive result. The results suggest that PHASTER was the more accurate prophage identification tool as it predicted the four LES phages that are known to be inducible (LES phages 2-5) as intact phages and LES phages 1 and 6 as questionable and incomplete, respectively. PHASTER also predicted genomic island 1 as an incomplete phage, which was a false positive. However, VirSorter predicted two genomic islands as phages and also only predicted prophage 5 out of the LES phages as category 1 (most confident predictions) and the others as category 2 (likely predictions). This means PHASTER is able to predict inducible phages as intact phages more accurately than VirSorter. When it comes to the accuracy of where the prophages are predicted in the genome, both tools did not predict any of the prophage insertion sites down to the base-pair. PHASTER was only 1bp away on one of the phages but both tools had similar levels of inaccuracy in predicting the ends of the phages. Therefore, overall, PHASTER was found to be the best phage prediction tool in 2016 for our needs. However this is not an extensive validation of both tools and only suits the needs of these projects at this time.

3.6 SUMMARY

By comparing six phage prophage identification tools, VirSorter and PHASTER were chosen as the most suitable for our data and our needs. PHASTER and VirSorter were then used to identify prophages with LESB58 to assess ease of use and output. Using the LESB58 genome, which contains 6 well defined prophages, accuracy was tested in VirSorter and PHASTER, with PHASTER proving to be the most accurate when identifying the known prophages in LESB58, both in completeness of genome and accuracy of the start and ends of the phage genome.

4. THE CARRIAGE, DIVERSITY AND TRANSFER OF PROPHAGES IDENTIFIED BETWEEN *PSEUDOMONAS AERUGINOSA* STRAINS WITHIN THE CHRONICALLY INFECTED LUNG

4.1 INTRODUCTION

In the previous chapter assessed two; informatics tools VirSorter and PHASTER were used to determine prophage regions within bacterial genomes and compared for function and utility using the well-studied LESB58 as a control. The outcome determined that both PHASTER and VirSorter could be used as tools, where the former offered utility in providing some level of prediction to inducible and remnant prophage regions with more confidence when comparing the inducible prophages from LESB58. This chapter uses PHASTER to characterise temperate phages that are carried as lysogens in bacterial populations in a greater diversity of bacterial populations. It was previously thought that *Pseudomonas aeruginosa* (Pa) infections present in the chronically infected lung had clonal populations, however, it is now known there can be more than one strain present at one time (Tingpej et al., 2007). With the presence of multiple sub-species of Pa in the chronic lung there is an additional compartment of genetic diversity that offers positive evolutionary selection potential, some of which may be carried by phages that could be shared between populations. For this reason, this carriage of phages between populations needs to be studied to identify the level of phage carriage, to try and determine the potential that horizontal gene transfer has between bacteria in the lung. To date, no study has focussed on the carriage of phages within lung populations of Pa, with no scientific rationale or hypothesis for the impact they may have. Having the ability to work between sub-populations identified from colony variants isolated from the lung of the same patient at the same time-point, may illustrate phages that are more frequently found or can infect between sub-populations within the lung environment. This chapter represents a retrospective study, looking directly at the genomic differences between Pa variants/strains isolated in the chronically infected lung of a patient that show co-colonisation in the lung of BR patients. PHASTER was used to identify prophages within Pa colony variants that were isolated from the same BR patient sputum sample. The prophage genomes carried were then be compared to the co-colonising strain(s). It has been previously reported that temperate

phages in *Pa* drive adaptive evolution (Davies et al., 2016b), therefore analysing and comparing temperate phages from chronically infected patients may give an insight into the role of temperate phages in chronic *Pa* lung infections. It may also determine whether these phages, that drive evolution, are inducible and able to transfer within the lung environment.

4.1.1 Non-clonal bacterial populations within chronically infected lungs

It is important to note that up until 20 years ago it was believed that the majority of CF or BR patients were colonised by clonal *Pa* strains independently (Jelsbak et al., 2007, O'Carroll et al., 2004). However, studies have shown phenotypic differences of proposed clonal variants that have been frequently observed, including increased antibiotic resistance, (Ratjen, 2009), altered metabolism and attenuated virulence (Hoiby et al., 2010b).

The focus of these studies is usually on the core genome of the *Pa*, with the accessory genome, including prophages and other transposable genetic elements, usually overlooked. This could be due in part to the high number of genes in bacteriophages, and horizontal gene transfer (HGT), that are small in size with no associated function at gene or protein level. This could be an important area to study, as there is limited distinction in the variation when comparing the core genomes (Winstanley et al., 2016). Therefore, the study of the accessory genome of these bacteria may help to uncover their biology, function and adaptive evolution in the lung. Co-infection with two different strains of *Pa* can also occur, rather than just genetic diversification within the lung (Aaron et al., 2004). When there are multiple strains in an environment, HGT can take place between the different strains, which may drive evolution within an environment such as the lung, by providing genes or plasmids that encode a selective advantage.

4.1.2 *P. aeruginosa* genomes from BR patients

The panel of genomes analysed in this study were chosen from a much larger library of *Pa* isolates from a national and international study by Hilliam et al. (2017), with access provided as part of a collaboration with the University of Liverpool. The chosen data consisted of genomes from 13 co-colonising *Pa* isolates isolated from six of the BR patients. The full study

by Hilliam et al (2017) contained a panel of 189 Pa isolates obtained from the sputa of 91 BR patients (Hilliam et al., 2017). The chosen isolates showed multiple morphological variants when cultured from a single time point, and from genomic analysis had differing MLST (Multi-Locus Sequence Typing) types, implying co-colonisation by different strains (Hilliam et al., 2017). From the six patients, five showed two colony morphologies and one showed three (Table 2.1). The morphological variation seen between isolates could not be explained by the differences in the core genome of the bacteria (Hilliam et al., 2017). These findings support the hypothesis that these differences may arise from the accessory genome, specifically the temperate phages studding the bacterial chromosome, which may link to the alternate phenotypes.

4.1.3 Carriage of prophages by *P. aeruginosa*

Prophages are a frequent feature of Pa sequenced genomes (Silby et al., 2011) and can contribute to bacterial fitness and virulence by the carriage of additional genes or modification of bacterial genes (Davies et al., 2016b). Studies have shown that the carriage of prophages can promote the lysis of competitors and also support isogenic-phage superinfection (McCallum et al., 2001). As the carriage of prophages by Pa can confer a fitness as a selective advantage, this could explain the high numbers of prophages and why they are observed in Pa strains, particularly in clinical strains such as LES (Liverpool epidemic strain), which show increased virulence when harbouring the LES prophages (Winstanley et al., 2009, Davies et al., 2016a). Phage conversion has also been associated with the evolution of fundamental clinical phenotypes in Pa, such as the mucoid phenotype (Miller and Rubero, 1984), but whether this is linked to lysogeny or driven by MOI is unclear. Busby et al. (2013) suggested that isolates that have a higher proportion of phage-related genes in their genomes are shown to be more pathogenic in *E.coli* (Busby et al., 2013). Many highly pathogenic bacteria other than Pa also carry multiple prophages e.g. *E.coli* O157:H7 (Hayashi et al., 2001b, Winstanley et al., 2009). This may suggest that prophages play a role in bacterial pathogenicity through carriage of function in the cargo genes encoded by these genetic elements. Therefore, it is likely that the carriage of prophages both support selection of Pa in the chronically infected lung and contribute to the adaptation of Pa in that

environment.

4.1.4 Temperate phage insertion as a source of genetic variation for evolution

As part of infection and conversion, temperate phages integrate into the bacterial genomes, therefore many have evolved ways to minimise the disturbance to the bacterial genome or its function that could be caused by integration. One way of minimising disturbance is by targeting highly conserved insertion sites in regions that do not disrupt essential host function, such as near to or within tRNA genes (Bobay et al., 2013). Temperate phages can even carry gene regions into insertion sites, permitting the maintenance of the coding sequence (Campbell, 1992). However, other transposable phages, such as Mu-like phages of *Pa*, insert into the bacterial genome at seemingly random sites (Toussaint and Rice, 2017), which is an approach they use to replicate before induction and lysis. Transposable phages are capable of copying and integrating into other sites in the genome, thereby causing large scale structural changes in the bacterial genome, as well as high numbers of gene mutations and inhibition of function due to transposition into functional genes in a manner analogous to transposon mutagenesis (Toussaint and Rice, 2017, Rehmat and Shapiro, 1983). This increase in genetic variation may be beneficial to the host during adaptation to new environments, due to the increased likelihood of a mutation accelerating evolution (Chao and Cox, 1983). In a recent study looking at evolution in populations of *Pa*, the bacteria adapted faster to a biofilm environment when co-cultured in ASM with temperate phages containing the transposable phage LES ϕ 4 (Davies et al., 2016b). This suggests that mutations caused by phages leads to increased rates of adaption to evolutionary pressures, indicating that phages play a significant role in bacterial adaptation (Davies et al., 2016b). This relates very closely to our study as evolution experiments by Davies et al. (2017) were modelled on *Pa* in the CF lung environment. They used ASM (artificial sputum media) mimicking lung sputum, which is very similar to the chronically infected BR lung, where *Pa* must rapidly adapt to the lung environment. This shows the significance of temperate phages, especially transposable phages, and the importance of genome insertion in chronic infection, where extensive genetic diversification and evolution can occur over the course of a chronic infection. This leads to two, possible non-exclusive hypotheses; (i)

do the bacteria receive function from the temperate phage (ii) do they mutate by phage transposition?

4.1.5 Spread of of temperate phages between Pa in a co-infected lung environment

Temperate phages have the ability to induce from their chromosomal integration in the bacterial genome via the lytic life cycle. On lysis of the bacterial cell, these phages are released into their direct environment and may infect nearby cells. It has been shown that free phages are present in the sputum of CF patients. These phages were thought to be lytic after testing their infectivity, as 40% showed a positive plaque assay when extracted directly from CF sputum (Fothergill et al., 2011). However, they are just as likely to be temperate phages that have been induced from bacteria the in the lung, such as Pa (James et al., 2015). In the environment of the chronically infected lung, where antibiotics are administered into the lungs, the potential of inducing temperate phages is higher due to the ability of some antibiotics to act as chemical inducing agents (Fothergill et al., 2011). Once these temperate phages are free in the lung they may have the ability to infect a new host, which may be a different strain if there is a co-infection. Phages have been linked to promoting horizontal gene transfer in this way (Blahova et al., 1994) between strains of the same species, but also to closely related species, and have been associated with virulence in other bacterial hosts (Brussow et al., 2004). This highlights the need for research into the prevalence of transfer of temperate phages in Pa in the lung environment and the effects it may have. This study looks at patients co-infected with multiple strains of Pa by analysing the sequenced bacterial genomes, attempting to reveal the carriage, diversity and transfer of temperate phages that occurs between the Pa isolates from the same lung environment.

4.2 AIMS

The aims of this study were to determine the differences and similarities between prophages carried in colony variants of Pa isolated from the same patient's lungs at the same time-point. Other aims were to ascertain if phages transferred across strains in the lung environment and top determine whether phage carriage is linked to the colony variant phenotype.

4.3 OBJECTIVES

The objectives of this research were to compare the putative prophages in these co-colonising isolates. Objectives included determining the following -

- Ascertain the number of prophage regions per bacterial genome.
- Identify the attB site of prophage integration assess whether this would impact on cellular function and lead to phenotypic alterations.
- Determine the levels of heterogeneity between phage regions identified.
- Determine whether phages are shared between isolates taken from the same patient.
- Compare prophage regions from all against all analysis of identified prophage region.

4.4 RESULTS

4.4.1 Identification of prophages within Pa genomes

In chapter 3, a number of phage identification tools were compared with PHASTER being chosen as the most fit for *purpose*, as it was able to identify the LES phages previously studied by the Winstanley group in Liverpool. Here, PHASTER was used to identify the prophage regions in 13 Pa genomes isolated from six BR patients. Each patient carried at least two colony morphologies when isolated from their sputa. A total of 72 phages were identified within the 13 bacterial genomes. These were either classed as intact (27), questionable (16) or incomplete (remnant) (29) phage genomes. (Figure 4.3.1). There was at least one prophage region identified in 12 out of 13 isolates. The most prophage regions identified in an isolate was 10, in isolates C125 and C86 from two unrelated patients. There were only three that did not contain any intact phages A100, C109 and C13. The prophages classed as intact are the most likely to be inducible, as they contain the structural and replication genes that enable a phage to proliferate and lyse from the cell. Without experimentation it is impossible to tell which phages are truly viable and infective. However, prophages classed as intact provides a strong indication of whether a prophage is putatively mobilisable and therefore able to transfer to another Pa genome.

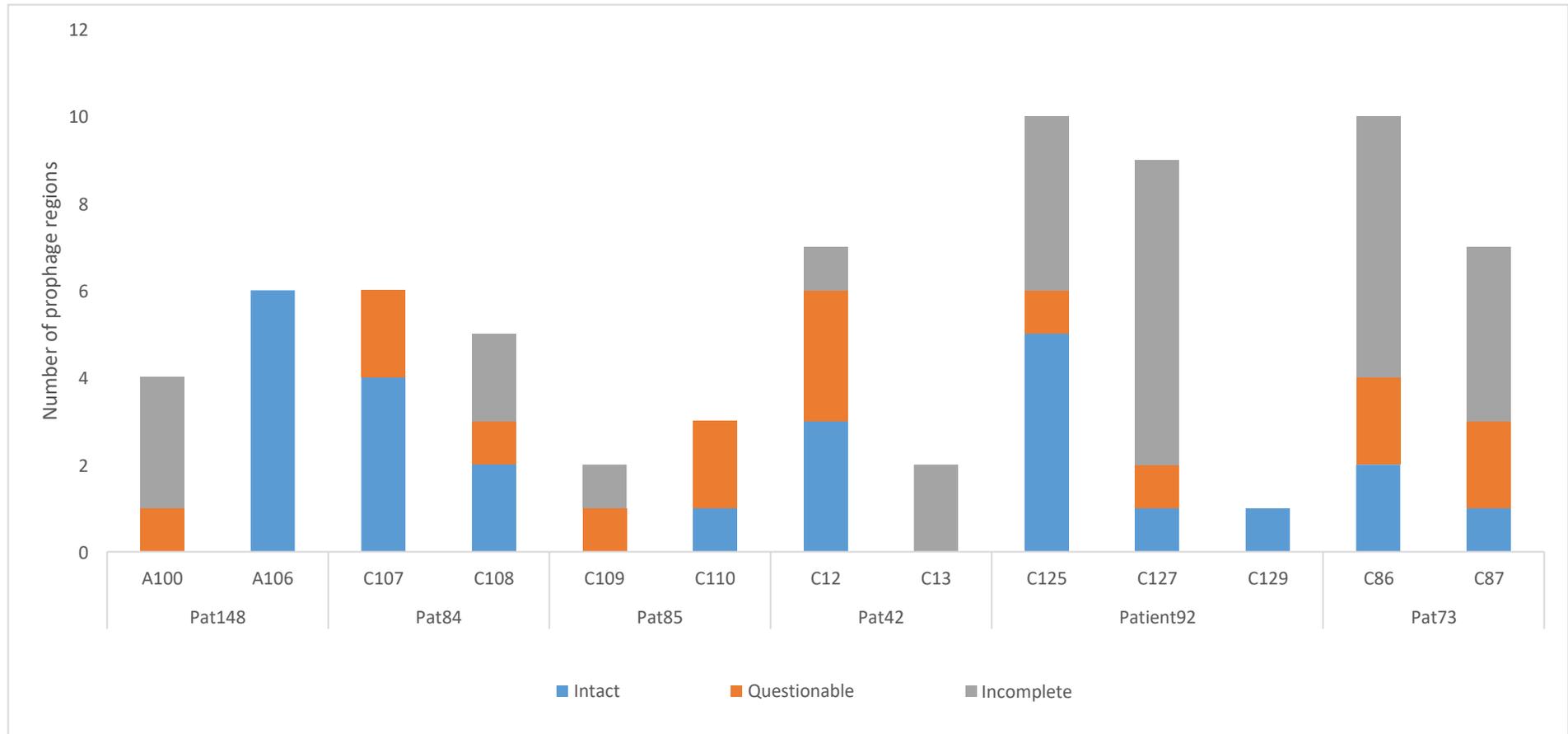


Figure 4.1. The number of prophages predicted using PHASTER in 13 Pa genomes. Numbers of intact (blue), questionable (orange) and incomplete (grey) prophages predicted in each Pa isolate genome from the same BR patient.

4.4.2 Integration points of prophage

As PHASTER predicts the attB location of the prophages within the bacterial genome, these are also the integration sites and can be used to determine bacterial chromosomal location. As shown in section 3.3.1.1, which analysed the accuracy of PHASTER prediction of the prophage, it is not entirely accurate at predicting where the start and end of the phage genomes are within the bacterial genome. However, it is the best way to predict the integration site without experimentation, by sequencing the bacterial and the phage genome induced from it and see where it fits within the bacterial sequence. If it were of further interest, it would also be possible to pinpoint the attP-attB junction via manual inspection of sequence. All 72 phages were analysed using bioinformatics approaches to ascertain whether the phage was inserted into a bacterial gene and whether it may have an impact on cellular function, by manually observing if the prophage boundaries map across the bacterial gene. The results showed that 30% of the predicted prophages in the cohort integrated into a known gene, with 7 being intact phage, 6 being questionable and 9 being incomplete. Only one isolate (C86) had phages that did not have a phage that integrated into a bacterial gene. In 6 cases, phages integrated into Indole-3-glycerol phosphate synthase, which was the only case where multiple phages integrated into the same gene, other than 'protein of unknown function/ hypothetical protein'. Other genes that were disrupted by the integration of phages were involved in secretion systems and extracellular matrixes, such as 'Carboxyl-terminal protease', 'Type I secretion system, outer membrane component LapE' and 'Extra cellular matrix protein PsIE'. These are all important in pathogenicity, being associated with biofilm formation. Interestingly, one phage in A100_1, classed as questionable by PHASTER, integrated into a transposase. This may mediate or catalyse the integration, or may limit further phage integration and act as restriction method. Table 4.1 shows the results taken from analysing each of the insertion points that were predicted by PHASTER.

Numerous temperate phages integrate near to or into tRNA genes. However, only one phage appeared to integrate downstream of bacterial tRNA genes. 12 out of 13 isolates were predicted to have downstream impact on gene function. Collectively, these data suggest that even from this small cohort number, many prophages likely impact on the cellular

function of the bacterial host by integration alone, even before integrated prophage genes are taken into consideration.

Table 4.1 The integration positions of prophages identified from Pa genomes

<i>Patient</i>	<i>Isolate_phage</i>	<i>Phage class predicted</i>	Closest phage relative	Intragenic insertion	<i>Integration location</i>
148	A100_1	Questionable	F10	Y	transposase
	A100_2	Incomplete	YMC11	N	
	A100_3	Incomplete	Strep_Jay21	N	
	A100_4	Incomplete	Pf1	N	
	A106_1	intact	YMC11	N	
	A106_2	intact	phiCTX	Y	Indole-3-glycerol phosphate synthase
	A106_3	intact	F10	N	
	A106_4	intact	D3	N	
	A106_5	intact	Mu	Y	Extra cellular matrix protein PsIE
	A106_6	intact	phiCTX	N	
42	C12_1	intact	YMC11	N	
	C12_2	questionable	F10	Y	Cysteine desulfurase CsdA-CsdE sulfur acceptor protein CsdE
	C12_3	intact	ECO1230	N	
	C12_4	intact	PM105	Y	Carboxyl-terminal protease
	C12_5	questionable	Pf1	Y	Protein of unknown function
	C12_6	incomplete	YMC11	N	
	C12_7	questionable	Pf1	N	
	C13_1	incomplete	Phi297	Y	Protein of unknown function
	C13_2	incomplete	YMC11	Y	Indole-3-glycerol phosphate synthase
84	C107_1	intact	F10	Y	Formiminoglutamase
	C107_2	intact	PM105	N	
	C107_3	intact	YMC11	N	
	C107_4	questionable	Pf1	N	
	C107_5	intact	JBD88a	N	

	C107_6	questionable	D3	Y	tRNA- (ms[2]io[6]A -hydroxylase
	C108_1	incomplete	PMG1	N	
	C108_2	intact	F10	N	
	C108_3	incomplete	Phi297	N	
	C108_4	questionable	Pf1	N	
	C108_5	intact	phiCTX	N	
85	C109_1	questionable	Pf1	N	
	C109_2	incomplete	YMC11	Y	Indole-3-glycerol phosphate synthase
	C110_1	intact	Ep3	Y	Indole-3-glycerol phosphate synthase
	C110_2	questionable	F10	Y	Transcriptional activator of acetoin dyhydrogenase operon AcoR
	C110_3	questionable	Pf1	N	
92	C125_1	intact	PM105	N	
	C125_2	incomplete	F10	N	
	C125_3	incomplete	F10	N	
	C125_4	incomplete	H70	Y	Protein of unknown function
	C125_5	incomplete	Pf1	Y	Chemotactic transducer
	C125_6	intact	YMC11	N	
	C125_7	questionable	phiO19P	N	
	C125_8	intact	PM105	N	
	C125_9	intact	YMC11	N	
	C125_10	intact	B3	Y	mobile element protein
	C127_1	intact	YMC11	Y	Indole-3-glycerol phosphate synthase
	C127_2	incomplete	F10	N	
	C127_3	incomplete	F10	N	
	C127_4	questionable	Tr60	Y	Protein of unknown function
	C127_5	Incomplete	PAJU2	Y	Type I secretion system, outer membrane component LapE
	C127_6	incomplete	Pf1	N	
	C127_7	incomplete	Pf1	N	
	C127_8	incomplete	F10	Y	Protein of unknown function
	C127_9	incomplete	Ab30	N	

	C129_1	intact	PPpW	Y	Indole-3-glycerol phosphate synthase
73	C86_1	questionable	F10	N	
	C86_2	incomplete	F10	N	
	C86_3	incomplete	F10	N	
	C86_4	incomplete	Pf1	N	
	C86_5	incomplete	Phi297	N	
	C86_6	incomplete	Pf1	N	
	C86_7	intact	Phi297	N	
	C86_8	questionable	Ab30	N	
	C86_9	intact	YMC11	N	
	C86_10	incomplete	F10	N	
	C87_1	incomplete	F10	N	
	C87_2	incomplete	Ectoca siliculosus virus	Y	hypothetical protein
	C87_3	incomplete	Phi297	Y	Molybdenum cofactor biosynthesis protein MoaA
	C87_4	questionable	Phi297	N	
	C87_5	questionable	Pf1	N	
	C87_6	incomplete	phiNITI	N	
C87_7	intact	Ep3	N		

4.4.3 Comparison of phage carriage across each isolate

4.4.3.1 Comparison of identified prophages using MAUVE

The phages of each Pa isolate were concatenated into multi-fasta files containing all phage genomes from each Pa isolate. Initially, MAUVE was used to visualise the phage comparisons, which uses ClustalW to compare these phage regions at the nucleotide level, with the concatenated phages from one isolate being the reference and the other isolate's phages from the same patient being the query. Therefore, as there were 6 patients, 6 comparisons were completed in a pairwise manner.

The results are displayed in figures 4.2a - 4.2f. The comparison between phages carried between morphotypes in patient 42 are shown in figure 4.2a. It shows that isolate C12 has 8 phages and C13 has only 2 phages. Only one phage appears to have high similarity, phage number 7 of C12 (C12_7) and phage number 2 of C13 (C13_2). They are characterised differently by PHASTER, with phage C12_7 characterised as questionable and C13_2 as incomplete. This indicates that there are features missing from C13_2 that are present in C12_7, even though their genomes are highly similar. Identification of similar phages in each morphotype may mean that these are the same bacterium that, over time, have evolved separately in the lung. It may also mean that this phage has transferred from one strain to the other, which would be the first report of this in the lung environment. This shows the greater need to determine carriage of phages across a wider cohort of Pa to give a better understanding of which phages are commonly carried and which may be advantageous to the Pa host. The comparison between patient 73's Pa morphotypes, shown in 4.3.3b, shows that isolate C86 and C87 had 6 phages with some homologous regions within their genomes, most with only part of another phage, which suggests expansion or attenuation of the phage genome or it could also indicate a prior recombination event. The results for patients 84 and 148, shown in 4.2c and 4.2f, respectively, show similar results to patient 73. These results are useful to identify and target specific phage genomes that are similar. The results for patient 85, shown in 4.2d, shows that isolates C109 and C110 have 2 phages that have distinct similarity at their boundaries. However, the central region of the

phage genome is very different, suggesting evolution or possibly the same type of phage but with rearrangement. Patient 92's phage genome comparison is shown in 4.3.3e and is slightly different from the others, as there were 3 isolates collected and therefore 3 phage complements to compare. Isolate C125 (on the top), C127 (in the middle) and C129 (on the bottom) only contained one phage. The one phage that was found in C129 was shown to be highly similar to an intact phage in both C125 and C127. C129's phage genome was 14kb smaller, possibly showing some expansion of attenuation that was mentioned previously.

MAUVE is an effective tool in comparing two sequences. However, when comparing multiple sequences at the same time the visualisations are less useful as pairwise. Being able to compare each phage across the whole panel of phages from all patients is important as it allows the study of carriage. Therefore, other visualisation methods were necessary to show in more detail, which parts of the phage genomes were similar to each other and which were different.

Figure 4.2a

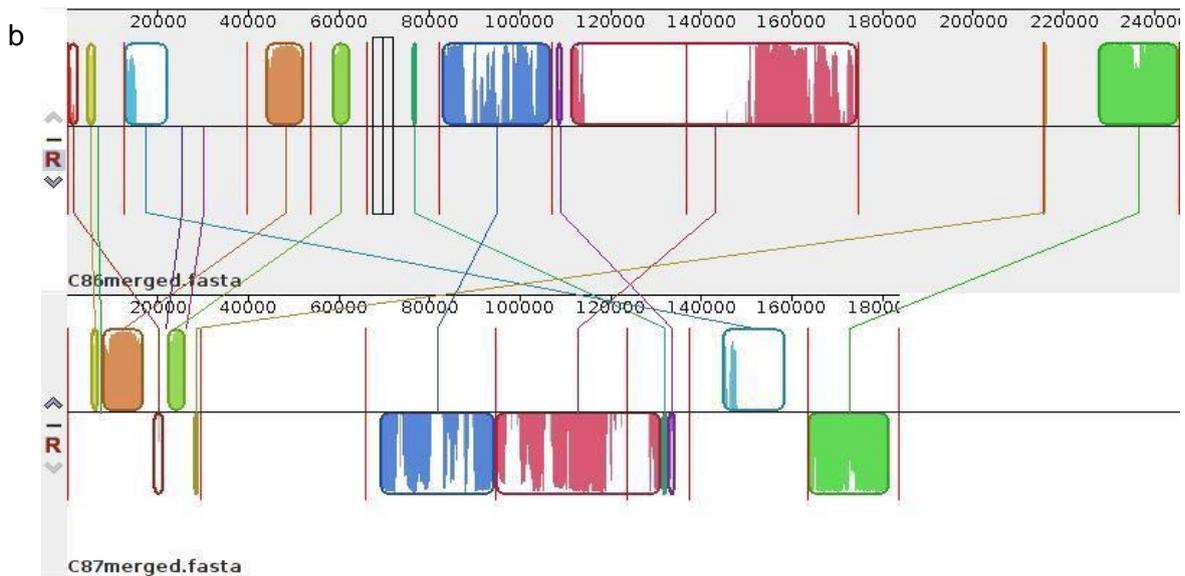
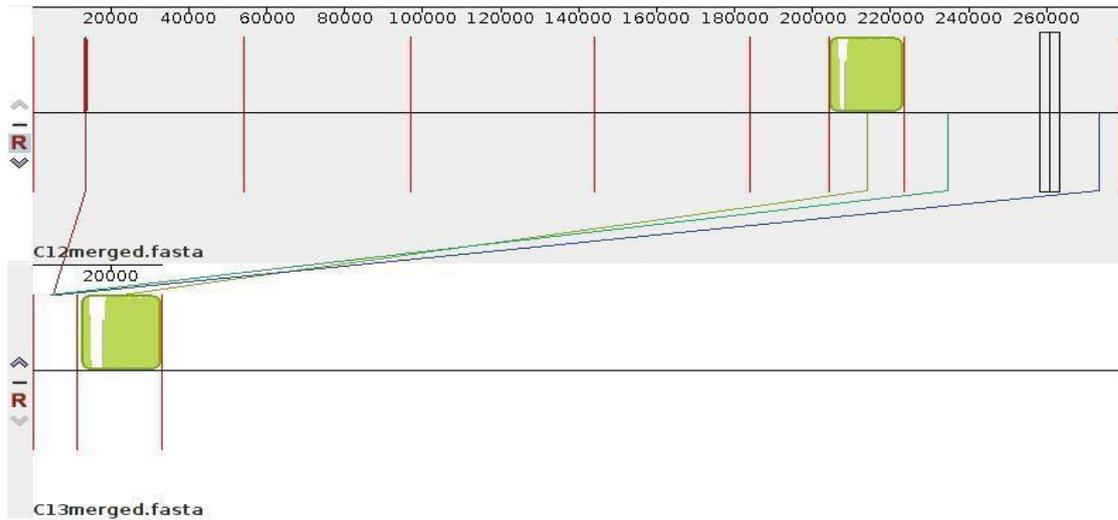


Figure 4.2a-b. MAUVE comparisons of prophage regions. Prophages from Pa isolates with differing phenotypes, taken at the same time point from the same patient, comparing between concatenated phages from isolates from the same patient. The vertical red lines show where each phage genome starts and ends. The block above or below the horizontal line dictates which strand the gene is on. The filled in colour shows regions of nucleotide similarity. The more colour and less white that is shown in a region suggests a higher percentage similarity in that region.

a) Patient 42, comparing the phage complements of isolates C12 and C13.

b) Patient 73, comparing the phage complements of isolates C86 and C87.

Figure 4.2c

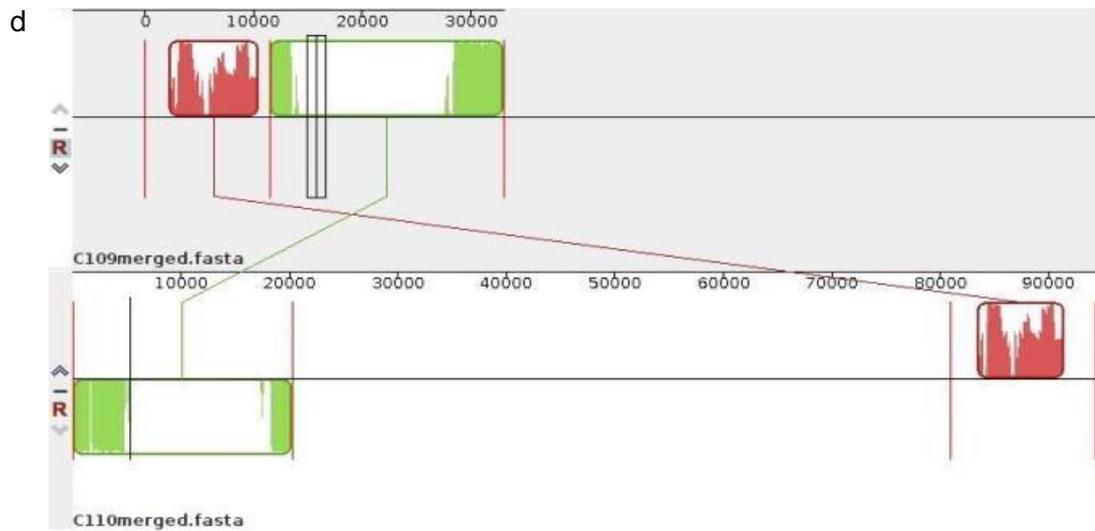
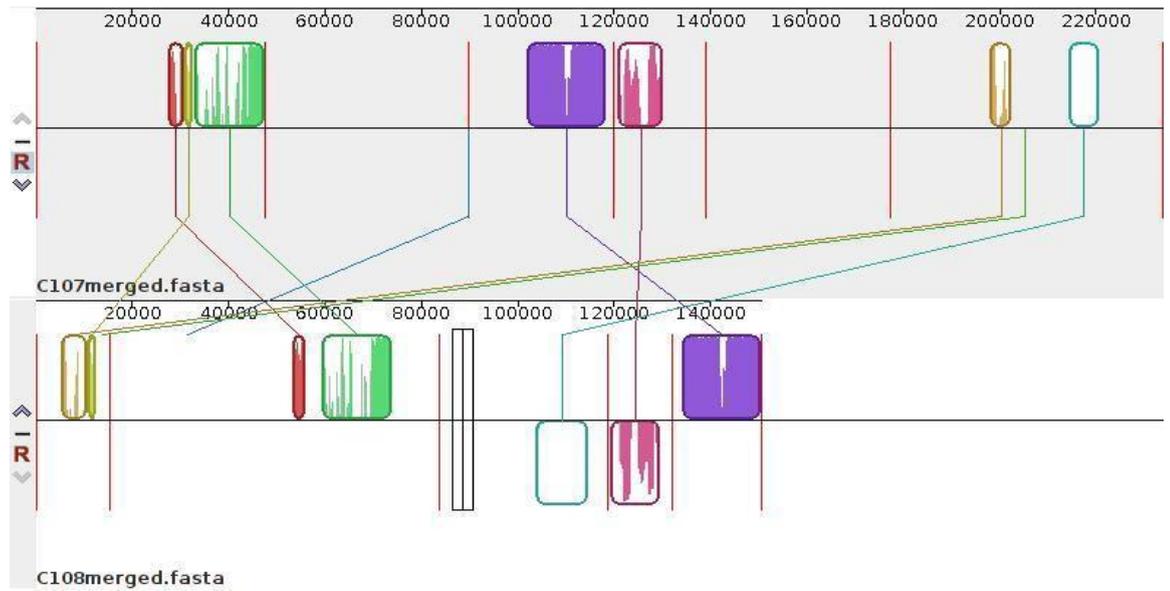


Figure 4.2c-d. MAUVE comparisons of prophage regions. Prophages from Pa isolates with differing phenotypes taken at the same time point from the same patient comparing between concatenated phages from isolates from the same patient. The red lines show where each phage genome starts and ends. The block above or below the horizontal line dictates which strand the gene is on. The filled in colour shows regions of nucleotide similarity. The more colour and less white that is shown in a region suggests a higher percentage similarity in that region.

c) Patient 84, comparing the phage complements of isolates C107 and C108.

d) Patient 85, comparing the phage complements of isolates C109 and C110.

Figure 4.2e

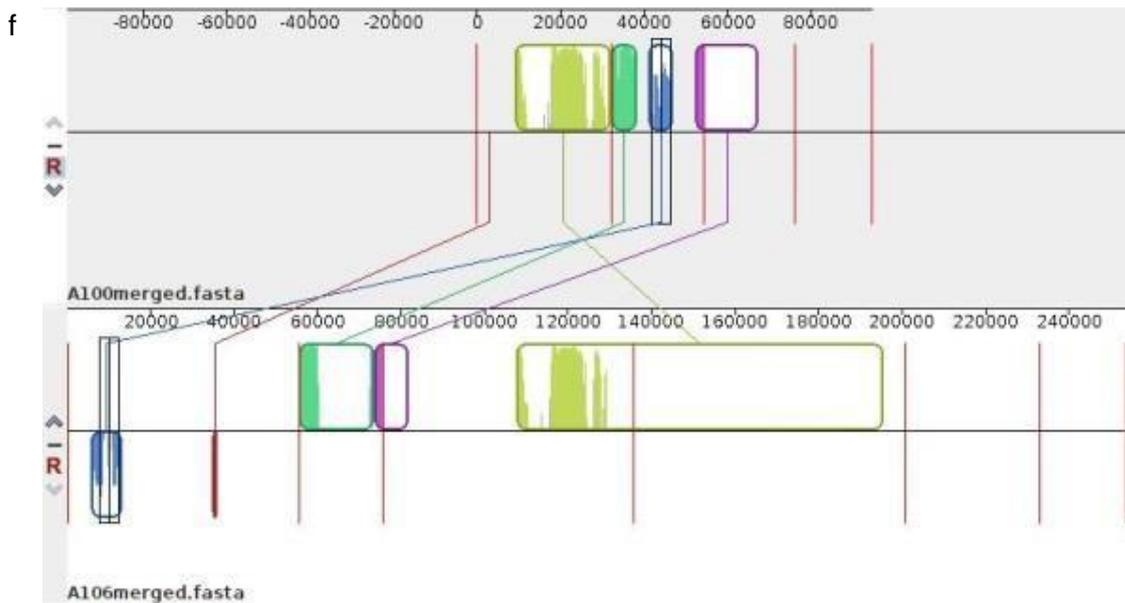
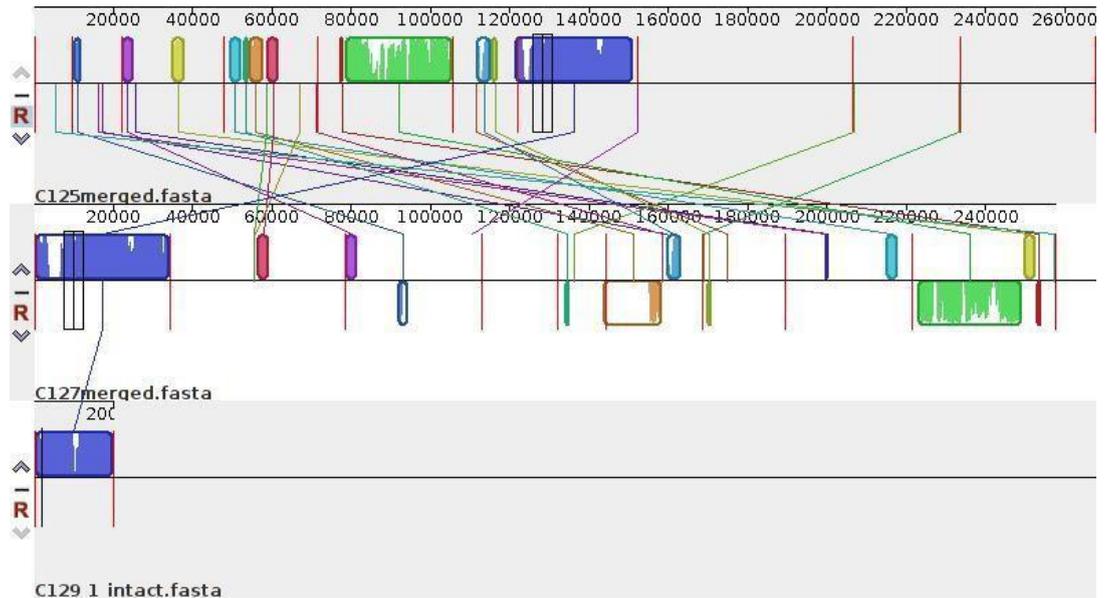


Figure 4.2e-f. MAUVE comparisons of prophage regions. Prophages from Pa isolates with differing phenotypes taken at the same time point from the same patient, comparing between concatenated phages from isolates from the same patient. The red lines show where each phage genome starts and ends. The block above or below the horizontal line dictates which strand the gene is on. The filled in colour shows regions of nucleotide similarity. The more colour and less white that is shown in a region suggests a higher percentage similarity in that region.

e) Patient 92, comparing the phage complements of isolates C125, C127 and C129, and has three rows.

f) Patient 148, comparing the phage complements of isolates A100 and A106.

4.4.3.2 Comparison of intact phages using MUMmer and Circos

To improve visualisation of pairwise genome comparison, other methods were needed. An approach used here had been previously used in other phage genome comparison studies (Smith et al. 2012). This pairwise comparison approach used MUMmer to compare the genomes against one another and the results visualised with Circos. The NUCmer command in MUMmer was used to generate nucleotide alignments using a compare to reference approach. The parameters were set at the default of a comparison of a minimum of 20 basepairs and 95% identity. Circos allows visualisation when comparing one region to multiple tests based on the pairwise alignment generated in NUCmer, allowing the comparisons to be layered and seen in the same image. This enables the visualisation of the comparison of all phages from one Pa genome against all the phages in the other, from each patient.

Figure 4.3 shows an example of this analysis, illustrating phages that were similar or had comparable regions at the nucleotide level. This is important in phages as they are modular in genome construction and thought of as mosaic in nature. Figures 4.4-4.9 shows Circos comparisons of the phage genomes from each of the Pa isolates for each patient, with a figure shown per patient. Figures are divided into sections A, B and C: (A) is colour coded in red and blue, with (A) representing the reference phage genomes from one isolate and blue representing the query phage genomes from the second isolate, from the same patient. Where three isolates are present, in figure 4.9, the third isolate's phage genome is coloured green. Images display the similarities between carriage of all phages between patient morphotypes. (B) shows the same comparison as (A) but the colours were altered so that the colours represent if the phage genomes were classed as intact (red), questionable (orange), or incomplete (yellow). This offers a better way of determining if the phage was mobilisable and therefore transferred between isolates. Then (C) shows all the phage genomes from that patient versus all the other phages to see difference in carriage between other patients and to see how similar the phage regions were in the cohort of 72 phages. This may begin to illustrate phage regions that are important in the chronic lung if found in

multiple patient isolates.

From all the patient results in figures 4.4-4.9, all isolates showed some similarity between one or more of the prophage genomes. In some cases, such as patients 42 and 85 (figures 4.5 and 4.7), these regions of similarity were small, suggesting that there are only a few genes that are similar. However, in other cases such as patients 73, 84 and 92 (figures 4.4, 4.6 and 4.9), there is extensive similarity between the other phages found in the comparator isolate from the same patient. This is shown in the thickness of the ribbons and number of ribbons between the phage genomes in the Circos images. These phages appear to be highly similar, however, they have differing genome lengths, which means the percentage of coverage that is homologous could be low. The phage genome lengths, percentage query coverage and percentage identity between each phage and its most similar phage from the comparator isolate from the same patient, are shown in table 4.2. There are regions for possible expansion or reduction to the prophage genomes, as a full genome can be homogeneous to part of the compared prophage genome, seen in figure 4.4 for patient 73, which compares phage C86_9 and C87_7. The prophage C86_9 has a much larger genome than C87_7, showing possible expansion or incorrect estimation of the insertion site on one of the phages. Table 4.2 shows genome lengths of C86_9 is 30.2kb and C87_7 is 20.2Kb in length, showing an expansion or restriction of 10Kb. The results in table 4.2 also show that 88% of C87_7's genome was homologous with C86_6, whereas only 58% of the C86_9 genome is homologous with C87_7. These similarities between the prophage genomes could illustrate conservation of a phage in the lung environment, especially for the phages that have a high percentage coverage and identity and are classified as intact phages. Phages such as C125_6 and C127_1, which have 85% coverage compared to the larger genome, suggest possible phage transfer between isolates, together with both being classified as intact and therefore putatively mobilisable. There were regions in the incomplete/remnant prophages, shown in yellow in B figures 4.4-4.9, that showed similarity between other incomplete phage genomes from different strains of Pa. This suggests that there are conserved regions of phages that are or were once mobilisable and transferred between each other, such as between C86_5 and C87_3 in figure 4.4b.

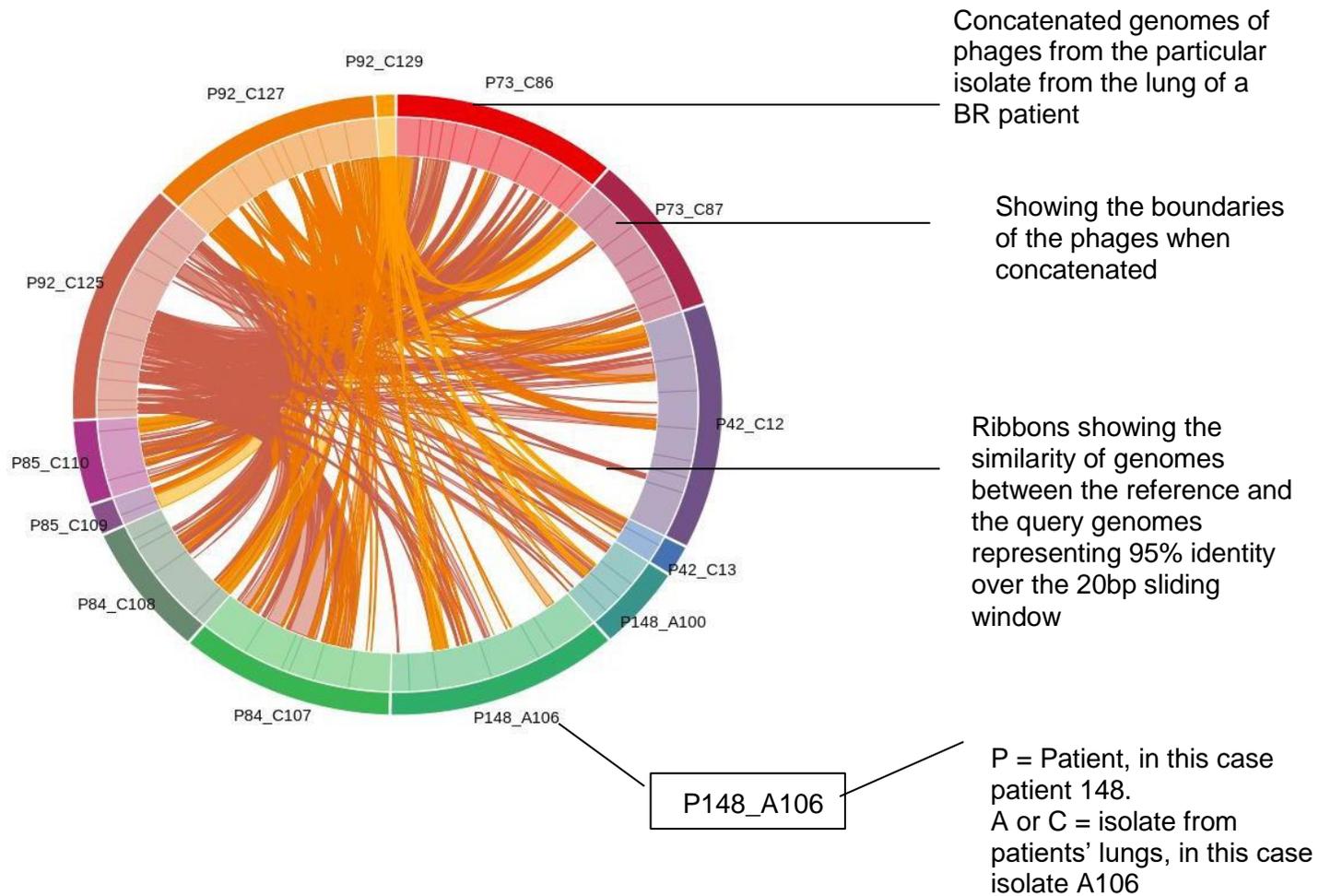


Figure 4.3 – An example of a Circos diagram - comparing all prophages from the 3 isolates from patient 92, compared against all of the other prophages from the other isolates in the cohort. The outer ring and the colours of the ribbons represent the isolate that the prophages came from. The inner ring is sectioned to show the start and end point of the concatenated phages for that particular isolate.

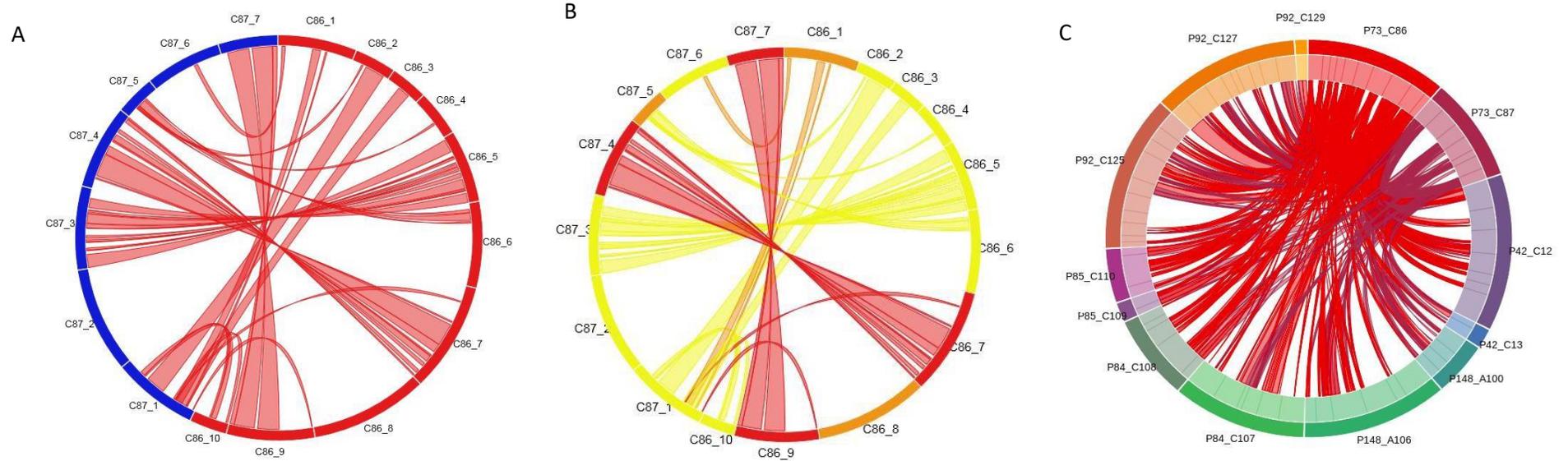


Figure 4.4a-c Circos comparison of phage genomes from Pa isolates C86 and C87 from patient 73. A) Circos plot showing the comparison between phage genomes from 2 isolates; C86 (reference) in red with 10 phages and C87 (query) in blue with 7 phages, with ribbons between them showing homology. B) Circos plot showing the same comparisons as (A) but coloured to show phage prediction with intact phages coloured red, questionable coloured orange and incomplete coloured yellow. C) Circos plot comparing all prophages from the isolates of both patient 73 Pa isolates compared against all of the other prophages from the other isolates in the cohort (explained in figure 4.3).

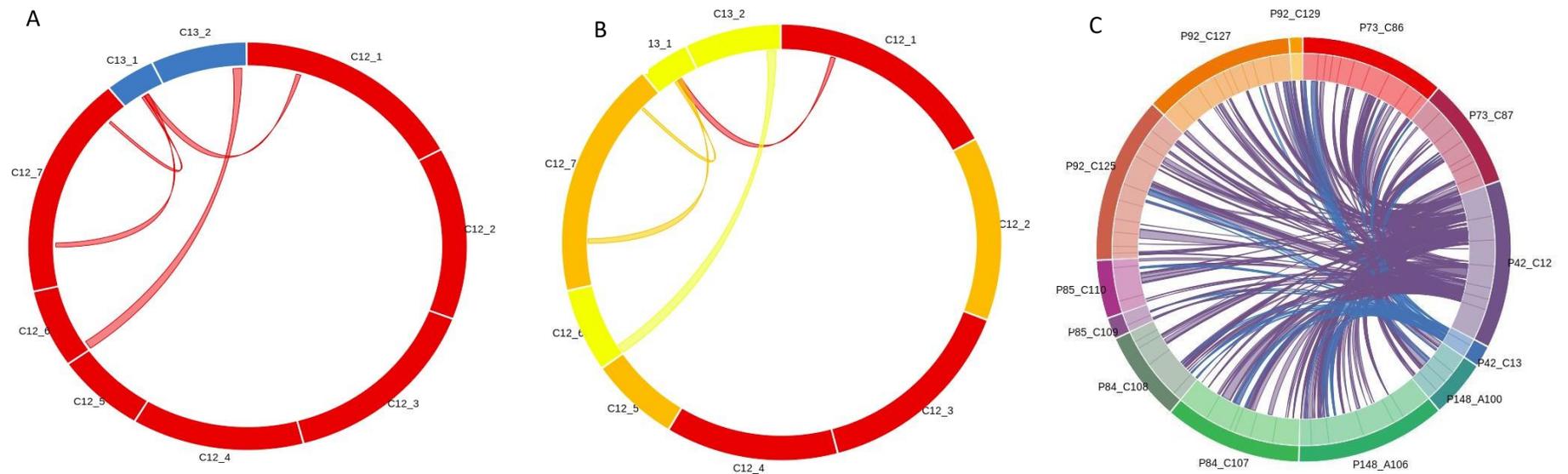


Figure 4.5a-c Circos comparison of phage genomes from Pa isolates C12 and C13 from patient 42. A) Circos plot showing the comparison between phage genomes from 2 isolates; C12 (reference) in red with 7 phages and C13 (query) in blue with 2 phages, with ribbons between them showing homology. B) Circos plot showing the same comparisons as (A) but coloured to show phage prediction with intact phages coloured red, questionable coloured orange and incomplete coloured yellow. C) Circos plot comparing all prophages from the isolates of both patient 42 Pa isolates compared against all of the other prophages from isolates in the cohort (explained in figure 4.3).

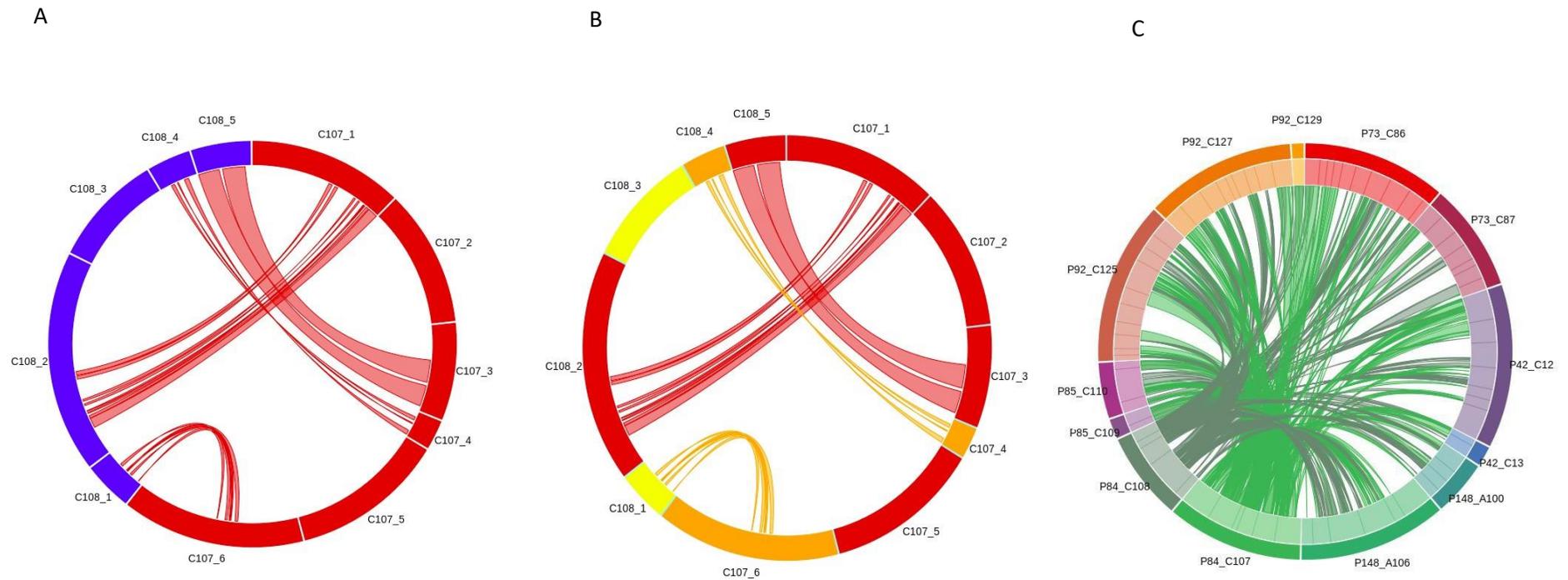


Figure 4.6a-c Circos comparison of phage genomes from Pa isolates C107 and C108 from patient 84. A) Circos plot showing the comparison between phage genomes from 2 isolates; C107 (reference) in red with 6 phages and C108 (query) in blue with 5 phages, with ribbons between them showing homology. B) Circos plot showing the same comparisons as (A) but coloured to show phage prediction with intact phages coloured red, questionable coloured orange and incomplete coloured yellow. C) Circos plot comparing all prophages from the isolates of both patient 84 Pa isolates compared against all of the other prophages from isolates in the cohort (explained in figure 4.3).

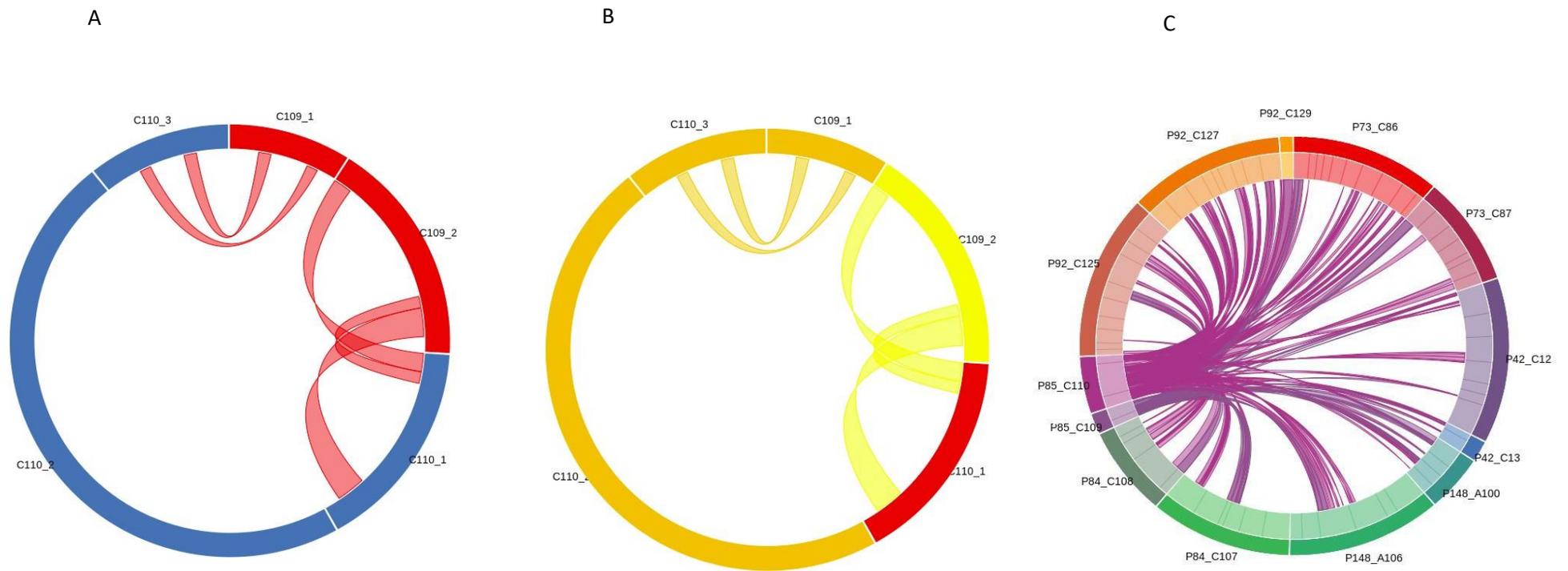


Figure 4.7a-c Circos comparison of phage genomes from Pa isolates C109 and C110 from patient 85. A) Circos plot showing the comparison between phage genomes from 2 isolates; C109 (reference) in red with 2 phages and C110 (query) in blue with 3 phages, with ribbons between them showing homology. B) Circos plot showing the same comparisons as (A) but coloured to show phage prediction with intact phages coloured red, questionable coloured orange and incomplete coloured yellow. C) Circos plot comparing all prophages from the isolates of both patient 85 Pa isolates compared against all of the other prophages from the other isolates in the cohort (explained in figure 4.3).

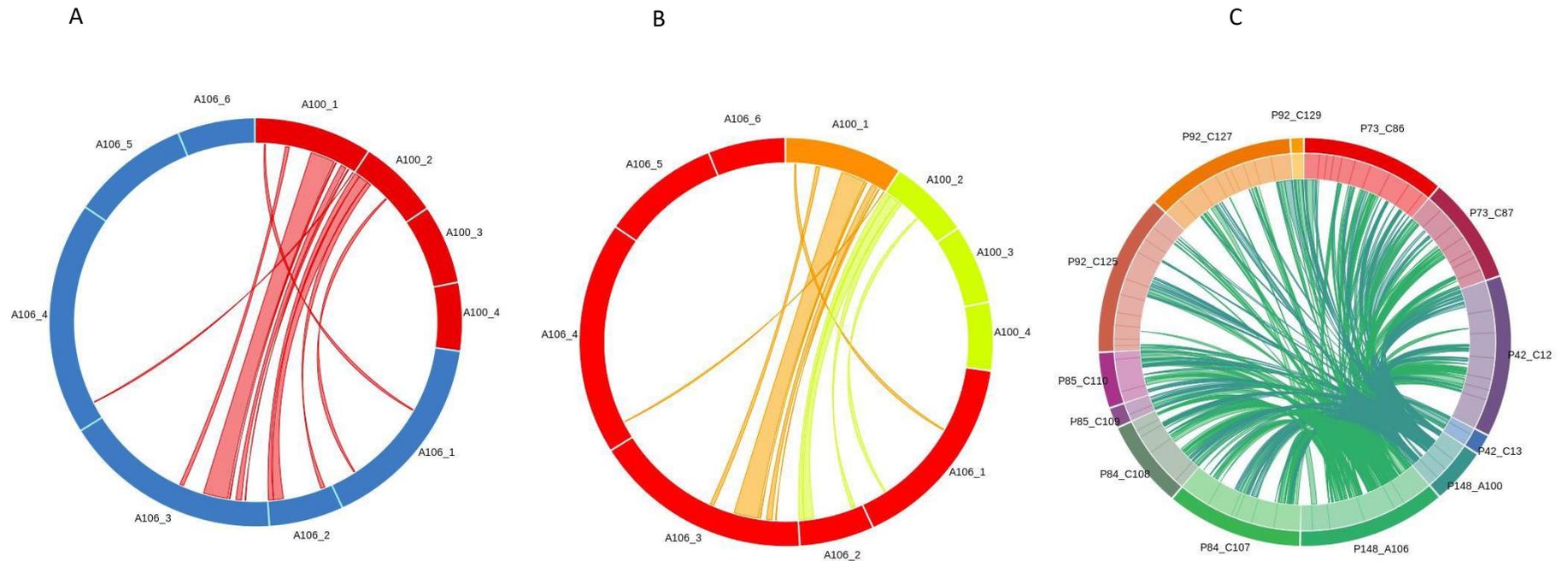


Figure 4.8a-c Circos comparison of phage genomes from Pa isolates A100 and A106 from patient 148. A) Circos plot showing the comparison between phage genomes from 2 isolates; A100 (reference) in red with 4 phages and A106 (query) in blue with 6 phages, with ribbons between them showing homology. B) Circos plot showing the same comparisons as (A) but coloured to show phage prediction with intact phages coloured red, questionable coloured orange and incomplete coloured yellow. C) Circos plot comparing all prophages from the isolates of both patient 148 Pa isolates compared against all of the other prophages from the other isolates in the cohort (explained in figure 4.3).

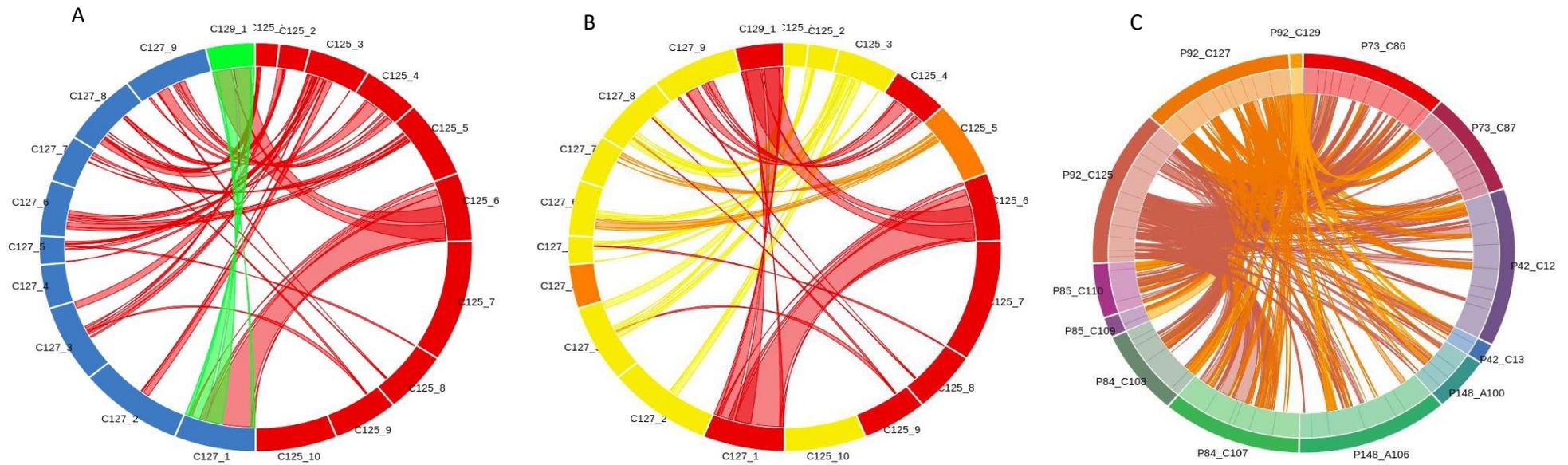


Figure 4.9a-c Circos comparison of phage genomes from *Pa* isolates C125, C127 and C129 from patient 92. A) Circos plot showing the comparison between phage genomes from 3 isolates; C125 (reference) in red with 10 phages and C127 (query) in blue with 9 phages and C129 (second query) in green with 1 phage, with ribbons between them showing homology. B) Circos plot showing the same comparisons as (A) but coloured to show phage prediction with intact phages coloured red, questionable coloured orange and incomplete coloured yellow. C) Circos plot comparing all prophages from the isolates of both patient 92 *Pa* isolates compared against all of the other prophages from the other isolates in the cohort (explained in figure 4.3).

Table 4.2 The length of each phages genome, percentage identity and coverage of each phage to the most similar other phage from the comparison isolate from the same patient.

Patient	Isolate	Phage	Completeness	Predicted size	Most similar phage from second isolate	% Identity	% Query cover	Most similar phage from third isolate	% Identity	% Query cover
P148	A100	A100_1	questionable	32.3Kb	A106_3	95%	36%			
		A100_2	incomplete	22Kb	A106_2	98%	36%			
		A100_3	incomplete	21.6Kb	N/A	N/A	N/A			
		A100_4	incomplete	18.6Kb	N/A	N/A	N/A			
	A106	A106_1	intact	55.5Kb	N/A	N/A	N/A			
		A106_2	intact	20.1Kb	A100_2	98%	39%			
		A106_3	intact	60.1Kb	A100_1	95%	19%			
		A106_4	intact	64.7Kb	N/A	N/A	N/A			
		A106_5	intact	32.5Kb	N/A	N/A	N/A			
		A106_6	intact	20.8Kb	N/A	N/A	N/A			
P42	C12	C12_1	intact	54Kb	N/A	N/A	N/A			
		C12_2	questionable	42.7Kb	N/A	N/A	N/A			
		C12_3	intact	47.3Kb	N/A	N/A	N/A			
		C12_4	intact	39.7Kb	N/A	N/A	N/A			
		C12_5	questionable	20.5Kb	N/A	N/A	N/A			
		C12_6	incomplete	19.1Kb	N/A	N/A	N/A			
		C12_7	questionable	7.5Kb	N/A	N/A	N/A			
	C13	C13_1	incomplete	11.2Kb	N/A	N/A	N/A			
		C13_2	incomplete	22Kb	N/A	N/A	N/A			
P73	C86	C86_1	questionable	27Kb	N/A	N/A	N/A			
		C86_2	incomplete	14Kb	C87_1	92%	59%			
		C86_3	incomplete	12.6Kb	C87_1	95%	35%			
		C86_4	incomplete	15.6Kb	N/A	N/A	N/A			
		C86_5	incomplete	25Kb	C87_3	96%	70%			
		C86_6	incomplete	29.7Kb	C87_5	96%	14%			
		C86_7	intact	38Kb	C84_4	98%	50%			

		C86_8	questionable	40.5Kb	N/A	N/A	N/A			
		C86_9	intact	30.2Kb	C87_7	96%	58%			
		C86_10	incomplete	12.6Kb	C87_1	93%	24%			
	C87	C87_1	incomplete	29.4Kb	C86_2	92%	28%			
		C87_2	incomplete	36.5Kb	N/A	N/A	N/A			
		C87_3	incomplete	28.7Kb	C86_5	96%	61%			
		C87_4	questionable	29.2Kb	C86_7	98%	66%			
		C87_5	questionable	13.4Kb	C86_6	96%	33%			
		C87_6	incomplete	26.2Kb	N/A	N/A	N/A			
		C87_7	intact	20.2Kb	C86_9	96%	88%			
P84	C107	C107_1	intact	47.5Kb	C108_2	98%	18%			
		C107_2	intact	42.2Kb	N/A	N/A	N/A			
		C107_3	intact	30.2Kb	C108_5	99%	50%			
		C107_4	questionable	19Kb	C108_4	98%	18%			
		C107_5	intact	38.3Kb	N/A	N/A	N/A			
		C107_6	questionable	56.5Kb	N/A	95%	N/A			
C108	C108_1	incomplete	15.2Kb	C107_6	95%	15%				
	C108_2	intact	68.6Kb	C107_1	98%	13%				
	C108_3	incomplete	34.8Kb	N/A	N/A	N/A				
	C108_4	questionable	13.4Kb	C107_4	98%	26%				
	C108_5	intact	18.4Kb	C107_3	98%	82%				
P85	C109	C109_1	questionable	11.5Kb	C110_3	98%	27%			
		C109_2	incomplete	21.5Kb	C110_1	98%	36%			
	C110	C110_1	intact	20.5Kb	C109_2	98%	39%			
C110_2		questionable	60.6Kb	N/A	N/A	N/A				
C110_3		questionable	13.4Kb	C109_1	98%	23%				
P92	C125	C125_1	incomplete	12.4Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C125_2	incomplete	25.7Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C125_3	incomplete	23.6Kb	C127_3	85%	13%	N/A	N/A	N/A
		C125_4	intact	34.3Kb	C127_9	90%	71%	N/A	N/A	N/A
		C125_5	questionable	16.3Kb	C127_6	89%	27%	N/A	N/A	N/A
		C125_6	intact	30.2Kb	C127_1	98%	94%	C129_1	99%	55%
		C125_7	intact	54.3Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C125_8	intact	27.1Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C125_9	intact	34.1Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C125_10	incomplete	9.5Kb	N/A	N/A	N/A	N/A	N/A	N/A
	C127	C127_1	intact	34.1Kb	C125_6	98%	85%	C129_1	99%	55%

		C127_2	incomplete	44.2Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C127_3	incomplete	34.4Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C127_4	questionable	19Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C127_5	incomplete	12.1Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C127_6	incomplete	24.4Kb	C125_5	89%	18%	N/A	N/A	N/A
		C127_7	incomplete	20.7Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C127_8	incomplete	32.1Kb	N/A	N/A	N/A	N/A	N/A	N/A
		C127_9	incomplete	36Kb	C125_4	90%	67%	N/A	N/A	N/A
	C129	C129_1	intact	20.1Kb	C125_6	99%	84%	C127_1	99%	93%

N/A – If there was under 10% coverage between the phage and any other phage from the comparison isolate from the same patient

4.4.4 Induction, DNA extraction and sequencing of phages from isolates to unequivocally identify the mobilisable phages

All 13 isolates were induced by exposure to norfloxacin (section 2.5.1). The lysate was processed to extract the DNA of any phages that had been induced, using DNase to remove high levels of background bacterial chromosomal DNA, with heat inactivation rupturing the capsid to present the genomes of the phages. These were then sequenced by NU-OMICS as per method in 2.8. Of the isolates, six out of 13 had detectable phages present when sequencing data was analysed. Seven isolates did not demonstrate inducible phages, whether this was due to a lack of mobilisable phages in that isolate or due to resistance to NFLX meaning they could not be induced using that method. However, four of the seven isolates had low sequence data (Not enough DNA that could be sequenced to form a genome) and therefore it was not possible to determine whether any phages were present. A summary table of the sequencing outcomes is shown in table 4.3 and shows that there were only 2 patients (patient 73 and patient 92) that demonstrated presence of phages in all their isolates and therefore their phages could be compared. The individual phages that could be re-assembled were mapped back against the bacterial genome data, using local pairwise BLAST to establish if they had been identified by PHASTER.

Each of the phages that were sequenced from isolates of patients 73 and 92, which had phages present in both/all three of their isolates, were then compared against the other phages from the other isolates from that patient by BLASTn (Table 4.4). From the results, the phages from patient 73's isolates that had the highest percentage identity and coverage were C86a and C87a. This high percentage similarity strongly suggests that they are the same phage. These phages also had a high similarity to C86c, identified by PHASTER. The phages in patient 92's isolates that had the highest percentage identity and highest percentage coverage were C125a and C127a. However, the phage sequenced in the third isolate (C129) was not similar to any phages from the other two isolates. As both patients 73 and 92 have isolates of differing morphologies that harbour

highly similar inducible phages, they are likely to have been transferred between isolates while co-colonising the chronically infected BR lung.

Table 4.3 Phage lysate sequencing results of phages induced from Pa strains isolated from BR patients

Patient	Isolate induced	SAMPLE	No mean high coverage contigs = potential Phages	Name of phage	Phage length	BLASTn-viral refseq (NCBI)
42	C12	LDH8	N/A- low seq data			
	C13	LDG4	N/A- low seq data			
73	C87	LDG5	1	C87a	27367	PAJU2
	C86	LDG6	3	C86a	39615	PAJU2
				C86b	36993	PhiCTX
				C86c	16213	YMC11/02/R656
148	A100	LDG7	N/A- low seq data			
	A106	LDG8	potentially chimeric could not resolve			
84	C107	LDH1	No phage identified			
	C108	LDH2	No phage identified			
85	C109	LDH3	1	C109a	50764	Phi297
	C110	LDH4	N/A- low seq data			
92	C125	LDH5	1	C125a	55893	PAJU2
	C127	LDH6	2	C127a	55891	PAJU2
				C127b	23501	PhiCTX
C129	LDH7	1	C129a	50102	Phi297	

Table 4.4 BLASTn comparison of sequenced phages induced from Pa strains isolated from BR patients

Patient	Isolate	Phage	Size (Kb)	Most similar phage from 2nd isolate	% Identity	% Query cover	Most similar phage from 3rd isolate	% Identity	% Query cover
P73	C86	C86a	27367	C87a	100%	99%			
		C86b	39615	N/A	N/A	N/A			
		C86c	36993	C87a	100%	99%			
	C87	C87a	C86a	16213	100%	66%			
			C86c	100%	27%				
P92	C125	C125a	55893	C127a	100%	100%	N/A	N/A	N/A
	C127	C127a	55891	C125a	100%	100%	N/A	N/A	N/A
		C127b	23501	N/A	N/A	N/A	N/A	N/A	N/A
	C129	C129a	50102	N/A	N/A	N/A	N/A	N/A	N/A

N/A – If there was under 10% coverage between the phage and any other phage from the comparison isolate from the same patient

4.5 DISCUSSION

The aim of this chapter was to compare phage carriage between Pa strains isolated at the same time-point from the same patient but which exhibited different colony morphologies. The chronically infected lung was previously thought to be clonal but morphological difference has been noted between colonies of Pa isolated from the same patient at the same timepoint. By looking at the genomes of the multiple Pa isolates that were isolated from the same patient it was confirmed that they were non-clonal strains by being different MLST types, this was also done as part of the Hilliam et al (2017) paper when comparing the core genomes of these bacteria (Hilliam et al., 2017). This data allowed the focus in this study to be on comparison of the accessory genome, specifically integrated temperate phages shared in the bacterial morphotypes genome. A key aim was to try to determine whether these phages transfer between the multiple shared isolates, possibly spreading genes that could offer a selective advantage within the lung microenvironment.

4.5.1 Phenotypic differences in Pa from the chronically infected lung

Other studies have seen transition? from mucoid to non-mucoid through environmental pressure, without change in their genome (Mathee et al., 1999) possibly through epigenetic regulation. Typically, the core genome of a bacterium is investigated for markers of adaptation, diversification and evolution. For example, in the first study of these isolates Hilliam et al. (2017) did not focus on the accessory genome in any way (Hilliam et al., 2017). There have been previous studies investigating the phenotypic diversity among Pa isolates between sputum samples, due to little being known about the underlying genetic diversity. Williams et al (2015) revealed extensive heterogeneity within Pa populations, including the co-existence of multiple divergent lineages in the CF lung (Williams et al., 2015), which also supports this study and the focus on the accessory genome. However, temperate phages were not looked at in depth as a possible cause of the genetic diversity and phenotypic changes in these studies.

4.5.2 Bacteriophage identification and integration

By identifying the temperate phages in the Pa genomes using PHASTER (section 2.11.1) and studying the integration sites of these prophages, it was established that 30% of prophages integrated into a bacterial gene. This has the potential to lead to phenotypic changes due to the interruption of a gene. In 6 cases, phages integrated into 'Indole-3-glycerol phosphate synthase' (IGPS), which was the only case where multiple phages integrated into the same gene, with 3 out of 6 being YMC11-like phages. YMC11 phages are Pa phages part of the *Siphoviridae* family. YMC11/02/R656 has been shown to be present in multidrug resistant strains of Pa from infected sputa and has shown to lower the virulence of Pa in the *Galleria mellonella* virulence model (Jeon and Yong, 2019). This suggests that this integration site may be phage specific. There is evidence that the deletion of the gene for IGPS reduces the fitness of Pa strain PAO1 in synthetic cystic fibrosis sputum (Gerth et al., 2012). Other genes that were disrupted by the integration of phages were involved in secretion systems and extracellular matrixes, such as 'Carboxyl-terminal protease', 'Type I secretion system, outer membrane component LapE' and 'Extra cellular matrix protein PsIE'. All of these gene products contribute to the fitness of Pa, of particular relevance is the exopolysaccharide PsIE, a key biofilm matrix component that initiates attachment to maintain biofilm architecture. Therefore, the disruption and knocking out of PsIE by the phage integration limits biofilm formation (Overhage et al., 2005), which in turn results in fewer vegetative bacterial cells for the phages to infect. As well as disrupting gene function, the integration of phages creates genomic variation to enable adaptation and evolution (Williams, 2013). Other than disrupting the bacterial genome, the phages add their own genes into the genomes of the bacteria that can alter the bacterial metabolism directly, such as providing virulence factors in the way of a toxin like CTX and STX toxins in *E.coli* and *Vibrio* species (Smith et al., 2012, Wagner and Waldor, 2002). There is also a positive correlation between lysogeny in Enterobacteria and pathogenicity (Bobay et al., 2013).

Integration of these phages can alter bacterial physiology. Therefore, it was also important to determine the diversity of the phages between isolates from the same patient to see if transfer of temperate phages occurs between isolates. By comparing and mapping phages

across all isolates from BR patients, this may illustrate phage presence frequency and whether they are conserved. It may also illustrate whether they integrate into the same location on the bacterial chromosome, which was seen here by YMC11-like phages integrating into Indole-3-glycerol phosphate synthase.

4.5.3 Prophage carriage and comparison between co-colonising Pa isolates

The Hilliam et al (2017) study determined that the co-isolated strains are not clonal.

Therefore, unsurprisingly, a different complement of phages was seen in each isolate from the same patient. At least one phage was identified in every isolate with a maximum of 10 phages with varying levels of completeness identified, observed for two of the isolates (C125 and c86). This number of predicted viable phages is high compared to other Gram-negative bacteria, with the number of prophages in bacterial genomes being highly variable. Many species of bacteria do not form lysogens (Fouts, 2006, Roux et al., 2015b), whereas other species, such as *Pseudomonas* isolates, usually form lysogens and frequently carry multiple prophages (James et al., 2015). This is due to the lack of restriction-modification systems found in their genomes, as they are the only feature discovered as yet that mediates phage infection (Oliveira et al., 2014). This makes C129 the only isolate that carries only a single prophage. This is an isolate of interest as it suggests it has a restriction-modification system and could be studied in further detail.

All of the prophage genomes identified were extracted and compared using BLASTn. Firstly, phage carriage between isolates from the same patient were compared, showing prophages with high percentage homogeneity to prophages from the comparable isolate, suggesting possible transfer of temperate phages. This was anticipated as induction of temperate phages and reinfection is likely, especially in a patient that is exposed to antibiotics on a long term basis (Tariq et al., 2019), such as in the case of a chronic infection, where lytic activity of temperate phages has been previously reported (James et al., 2015). Inflammation seen in a chronic lung infection also has the ability to induce phages through the inflammation response and the release of cytokines, seen in *Salmonella* in the gut, to boost the transfer of phages (Diard et al., 2017). Conversely, Pf phages from Pa have been shown to alter the progression of the inflammatory response, possibly to remain within the

lysogenic cycle (Secor et al., 2017).

These comparisons between the phages from the isolates from the same patient were visualised using Circos to clarify the parts of the genome that were similar and how similar the genomes were. The class of phage, whether it was intact, questionable or incomplete, was taken into consideration. If the prophage was classed as intact by PHASTER it was the most likely to be inducible as it has an intact genome and all the phage's essential genes for infection, conversion and replication. The phages that had high levels of similarity throughout the whole genome, and were classed as intact, suggest strongly that these phages are inducible and transferred while in the lung environment. This was then further supported by experimental work to induce and isolate the temperate phages from the Pa isolates. These isolates were then sequenced and the phage DNA was compared to one another. The DNA sequencing of the induced phages was only successful in two out of six patients. However, the results showed the phages that were induced from non-clonal isolates from the same patient were nearly genetically identical, with only a few base pair changes, strongly suggesting that these phages were transferred while in the lung environment. This also supports the hypothesis that antibiotic therapy may cause the induction of temperate phages and lead to transfer into other strains within the chronic environment. It is therefore possible that transferring genes may give antibiotic resistance or increase bacterial fitness and play a role in Pa chronicity. This hypothesis could also be tested in ASM co-culture experiments to see if the same temperate phages transfer.

4.6 SUMMARY

This study has shown the carriage of prophages in Pa genomes isolated from the chronically infected lung of BR patients co-colonised with multiple strains of Pa. The integration sites of these phages may knock-out bacterial genes where they integrate, which may change the phenotype of the Pa. There is also evidence of transfer of prophages between strains of Pa, which is the first time it has been reported to happen in the chronic lung environment.

5. WHAT ROLE DO TEMPERATE PHAGES PLAY IN SUBVERTING THE METABOLIC FUNCTION OF THEIR *P. AERUGINOSA* HOST?

5.1 INTRODUCTION

Chapters 3 and 4 mostly focussed on the identification and genomic diversity of temperate phages in a small cross-sectional panel of *Pseudomonas aeruginosa* (Pa). How the diversity in prophage genomes translates to alteration of bacterial cell function is still an understudied area in phage-bacterial host interactions. This is made more complex by polylysogeny and the high numbers of bacteriophage genomes that are present with high numbers of genes of small size with no or limited associated function. Approaches outside of genomics are needed to try to align function to conversion and lysogeny. A focus is needed to try and understand the association of lysogeny to environmental evolutionary adaptation of the bacterial host. How subversion of cell function by temperate phages plays a role in microbial infection and disease progression is not fully understood. Classically, they do alter cell function, in part, by carrying specific genes that promote positive evolutionary selection for the lysogen e.g. antibiotic resistance. It has been reported that temperate phages in Pa have the ability to enhance pathogen fitness in chronic lung infection (Davies et al., 2016a). It is valuable to know how this takes place and how can we understand what changes metabolically during lysogeny, and if these changes have the ability to enhance the fitness and alter the pathogenicity or virulence of Pa. To focus on the metabolism of the lysogen, comparative metabolite profiling was used to identify any phage-mediated changes in bacterial cell metabolism. The initial focus of this work was targeted at the well-characterised LES (Liverpool epidemic strain) phages as both individual and in-complex phages integrated into the genome of Pa host PAO1 are available (Winstanley et al., 2009). Previous studies by this group have illustrated that prophages induced from Pa from chronic lung infection accrue genes with specific function over time. Following these studies on the LES phages, the same methods were used to examine the metabolic changes to lysogens of PAO1 carrying a full complement of phages induced from clinical Pa isolates from patients at different stages of disease progression in CF and BR.

5.1.1 Effects temperate phages can have on their bacterial hosts

Temperate phage carriage is often associated with increased bacterial virulence (Waldor, 1998), where prophage carriage has been increasingly highlighted by the rise in the use of genome sequencing over the last 20 years, supported further by transcriptomics, allowing greater characterisation of the subtle ways in which prophages contribute to bacterial pathogenicity. Temperate phage infection and subsequent expression of phage-encoded genes by the host can have significant effects on a bacterial phenotype. These phage-encoded genes are usually not directly involved in viral replication and may provide a benefit to their bacterial host, classed as genes of 'moronic' function or moron genes (Brussow et al., 2004). These moron genes can benefit the host in a number of ways, by enhancing their virulence, either by encoding toxins e.g. Shiga toxin (Plunkett et al., 1999) or indirectly, and by enhancing the biological fitness (Davies et al., 2016a, Plunkett et al., 1999) of bacteria during infection (Hacker and Carniel, 2001). For instance, the majority of the genetic difference in bacterial genomes are associated to mobile genetic elements, notably phages in strains of *E.coli* denote whether they are an avirulent or virulent strain as temperate phages carry the Shiga toxin gene (Hayashi et al., 2001b, Ohnishi et al., 2002, Wagner et al., 2001). In *Pa* temperate phages can enhance their host by adding function that may offer a selective advantage within the environment of the chronically infected lung. They are able to manipulate cellular gene expression (Donnelly et al., 2015) and metabolism (Chevallereau et al., 2016). They can also drive genomic differences in their *Pa* host through integration into the chromosome. The effects of temperate phage on their hosts are becoming increasingly recognised. However, there needs to be further focus as the literature on the metabolic impact of temperate phage on their host is limited.

5.1.2 LES bacteriophages

The six temperate LES phages from the genome of the *Pa* isolate LESB58 (details of the phages shown in table 5.1) (Winstanley et al., 2009) are well defined and their effects on their *Pa* host have been studied previously (Fothergill et al., 2011, Ashish et al., 2012, James et al., 2012, James et al., 2015, Davies et al., 2016a), but not at a metabolic level. Davies et al (2016a) demonstrated that the temperate LES phages can enhance the fitness

of their Pa host. To demonstrate this, PAO1 lysogens were created using each of the mobilisable LES phages alongside a combination of PAO1 polylysogens with LES phages in a defined complex. These lysogens and polylysogens were tested for competitive fitness in a rat chronic lung infection model. The results showed temperate LES phages improved the competitiveness of lysogens against phage-susceptible populations in chronic lung infection (Davies et al., 2016a). These studies suggest that the temperate LES phages influence and enhance multiple stages of infection, altering the fitness of the host bacteria in the chronic lung environment. However, it is unclear how the LES phages effect the bacterial host to provide this enhanced fitness, which is something that the metabolomic analysis may be able to explain.

Table 5.1 Prophages identified in *Pseudomonas aeruginosa* strain LESB58. Data from Winstanley *et al.* (2009).

LES prophage	Characteristics	Number of genes	Related phages in reference strain PAO1	Known related phages most – like???	Phage morphology
φ1	Defective prophage, predicted to encode pyocin R2	19	Defective prophage gene cluster encoding pyocin R2	N/A	Pyocin gene clusters predicted to have evolved from phage tail genes
φ2	Active prophage, encodes integrase	44	None	F10	<i>Siphoviridae</i>
φ3	Active prophage, encodes integrase	53	None	F10	<i>Siphoviridae</i>
φ4	Active prophage, encodes transposase	48	None	D3112	<i>Siphoviridae</i>
φ5	Active prophage, encodes integrase	65	None	D3 (transposable phage)	<i>Siphoviridae</i>
φ6	Active prophage, encodes integrase	12	Pf4 (filamentous phage implicated in biofilm dispersal)	Pf1 (family: <i>Inoviridae</i>)	Filamentous phage

5.1.3 Metabolomics to analyse prophage-host interactions

It is widely known that bacteriophages carry small genes, some less than 200-300bp, that have no known function when studied at the genetic and amino acid level. These factors make it difficult to align function to phage conversion. A study in *E.coli* reported that around 65% of the genes carried by Shiga toxin encoding bacteriophages could not be related to a known cellular function (Smith et al., 2012). These studies see similar levels of phages and similar gene carriage in Pa temperate phages. Metabolomics approaches can be used to further understand phage host interactions and the significance of environmental evolutionary adaptation that may be driven by phages that have evolved in that environment. From the moment of phage infection, the virus transfers auxiliary metabolic genes (vAMGs), which can enhance the metabolic potential of the host during the infection process (Rosenwasser et al., 2016). An example of this outside of Pa is the carriage of photosynthetic genes in SPM-2, a phage that infects Cyanobacteria, which carries a gene that increases cell energy, allowing for lytic phage infection to occur (Shan et al., 2008). Phage-mediated metabolic changes in bacteria are hypothesised to markedly influence global nutrient and biochemical cycles, mainly by microbial mortality following infection, impacting microbial diversity. Phages exploit the resources of the host cells during lytic infection to drive viral replication, such as by intervening roles in phosphate and nitrogen metabolism, (Sullivan et al., 2010) and other metabolic pathways (Falkowski et al., 2008, Suttle, 2007).

However, experimental data on the impact of phage infection and temperate phage carriage on bacterial metabolism are still limited. Advances in metabolomics, enabled by the progress in mass spectrometry, has allowed analysis with high metabolite coverage and at multiple time points during infection (Timischl et al., 2008). The first study to use metabolomics approaches to investigate the phage infection of an environmentally relevant marine bacterium *Sulfitobacter* sp. 2047 was by Ankrah et al. (2014). Their results showed significant physiological differences between phage-infected and non-infected cells and speculated whether this response could be widespread in different phage-host systems (Ankrah et al., 2014).

Metabolomics links phenotypic and genotypic studies in order to generate a better overview of the effects a certain condition has on a cell (Weckwerth, 2003). Most of the metabolomics research into phage-host interaction in *Pa* up to now has been focused on lytic phages (De Smet et al., 2016, Chevallereau et al., 2016, Zhao et al., 2017) and temporal change to host metabolism between initial infection and lysis of the host cell. De Smet et al. (2016) used PAO1 reference strain and infected it during exponential growth with six different phages to determine if phage infection results in a metabolic 'universal' host response or if different phage species gave different metabolic changes. From LC-MS analysis, 118 metabolites were observed and phage infection had a significant influence on the concentration of 24.5% of these metabolites. It showed that lytic phages, as part of active infection, provoke an increase in pyrimidine and nucleotide sugar metabolism, probably linked to nucleic acid replication. However, only 2.4% of these changes were common to all investigated phages. These results show that the metabolomic changes mediated by lytic phage infection are not universal and differ between the individual phages tested (De Smet et al., 2016). There have been few studies looking at the metabolomics of lysogens that carry prophages after established lysogenic phage infection, despite the benefits of prophage carriage being well published. One study looking at the metabolomics of lysogens focused on the shiga toxin-prophage ϕ 24B on its *E.coli* host, which showed phage-mediated control of *E.coli* biotin and fatty acid synthesis and is rate limiting to cell growth. Also, through ϕ 24B integration, the lysogen gains increased antimicrobial tolerance to chloroxylenol and 8-hydroxyquinoline, providing validation into the method of metabolomics to look at the possible impact of integration of a prophage on the lysogen and the possible benefits they can give. Here, an untargeted metabolomic approach is taken that offers a semi-quantitative result of the possible metabolites present and their abundance with/without the integration of different prophages (Chokkathukalam et al., 2014)

5.1.4 Changes in virulence caused by temperate phages in *P. aeruginosa* in chronic lung infections

Temperate phages have been implicated in the pathogenesis of Pa lung infection in recent studies (Burgener et al., 2019, Tsao et al., 2018) and may aid the establishment of chronic infection. Filamentous phages (Pf) have been readily found in chronically infected lungs of CF patients and are associated with increased biofilm formation (Burgener et al., 2019). They have been shown to promote the formation of biofilms in Pa by spontaneously assembling into a liquid crystalline structure when alongside polymers found in CF sputum. These liquid crystals also insulate against antibiotic entry by stopping diffusion through biofilms. This can drive antibiotic tolerance against the most common anti-pseudomonas antibiotics, causing an increase in the virulence of the Pa (Secor et al., 2015). Therefore, Pf phages are likely to play a role in the pathogenesis of Pa chronic lung infections. They have also been shown to increase the virulence of Pa in animal models (Rice et al., 2009, Secor et al., 2017).

A recent study by Tsao et al (2018) considers the association between temperate phage genes ('phage morons') in Pa and the reduction in virulence of Pa. They identified moron genes that inhibit twitching and swimming motilities and observed an inhibition in the production of virulence factors in Pa, such as rhamnolipids and elastase (Tsao et al., 2018). This possibly allows the Pa to remain in the lungs for longer without detection and clearance by the immune system. Importantly, this demonstrates the scope of phage-mediated phenotypic changes that can be shown experimentally. These phenotypic changes provoked by the prophage are likely to be caused by a metabolic shift in the Pa, which may be picked up by metabolomics and allow the mechanics to be researched further.

5.2 AIMS

- Investigate how the temperate LES phages subvert the host bacterial cell's metabolism, using individual and polylysogens of PAO1.
- Investigate how temperate phages from clinical Pa isolates from CF and BR patients at different disease progressions effect the metabolism of PAO1 lysogens.

- Determine if clinically related phages induced from CF and BR patient Pa samples have the ability to change the virulence of PAO1.

5.3 OBJECTIVES

- Using LC-MS and metabolomics analysis, investigate how phages change the bacterial metabolism of PAO1 in Artificial Sputum Media (ASM) to better represent the environment of lung sputa. Firstly, LES phage lysogens of PAO1 will be investigated and then PAO1 that have been infected with clinical phages induced from CF and BR patient Pa samples. Using a *Galleria* virulence model, compare the infection of PAO1 with PAO1 that have been infected with clinical phages induced from CF and BR patient Pa samples.

5.4 RESULTS

5.4.1 Metabolomics results for PAO1 and LES phage lysogens

The lysogens of the three mobilisable LES phages infected into PAO1 were made by collaborators at the University of Liverpool. These lysogens were PAO1 containing LES2, 3, and 4 separately, and these were then created as polylysogens, LES3+4 and LES2+3+4. The lysogens LES2, LES3, LES4, LES3+4, LES2+3+4 and PAO1 (without added phages as a control) were grown in ASM for 3 days (see method section 2.3.2.2), carried out in triplicate for each lysogen. This method was previously reported by Kirchner et al (2012) to look at bacterial growth of biofilms by PAO1. It also has greater relatedness to the conditions within the chronically infected lungs compared to liquid media LB. The samples were prepared and extracted for LC-MS processing using the method in section 2.10.1. The samples were run on a Dionex 3000 Ultra High Pressure Liquid chromatography (UHPLC) system connected to the Q-Exactive classic high resolution mass spectrometer system by the metabolomics team at Northumbria University (see section 2.10.2 for detailed methods).

The output spectra identified 4768 metabolites. To streamline to the most significant metabolites, analysis of the dataset was performed using the Statistical analysis software of MetaboAnalyst, an online metabolomics informatics tool (using method in section 2.13), from which the dataset is uploaded and the parameters set to Peak Integrity Table with the format (unpaired). Data filtering was performed using Non-parametric relative standard deviation (MAD/median), with normalisation achieved by log transformation and pareto scaling. The resulting analysis path facilitated both Univariate, Chemometric as well as clustering analysis. By using the Partial Least Squares - Discriminant Analysis (PLS-DA), this gave the Variable Importance in Projection (VIP) score and any metabolite with a VIP score under 1 was discarded. The VIP is a weighted sum of squares of the PLS loadings, considering the amount of explained Y-variation in each dimension. Therefore, the higher the VIP score, the more significant the metabolite change in concentration. The 166 metabolites that were deemed to be important in explaining the difference between samples were compared to the PAMDB (*Pseudomonas aeruginosa* metabolomic database) using their molecular mass (Mr) to give putative compound/metabolite names replacing the M numbers generated by the LC-MS.

The significant metabolites were then analysed by using the Statistical analysis mode of MetaboAnalyst in the same way as described previously, but with fewer metabolites and compound names. Of the 166 metabolites, 123 (74%) were recognised by MetaboAnalyst. The compound names that were not recognised were denoted by N/A, instead of a compound name. The multivariate data or clustering analysis was studied from the output of the statistical analysis, as it compared all the lysogen groups and is shown on a principal component analysis (PCoA) plot in figure 5.1. The PCoA for each comparison has a corresponding heat map defining the upregulation or downregulation of those metabolites that confer the differences seen in the PCoA (Figure 5.2b). Data is also shown as a corresponding dendrogram (Figure 5.2a), illustrating how the lysogens cluster based on similarity of their metabolism. The PCoA plot shown in figure 5.1 demonstrates that the metabolism of PAO1 (without infection by LES phage) and the phage derived metabolism of each of the LES lysogens are markedly different. This is illustrated by the separation

between lysogen and PAO1 datapoints/nodes.

All lysogen datapoints are tightly clustered and therefore intra-variability is limited, showing the data's reproducibility. The points for lysogens LES3, LES4 and LES3+4 are clustered together, showing that LES3 and LES4 evoke similar phage-mediated metabolic responses in their PAO1 host when infected separately or in complement. LES2 and LES2+3+4 lysogens show the greatest impact on the bacterial cell metabolism in the PCoA (figure 5.1). The corresponding heatmap in figure 5.2b displays differential abundances of each of the metabolites that were classed as significant in describing the difference between PAO1 and the lysogens. It shows multiple clusters of metabolites that are highly up regulated in PAO1 and are down regulated by all of the lysogens, including homocarnosine, L-ornithine and riboflavin, and vice versa in many metabolites.

Place initial figures (5.1, 5.2) here, otherwise the reader has to keep skipping forward and backwards.

To investigate altered metabolites in further detail, each of the LES lysogens were compared individually to PAO1, as shown in figures 5.3-5.7. The PCoAs in figures (a) 5.4-5.7 confirm that there is a phage-mediated alteration in metabolism between PAO1 and each of the LES lysogens. The dendrograms shown in figures (b) 5.3-5.7 show that the LES lysogens cluster separately from PAO1 samples, showing a distinct difference between the lysogen metabolisms and the controls. In addition, the sample replicates cluster, which demonstrates reproducibility. Comparing these separately allows direct comparison to see how each of them alter the regulation in turn, comparing Mass Spectra to the control, and is shown in the heatmaps in figures (c) 5.3- 5.7 (these are shown in a larger view in appendix E figures 1a-e). As there were a high number of metabolites, it was necessary and informative to use pathway analysis to determine which metabolic pathways of the bacterial host are altered by lysogeny.

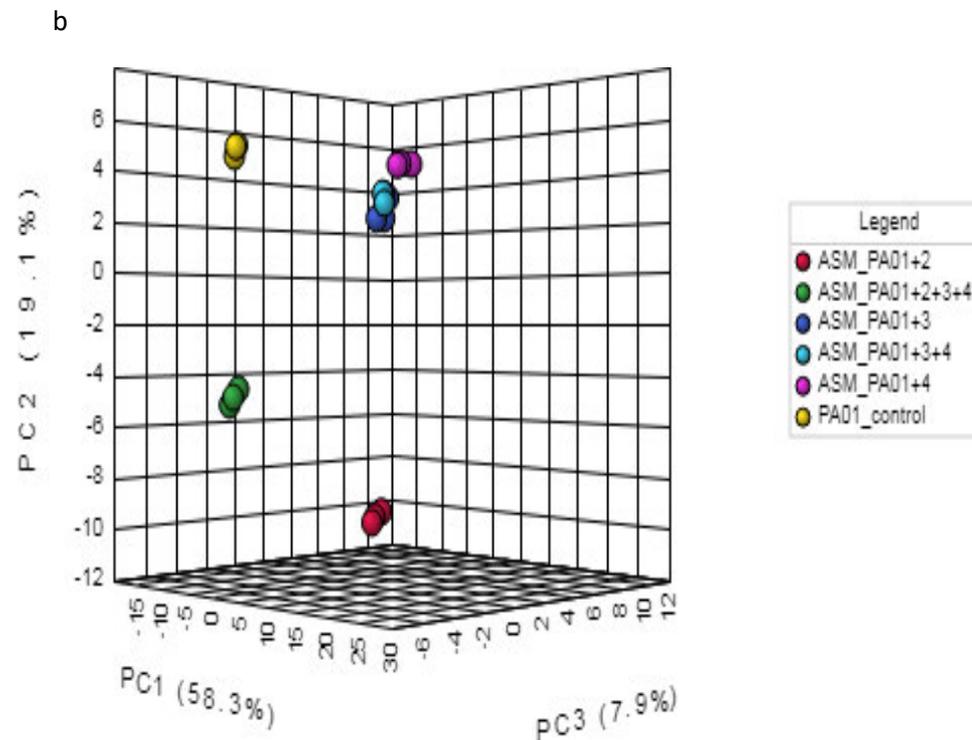
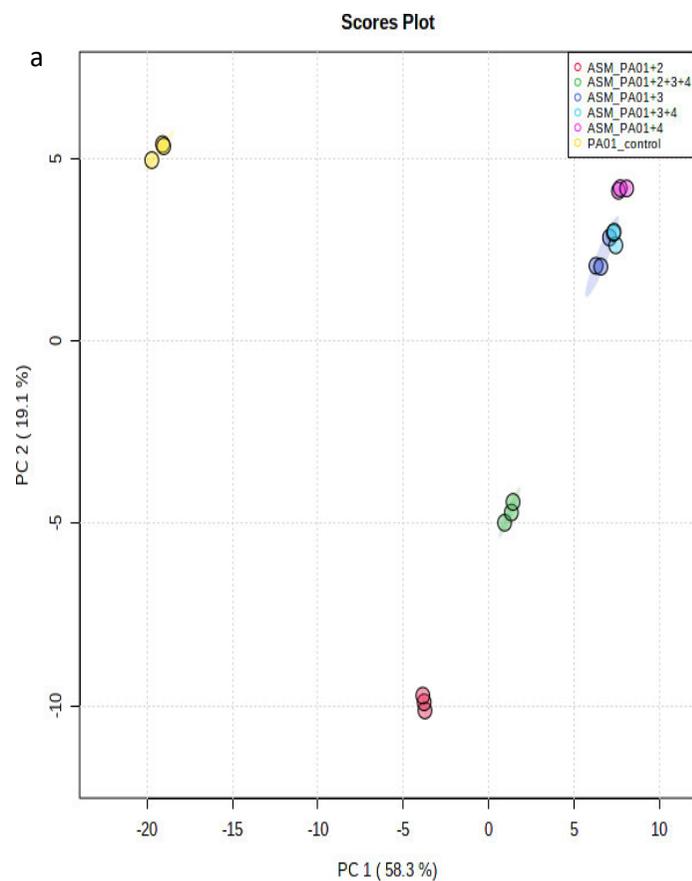


Figure 5.1 – PCoA analysis of all LES lysogens’s metabolomic profiles compared with one another and the control. (a) 2D PCoA plot. (b) 3D PCoA plot with each node representing the metabolic signature results (significant metabolites only) of each LES lysogen and the PA01 control grown in ASM and ran in negative mode, showing the dissimilarity between the metabolisms of each lysogen.

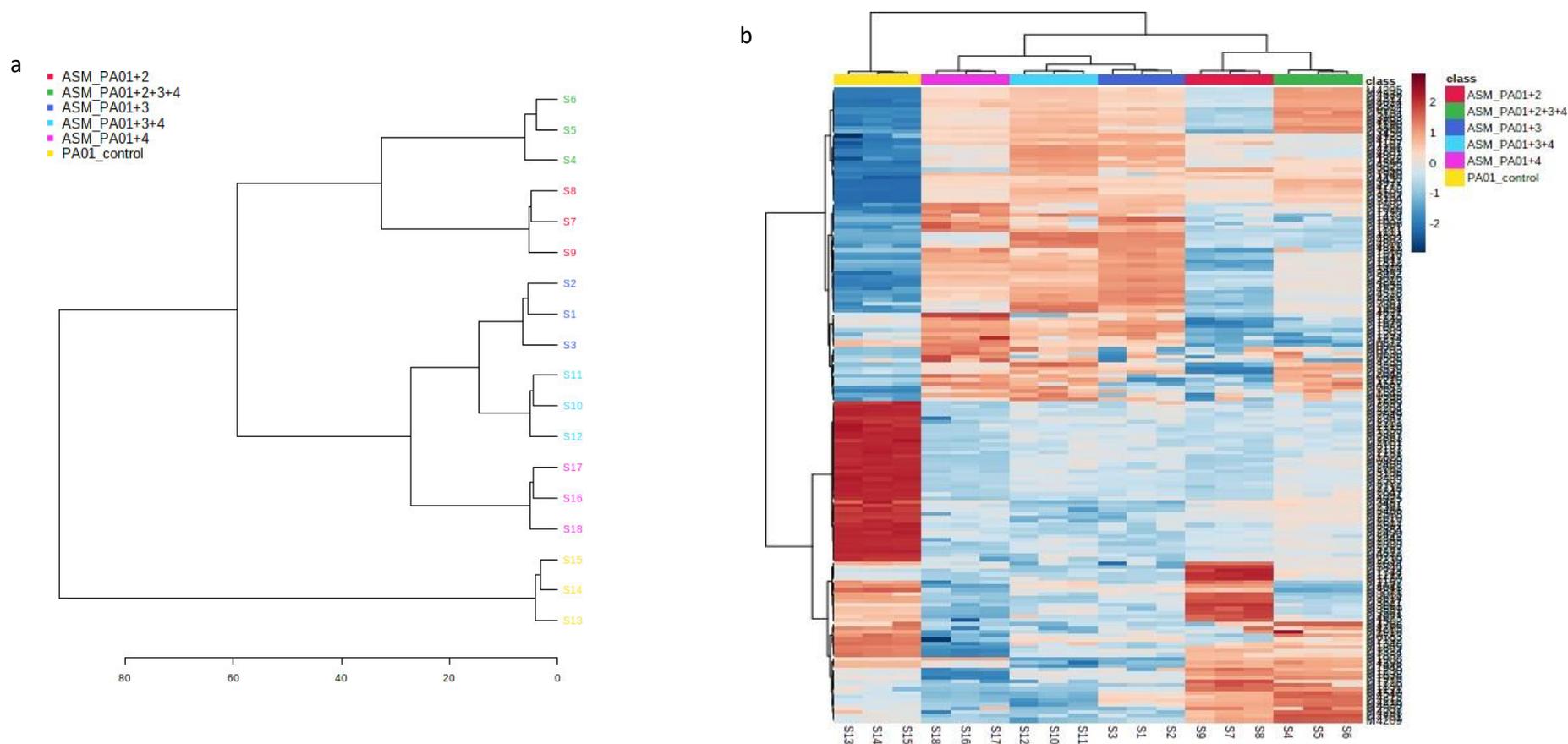


Figure 5.2 – Dendrogram and heatmap of all the LES lysogens' metabolomic profiles compared with one another and the control (a) Hierarchical clustering result shown as a dendrogram (distance measure using Euclidean and clustering algorithm using ward D), showing all samples of LES lysogens and control compared. (b) Clustering result shown as heatmap representing the up and down regulation of each significant metabolite between the lysogens and the control. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

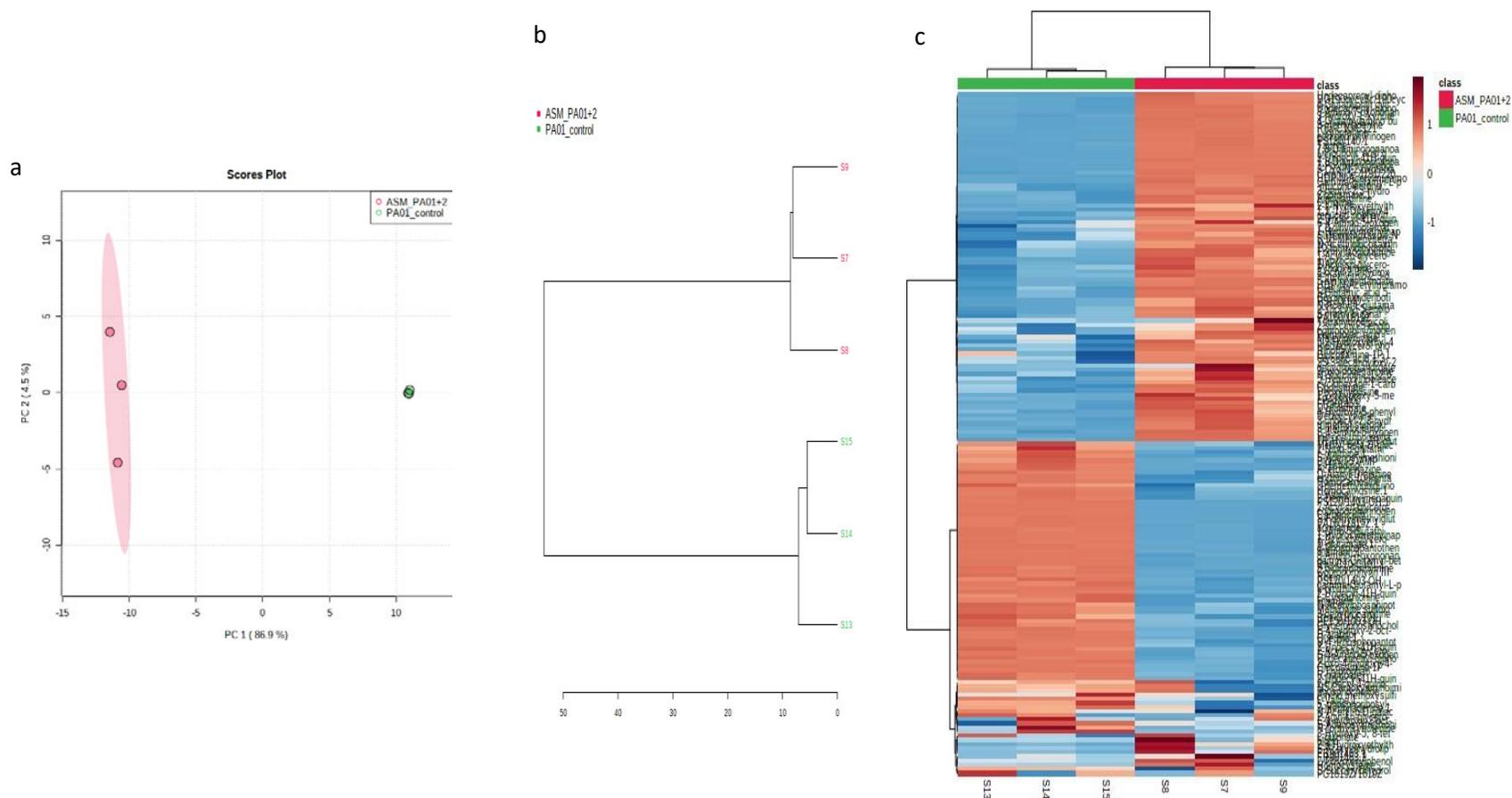


Figure 5.3 – PCoA, dendrogram and heatmap of the LES2 lysogens metabolomic profiles compared to the control. (a) 2D PCoA plot with each node representing the MS results (significant metabolites only) of each LES-2 lysogen and the PAO1, showing the dissimilarity between the metabolism of the LES-2 lysogen and PAO1 without the prophage. (b) Hierarchical clustering result shown as dendrogram, showing LES-2 lysogen and the control compared. (c) Clustering result shown as heatmap, representing the up and down regulation of each significant metabolite between the LES-2 lysogen and the control. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample (S) numbers are associated to is detailed in appendix F.

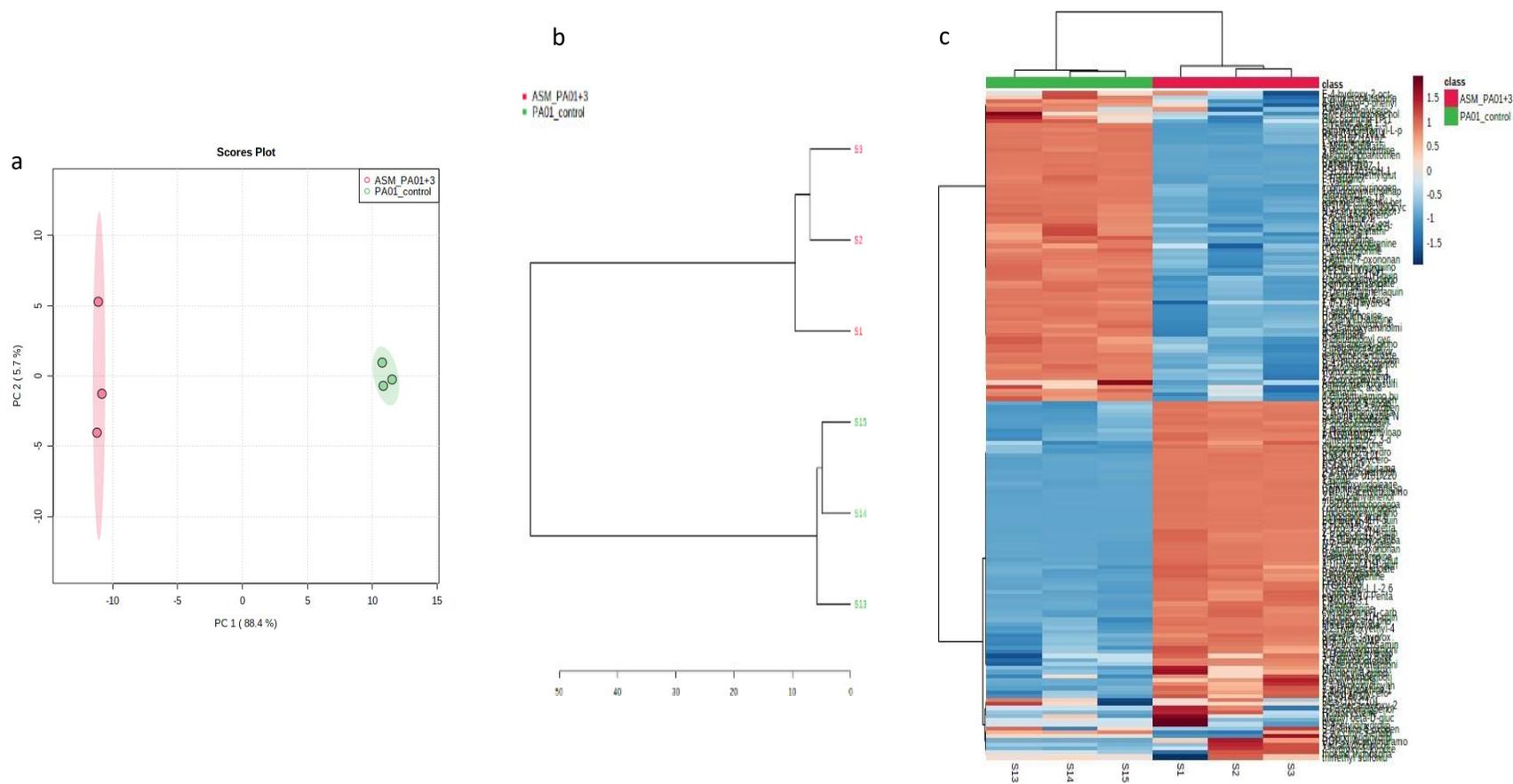


Figure 5.4– PCoA, dendrogram and heatmap of the LES3 lysogens metabolomic profiles compared to the control (a) 2D PCoA plot with each node representing the MS results (significant metabolites only) of each LES-3 lysogen and the PAO1, showing the dissimilarity between the metabolism of the LES-3 lysogen and PAO1 without the prophage. (b) Hierarchical clustering result shown as dendrogram, showing LES-3 lysogen and the control compared. (c) Clustering result shown as heatmap, representing the up and down regulation of each significant metabolite between the LES-3 lysogen and the control. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

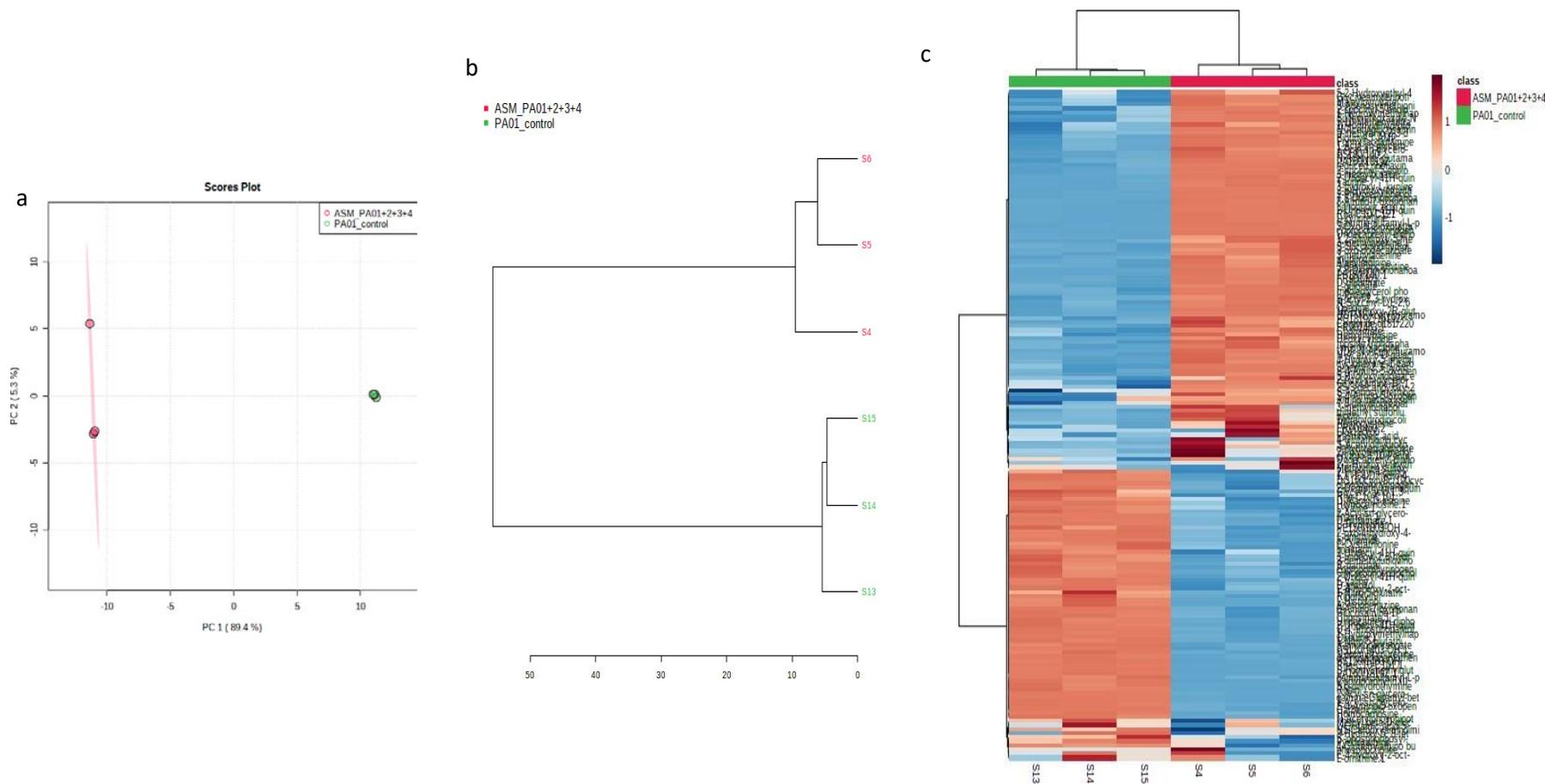


Figure 5.6 – PCoA, dendrogram and heatmap of the LES2+3+4 lysogens metabolomic profiles compared to the control (a) 2D PCoA plot with each node representing the MS results (significant metabolites only) of each LES-2+3+4 polylysogen and the PAO1, showing the dissimilarity between the metabolism of the LES-2+3+4 lysogen and PAO1 without the prophage. (b) Hierarchical clustering result shown as dendrogram, showing LES-2+3+4 polylysogen and the control compared. (c) Clustering result shown as heatmap, representing the up and down regulation of each significant metabolite between the LES-2+3+4 polylysogen and the control. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

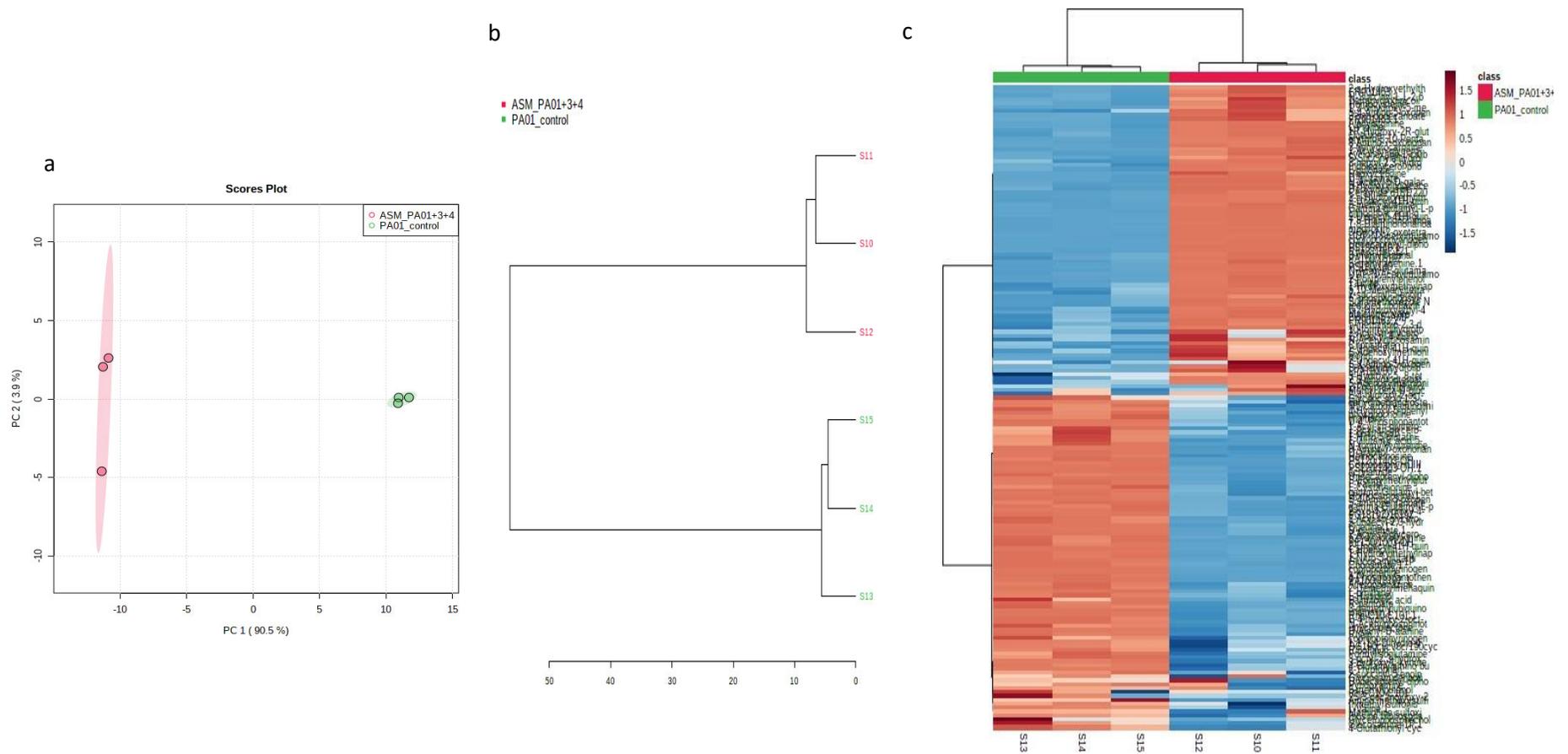


Figure 5.7– PCoA, dendrogram and heatmap of the LES3+4 lysogens metabolomic profiles compared to the control a) 2D PCoA plot with each node representing the MS results (significant metabolites only) of each LES-3+4 polylysogen and the PAO1, showing the dissimilarity between the metabolism of the LES-3+4 polylysogen and PAO1 without the prophage. (b) Hierarchical clustering result shown as dendrogram, showing LES-3+4 polylysogen and the control compared. (c) Clustering result shown as heatmap, representing the up and down regulation of each significant metabolite between the LES-3+4 polylysogen and the control. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F

5.4.1.1 Pathway analysis and metabolite concentration comparison of LES lysogens

The pathway analysis was performed again using MetaboAnalyst, using data generated from a metabolic spectra search and metabolites that were identified comparing to PAMDB (74%). Metabolites were then narrowed down to only metabolites with associated Kyoto Encyclopedia of Genes and Genomes (KEGG) IDs (54%). This increased the likelihood of MetaboAnalyst being able to map the metabolites to a pathway. The impact results from each of the pathways compared to the control were then concatenated to give the heatmap shown in figure 5.9. These results show significant phage mediated metabolic pathway differences in the lysogens when compared to the control (p value <0.05). The impacts of each of the LES phages on their bacterial host's metabolic pathways are relatively similar, apart from LES-2. The pathways that were impacted the most by the addition of the LES phages (top 6) were 'Carbapenem biosynthesis', 'Taurine and hypotaurine metabolism', 'D-Glutamine and D-glutamate metabolism', 'Biotin metabolism', 'One carbon pool by folate' and 'Pantothenate and CoA biosynthesis'. The exception is LES-2, which was shown to not have a highly significant impact on the 'Taurine and hypotaurine metabolism', whereas the other LES lysogens did. The LES-2 lysogen also showed altered sulphur biosynthesis, which was not seen in the other lysogens. The other 5 pathways had similarly high levels of impact across all 5 LES lysogens. The effects these pathway changes may have on the Pa host are discussed further in section 5.5.

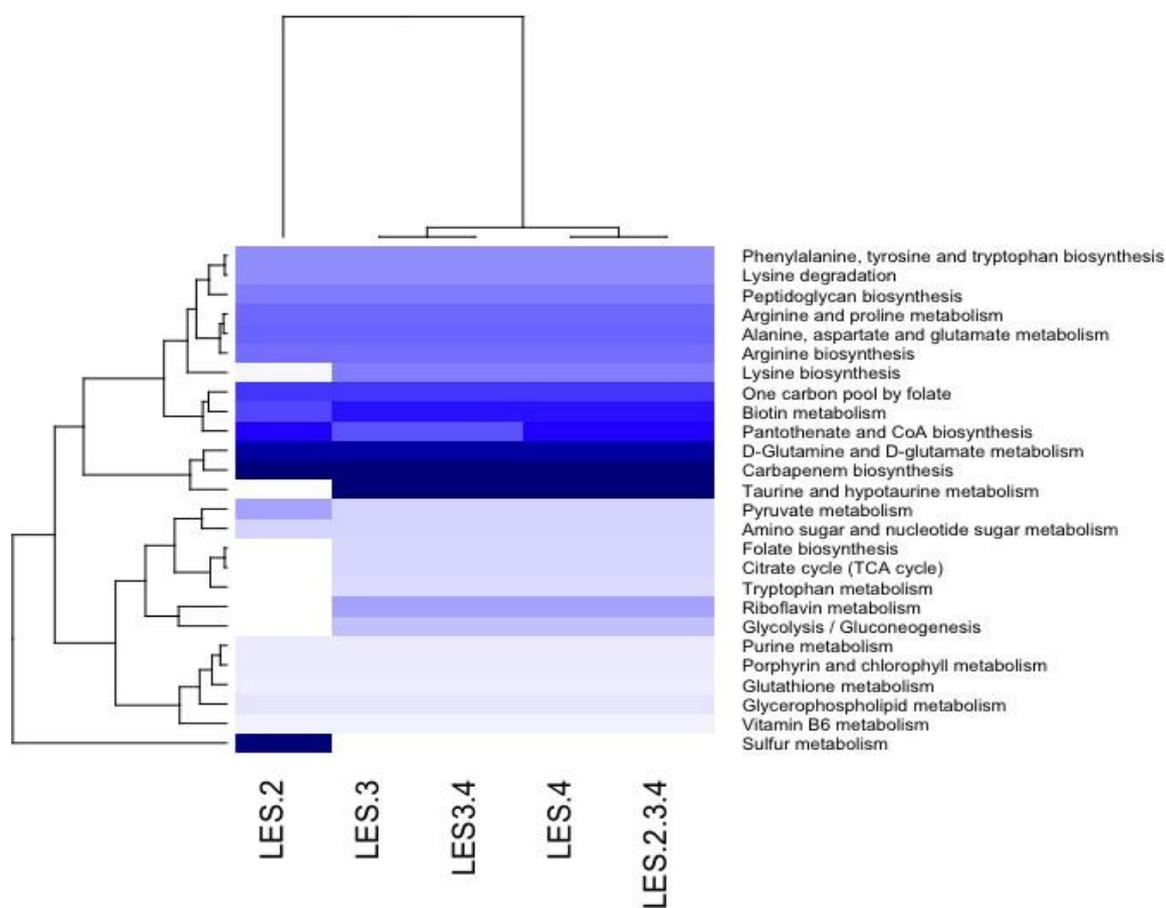


Figure 5.9 - Pathway analysis of LES phage lysogens. A heatmap showing the phage mediated bacterial host metabolic pathway 'impact' of LES phage/s compared to the uninfected PAO1 control. Darker blue represents more significant changes to the pathway i.e. more metabolites are changed within a pathway or a significant metabolite to that pathway has been affected compared to the control. These data was collated from MetaboAnalyst pathway analysis tool on each lysogen.

5.4.2 Metabolomics results for lysogens of PAO1 and clinical *P. aeruginosa* lysogenic phages from BR and CF patients

This work builds on earlier research carried out by our research group, with data published on the characterisation of 20 lysogens of PAO1 (Tariq et al., 2019), where phages were induced from Pa from four clinical groups based upon the aetiology of their isolation; 'Adult CF' (>15 years of age), 'Paediatric CF' (<15 years of age), '>10 BR (>10 years since date of diagnosis)' and '<10 BR (<10 years since date of diagnosis)'. The aim was to compare lysogens depending on aetiology of the infection of the bacteria the phages were induced from. PAO1 lysogens were created that contained single or multiple prophages induced from a single Pa culture from one of these groupings. All 20 were sequenced for confirmation of integration and phage genome type (Tariq et al., 2019). These previous studies work under the hypothesis that evolution in the lung of both the bacteria and predating phage will be exposed to high amounts of selective pressure and will therefore undergo rounds of adaptation and evolution driven by the lung environment. Some of these genetic traits, which cause metabolic shift(s), may be carried by these phages and having aetiology-based function that directly links to the pathophysiology of disease. This study uses the same approach in the preparation and analysis for the metabolite data as was used for the metabolomics analysis of LES lysogens.

The output from the LC-MS gave a Peak Integrity Table identifying 2475 metabolites, which were then streamlined to the most significant metabolites as previously described (section 5.4.1). Reduction techniques based on quality matrices gave 546 significant metabolites, significantly more than were identified for LES lysogens. These were plotted on a PCoA (figure 5.10) using the statistical analysis mode on MetaboAnalyst, which displayed very little dissimilarity between the groups of lysogens and the control. To attempt to identify which metabolites had been changed between groups and the control, reduction to only the most significant metabolites was necessary to detect the shift in metabolism better between the groups. This was achieved by running the statistical analysis mode of MetaboAnalyst on each group and comparing it to the control and taking the top 100 most significant

metabolites from each, which were concatenated, and subsequently yielded 142 metabolites. These were the most significant metabolites associated with the metabolic difference between the uninfected bacterium PAO1 and each lysogen.

Both the full list of 546 significant metabolites and the streamlined list of 142 metabolites were ran through MetaboAnalyst's statistical analysis mode again to compare the outcome of streamlining the metabolites further. Multivariate data analysis comparing all the lysogen groups for the full and streamlined sets of significant metabolites are shown on a PCoA plot in figure 5.10a and figure 5.11a, respectively. From these PCoA plots (Figure 5.10a), there is very little separation between the nodes from each group and the control, whereas in figure 5.11a the nodes that represent the control PAO1 without phages cluster separately, showing that there is a change in the streamlined metabolites between the control and the lysogens. However, it also shows there is a limited difference in metabolism between the groups of lysogens when all the significant metabolites are considered. It is interesting that there may be a core phage-derived control of metabolism and further tools maybe needed to identify into the unknown metabolites present. Each PCoA plot has a corresponding heatmap that shows the specific metabolites that are up or down regulated in each sample from each clinical group. They are shown in figures 5.10c and 5.11c with a corresponding dendrogram shown in figure 5.10b and 5.11b, which illustrates how the samples from each of the lysogen groups cluster together based on their metabolism. Alongside the PCoA, this also displays that the samples within different clinical groups cluster with one another, showing similar metabolisms. This is unlike the obvious separation seen in the LES lysogen groups. This may be due to the phages, that were integrated into the PAO1, being from clinical isolates with 5 different phage/s per clinical group (each with an $n=5$), rather than a single phage change (with an $n=3$) in the LES lysogens, which means there were more variables to take into consideration. However, figure 5.11 (the streamlined metabolite group) shows an obvious cluster of the PAO1 (control) samples away from the lysogens (Figure 5.11d), which shows the average of all samples from each group, confirming the differences in the metabolites regulation between the control and the lysogen groups.

To look at this in further detail, each of the lysogen groups were compared separately

against PAO1, comparing the streamlined metabolites (Figures 5.12-5.14). The PCoAs in figures 5.12a, 5.13a, 5.14a and 5.15a confirm that there is a marked change in the metabolism between PAO1 and each of the lysogen groups. The nodes in the PCoA plots were not as tightly clustered as the nodes for the LES lysogens replicates as these groups contained 5 different lysogens, rather than biological replicates of the same lysogen. The dendrograms shown in figures 5.12b, 5.13b, 5.14b and 5.15b show that the lysogens cluster separately from PAO1 samples, showing a distinct difference between the lysogens metabolisms and the controls. However, the results in figure 5.12b, which show lysogens in the Paed_CF clinical group compared to the control, show that the lysogens cluster in two groups either side of the control samples, which could suggest there were two separate types of phage that changed the metabolism in different ways within the group.

Comparing these separately allows for direct comparison between PAO1 and each lysogen/polylysogen within the clinical groups to see how each of them regulate metabolites identified by MS compared to the control, as shown by the heatmaps in figures 5.12c, 5.13c, 5.14c and 5.15c (these are shown in a larger view in appendix E figures 2-4). In the heatmaps in figures (c) the majority of lysogens in each group regulated each of the metabolites in a similar way. However, it is apparent that some of the samples/lysogens within the clinical groups regulated the metabolites in opposing ways, such as sample S24 in the adult CF group (figure 5.12). This shows that different temperate phages isolated from Pa strains from patients with the same clinical conditions can affect the Pa metabolism differently. If a lysogen within a group altered a particular metabolite of interest significantly, it can easily be identified here, and further work can be carried out to investigate it in more detail. However, this study was not looking for differences within the groups of lysogens but focusing on whether there were changes of PAO1 in the metabolism due to the addition of clinical prophages, and determining the functions they might have on the Pa host.

As there were a high number of metabolites it was necessary and informative to perform pathway analysis to determine which pathways the significantly changed metabolites fit into.

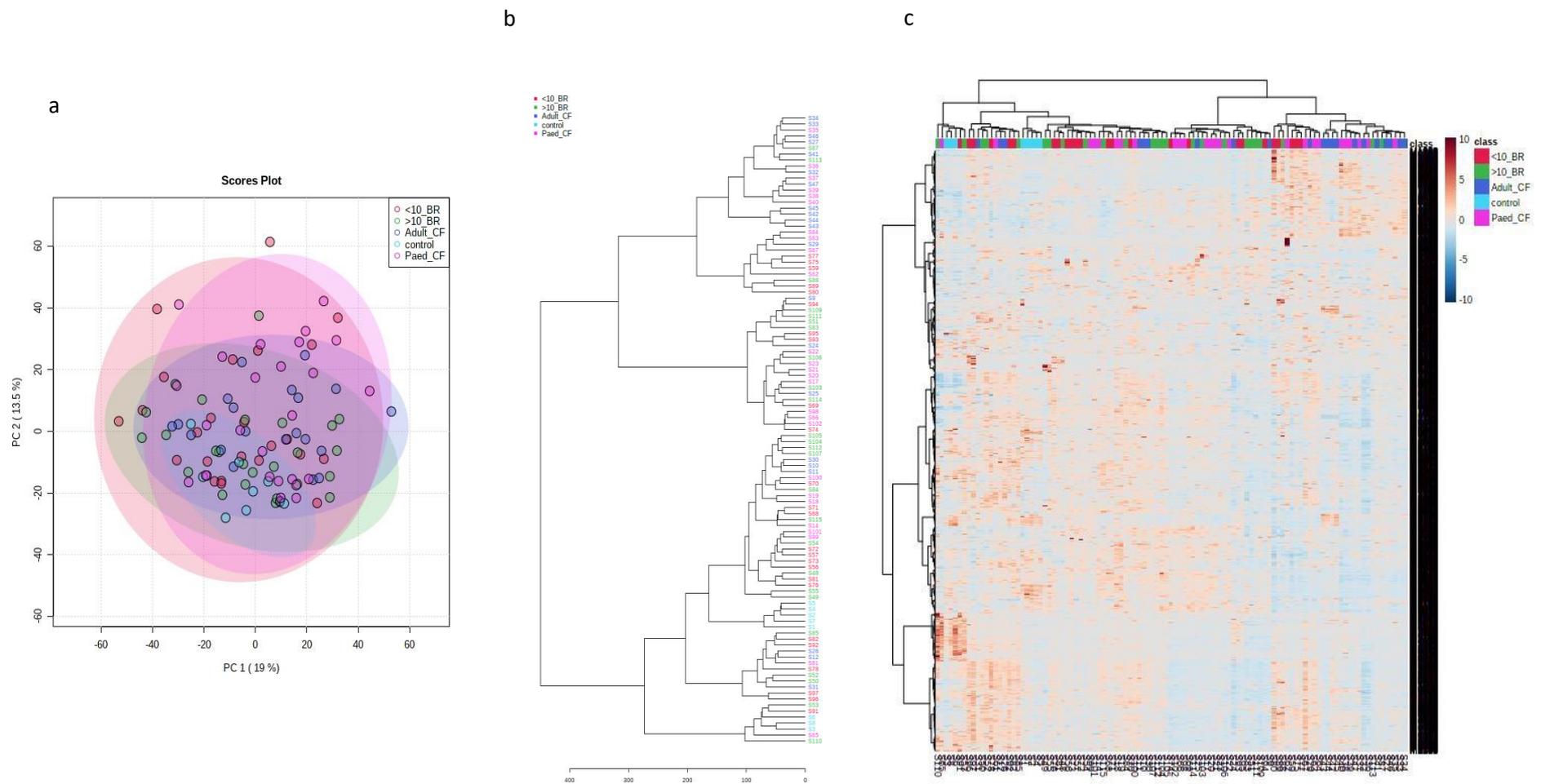


Figure 5.10 – Metabolomics results of all 546 metabolites for all lysogens from each of the four clinical groups compared with one another and the control. (a) 2D PCoA plot with each node representing the MS results for the significant metabolites only (546 metabolites total) of each lysogen and the PAO1 control grown in ASM and run in negative mode, showing the dissimilarity between the metabolisms of each lysogen. (b) Hierarchical clustering result shown as dendrogram (distance measure using euclidean, and clustering algorithm using ward D), showing all samples of lysogens from each group and control compared. (c) Clustering result shown as heatmap (distance measure using euclidean, and clustering algorithm using ward D), representing the up and down regulation of each significant metabolite between the lysogens and the control.

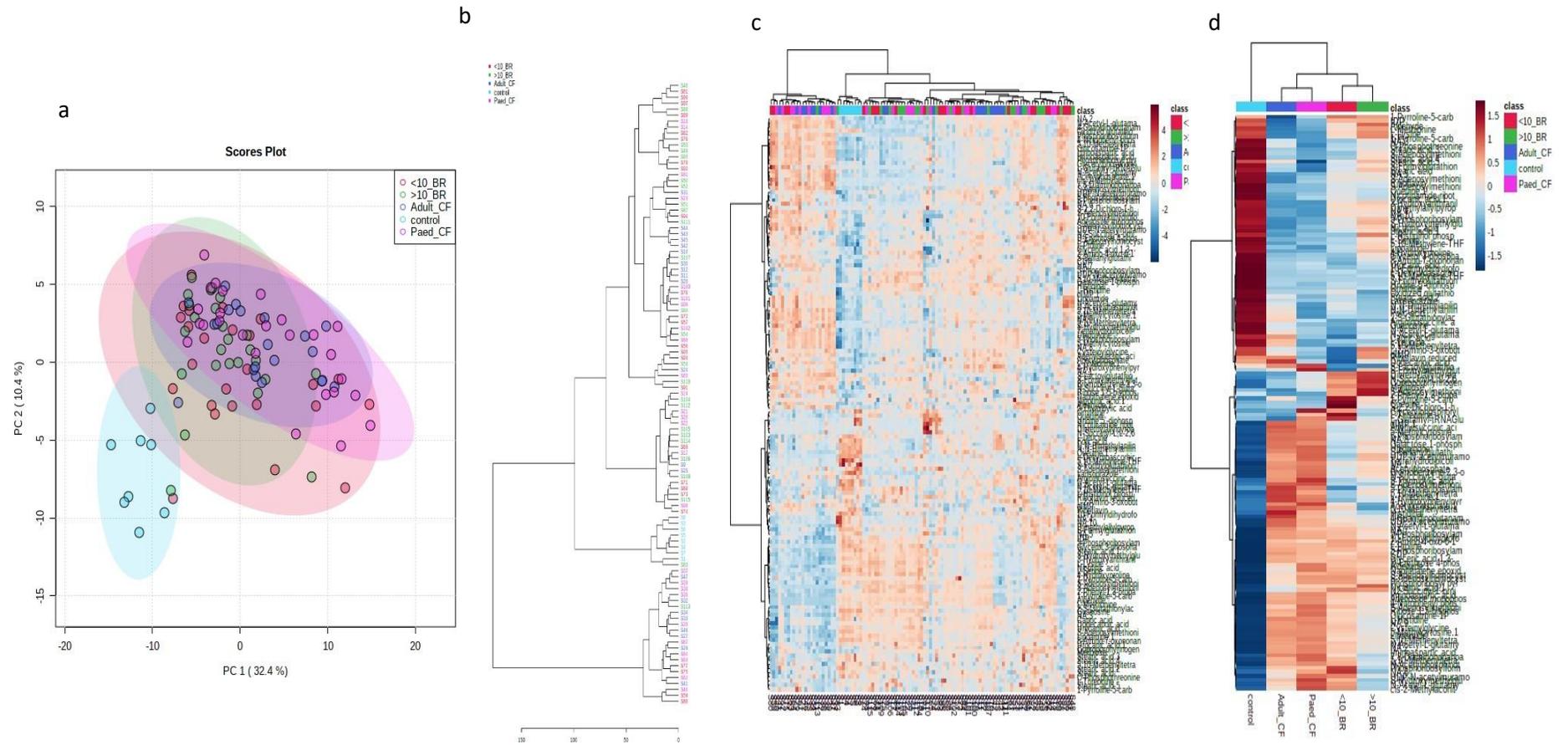


Figure 5.11 – Metabolomics results of 142 streamlined metabolites for all lysogens from each of the four clinical groups compared with one another and the control. Narrowed down to only the most significant metabolites. (a) 2D PCoA plot with each node representing the MS results for the top 100 most significant metabolites when each group was compared to the control (142 metabolites total) of each lysogen and the PAO1 control grown in ASM and ran in negative mode, showing the dissimilarity between the metabolisms of each lysogen. (b) Hierarchical clustering result shown as dendrogram (distance measure using euclidean, and clustering algorithm using ward D), showing all samples of lysogens from each group and control compared. (c) Clustering result shown as heatmap (distance measure using euclidean, and clustering algorithm using ward D), representing the up and down regulation of each significant metabolite between the lysogens and the control. (d) The average score of the samples from the clade, and for the control, was taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

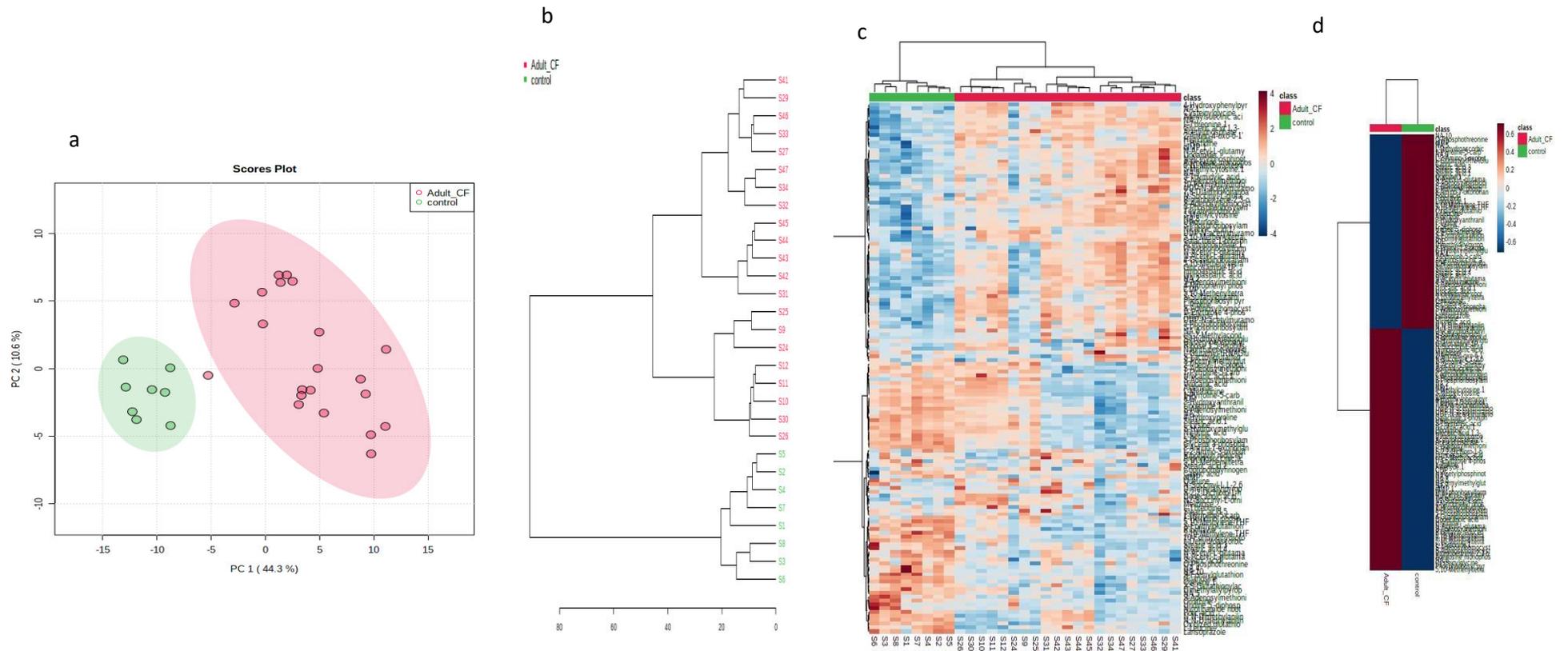


Figure 5.12 – PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from the adult_CF clinical group (lysogens containing prophages from Pa samples from adult CF patients) compared with the PAO1 control. (a) 2D PCoA plot with each node representing the MS results for the top 100 most significant metabolites from each group concatenated (142 metabolites total) for the lysogens from the adult_CF clinical group and the PAO1 control grown in ASM and ran in negative mode. (b) Hierarchical clustering result shown as dendrogram (distance measure using euclidean, and clustering algorithm using ward D), showing all samples of adult_CF lysogens and control compared. (c) Clustering result shown as heatmap (distance measure using euclidean, and clustering algorithm using ward D), representing the up and down regulation of each significant metabolite between the lysogens and the control. (d) The average score of the samples from the group and for the control were taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

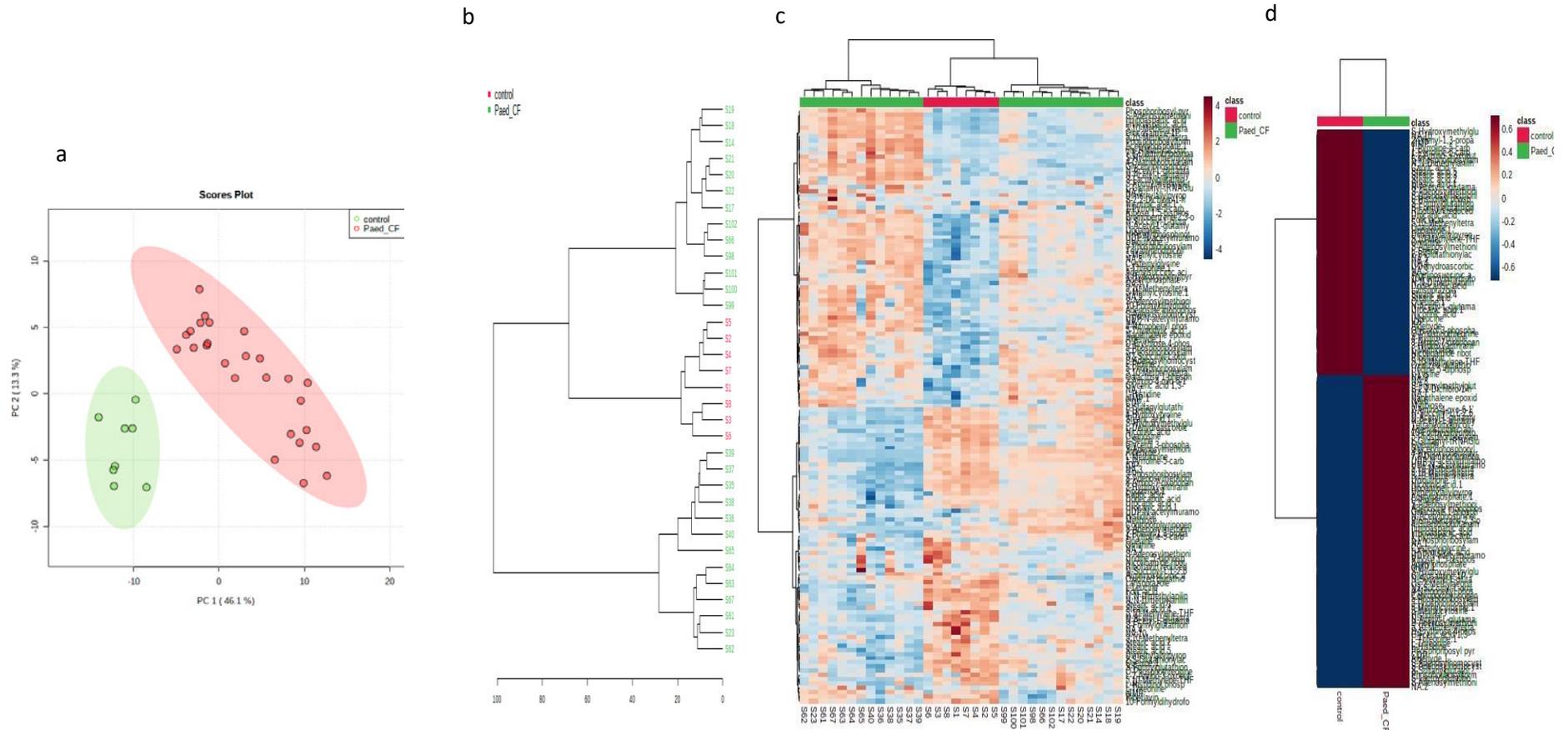


Figure 5.13 – PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from the paed_CF clinical group (lysogens containing prophages from Pa samples from adult CF patients) compared with the PAO1 control. (a) 2D PCoA plot with each node representing the MS results for the top 100 most significant metabolites from each group concatenated (142 metabolites total) for the lysogens from the paed_CF clinical group and the PAO1 control grown in ASM and ran in negative mode. (b) Hierarchical clustering result shown as dendrogram (distance measure using euclidean, and clustering algorithm using ward D), showing all samples of paed_CF lysogens and control compared. (c) Clustering result shown as heatmap (distance measure using euclidean, and clustering algorithm using ward D), representing the up and down regulation of each significant metabolite between the lysogens and the control. (d) The average score of the samples from the group and for the control were taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

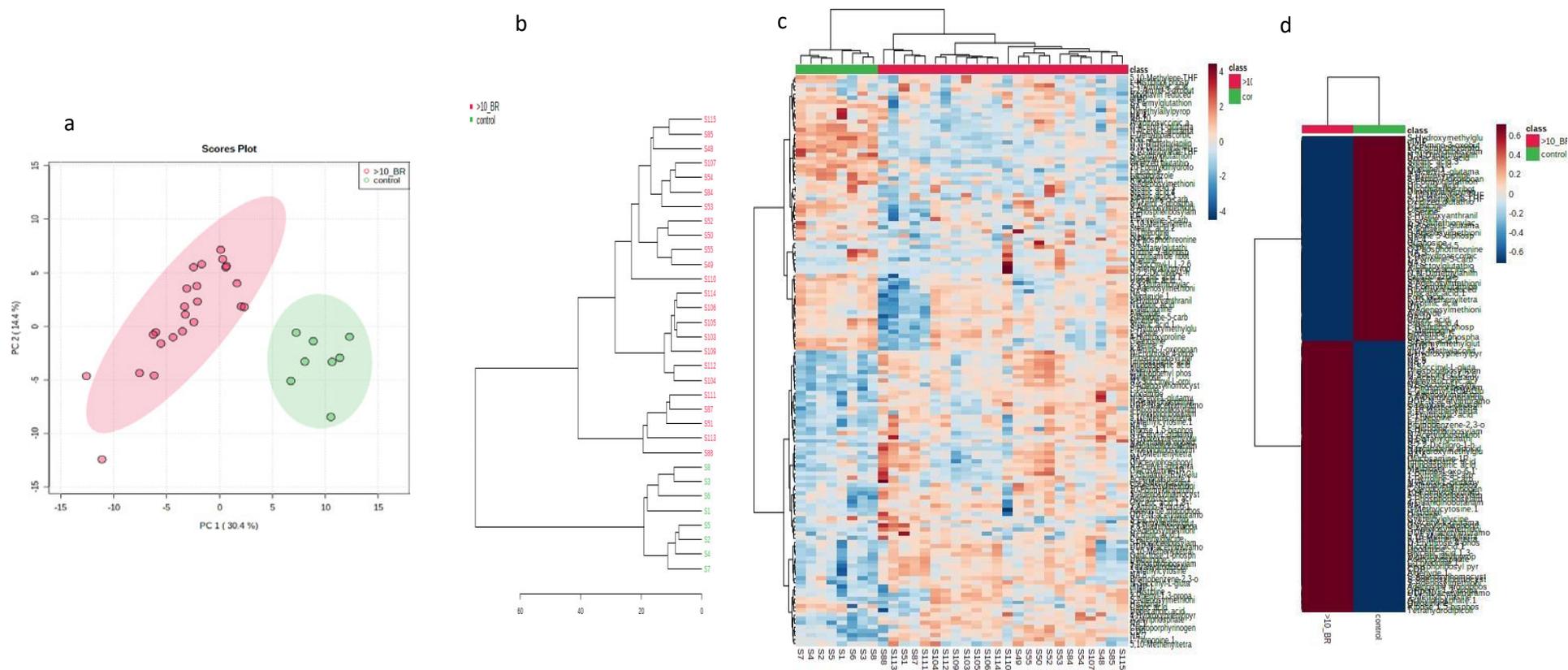


Figure 5.14 – PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from >10 BR_CF clinical group (lysogens containing prophages from Pa samples from bronchiectasis patients diagnosed >10 years prior) compared with the PAO1 control. (a) 2D PCoA plot with each node representing the MS results for the top 100 most significant metabolites from each group concatenated (142 metabolites total) for the lysogens from the >10_BR clinical group and the PAO1 control grown in ASM and ran in negative mode. (b) Hierarchical clustering result shown as dendrogram (distance measure using euclidean, and clustering algorithm using ward D), showing all samples of >10_BR lysogens and control compared. (c) Clustering result shown as heatmap (distance measure using euclidean, and clustering algorithm using ward D), representing the up and down regulation of each significant metabolite between the lysogens and the control. (d) The average score of the samples from the group and for the control were taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

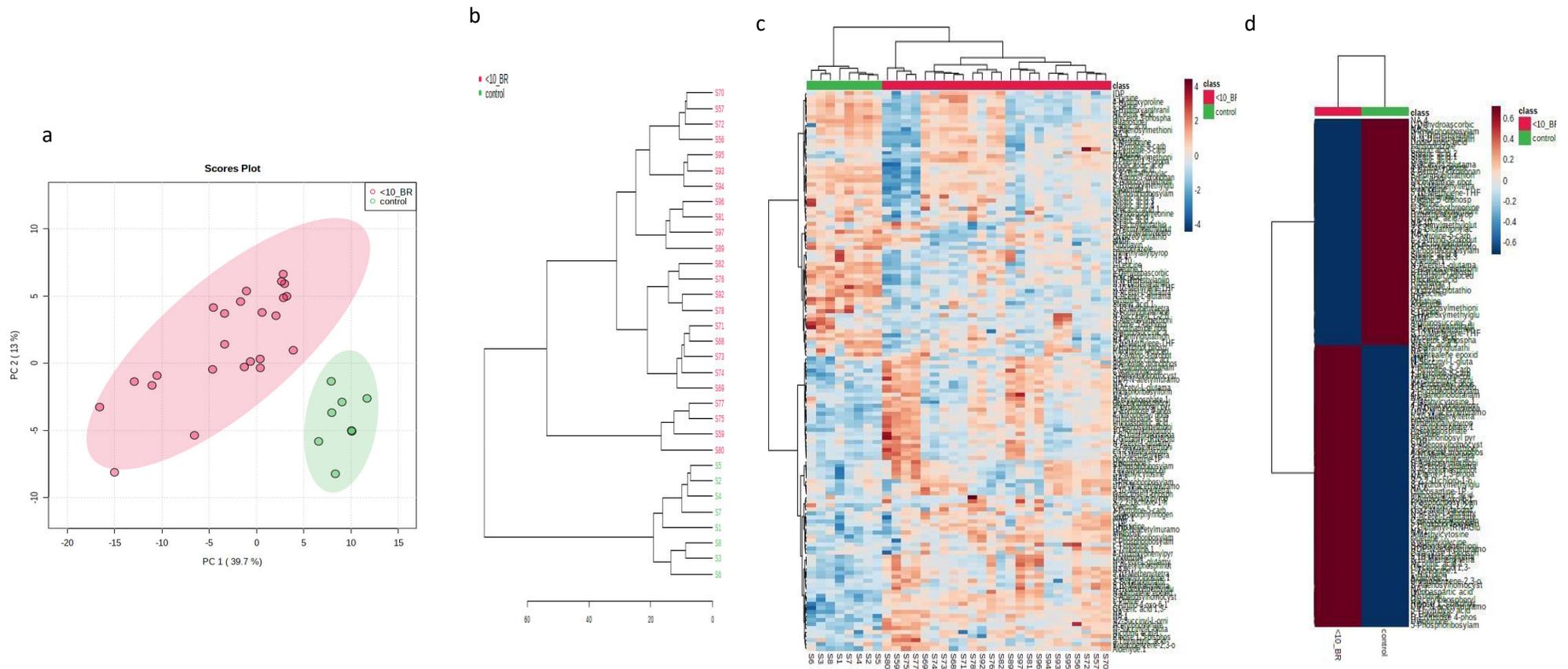


Figure 5.15 PCoA, dendrogram and heatmaps for the metabolic profiles of lysogens from <10 BR_CF clinical group (lysogens containing prophages from Pa samples from bronchiectasis patients diagnosed <10 years prior) compared with the PAO1 control. (a) 2D PCoA plot with each node representing the MS results for the top 100 most significant metabolites from each group concatenated (142 metabolites total) for the lysogens from the >10_BR clinical group and the PAO1 control grown in ASM and ran in negative mode. (b) Hierarchical clustering result shown as dendrogram (distance measure using euclidean, and clustering algorithm using ward D), showing all samples of >10_BR lysogens and control compared. (c) Clustering result shown as heatmap (distance measure using euclidean, and clustering algorithm using ward D), representing the up and down regulation of each significant metabolite between the lysogens and the control. (d) The average score of the samples from the group and for the control were taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. Full list of metabolites can be seen in a larger view in appendix E. The lysogen the sample/S numbers are associated to is detailed in appendix F.

5.4.2.1 Pathway analysis of CF and BR phage PAO1 lysogens differentiated into 4 groups, stratified by age or since date of diagnosis

The pathway analysis between the lysogens in each clinical group compared to PAO1 was performed using MetaboAnalyst, again using PAMDB, of which 92% were recognised by MetaboAnalyst for comparison. This was then reduced to metabolites with associated KEGG IDs (81% of total 142 metabolites). The impact scores for each pathway compared to the control were used to create a heatmap (Figure 5.16). These results show the phage mediated metabolic pathways that were changed significantly in the lysogen (p value <0.05) compared to the PAO1 control, with 26 pathways being significantly impacted by one or more of the lysogen groups compared to the control. Of particular interest was the high level of similarity in pathways between clinical groups (figure 5.16), which may be beginning to show a core impact of phage conversion by Pa phages in the lung environment. The pathway results from MetaboAnalyst only show if the pathway has been impacted (metabolites within that pathway have changed significantly up or down in concentration when the lysogen is compared to the control). Therefore, the metabolites within each pathway need to be referred back to the heat maps in figures 5.11-5.15 to determine if the metabolites are positively or negatively impacted by the addition of the phage. However, unless the majority of the metabolites present in the pathway are up or down regulated, it is difficult to determine if the pathway itself is up or down regulated. The six pathways that were impacted the most by the addition of these phages were: 'Novobiocin biosynthesis', 'One carbon pool by folate', 'Arginine and proline metabolism', 'Arginine biosynthesis', 'Glycerophospholipid metabolism' and 'Biotin metabolism'. The effects these pathway changes may have on the Pa host is discussed further in section 5.5.

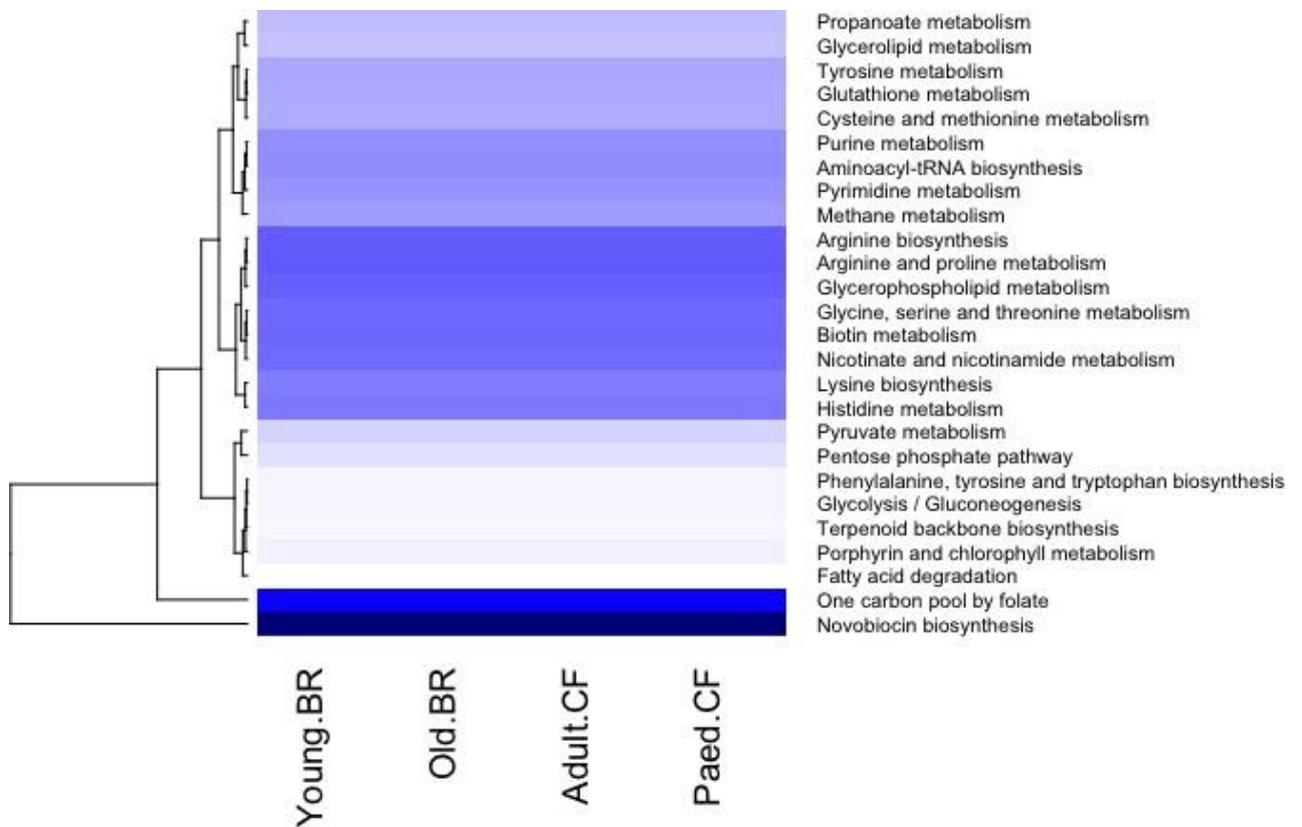


Figure 5.16- Pathway analysis of lysogens containing clinical *Pa* phages from four clinical groups. A heatmap showing the phage mediated bacterial host metabolic pathway 'impact' of clinical phage/s in each clinical group compared to the uninfected PAO1 control. Darker blue represents more significant changes to the pathway i.e. more metabolites are changed within a pathway or a significant metabolite to that pathway has been affected compared to the control. This data was collated from MetaboAnalyst pathway analysis tool on each lysogen.

5.4.3 Virulence testing of PAO1 lysogens containing the clinical *P.aeruginosa* phages from BR and CF patients

The twenty CF and BR phage lysogens of PAO1 from section 5.4.2 were taken forward and used to infect *Galleria* moth larvae (for method see section 2.8.1) to test the virulence in comparison to the PAO1 control (carried out by our collaborators at the University of Wroclaw with specific acknowledgement to Dr Tomasz Olszak). Figures 5.17-5.21 show the virulence results, demonstrated by larvae kill curves after being injected with an equal CFU of lysogen bacterial cells. These figures are then grouped into the four groups, with each figure having graphs a-e for the 5 different lysogens in the group. These virulence results showed the addition of temperate CF and BR phage PAO1 lysogens into a *Galleria* virulence model. Within 24 hours the wild-type PAO1, with no additional temperate phages, killed 100% of the *Galleria* larvae. However, all of the CF and BR lysogens of PAO1 had at least 10% surviving 24 hours after infection, with 18 out of 20 of the lysogens having 10% of larvae still surviving after 72 hours with all graphs showing a p value of 0.0488 or less. The two lysogens that had the lowest survival and therefore the phages that evoked the least reduction in virulence on their host were BR299 (< 10 year Bronchiectasis) and BR313 (>10 years Bronchiectasis) (figures 5.20c and 5.21d), where 100% larvae were dead 48 hours post-infection with both lysogens containing a F10-like phage (Tariq et al., 2019). The lysogen with the highest survival and therefore the phages evoking the highest reduction in virulence on its host was CF52 (Adult Cystic Fibrosis) (figure 5.9b), which had 50% of *Galleria* surviving 72 hours post-infection. There was a slightly higher average survival 72 hours after infection in the lysogens in the later stage CF group than the early CF stage group (but this was not statistically significant). The BR data showed the opposite result with the early-stage data showing slightly increased survival.

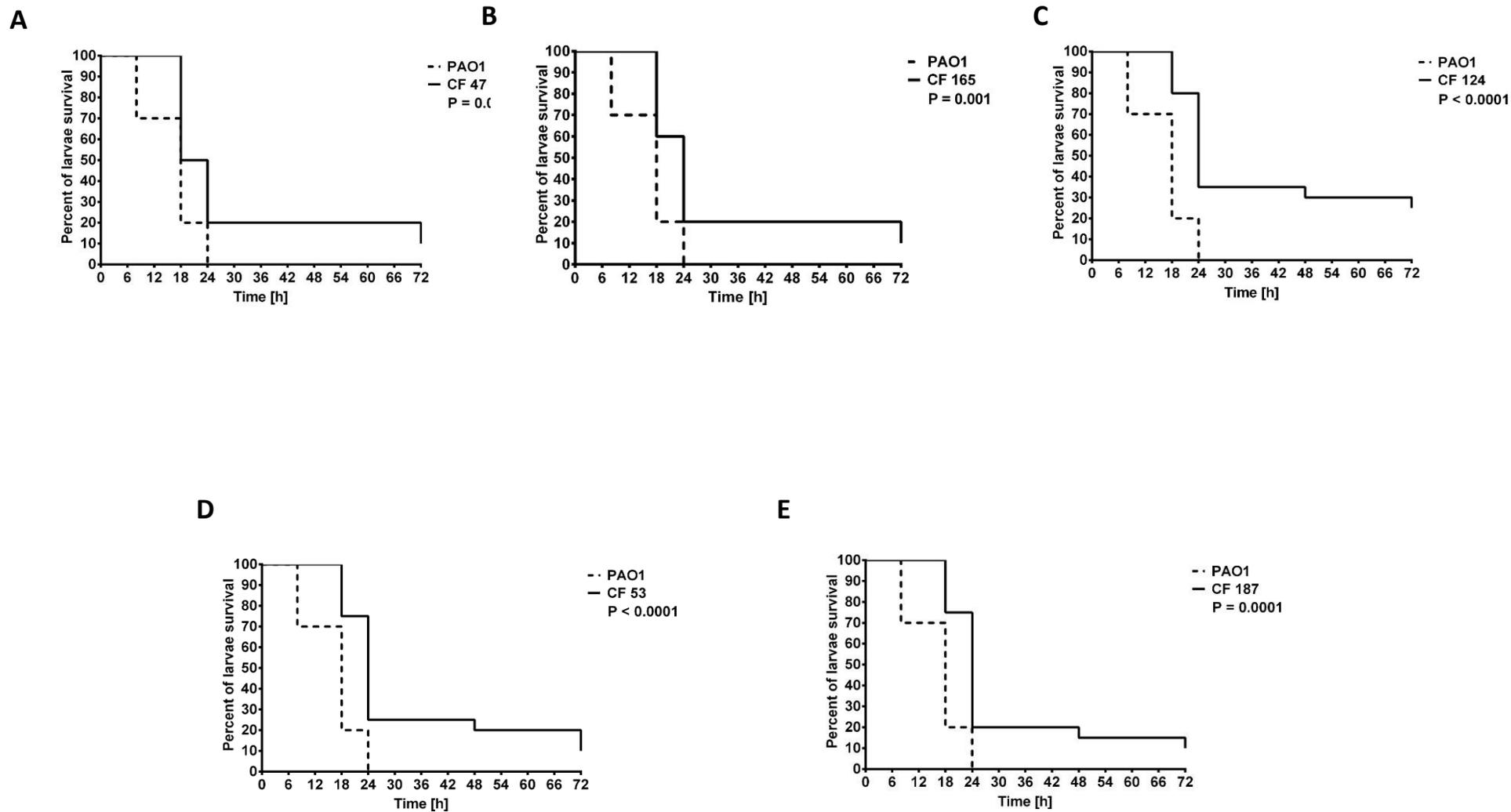


Figure 5.17a-e Survival curves Virulence results for *Galleria* model tests against 5 PAO1 lysogens clinical group - early stages of CF (aged under 16 years).

Each containing phages from Pa clinical isolates induced from Pa samples from patients with early stages of CF (aged under 16 years).

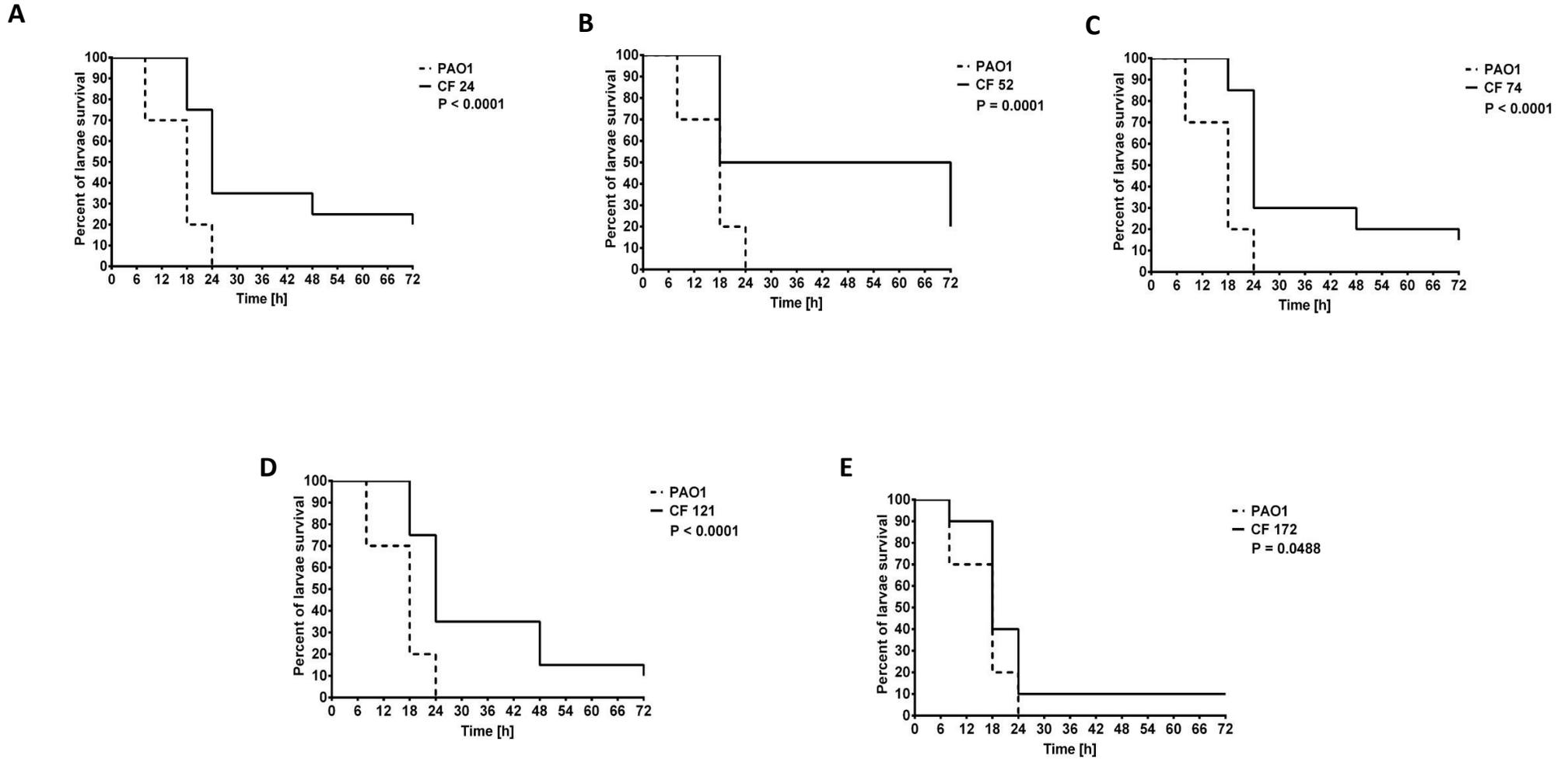


Figure 5.18a-e Virulence results for galleria model tests against 5 PAO1 lysogens clinical group - later stages of CF (aged over 16 years).
 Each containing phages from Pa clinical isolates induced from Pa samples from patients with later stages of CF (aged over 16 years).

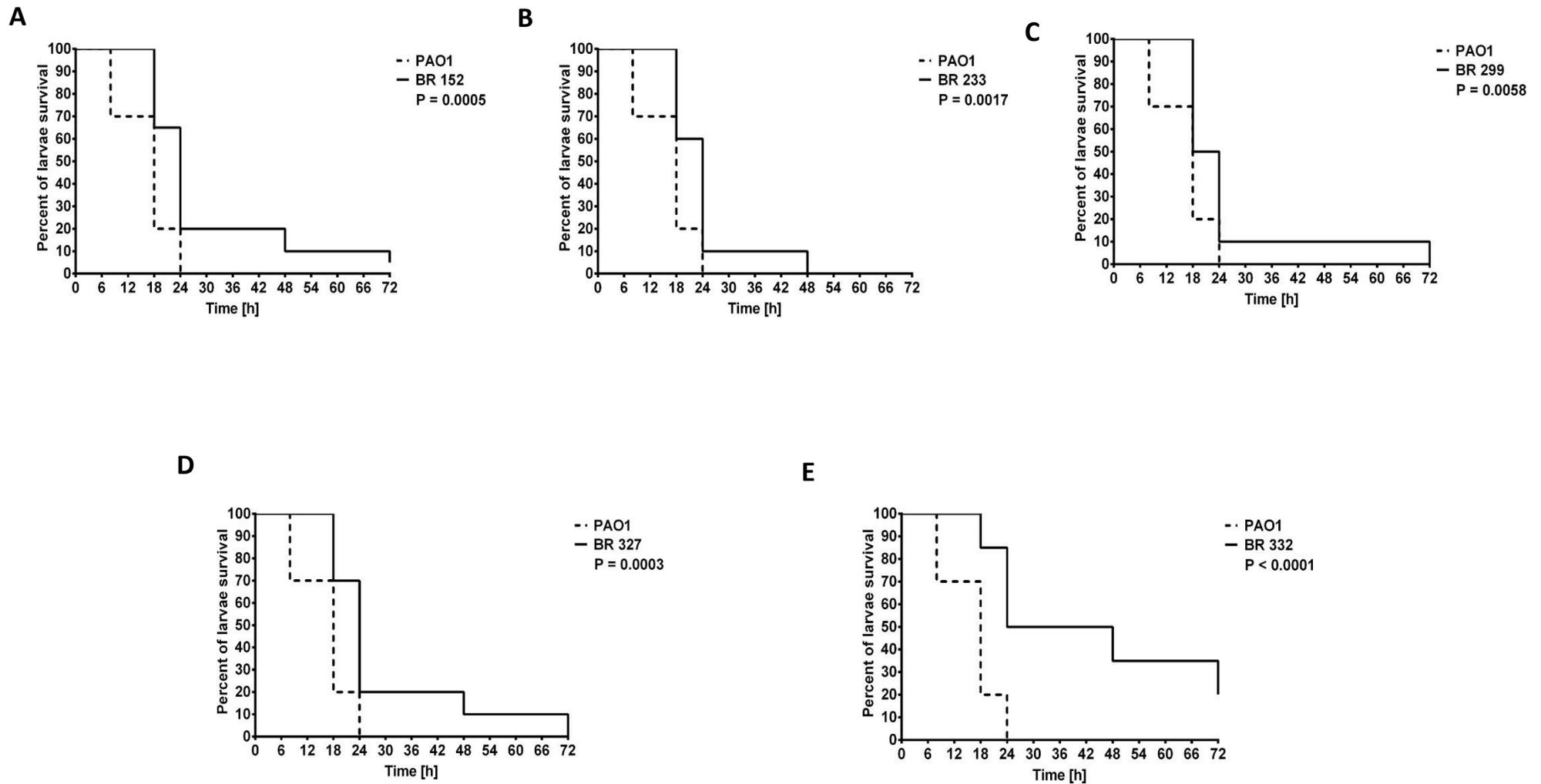


Figure 5.19a-e Virulence results for galleria model tests against 5 PAO1 lysogens clinical group - early stages of non-CF-BR (diagnosed >10 years before). Each containing phages from Pa clinical isolates induced from Pa samples from patients with early stages of non-CF-BR (diagnosed >10 years before).

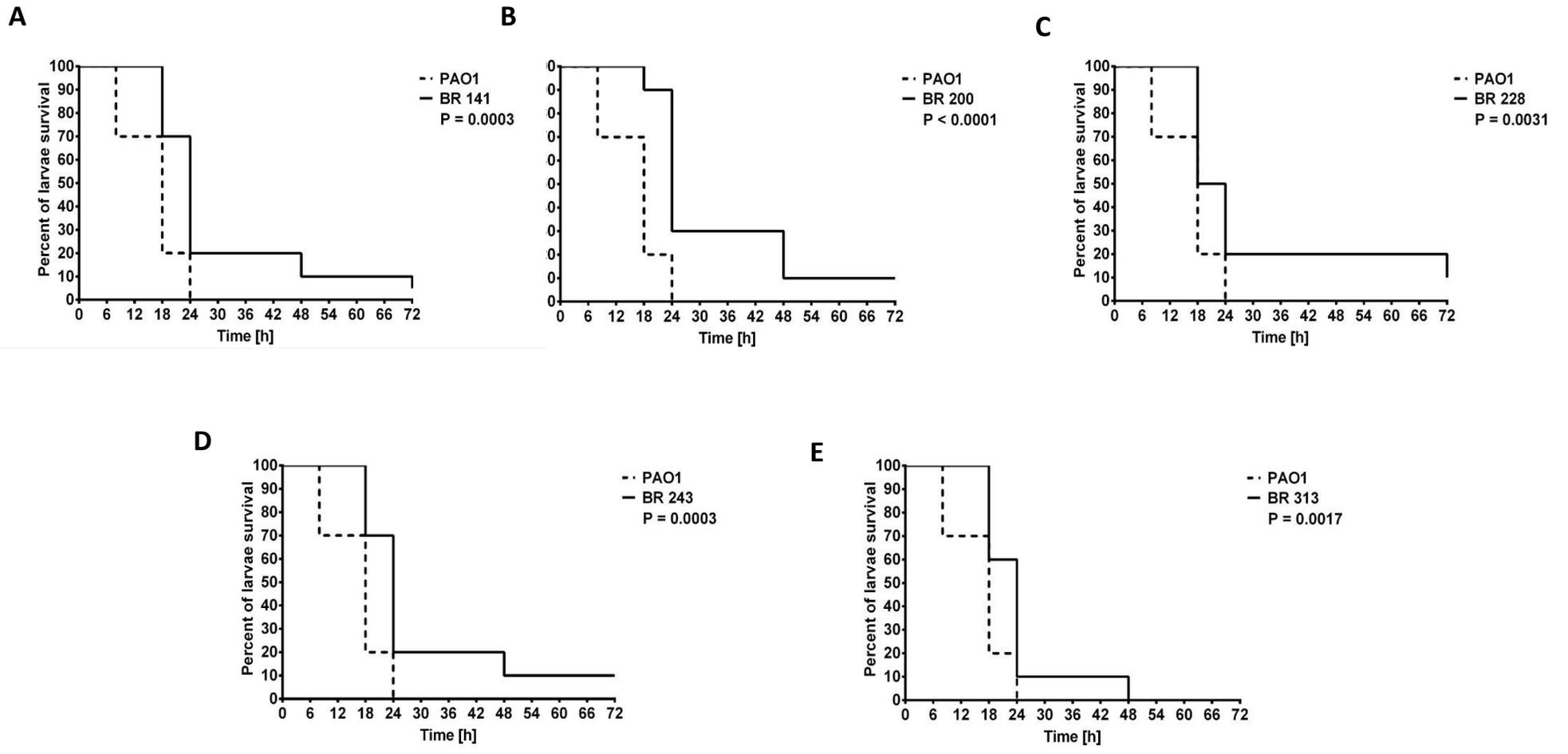


Figure 5.20a-e Virulence results for Galleria model tests against 5 PAO1 lysogens clinical group - late stages of non-CF-BR (diagnosed <10 years before). Each containing phages from Pa clinical isolates induced from Pa samples from patients with late stages of non-CF-BR (diagnosed <10 years before).

5.5 DISCUSSION

5.5.1 Subversion of the PAO1 metabolism caused by LES prophages

This study aimed to investigate the effect of different temperate phage carriage on the metabolism of the host Pa bacteria, and to compare the virulence of those lysogens to the wild-type control Pa without the addition of temperate phages. This would enable us to begin to unravel what the metabolic alterations caused by prophage carriage metabolically, and infer which functions have an impact on the virulence of the bacteria. This was undertaken using phages induced from two small cross-sectional studies of clinically relevant phages. Firstly, phages from the Liverpool Epidemic Strain (LES) of Pa and secondly, phages from a previous study undertaken by our research group focusing on CF and BR patient isolates from chronic respiratory disease.

The metabolomics results for the LES phages enabled conclusions to be made about how the LES phages differ from one another in terms of how they affect their host Pa separately and in combination. The LC-MS results identified 4768 features, demonstrating the sensitivity of the LC-MS to be able to identify such a high number of metabolites. From these metabolites 166 were shown to be significant, where the abundance of these metabolites is significantly altered compared to the other lysogen groups or the PAO1 wild-type. Therefore, 3.4% of the identified metabolites had alterations mediated by the LES temperate phages. This seems low in comparison to other studies looking at metabolic changes in phage infection from lytic phages, which showed 24.5% of the metabolites had been affected when the phage was added (De Smet et al., 2016). However, in that study only 918 metabolites were identified. In addition, it is expected that lytic phages would have a more significant effect on the bacterial host's metabolism than a lysogenic phage. It also illustrates the subtlety of the phage control between lytic and lysogenic infection. The hit and run lytic infection stimulates high rates of changes during infection, where the temperate phages stimulate a continual low level impact on the host cell. The features that were shown to be significant were then compared to the PAMDB to give putative compound names to each metabolite. By comparing to the PAMDB, it gave the most reliable result as it is only compared against

metabolites known to be found in Pa.

There are drawbacks to metabolomics work, the main one being the lack of specific metabolite databases to search against and other Pa bacterial metabolomes to compare to. While metabolomics in humans has been used since 1970 (Cunnick et al., 1972), recent improvements in the resolution of mass spectrometry has made this approach more available for bacterial metabolomics, especially in terms of the metabolic effect that phages have on their host bacteria. This study has given an insight into what is achievable, alongside current drawbacks to this approach. It has also given specific pathways for further experimental study, which may not have been thought to be phage-mediated or phage-controlled previously.

The statistical analysis of the Peak Integrity Table from the MS results displayed in the PCoA in figure 5.1 shows that there was a difference in metabolism between the 5 LES lysogens. This was then observed in further detail in the heatmap in figure 5.2. This demonstrated that the LES 3 and 4 separately, in combination with each other and in combination with LES2, impact the metabolites in similar ways. However, LES2 has a very different metabolic profile. These results are supported by previous studies investigating the LES phages that show how they have different infection properties (James et al., 2012) and therefore effect the hosts in different ways and also how the LES phages increase competitive fitness during lung infection at different rates (Davies et al., 2016a). By looking at the genomes of LES2 and LES3, they have a high level of heterogeneity with 82.2% shared identity (Winstanley et al., 2009). This suggests that there is only a small part of LES2's genome that is responsible for the metabolic changes it evokes on its host that are different to LES3. This also shows how similar phages can subvert the same host in different ways.

All LES? lysogens showed a significant change in 166 of the identified metabolites in comparison with the PAO1 control. The changes the phages are evoking could be advantageous to the bacteria, given the symbiotic relationship between the bacteria and the phage. By looking more closely into these metabolites it can explain how temperate phages may play a role in enhancing bacterial fitness and changing the virulence of the Pa host,

which has not been investigated previously.

To understand the Pa lysogen metabolome further, pathway analysis was performed (figure 5.9). The top pathways that are impacted by the addition of LES phages are listed in 5.4.1.1. The carbapenem biosynthesis pathway was highly impacted in all the lysogens, and by cross referencing to the heatmap in figure 5.3b, the metabolites associated with carbapenem pathway upregulation show that the pathway was upregulated by the addition of phages. Upregulating this pathway suggests that it would allow synthesis of carbapenems and therefore gives the ability to bacteria to tolerate or be resistant to carbapenem antibiotics. This would be beneficial to the prophage, as this would enable the Pa to survive an environment containing carbapenem antibiotics, which are commonly used to treat Pa infections. This hypothesis could be tested experimentally by assaying for resistance. Taurine and hypotaurine metabolism are also highly impacted by the addition of the LES phages. It is reported to be an antioxidant and membrane stabiliser in *Pseudomonas putida* (Bojanovič et al., 2017), which could also be its role in Pa. However, if down regulated it could destabilise the cell membrane. This may offer a selective advantage that links to cell surface differences such as immune detection, antibiotic tolerance or biofilm formation. D-Glutamine and D-glutamate pathway metabolism was seen to be upregulated in all the lysogens compared to the control. This was expected as this pathway is used to make new proteins, which is necessary to make phage proteins.

Biotin metabolism has been shown to be upregulated with the addition of prophages in *E.coli* (Holt et al., 2017) and it is shown to be upregulated here in all lysogens. Biotin metabolism is intrinsically linked to the fatty acid synthesis pathway (Lin et al., 2010) and, in the presence of antimicrobials, it may be that prophage carriage promotes broad-range antimicrobial tolerance by increasing cell wall lipids, which in turn alters membrane fluidity lowering ease of access to the cell wall by B-lactams or by blocking transport of intracellular targeting antibiotics. This could also be the case in Pa and would be beneficial to the bacteria and the prophage. Altered tolerance to clinically relevant antibiotics was seen at sub-inhibitory concentration in Pa by Tariq et al. (2019), although prophage carriage also offered increased sensitivity to some antibiotics.

The one carbon pool by folate pathway has been shown to support multiple physiological processes and was impacted significantly by the addition of LES phages . These include biosynthesis of purines, thymidine and amino acid homeostasis of glycine, serine, and methionine. This pathway is upregulated in all the lysogens due to purine and thymidine being necessary for the prophage, as they play key roles in the construction of DNA and RNA in phages, as well as bacteria. Pantothenate and CoA biosynthesis was upregulated in all the lysogens and is associated with production of fatty acids, like biotin metabolism (Leonardi and Jackowski, 2007). Pantothenate (vitamin B5) is the key precursor for the biosynthesis of coenzyme A (CoA). CoA is an essential cofactor for cell growth and is involved in many metabolic reactions, including the synthesis of phospholipids, synthesis and degradation of fatty acids, and the operation of the tricarboxylic acid cycle. This suggests again that the prophages can alter the membrane fluidity of their Pa host, meaning these prophages have highly impacted two pathways involved in fatty acids and lipids, which are characteristics that can be experimentally tested with further work.

If the pathways are looked at together, some comparisons can be made as some of the impacted pathways affect similar areas of the Pa metabolism. Biotin metabolism, peptidoglycan biosynthesis, pantothenate and CoA biosynthesis and glycerophospholipid metabolism and are all involved in fatty acid metabolism, which is associated to the cell membrane. With all these pathways being significantly impacted, it suggests that the cell membrane is affected by temperate phage integration. There have been studies showing prophages upregulating the biotin pathway in *E.coli*, which is rate limiting to cell growth, whilst also promoting a broad-range of antimicrobial tolerance by increasing cell wall lipids (Holt et al., 2017). The LES phages could have a similar effect on the biotin and other fatty acid associated pathways, as seen in *E.coli*. However, further experimental evidence would be required, such as carrying out a biotin assay and antibiotic testing, to indicate if there is antibiotic tolerance gained. If this is the case, it would confirm that temperate phages can play a significant role in antibiotic tolerance, without the need to carry antimicrobial resistance genes on their genome.

The pathway analysis using MetaboAnalyst has some shortcomings, as comparisons could

only be made to metabolic pathways that are within its database. Therefore, the metabolism of *Pseudomonas putida* was used as it represents the closest relative to Pa available. As *P. putida* is a plant pathogen and is found in soil, the selective pressures would be different to Pa and the metabolism might be quite different. Also, to enable matching to a pathway, only metabolites that were associated with a KEGG number were taken forward for the pathway analysis, as those not associated to a KEGG pathway were not recognised and therefore not considered. This narrowed down the number of metabolites but may have put some level of bias into the results from the pathway analysis, as only metabolites that were recognised could be used. It does illustrate with the reduction in data between lysogen and wild-type PAO1 that there is a lot more to unveil from this data.

5.5.2 Effects of temperate phage communities from Pa isolated from different aetiologies

The other results in this study were regarding the PAO1 lysogens carrying full complements of phages induced from Pa from four clinical groups at differing stages of disease progression. This meant that the metabolomics results were the effects of mainly polylysogeny, where the phages had evolved together in the respiratory tract, which may better illustrate both the multiple phage carriage and altered metabolism in synergy. Some phages, like Acr phages, give cooperative infection with one phage offering sensitivity to the second phage, possibly by blocking a phage resistance mechanism, such as CRISPR-Cas in the case of Acr phages (Bertani, 1971; Chevallereau et al., 2020,). The lysogens and the control were also cultured in ASM media (Sriramulu et al., 2005, Kirchner et al., 2012) to mimic the environment of the chronically infected lung. ASM has been used previously in Pa and phage research (Kirchner et al., 2012, Davies et al., 2016b, Davies et al., 2017) and in some recent metabolomic analysis of Pa biofilms (Depke et al., 2020). There has also been metabolomic analysis showing different metabolic shifts when Pa is grown as a biofilm (Gjersing et al., 2007), that occurs when grown in ASM. Using ASM makes the environment more similar to the clinical environment these phages were isolated from, in the hope that the phages would cause the same effects to PAO1 as they do in their original Pa host strain. The limitation of ASM compared to *in vivo* models is there is no fluctuation in the

components during infection, such as mucin levels in the CF lung (Hill et al., 2018). There is also no interaction with immune cells that would normally play a role during infection and may influence how the function of the Pa.

The LC-MS analysis was able to identify 2475 features, of which 546 were significant, which was higher than seen in the LES phages even though there were a lower total number of features identified. Due to the high numbers of significant metabolites, this was narrowed down further to observe the shift in metabolism between the different groupings of phages. This was demonstrated to be effective when comparing the PCoA plots in figures 5.10a and 5.11a where all of the significant features and the streamlined set of 142 are shown, respectively. The statistical analysis on the abundances of the metabolites from the MS results showed that the 4 groups of phages from differing aetiologies subvert the PAO1's metabolism in a similar way to each other (figure 5.11) and all groups significantly changed the metabolism in comparison to the PAO1 control. These results suggest there may be a core metabolic shift that is caused by the insertion of Pa phages from the lung environment, as that is the commonality between these sets of 20 phages. The phages were from different aetiologies but have all evolved in the chronic lung for different periods of time during disease progression. This also suggests that these core metabolites are changing, which may be giving the Pa an advantage within the lung.

For pathway analysis, only the metabolites that were associated to KEGG numbers were taken forward. As said previously, this adds a degree of bias, which may account for the similarity in the levels of impact between the lysogen groups shown in figure 5.16, which show that all lysogen groups affected each of the pathways at the same level. This also may be because the phages in each of the groups were similar and there was not a distinct difference between the groups of phages, or it could be because only the most significant metabolites were taken into consideration. However, these may be the most impacted pathways and could be the core pathways effected by these prophages and possibly other prophages associated with the chronically infected lung environment. The impact the prophages have on these pathways may be advantageous to the Pa host and may play a role into why it is able to cause a chronic infection.

The pathways that showed the greatest change are mentioned in section 5.4.2.1 (Figure 5.16) and are considered in detail here. The most changed pathway in all lysogen groups was the Novobiocin biosynthesis pathway, meaning that more novobiocin could be synthesised in the lysogens compared to the control, which has also been seen in *Streptomyces* (Steffensky et al., 2000). Novobiocin belongs to the amino coumarin antibiotics and targets bacterial DNA gyrase by interacting with the N-terminal 24-kDa subdomain of the gyrB subunit, inhibiting DNA gyrase (Pravin and Shahul, 2012). This is interesting as this could kill the bacteria if enough was made, which would not be beneficial to the prophage, or it could provide a competitive advantage to the cell producing it, if secreted. Therefore, the purpose of this may be to slow down the replication and growth rate of the bacteria, which would be beneficial if in low nutrient environments. It is also possible that this pathway is involved in catabolism of novobiocin rather than synthesis. Arginine and proline metabolism is highly impacted and upregulated in all the clinical lysogen groups. This pathway synthesises amino-acids from glutamate, and has shown to be upregulated in PAO1 by different growth conditions in a previous metabolomics study (Frimmersdorf et al., 2010). Amino-acid synthesis is also necessary for the prophage to build their own proteins. The arginine biosynthesis pathway, similar to the last pathway mentioned, leads to the production of arginine. It is synthesised from citrulline in arginine and proline metabolism by the sequential action of the cytosolic enzymes arginine succinate synthetase and arginine succinatelyase. This pathway is highly impacted in all the lysogen groups, as this pathway synthesises the amino-acid arginine, which is necessary for the prophages to build their own proteins. Glycerophospholipid metabolism is highly impacted in all the lysogen groups, in line with the upregulation of the biotin pathway. Both of these pathways are associated with fatty acid synthesis and homeostasis (Kondakova et al., 2015). Pa glycerophospholipids support structural functions and play important roles in bacterial adaptation to environmental modifications (Kondakova et al., 2015). Therefore, by prophages affecting the regulation of these pathways, it will affect the level of bacterial adaption, which has been shown to be increased in Pa by the addition of prophages (Davies et al., 2016b) and would be advantageous to both bacteria and prophage.

Purine and pyrimidine metabolism pathways are also impacted in all clinical lysogen groups as highly as the others mentioned. However, these pathways are of interest as phages have been shown to subvert purine and pyrimidine synthesis to aid viral construction and proliferation in lytic infection (De Smet et al., 2016, Friedman and Gots, 1953). It has also been shown through metagenomic analysis that well adapted prophages of Pa in the lung carry genes that are involved with purine, pyrimidine and different phosphate utilisation (Tariq et al., 2015), which was supported by another study that showed significant changes in RNA metabolism during bacteriophage infection of Pa (Chevallereau et al., 2016). This gives support to this metabolomics result and gives reason for the phages benefiting from the subversion of this pathway. To enable this data to be published, the second MS step must be completed or experimental evidence for some of these pathways has to be completed. However, this data gives target metabolites and pathways of interest to search and confirm.

If the that were seen to be impacted by both set of metabolomics data for the LES phages and for the clinical phage pathways are compared , there are common pathways that are seen to be significantly impacted in both data sets. This suggests that these pathways are associated with core changes resulting from the presence of these prophages in Pa. These would be the first pathways and associated metabolites to investigate when performing MS-2 (the second mass-spectrometry step to ID the metabolites of intererest by further fragmentation) the or by testing experimentally with associated functional assays such as a biotin assay.

5.5.3 Temperate phage effects of the virulence of Pa

The last part of this study was to look at the virulence of these same lysogens containing phages from clinical Pa samples from CF and BR patients at different levels of disease progression. These virulence results showed that the addition of temperate phages from clinical Pa isolates significantly lowers the virulence of the PAO1 in a *Galleria* model. Reductions in virulence due to a prophage has been reported before with Pf phages (Rice et al., 2009). Pf phages also alter the inflammatory response due to this loss of virulence (Secor et al., 2017), which is beneficial to the Pa host. This is the first time, to the

best of our knowledge, that a decrease in virulence is being associated to a much larger diversity of non-filamentous inducible phages, suggesting this characteristic evoked by prophage integration is more widespread than reported.

There was a slightly higher average survival 72 hours after infection in the lysogens in the later stage CF group than the early CF stage group. This may suggest that the phages induced from the later stage have evolved to make their host less virulent. However, this result was not significant and was not seen between the early and late stage BR results that showed the opposite outcome, with the early stage showing slightly increased survival. The number of bacteria harvested at time of larval death was similar and not statistically significant.

Interestingly, the pathway results showed the lysogens to impact the pathways at the same level, independent of the clinical group the phage was from. However, from the heatmaps in figure 5.11-5.14, variation between how the different lysogens within the same group affect each of the metabolites is shown. Here we see the difference in the virulence between each of these lysogens even with lysogens with phages from the same clinical group, as supported by the heatmaps in figures 5.11-5.14. This could mean that metabolites with a lower significance, which evoke smaller changes, have a larger impact on Pa virulence than first thought and all metabolites of significance should be looked at in more detail, rather than looking simply at the most significant metabolites. With more experimentation it would be possible to identify which metabolites are responsible for lowering the virulence of PAO1 and this would shed some light onto how the Pa may avoid the human immune system and cause Pa infections to become chronic.

5.6 SUMMARY

This chapter demonstrates that untargeted metabolomics can be used to identify the subversion caused by the addition of prophages into a common Pa host, by firstly looking at the addition of the well-defined LES phages to PAO1 singularly and in combination as polylysogens. This demonstrated that the different LES phages cause a significant metabolic shift in comparison to the control. The single LES2 phage also evokes a different

change in metabolism compared to LES3 and LES4. When combined with the other phages, the metabolism of LES2+3+4 becomes more similar to that of LES2. This was paired with pathway analysis of the significantly changed metabolites, which has highlighted pathways that are impacted significantly to provide a target for further MS-2 analysis or experimental confirmation of metabolite change. A number of pathways were impacted in all lysogens, suggesting they might be a common feature of Pa lysogenic infection and were associated with lipids and cell wall stability.

The metabolic effects on PAO1 from integration of full complements of phages from differing aetiologies, showed that they all had a core metabolic shift that was not dependent on aetiology. This showed that 20 different phage communities all subvert the Pa host in the same way, which is exciting as it shows there is evidence of a uniform change that takes place in Pa in the lung environment. If these metabolites and associated pathways (that showed the same result) can be confirmed, it would allow greater understanding into the role of prophages in chronic Pa lung infection.

The virulence of some lysogens, as mentioned above, containing clinical phage communities, were tested using the *Galleria* model, which showed that all the lysogens significantly reduced the virulence of PAO1, suggesting that possibly some of the core metabolites that were subverted are the cause of this reduction in virulence.

6 GENOME COMPARISON OF TEMPERATE PHAGES FROM THE INTERNATIONAL *PSEUDOMONAS* CONSORTIUM DATABASE.

6.1 INTRODUCTION

There has been a sizeable amount of work directed at the prophages of the Liverpool epidemic strain (LES) of Pa, due to its transmissibility and infectivity (Winstanley et al., 2009, Fothergill et al., 2011, James et al., 2015, Davies et al., 2016a, Davies et al., 2016c). Yet there is little known from larger genomic datasets available whether these bacteriophages are a good representation of temperate phages in Pa or are specifically lung associated Pa temperate bacteriophages. The aim of the research in this chapter was to take the approaches developed in chapters 3, 4 and 5 to investigate a larger repository of Pa genomes focusing on prophage diversity. In collaboration with Professor Roger Levesque, the lead investigator of the International *Pseudomonas* Consortium Database (IPCD), a repository of >1500 Pa genomes that have been sequenced by Illumina sequencing. This is a well-curated repository, most isolates have been sequenced and assembled, usually to under 100 contigs per genome (Freschi et al., 2015b). It is important to note that even though most of the repository is built from international isolates linked to respiratory infection (mainly CF), it also contains other clinical, environmental and animal samples. This offers a greater opportunity to compare lung related phage carriage and diversity between environments.

6.1.1 The International *Pseudomonas* Consortium Database – IPCD

The IPCD is the largest collection of Pa isolates and genomes sequenced to date, held at the University of Laval. The IPCD was started in 2014 by an international community of research scientists from 31 institutions globally, offering their *Pseudomonas* isolates to the collection (Freschi et al., 2015b). The aim of the consortium was to create a panel to represent the maximal genomic diversity of Pa. The IPCD currently (at the time of writing August 2020) contains 1763 Pa isolates, spanning as period of 125 years, covering 35 countries and 5 continents. The genomes of 1165 of the panel have been sequenced,

annotated and are well curated. The IPCD differs from other Pa repositories (Pirnay et al., 2002, Wiehlmann et al., 2007) as it is not solely focused on clinical strains. However, within the panel there is still a strong emphasis on Pa from CF lung infections. The IPCD has been used for a Pa pan-genome study using the whole database looking into how the IPCD may provide new insights on population structure, horizontal gene transfer, and pathogenicity of Pa (Freschi et al., 2019). It has also been used to make comparisons between Pa isolated from different geographical locations (Subedi et al., 2018). It is possible to request access to isolates from a specific source or location to study, for example; Pa isolated from companion animals, to study shared reservoirs of resistance (Scott et al., 2019). However, there has not been any studies into surveying the prophages within these Pa isolates. The range of isolates enables a wider cross-sectional examination of global temperate phage diversity and is independent of isolation environment.

6.1.1.1 Why are pan-genome studies of *Pseudomonas aeruginosa* important?

Research on the genetic diversity of microbial populations is vital to the understanding of the evolution, ecology and epidemiology of infectious diseases caused by a particular bacterium. However, most studies on bacterial chromosomal diversity are usually focused on strains of bacteria that cause human infections (Wiehlmann et al., 2007; Pirnay et al., 2009). This causes only a partial view of the diversity of a bacterial species such as Pa, as well as the distribution of antibiotic resistance genes and mobile genetic elements (MGEs) (Perry and Wright, 2013) by not also looking at environmental strains. Members of the genus *Pseudomonas* colonise a variety of environments, shown in their adaptable metabolic capacity and ability to manage fluctuating environmental conditions (Palleroni, 1992). The metabolic adaptability of Pa aids bacterial survival in clinical environments, such as the CF lung, but also a variety of environmental niches such as sea water (Silby et al., 2011). This shows the importance in examining the genomes not just of clinically relevant Pa isolates but environmental isolates, as this may give an insight into the evolutionary history of Pa.

The first pan-genome study using IPCD identified 5 phylogenetic groups based on the core genome (Freschi et al., 2019). Comparisons showed that variation among isolates was partly linked to this population structure. They also detected a total of 3,010 plasmids (putative complete and fragmented), some of which contained resistance or virulence genes (Freschi et al., 2019). Thus, showing that HGT may play an important role in the evolution and virulence of Pa and confirms the importance of research into MGE, such as phages in Pa.

6.1.2 The carriage of virulence factors by phages

Phages are responsible for transduction of a wide range of different virulence factors into a bacterial cell. These virulence factors can include bacterial toxins, such as the Stx toxin in *E.coli* (Strauch et al., 2008), and cell surface proteins that mediate bacterial attachment, illustrated in *Streptococcus mitis*, where phage encoded virulence factor aids bacterial platelet adherence in bacteremia infection (Bensing et al., 2001). Virulence factors that are carried on the phage genome can be accrued by mispackaging or through recombination within the cell (Frobisher, 1927). Some phages encode regulatory factors that show more subtle changes in the metabolism of the bacteria that leads to a fitness advantage. A good example of this is pckA gene regulation in *E.coli* required for gluconeogenesis, which is regulated directly by the principal repressor of many different temperate prophages, the cI protein. cI binds to the regulatory region of pckA, thereby shutting down pckA transcription. Thus, down-regulation of the host pckA pathway, which in turn increases lysogen fitness by lowering the growth rate of lysogens in energy-poor environments (Chen et al., 2005). They can also increase expression of virulence genes not encoded by the phage. For example, temperate phage have been shown to increase the expression of protein M in group A *Streptococci*, a protein present on the bacterial genome that when promoted by the phage allows the bacteria to be anti-phagocytic (Spanier and Cleary, 1980). In section 1.5.5 there are more details into the virulence factors carried by Pa and their effects. These more subtle regulatory changes are harder to notice than a toxin gene on the genome of the prophages. Therefore, these cannot be observed solely from the genetics of the phage and may be

overlooked. They may also change depending on the strain the phage infects. This is where metabolomics is useful to see the effects of prophages on the metabolism of the bacteria, giving a better insight than the genetics alone (this is covered in chapter 5).

6.2 AIMS

- Identify temperate phages from the sequenced Pa genomes in the IPCD, the largest data-mining assessment of temperate phages in Pa to date.
- Search determined phage genomes against the VFDB (virulence factor database).
- Compare the genomes of all the intact phages to determine the genotype and compare back to the aetiological data.
- Compare all the intact phage genomes to identify any prevalent phage types in Pa isolates from the IPCD.

6.3 OBJECTIVES

- Use PHASTER to identify each phage class (intact, questionable and incomplete) prophages within the IPCD Pa genomes.
- Use the VFDB to determine virulence factors or genes that may be advantageous to the Pa host within the intact prophage genomes identified.
- Compare the intact phage genomes using SaturnV (which uses protein-protein comparison in an 'all versus all' manner) to group similar prophages carried by Pa isolates from the IPCD and determine the diversity of temperate phages in Pa.

6.4 RESULTS

6.4.1 Prophage identification from *P. aeruginosa* genomes in the IPCD

As part of the study we completed the PHASTER search on 1030 Pa genomes. A total of 6499 phages were identified (see section 2.11.1 for method), 1854 intact (putatively mobile phages), 1358 questionable prophages and 3287 incomplete prophages (figure 6.1). Within the panel of bacteria there is a bias towards CF lung infection and other human

infection such as burns, UTIs and non-CF lung infections. The distribution and average number of phages per isolate was calculated (figure 6.2). Environmentally isolates of Pa carried the highest number of phages for all classes (intact, questionable, and incomplete), with a slightly higher number of intact prophages per isolate than those derived from human and animal hosts. However, even though the isolates with the highest number of overall phages (28 phages) was from an environmental isolate, the Pa isolate with the highest number of intact phages carried 8 and was isolated from a human host. A total of 175 of the 1030 Pa isolates (17%) had no intact phages identified and there was one isolate that had no phages at all. This isolate was from a soil sample in Canada and could be used for further study regarding restriction against phage DNA. However, of the 175 isolates that had no intact phages, 166 (95%) had one or more questionable phages identified which could also be putatively mobile

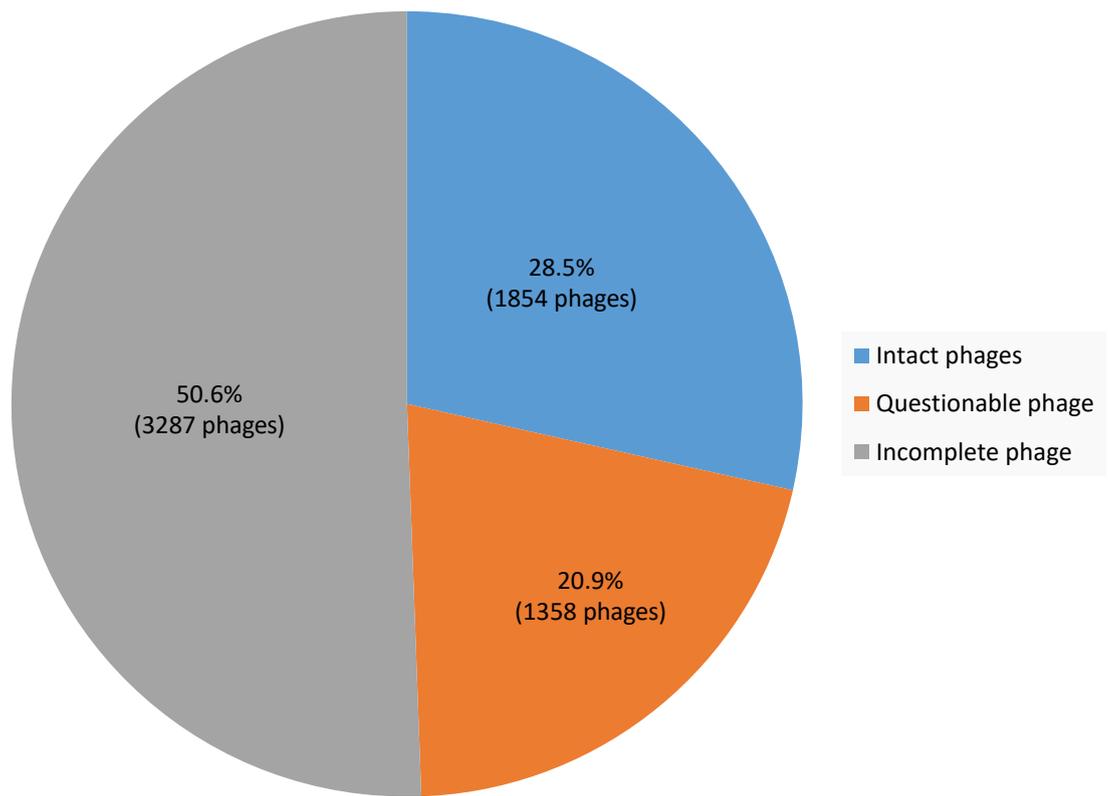


Figure 6.1 Distribution of prophages in Pa from IPCD identified by PHASTER (n=1030 strains). A total of 6499 phages were identified using PHASTER and classed as intact, questionable or incomplete.

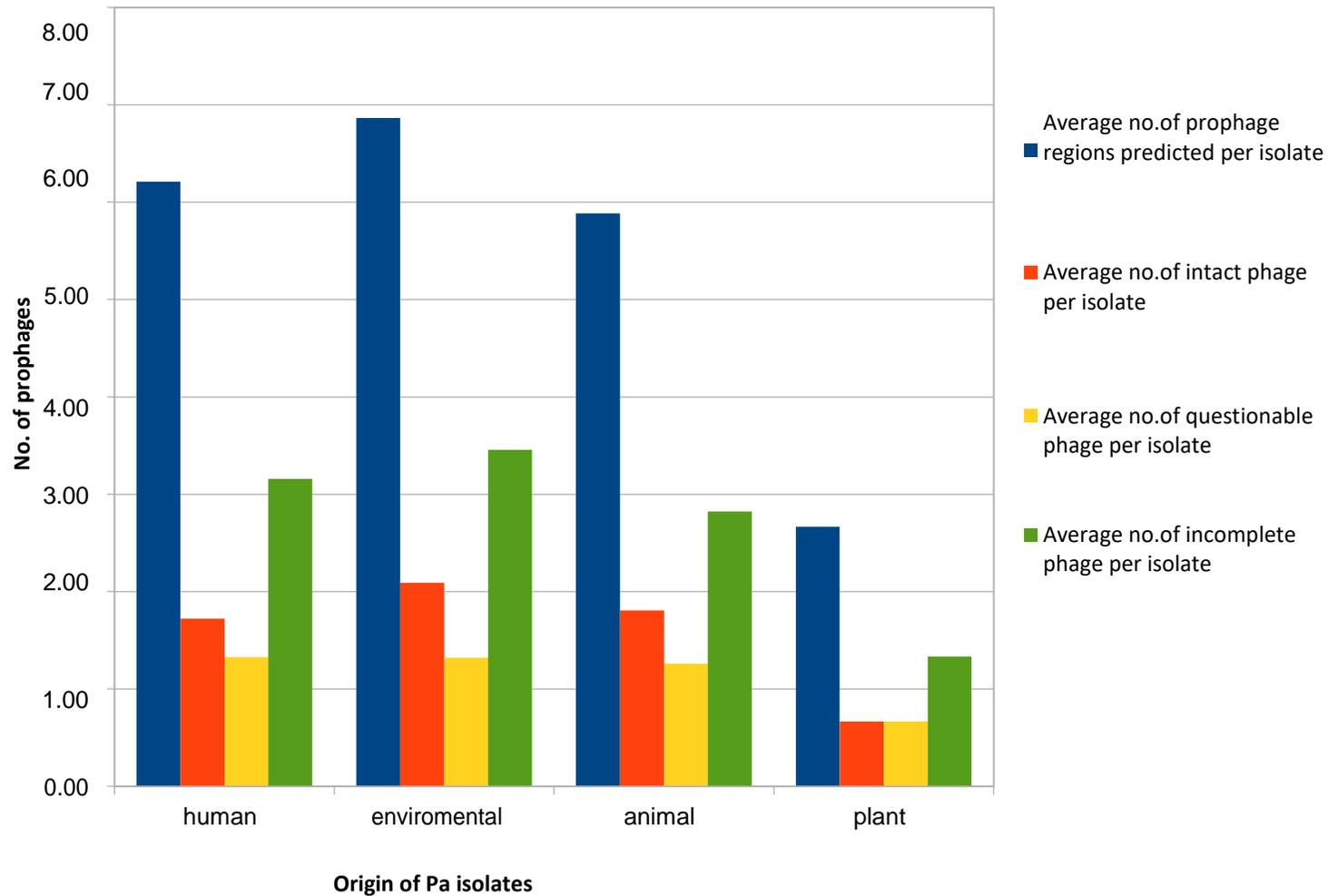


Figure 6.2 -The average number of prophages from each class. The average number of prophages in all classes identified Pa from each origin the Pa were isolated from (blue). Intact prophages (orange), questionable (yellow), and incomplete (green) per isolate depending on their origin. Data regrouped depending on the environment of isolation, whether human, environmental, animal or plant derived.

6.4.1.1 Distribution of putative virulence genes within the IPCD prophages panel.

Each phage genome region identified in the IPCD panel was compared in a pairwise manner to the virulence factor database (VFDB). From the 6499 prophages detected using PHASTER from the IPCD panel there were 2100 hits against the VFDB, ranging from 23-100% identity. Within the intact (putatively mobile) phages there were 524 hits, with 23 exhibiting 100% identity to virulence genes within the VFDB. Interestingly, only the intact phages had regions of 100% nucleotide sequence similarity to the database derived virulence genes. It was only these genes that were taken forward for further study as they are putatively mobile and share 100% sequence similarity. If these virulence genes were found in remnant phages, at high conservation, this would suggest that it is important for these positive environmental selection. The 23 phages that were identified as carrying these putative virulence gene regions matching the search criteria can be seen in in table 6.1. Of the 23 virulence genes identified with 100% identity and carried on phages that are deemed to be mobilisable, only 7 different virulence factor genes were found. The most common was *rpoN*, an RNA polymerase sigma-54 factor that was distributed between 10 of the 23 phages (however, these 10 phages are the same phage found in 10 different isolates). The 23 prophages were found in *Pa* isolated from a range of environments, including lung infections (n=7) and sea water (n=6). Some of the isolates, however, are the same phage and therefore the same virulence gene, due to some of the samples being taken from the same patient, such as VF_P10, VF_P11 and VF_P12, isolated from *Pa* from a UTI. This leads to a degree of bias as it appears there are more phages carrying virulence genes, but the phages may be duplicated in the cohort. Interestingly, there were phages that were the same (have >95% nucleotide sequence similarity) and were found within *Pa* genomes of isolates that were collected from different origins in different countries, which makes these phages very interesting as genome conservation may be linked to selective pressure. These 10 phages are identified in clade 6 and all possess a genome size of ~34kb, made more interesting by carriage of the RNA polymerase sigma factor virulence gene (*rpoS*). This also shows that

within the 6499 identified prophages there are phages that have sequence conservation, possibly caused by environmental selection pressures.

Table 6.1 – Prophage carriage of virulence genes identified with genome comparison to the VFDB.

Phage ID	Pa isolate ID	Isolation Site	Region length	% identity	Phage clade	Virulence gene name
VF_P1	331	Environment (CF patient home)	50.8Kb	100	Clade 4	(<i>rpoS</i>) RNA polymerase sigma factor <i>RpoS</i> [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P2	390	Bacteraemia	48.1Kb	100	Clade 4	(<i>rpoS</i>) RNA polymerase sigma factor <i>RpoS</i> [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P3	30	CF sputum	43Kb	100	Clade 5	(<i>waaA</i>) lipopolysaccharide core biosynthesis protein WaaP [LPS (VF0085)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P4	748	Sea water	43.5Kb	100	Clade 4	(<i>pvdE</i>) pyoverdine biosynthesis protein PvdE [Pyoverdine (VF0094)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P5	748	Sea water	43.5Kb	100	Clade 4	(<i>pvdF</i>) pyoverdine synthetase F [pyoverdine (IA001)] [<i>Pseudomonas aeruginosa</i> PAO1]

VF_P6	747	Sea water	43.5Kb	100	Clade 5	(pvdE) pyoverdine biosynthesis protein PvdE [Pyoverdine (VF0094)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P7	747	Sea water	43.5Kb	100	Clade 5	(pvdF) pyoverdine synthetase F [pyoverdine (IA001)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P8	755	Sea water	43.2Kb	100	Clade 5	(stp1) serine/threonine phosphoprotein phosphatase Stp1 [HSI-2 (SS179)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P9	753	Sea water	42.9Kb	100	Clade 5	(stp1) serine/threonine phosphoprotein phosphatase Stp1 [HSI-2 (SS179)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P10	1189	UTI	41.7Kb	100	Clade 5	(cupE2) hypothetical protein [CupE fimbriae (AI449)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P11	1190	UTI	41.7Kb	100	Clade 5	(cupE2) hypothetical protein [CupE fimbriae (AI449)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P12	1191	UTI	41.7Kb	100	Clade 5	(cupE2) hypothetical protein [CupE fimbriae (AI449)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P13	1269	Acute Infection	39Kb	100	Clade 4	(<i>rpoS</i>) RNA polymerase sigma factor <i>RpoS</i> [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]

VF_P14	669	Catheter	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P15	644	Non-CF Sputum	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P16	642	Non-CF Sputum	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P17	651	Non-CF Sputum	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P18	640	Non-CF Sputum	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P19	710	Non-CF Sputum	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P20	718	Blood	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P21	1053	Urine	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]

VF_P22	664	Hospital environment	34Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]
VF_P23	1377	CF throat Swap	34.6Kb	100	Clade 6	(rpoN) RNA polymerase factor sigma-54 [type IV pili (AI097)] [<i>Pseudomonas aeruginosa</i> PAO1]

6.4.1.2 Transmission electron microscopy (TEM) images of induced phages

To enable downstream study of whether the phages carrying these putative virulence gene regions could be mobilised from their bacterial host, attempts were made to induce the phages using norfloxacin (NFLX) at the University of Laval, one of the leading Universities in the world for imaging bacterial viruses and the University of the late Hans Ackermann. TEM was used to visualise phages present in the lysates from the induction. Figures 6.3a-o show TEM images that were taken of phages from phage lysates that showed a positive spot assay. 15 out of the 23 showed a positive spot assay, showing that there were active phages present in the lysate. The number below each TEM image corresponds to the IPCD ID for the isolate the phage was induced from. The phage/s in the images can not be confirmed to carry the virulence genes as there were multiple phages in each of the genomes that could be induced. All the images show phages in the lysate and fall within either the Myoviridae or Siphoviridae phage families in morphology and presence of putative contractile/non-contractile tails, respectively. From TEM imaging it is not possible to know whether these phages are carrying the virulence genes. These phages were not purified or sequenced to confirm the virulence gene due to time restrictions at the University of Laval. However, this could be done as further study as the phage DNA was extracted from each lysate, which could be used to confirm if the phages carrying the virulence genes were present in the lysate.

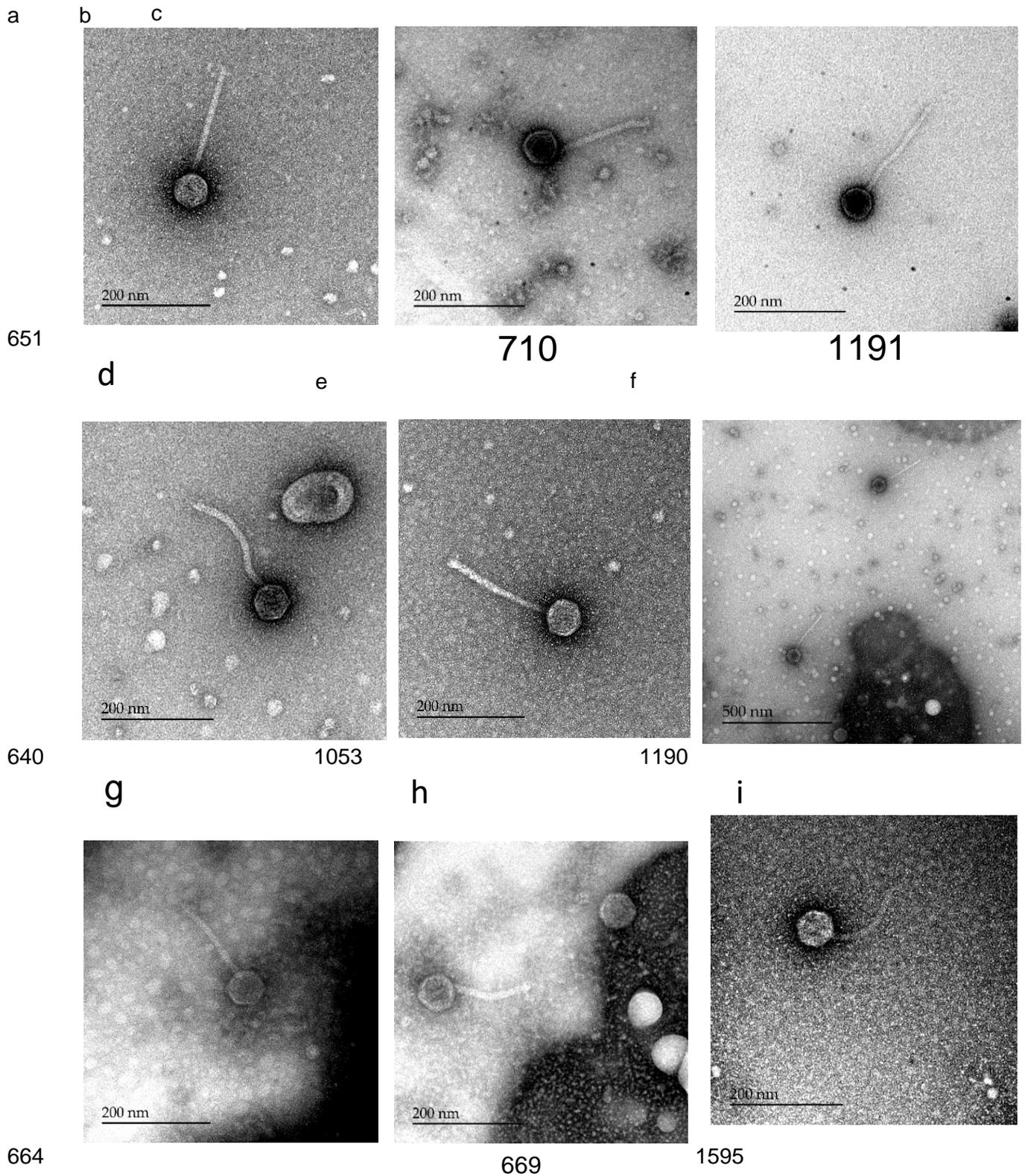


Figure 6.3a-i – Crude transmission electron microscopy (TEM) images of induced phages from ICPD Pa isolates stained with 2% v/v uranyl acetate. They contained phages that carry virulence genes. The Pa isolate ID in table 6.1 corresponds to the number under the image showing which isolate the phages were induced from.

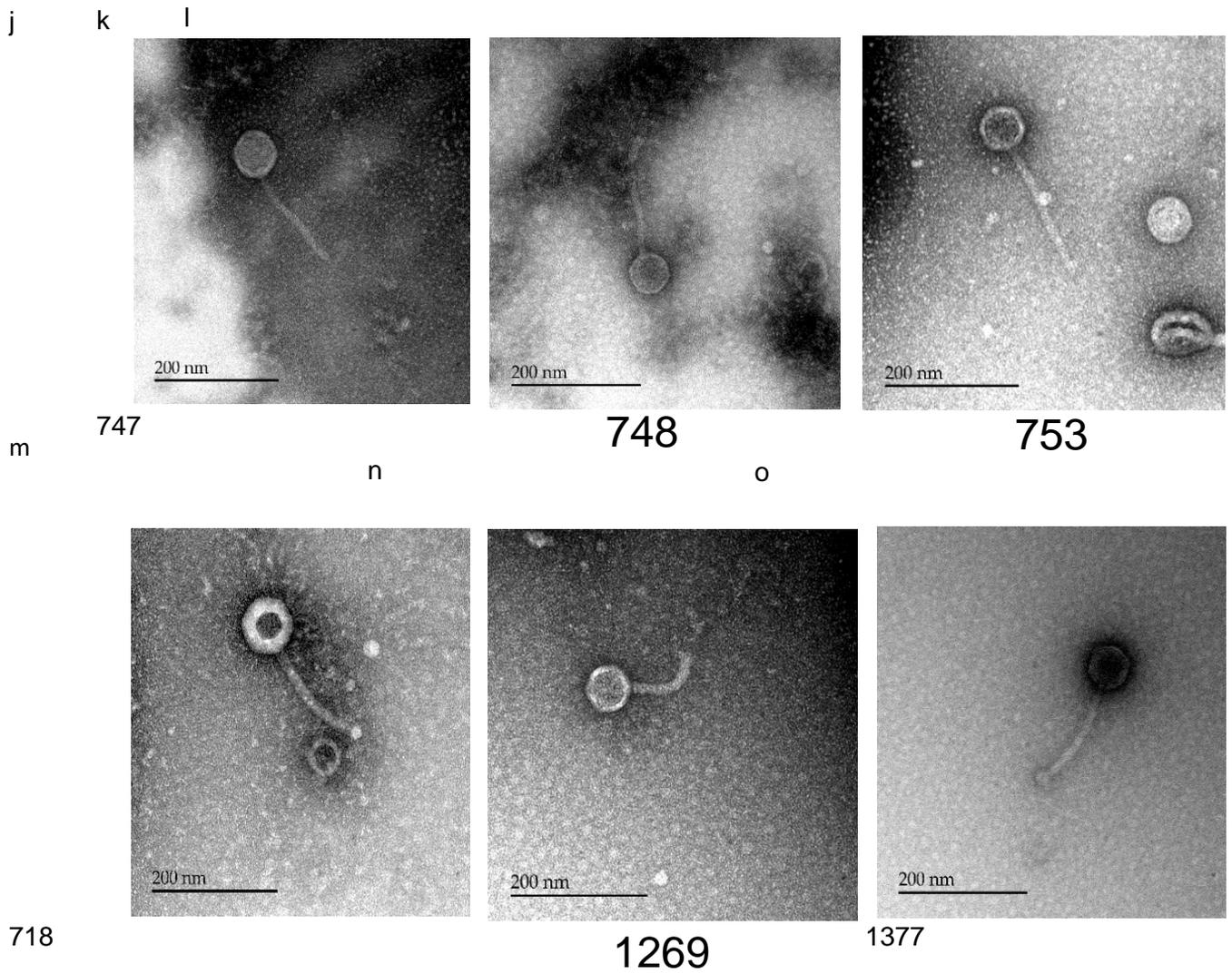


Figure 6.3j-o – Transmission electron microscopy images of induced phages from *Pa* isolates that are from the IPCD stained with 2% v/v uranyl acetate. They contained phages that carry virulence genes. The *Pa* isolate ID in table 6.1 corresponds to the number under the image showing which isolate the phages were induced from.

6.4.2 SaturnV diversity analysis of *Pseudomonas aeruginosa* prophage

As we delve deeper into differential carriage of genes by inducible prophages that may offer selective function for their receptive bacterial host, sometimes identifying this function at the nucleotide level can be difficult. This is because temperate bacteriophages are modular in genome organisation, yet their genes encoding similar function can be greatly heterogeneous (Smith et al. 2012). We therefore utilised the bioinformatics tool SaturnV (Freschi et al., 2015a). This tool translates each prophage's putative coding region to an amino acid sequence, comparing between amino acid sequences in the cohort in a pairwise manner, to cluster based on carriage of these phage gene encoded proteins between prophages. As these phages have been isolated from bacteria isolated globally, amino acid similarity and conservation in carriage between phages is important as it begins to determine important genes, either within the life cycle of the bacteriophage or that link to evolution within a specific environment.

The dendrogram derived from the SaturnV comparison for all the intact phages from the IPCD is shown in figure 6.4. It illustrates that these phages cluster into 8 main and identifiable clades with only a few outliers. These phages are grouped determined by protein similarity, by using BLAST to compare all the phage genomes in each clade against the NCBI virus database, giving the most common phage in the database to be the phages in each clade. This suggests that each clade of phages has specific and conserved proteins that differ from one another through their clustering. However, since analysing this data and going back to the dendrogram there appears to be many smaller clades within the larger ones suggesting even more diversity within genomes of Pa prophages. Prior to publication this data will be re-analysed to give a more accurate representation on the number of clades present to show the considerable amount of genomic diversity of prophage of Pa, which would be the largest and most diverse analysis of Pa prophages.

Figure 6.4 also illustrates the isolation source of the Pa isolate that the prophage was identified in by the colour of the node on the tree. There are many more phages from Pa

isolates from humans due to the IPCD originally being a collection of these respiratory infections. Therefore, there is a bias to Pa isolated from lung infections. From the tree in figure 6.4 we can see that phages from Pa from different sources tend to cluster with each of the clades, such as the cluster of purple (phages from environmental samples) in clade 1. This suggests that phages in Pa from similar environments are more similar than phages from a different environment. This again supports this work and the ability to study and define function derived from evolution in chronic respiratory disease.

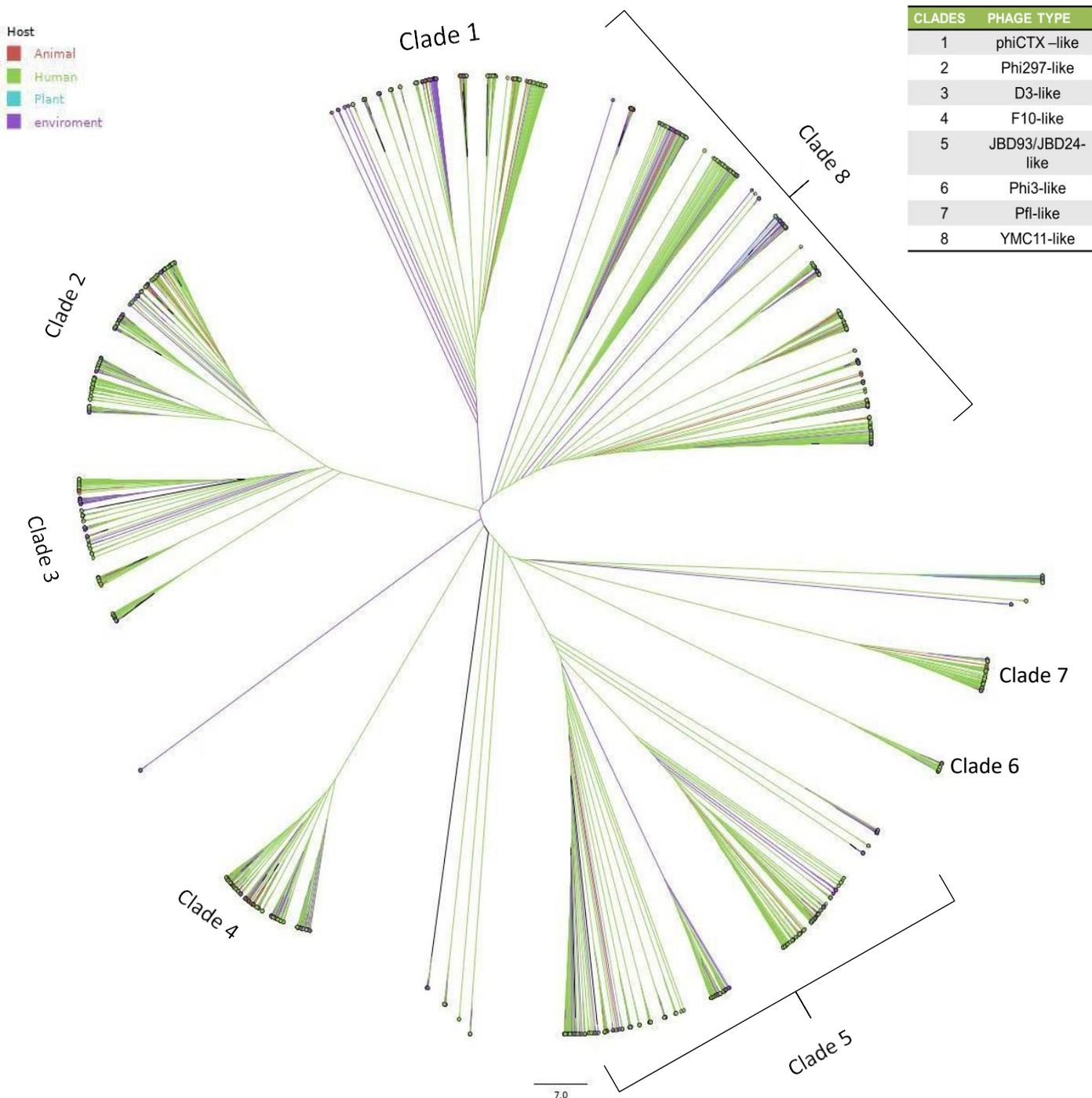


Figure 6.4 Comparison of the intact temperate phages identified in all the IPCD isolates using SaturnV. SaturnV was used to compare the proteins in all the intact temperate phages identified in all the IPCD isolates and clusters them depending on the presence or absence of proteins in the phage genomes. The colour of the node and branch shows the host in which the Pa that carried the phages, shown here, was derived (shown in the key on the left). The key on the right shows which phage in each clade was most similar to, when compared on the blast database. The cut off for SaturnV was >50% identity for the protein across >85% of alignment, covering the length to be considered as an ortholog protein.

6.5 DISCUSSION

6.5.1 Prophage Identification

The chapters in this thesis thus far have focused on carriage of phages across a small number of bacteria isolated from the lung. But how widespread and the complexity of bacteriophages carriage in *Pseudomonas aeruginosa* is yet to be determined, as the usual focus is the non-accessory backbone of the bacterial genome. The collaboration with Professor Levesque at the University of Laval and the research visit grant funded by the Microbiology Society allowed access to the 1030 IPCD genomes for us to data-mine for prophages. Our first findings revealed that there are a high number of prophages carried by Pa isolates, with higher numbers of incomplete prophages in the Pa genomes than intact or questionable prophages. This has been reported in other bacterial backgrounds, for example *E.coli* O157:H7 strain EDL933, 17 remnant or cryptic prophages are reported alongside the single inducible bacteriophage 933W (Plunkett et al, 1999). These incomplete remnant prophage regions or cryptic phages are thought to have been mobilisable viruses that, through mutation, have lost the ability to induce from their integration site. Over time the genome will lose areas that do not confer the bacterium a selective advantage. We originally hypothesised that clinical isolates may carry larger numbers of prophage regions due to the possible accrual of function that may be offered by infecting phages. These findings determine that the complexity in the numbers of prophage carriage is shared between both clinical and environmental isolates. Only one isolate out of 1030 had no prophages identified in its genome and the highest number of prophages in a single isolate was 28. This demonstrates just how much phages can impact the bacterial genome as prophages significantly contribute to the accessory genome and thus represent a substantial factor in the genomic diversity of Pa. Both of these bacterial isolates would be important for further study downstream of this research, as one is receptive to high numbers of phage infections that may be linked to an altered immunity to incoming phages and therefore, an attenuated resistance to phage infection (Bondy- Denomy et al., 2013, Pawluk et al., 2014). Secondly, studying why the other isolate does not carry any phage regions

at all and what is the exclusion and immunity mechanism associated with it, are further questions that could be asked.

This work focused only on intact prophage genomes, due to their putative ability to induce and transduce across other bacteria. However, the importance of the incomplete prophage genomes should not be overlooked, as these regions were most likely intact and over generations progressive deletion and evolution have selected any genes that could give their bacterial host an advantage and kept them within the genome (Canchaya et al., 2002, Rezaei Javan et al., 2019). As all the incomplete prophage genomes have been identified and are stored, this is something that could be investigated further in the future.

6.5.1 Comparison of origin of Pa host and prophage carriage

There are, on average, slightly more phages per isolate from environmental Pa samples than from human or animal hosts. This may be down to Pa samples from environments, such as sea or sewage, being more exposed to more phages due to the complexity of the microbial environment. Investigations of temperate phages in other bacteria showed that every genome of 482 pneumococcal strains contained prophages and some persisted for extended periods of time without changes at the nucleotide sequence level (Brueggemann et al., 2017). In *E.coli* O157, 18 prophages were identified. However, most of them were incomplete and so could not be induced and released from the host cell (Asadulghani et al., 2009). There has been little research comparing temperate phages between bacterial strains. However the numbers seen in Pa isolates in the IPCD are on par with those observed for other clinically relevant species. In the environment, such as sea and sewage, there is evidence of high numbers of lytic phages, some of which may actually be temperate phages that have been induced, which would concur with our results showing Pa environmental samples having slightly higher numbers of prophages than clinical samples.

6.5.2 Presence of virulence factors within Pa prophage genomes

It has been previously reported that bacteriophages can carry functional genes that may alter the pathogenic profile of the infected bacterium. Examples include; *stx*, *lom* and *bor* (Wagner et al., 2001, Vica Pacheco et al., 1997, Plunkett et al., 1999). With increasing numbers of well curated databases like VFDB, it is possible to complete pairwise comparisons offering a hit or positive ID of a virulence gene within the panel, which shows similarity as distant as 23% identity to a gene within the database. Due to the number of prophages identified in the IPCD, our findings and ability to test if all were viable as AMR genes, was limited to 100% sequence identity to the virulence gene. Of the 23 phages that were shown to carry a virulence gene with a 100% identity to the whole gene, only 7 different virulence genes were identified, with the most common being (*rpoN*) RNA polymerase factor sigma-54, which was found in 10 out of 23 phages. As mentioned previously (section 6.4.1.1), these 10 phages have high sequence similarity, which suggests they are the same phage. This makes this phage containing (*rpoN*) RNA polymerase factor sigma-54 relatively common, rather than the gene being common in different phages. RNA polymerase factor sigma-54 has been shown to increase bacterial pathogenesis by increasing motility of *Pa* and increased efficiency of the *Pa* being internalised by human host cells compared to the *rpoN* mutant in *Pa* (Totten et al., 1990, Plotkowski et al., 1994). Therefore, it would be beneficial to a *Pa* host to carry this phage. Transcription initiation is highly regulated in bacterial cells, allowing adaptive gene regulation in response to environment cues. RNA polymerase factor sigma-54 enables such adaptive gene expression through its ability to lock the RNA polymerase down into a state that is unable to melt out promoter DNA for transcription initiation.

Another virulence factor that was identified was (*rpoS*) RNA polymerase factor sigma-*RpoS*, which plays a key role in protecting *Pa* from different environmental stress conditions, including starvation, hyperosmolarity, oxidative damage and reduced pH (Jørgensen et al., 1999). The bacteria will probably have a copy of *rpoS*, therefore the phage carrying it could have two copies in the genome. This would probably increase its expression through gene

addition, which could mean that the lysogen is more reactive to stress and able to survive over Pa without the prophage. It could also mean that there is an increase in production of *rpoS* when it is stimulated by environmental factors and allows a faster response, which could give a selective advantage over the Pa without the prophage and only one copy of *rpoS*. Studies show that the survival of *rpoS* knock-out mutants is significantly affected, especially in the presence of antibiotics (Mata et al., 2017). A study that placed an additional copy of the *rpoS* gene into *Pseudomonas knackmussii* showed that the addition of a second *rpoS* copy, under control of its native promoter, resulted in an increase of the proportion of cells expressing the *P_{int}* and *P_{inR}* promoters. When it was replaced by a *rpoS*-mcherry fusion, a double-copy *rpoS*-mcherry-containing strain displayed twice as much mCherry fluorescence. This suggests that a double copy of *rpoS*, as seen in the lysogeny, would double the dose of the protein being regulated (Miyazaki et al., 2012), therefore, increasing the survival of the Pa giving the lysogen a selective advantage.

Other virulence factors identified, such as PvdE and PvdF, are involved in pyoverdine biosynthesis. It has been reported pyoverdine, that is secreted by Pa and regulated by PvdE and PvdF, regulates the production of exotoxin A, an endoprotease, as well as itself, which contributes to Pa causing disease (Lamont et al., 2002). As these genes are also present in the Pa genome, the presence of them in the prophage would also mean there would be two copies of the gene in the genome. These are both promoters, which may mean that the duplication of a promoter, or expression from a single promoter in situ, would increase the regulation on the protein, in this case pyoverdine, which would also lead to increased iron uptake from transferrin and lactoferrin. This could be tested experimentally to see if iron uptake increases in the lysogen compared to without the phage. A study showed that PvdE knock-out lost invasive ability toward HCECs (human corneal epithelial cells), therefore this also may suggest that an additional gene may aid the Pa's ability to invade human cells, which could also be tested on HCEC and on lung epithelial in further studies. The virulence factor WaaA lipopolysaccharide core biosynthesis protein WaaP was identified in a phage from CF sputum. It plays roles in lipopolysaccharide biosynthesis, which aids the Pa in the

evasion of macrophage-mediated killing (Elamin et al., 2017). This is beneficial in the environment of the Pa isolate that the phage was identified, as it was isolated from the CF lung.

The drawback of using the VFDB is that these are virulence factors found in bacteria rather than in phages. Therefore, it is likely that these genes have been picked up by the bacteria over time by recombination or mis-packaging of phage DNA and then kept over rounds of evolution. With more phages being sequenced and annotated it may be possible to create a phage virulence factor database that would be able to identify genes that may have subtly changed the metabolism of the bacteria.

6.5.3 Imaging of induced phages

The TEM images shown in figure 6.3 show that there are inducible phages that can be visualised in the 23 Pa isolates tested. The next step to verify that these phages, or phages within the mix of phages in the lysate that were used to create the TEM, contain the phage with the virulence factor. This can be done by purifying and sequencing each phage genome or simple approaches like PCR of the purified phage DNA for the incidence of these identified putative genes with virulence factors that were identified. If positive, these can be purified and made into lysogens of a common Pa strain to test how the virulence factor affects its host. This was unable to be completed due to time constraints. The phages that had hits of a lower percentage identity could also be looked into, especially if the VF identified is of specific interest. For example, if only part of the VF gene is seen in the phage genome, that part of the gene may be offering phage function.

6.5.4 Genetic diversity of Pa prophages identified from the IPCD

The diversity of the intact temperate phages shown in figure 6.4 demonstrates the high level of diversity seen in the temperate phages of Pa. There have not been any large investigations published studying the diversity of temperate phages (prophages) within Pa isolates, but many studies on lytic phages that infect Pa (Sepulveda-Robles et al., 2012, Essoh et al., 2015, Ha and Denver, 2018). These results, when compared to the lytic phage,

show that temperate phages in Pa are as diverse as the lytic phages in Sepulveda-Robles et al. (2012), as they too grouped into 8 taxonomic groups. However, this was done by comparing the phages using MUMmer against 40 tailed Pa phages that were reported in GenBank at the time. Also, they used a different method than SaturnV by calculating the sum of the identical regions divided by the length of the longest sequence and a break length of 60 nucleotides was used to determine the extent of the identical regions. Since the known species share 53% nucleotide sequence homology, that threshold was used to separate the new Pa phage sequences into species (Sepulveda-Robles et al., 2012).

Tariq et al. (2019) compared 105 Pa temperate phages specifically from Pa isolates from CF and BR patients (Tariq et al., 2019). However, all these phages were induced and sequenced, therefore confirming that they are intact and inducible rather than predictions of putative phages. Tariq et al also found 8 clades of phages, which suggests that even with a much larger cohort, diversity does not significantly increase. Of the 8 clades seen in this paper, 3 were the same as seen during this study, namely F10-like, PhiCTX-like (which is now known as *Citexvirus*) and Phi297-like. However, the other clades seen in the paper could possibly have phages they were more similar to if there were a higher diversity of phages, as seen in this study. This suggests it may be worth looking at each of the 8 clades individually as a cohort and trying to identify clusters within each of the clades of different phage types.

The way that the prophages clustered into the 8 clades, presented by SaturnV analysis, suggests they clustered mainly due to their number of shared orthologous proteins. Therefore, each clade of phages was able to be associated with a phage type, such as phiCTX-like shown in figure

6.4. Given the importance of phages in driving adaptive evolution of bacterial pathogens and Pa, these findings highlight a number of temperate phage types that are the most common in Pa from a very varied database (IPCD). This will enable further studies to focus on these in greater detail. The majority of the phage types have limited research into their

biology. However, some noteworthy examples that have reported characteristics are phiCTX-like phage (clade 1), which is a cytotoxin-converting phage and carries the CTX gene. Phi297 phages (clade 2) have been described as producing halo-morphologies in plaque assays and have a narrow host range (Burkal'tseva et al., 2011). The Phi297-like phages also show genome homology to D3 phages (clade 3), hence them clustering closely. The D3 phages (clade 3) are responsible for serotype conversion in their host Pa, which is very beneficial to the Pa in avoiding the human immune system (Kropinski, 2000). Clade 5 consisted of JBD93/JBD26-like phages. JBD93 and JBD26 phages contain genes for prophages that mediate defense against phage infection, with JBD93 phages being reported to contain an anti-CRISPR gene, which actively stop the Pa host's CRISPR system from working and allowing phage infection (Bondy-Denomy et al., 2013). This clade was the second largest clade with 311 phages, suggesting this type of prophage is common in Pa. In addition, it is possible that the inactivation of phage defense mechanisms in Pa, which allows the infection of new phages, is beneficial in some way as the group was so large. Pf1-like phages (clade 7) were a relatively small number of phages (78 in total). However, Pf phages are known to play a role in the pathogenicity of Pa in CF. They also increase the tolerance of the Pa host to antipseudomonal antibiotics. Therefore, more of these types of phages would have been expected with a database that is biased towards CF samples (Burgener et al., 2019).

The largest clade was clade 8, which were YMC11-like phages and have had little investigation. However, it has been reported that phages within this YMC11 group, specifically YMC11-R656 and YMC11-R1836, enhance the survival of *Galleria mellonella* larvae (Jeon and Yong, 2019), which we have also shown to be true in other types of phages in chapter 3.

Within these 8 clades of phages there was diversity, as well as differentiation between phages from different origins. In figure 6.4, phages from the environmental isolates mostly

cluster separately from the lung isolates, showing that there are parts of the phage genome that are associated to its environment. These are the genes that are of interest, especially if there are prevalent ones within phages from a certain environment. These genes are likely to be in the accessory genome of the phages. As we are comparing whole phage genomes, the clustering is based on the phages architecture and as most viruses have a conserved architecture of genes linked to its integration and replication strategies, this will make up most of these comparisons. Thus, within the clades, the diversity is more based on intragenic differences or the carriage of small numbers of different genes instead of the phages architecture.

Further work would be necessary to look at the diversity of the accessory genes in the prophages of each clade to identify prevalent genes of interest to focus on. If the core proteins were removed and the accessory proteins of the phages were compared this could enable identification of groups of phages to focus on and genes they have in common. As the accessory genome of prophages has the ability to adapt and evolve with its environment, these common genes may suggest they offer an advantage to the Pa from the specific environment they were isolated from. Also further analysis of the dendogram would enable a more accurate representation of the number of different groups/clades that are present as describes in the results, and give a better picture of the overall diversity of Pa prophages on a global scale.

This study supports all work on temperate phages in Pa, as it shows how common they are within the Pa genome from all environments. Also, some Pa phages carry virulence factors and therefore may impact the virulence of the Pa host. The following chapter looks into this further, using metabolomics to highlight the subtle changes that phages from each of the 8 clades can cause on the host's metabolism.

6.6 SUMMARY

Datamining of 1030 genomes for prophage genomes using PHASTER resulted in 6499 intact prophage genomes that were taken forward to analyse genetic diversity. The genetic diversity was analysed using Saturn V, which translated the prophage genomes and then

compared them at the protein level depending on the presence or absence of each protein. This gave 8 clades that were compared against the BLAST database to give their nearest phage type. Phages that were identified in Pa genome isolates from lungs were compared separately, but as there was such a bias to CF isolates in the IPCD there was not much difference. The phage genomes were also compared against the VFDB to identify any virulence factor genes that were carried by the phages. The results showed that there were many partial hits and 23 hits that have 100% identity to genes in the VFDB. The PA isolates that harboured these genes and were induced were imaged to show the presence of induced phages.

7 METABOLOMIC AND FUNCTIONAL COMPARISON OF LYSOGENS CONVERTED WITH PHAGES FROM THE INTERNATIONAL *PSEUDOMONAS* CONSORTIUM DATABASE (IPCD).

7.1 INTRODUCTION

7.1.1 Importance of understanding the effects of temperate phages on their bacterial host on a large scale

The previous chapter looked at the genomics and diversity between lysogenic phages found in the IPCD panel. The diversity illustrates 8 groups of these phages that are common in Pa (potentially numerous more after reanalysis on the data). There have been multiple studies focusing on Pa temperate phages' effects on the host, including focus on the LES phages and Pf phages. These were also considered within the research presented in this thesis, looking at the metabolic effects of the LES phages in chapter 5. There have been further studies looking at functionality and phage-encoded phenotypes driven by temperate phages from Pa isolated from specific sources, such as CF lungs (Tariq et al., 2019, James et al., 2015), which was also the focus in chapter 5. There have been some studies analysing the genetic diversity of phages in Pa, including chapter 6 of this thesis, but these are mostly on a small scale with limited phage host combinations and no further experimental work to assess the effect of these phages on the host (Ha and Denver, 2018). Thus far there have been no publications (to the best of our knowledge) assessing the metabolic impact of environmentally relevant and genetically diverse temperate phages in Pa. This is important as it could show that infection by genetically diverse temperate phages can cause different metabolic shifts in the same bacteria and whether this influences host fitness and or virulence in chronic respiratory disease. It may also show that there is core control of bacterial function that is beneficial for the phages when integrated as a prophage. This is valuable in determining if there are types of temperate phage or genes within phages that are key to facilitating enhanced fitness in the host and, in the case of Pa in lung infection, moving from colonisation and increasing the ability to adapt and evolve to allow chronic

infection of the lung.

7.2 AIMS

- Identify prevalent/common carriage of genes between phages determined to be taxonomically distinct within the IPCD panel of Pa. To also determine unique genes in each clade of phage diversity so that it can be used as a marker of lysogeny for isolation purposes.
- Create Pa lysogens with representative phages from each clade and confirm prophage presence by PCR using oligonucleotide primers selected for each clade.
- Determine if the phages from each clade alter the metabolic physiology of their Pa host. To compare between the phage groupings and to compare/contrast which metabolic pathways are subverted by phage conversion.

7.3 OBJECTIVES

- Use bioinformatic tool ROARY to ascertain which phage genes are the most prevalent in each clade of phages and which are not present in the other clades. Make primers to these specific genes to enable confirmation of integration when making lysogens.
- Select isolates from the IPCD that contain only a single intact phage representing each clade.
- Induce phages from Pa isolates from each clade that only carry one phage and test their host range on lab strains PAO1, PA14 and PAK to determine which is the best host to create lysogens for each of the selected phages.
- Infect the chosen Pa host with the phage to make a lysogen and check integration using PCR and primers for each clade.
- Grow each lysogen in ASM and prepare for metabolomics analysis to determine how phages from each clade effect the Pa host.

7.4 RESULTS

7.4.1 Characterisation of bacteriophage gene frequency using ROARY and subsequent selection of target species from the IPCD panel that carry only one predicted inducible phage from each clade.

ROARY is a protein-based search tool that translates the DNA sequence before pairwise 'all on all' comparison using the default parameters that were 8 threads that generates a core gene alignment using 8 threads and a minimum blastp percentage identity of 90%. The tool was sequentially run on each clade identified in Chapter 6. The aim was to ascertain which coding genes are carried at high frequency between phages in each clade, enabling the selection of a gene to be used to make primers for identification of phages from that specific clade by PCR. The results in table 7.1 show the top 10 genes carried at high frequency for each clade. The original annotations were completed by prokka. However, the genes of high prevalence were then run against the NCBI viral database to obtain a more accurate annotation for the genes. These genes were then compared to the genomes of phages from the other clades to make sure these genes were truly clade specific. Primers were then designed for each of these genes (shown in table 2.2).

7.4.1.1 Selection of phages from clades in the IPCD

Only prophages carried in Pa strains isolated from lung infections were selected, as the role of temperate Pa phages in lung infection was the main focus of this study. Phages were selected from each clade by also choosing only Pa isolates with one intact (putatively mobile) prophage. This was done so that during induction there was no need to differentiate between plaques and avoided a plaque purifications step. To maximise the chance of being able to select, induce and convert one of the bacterial hosts mentioned, 10 isolates (where possible) fitting the inclusion criteria for each clade were taken forward (shown in table 7.2.

Table 7.1 Most prevalent genes in phages from each clade from prophages of the IPCD isolates

Clade	Prokka Annotation	Viral database annotation	No. phages in clade with gene	No. phages in clade
Clade 1	hypothetical protein	baseplate assembly protein	194	214
Clade 1	hypothetical protein	putative Peptidoglycan domain protein	193	214
Clade 1	hypothetical protein	phage tail protein	193	214
Clade 1	hypothetical protein	putative Peptidoglycan domain protein	193	214
Clade 1	hypothetical protein	baseplate assembly protein	192	214
Clade 1	hypothetical protein	phage Tail Protein X family protein	191	214
Clade 1	hypothetical protein	phage major tail tube protein	190	214
Clade 1	Putative prophage major tail sheath protein	phage tail sheath protein	190	214
Clade 1	hypothetical protein	phage major capsid protein, P2 family	189	214
Clade 1	hypothetical protein	capsid-scaffolding protein	189	214
Clade 2	Carbon storage regulator	global regulatory protein	248	292
Clade 2	hypothetical protein	Hypothetical protein/ phage-encoded membrane protein	198	292
Clade 2	hypothetical protein	Hypothetical protein/ phage-encoded membrane protein	198	292
Clade 2	hypothetical protein	hypothetical protein	171	292
Clade 2	hypothetical protein	hypothetical protein	153	292
Clade 2	hypothetical protein	exonuclease	138	292
Clade 2	Protein RecT	hypothetical protein	138	292
Clade 2	hypothetical protein	Holliday junction resolvase	135	292
Clade 2	hypothetical protein	putative DNA segregation ATPase	135	292
Clade 2	hypothetical protein	hypothetical protein	131	292
Clade 3	hypothetical protein	major capsid protein	95	169
Clade 3	hypothetical protein	terminase large subunit	95	169
Clade 3	hypothetical protein	hypothetical protein	95	169

Clade 3	hypothetical protein	putative phage head-tail adaptor	94	169
Clade 3	ATP-dependent Clp protease proteolytic subunit	ClpP protease	94	169
Clade 3	hypothetical protein	putative head-tail connector protein	94	169
Clade 3	hypothetical protein	hypothetical protein	93	169
Clade 3	hypothetical protein	hypothetical protein	93	169
Clade 3	hypothetical protein	hypothetical protein	93	169
Clade 3	hypothetical protein	portal protein	89	169
Clade 4	hypothetical protein	hypothetical protein	164	219
Clade 4	hypothetical protein	hypothetical protein	158	219
Clade 4	hypothetical protein	hypothetical protein	158	219
Clade 4	hypothetical protein	hypothetical protein	151	219
Clade 4	hypothetical protein	hypothetical protein	140	219
Clade 4	hypothetical protein	hypothetical protein	120	219
Clade 4	hypothetical protein	hypothetical protein	117	219
Clade 4	hypothetical protein	hypothetical protein	110	219
Clade 4	hypothetical protein	hypothetical protein	110	219
Clade 4	hypothetical protein	Putative tail length tape measure protein	109	219
Clade 5	hypothetical protein	hypothetical protein	210	311
Clade 5	hypothetical protein	hypothetical protein	195	311
Clade 5	Membrane-bound lytic murein transglycosylase F	transglycosylase SLT domain protein	172	311
Clade 5	hypothetical protein	hypothetical protein/ putative holin protein	156	311
Clade 5	hypothetical protein	hypothetical protein/ endolysin	152	311
Clade 5	hypothetical protein	hypothetical protein	145	311
Clade 5	hypothetical protein	hypothetical protein	142	311
Clade 5	hypothetical protein	hypothetical protein	138	311
Clade 5	hypothetical protein	hypothetical protein	134	311
Clade 5	hypothetical protein	hypothetical protein	133	311
Clade 6	hypothetical protein	portal protein	23	30
Clade 6	hypothetical protein	serine peptidase	22	30
Clade 6	hypothetical protein	major capsid protein	22	30
Clade 6	hypothetical protein	putative terminase ATPase subunit	21	30
Clade 6	hypothetical protein	hypothetical protein	13	30
Clade 6	hypothetical protein	hypothetical protein	13	30
Clade 6	hypothetical protein	hypothetical protein	13	30

Clade 6	hypothetical protein	hypothetical protein	13	30
Clade 6	hypothetical protein	hypothetical protein	13	30
Clade 6	hypothetical protein	hypothetical protein	13	30
Clade 7	hypothetical protein	ssDNA binding protein	70	78
Clade 7	hypothetical protein	hypothetical protein	61	78
Clade 7	hypothetical protein	hypothetical protein	61	78
Clade 7	hypothetical protein	hypothetical protein	60	78
Clade 7	hypothetical protein	hypothetical protein	59	78
Clade 7	hypothetical protein	hypothetical protein	59	78
Clade 7	Tyrosine recombinase XerD	integrase	57	78
Clade 7	hypothetical protein	hypothetical protein	52	78
Clade 7	hypothetical protein	hypothetical protein	51	78
Clade 7	hypothetical protein	hypothetical protein	51	78
Clade 8	hypothetical protein	hypothetical protein	313	339
Clade 8	hypothetical protein	lytic enzyme	313	339
Clade 8	hypothetical protein	hypothetical phage protein	311	339
Clade 8	HTH-type transcriptional regulator PrtR	HTH-type transcriptional regulator PrtR	309	339
Clade 8	hypothetical protein	hypothetical protein	309	339
Clade 8	hypothetical protein	hypothetical protein	309	339
Clade 8	hypothetical protein	excisionase	308	339
Clade 8	putative protein Ybil	DksA-like zinc finger domain containing protein	308	339
Clade 8	hypothetical protein	hypothetical protein	299	339
Clade 8	hypothetical protein	putative major tail protein	275	339

7.4.2 Induction and spot assay to test infectivity of phages

The isolates containing the phages that were chosen, shown in table 7.2, were subject to NFLX induction (using the method in section 2.5.1) and spot assays were completed for each against three lab strains; PAO1, PA14 and PAK (using the method in section 2.5.4). The results in table 7.3 shows if there are phages induced from the original IPCD isolates can infect any of the three Pa lab strains to make a lysogen. PA14 was the most susceptible of the three strains and was also able to be infected by phages from the most clades, unlike PAO1 and PAK. The phages that showed to positively infect PA14 were then taken forward to form lysogens with PA14.

Table 7.2 - *P. aeruginosa* isolates selected from the IPCD for phage induction. Pa isolates that are from lung infection and contain only one intact phage from a particular clade identified using PHASTER, highlighted in yellow and orange are Pa isolates where prophages were successfully induced and made into lysogens in PA14.

Isolate ID	IPCD ID	Phage clade	Isolation location	Human Isolation site	Human Pathology
1	108	clade1	unknown	Sputum	Bronchiectasis
2	148	clade1	Sherbrooke	Sputum	Cystic fibrosis
3	1330	clade1	Quebec	Sputum	Cystic fibrosis
4	32	clade1	Montreal	Sputum	Cystic fibrosis
5	1323	clade1	Quebec	Sputum	Cystic fibrosis
6	1584	clade1	Chicoutimi	Sputum	Cystic fibrosis
7	177	clade1	Sherbrooke	Nasopharyngeal swab	Cystic fibrosis
8	1450	clade1	Saint-Roch-de-Achigan	infected lung	Cystic fibrosis
9	170	clade2	Sherbrooke	Sputum	Cystic fibrosis
10	1319	clade2	Quebec	Sputum	Cystic fibrosis
11	1627	clade2	Calgary (CACFC)	Sputum	Cystic fibrosis
12	1325	clade2	Quebec	Sputum	Cystic fibrosis
13	1632	clade2	Calgary (CACFC)	Sputum	Cystic fibrosis
14	164	clade2	Sherbrooke	Nasopharyngeal swab	Cystic fibrosis
15	355	clade2	Brisbane	Sputum	Cystic fibrosis
16	1592	clade2	Chicoutimi	Sputum	Cystic fibrosis

17	1572	clade2	Montreal	Sputum	Cystic fibrosis
18	922	clade3	Hobart	Sputum	Cystic fibrosis
19	948	clade3	Hobart	Sputum	Cystic fibrosis
20	971	clade3	Hobart	Sputum	Cystic fibrosis
21	976	clade3	Hobart	Sputum	Cystic fibrosis
22	1257	clade3	Melbourne	Sputum	Cystic fibrosis
23	1258	clade3	Hobart	Sputum	Cystic fibrosis
24	1360	clade3	unknown	Throat swab	Cystic fibrosis
25	1408	clade3	Hannover	Throat swab	Cystic fibrosis
26	1579	clade3	unknown	Sputum	Cystic fibrosis
27	103	clade3	Unknown	Sputum	Pneumonia
28	346	clade4	Brisbane	Sputum	Cystic fibrosis
29	940	clade4	Hobart	Sputum	Cystic fibrosis
30	73	clade4	Montreal	Sputum	Cystic fibrosis
31	1069	clade4	Nottingham	Sputum	Cystic fibrosis
32	74	clade4	Montreal	Sputum	Cystic fibrosis
33	1085	clade4	Nottingham	Sputum	Cystic fibrosis
34	1589	clade4	Chicoutimi	Sputum	Cystic fibrosis
35	1517	clade4	Rouyn-Noranda	Throat	Cystic fibrosis
36	1072	clade4	Nottingham	Sputum	Cystic fibrosis
37	128	clade5	Sherbrooke	Sputum	Cystic fibrosis
38	1333	clade5	Quebec	Sputum	Cystic fibrosis
39	1067	clade5	Nottingham	Sputum	Cystic fibrosis
40	1080	clade5	Nottingham	Sputum	Cystic fibrosis
41	1335	clade5	Quebec	Sputum	Cystic fibrosis
42	905	clade5	Dunedin	Sputum	Cystic fibrosis
43	1436	clade5	Lund	Throat swab	Cystic fibrosis
44	1629	clade5	Calgary (CACFC)	Sputum	Cystic fibrosis
45	1515	clade5	Rouyn-Noranda	Throat	Cystic fibrosis
46	644	clade6	Leiden	Sputum	Non-CF
47	1363	clade6	Munich	Throat swab	Cystic fibrosis
48	124	clade7	Sherbrooke	Sputum	Cystic fibrosis
49	810	clade7	Birmingham	Sputum	Cystic fibrosis
50	1328	clade7	Quebec	Sputum	Cystic fibrosis
51	1412	clade7	Hannover	Throat swab	Cystic fibrosis
52	1574	clade7	Montreal	Sputum	Cystic fibrosis
53	545	clade7	Seattle	Throat	Cystic fibrosis
54	1604	clade7	Chicoutimi	Sputum	Cystic fibrosis
55	159	clade7	Sherbrooke	Nasopharyngeal swab	Cystic fibrosis
56	715	clade7	Vancouver	Sputum	Cystic fibrosis
57	1451	clade7	Montreal-Nord	infected lung	Cystic fibrosis

58	43	clade8	Montreal	Sputum	Cystic fibrosis
59	112	clade8	Sherbrooke	Sputum	Cystic fibrosis
60	1308	clade8	Quebec	Sputum	Cystic fibrosis
61	1309	clade8	Quebec	Sputum	Cystic fibrosis
62	1307	clade8	Quebec	Sputum	Cystic fibrosis
63	1341	clade8	Quebec	Sputum	Cystic fibrosis
64	1370	clade8	Munich	Throat swab	Cystic fibrosis
65	1433	clade8	Hannover	Throat swab	Cystic fibrosis
66	1593	clade8	Chicoutimi	Sputum	Cystic fibrosis
67	29	clade8	Montreal	Sputum	Cystic fibrosis
68	1488	clade8	Rouyn-Noranda	Sputum	Cystic fibrosis
69	334	clade8	Brisbane	Sputum	Cystic fibrosis

Table 7.3 Prophages were induced from *Pa* isolates from the IPCD collection and tested for lysogeny against *Pa* strains PA01, PA14 and PAK by spot plaque assay. Red and green cells indicate the absence and presence of zones of clearing, respectively.

Clade	Isolate/Phage ID	PA01	PA14	PAK
1	S1	Green	Green	Green
1	S2	Red	Red	Red
1	S3	Red	Red	Green
1	S4	Red	Green	Green
1	S5	Red	Green	Red
1	S6	Red	Red	Red
1	S7	Green	Green	Red
1	S8	Green	Green	Red
2	S9	Red	Red	Red
2	S10	Red	Red	Red
2	S11	Red	Red	Red
2	S12	Red	Green	Red
2	S13	Red	Red	Green
2	S14	Red	Red	Red
2	S15	Red	Green	Red
2	S16	Red	Green	Red
2	S17	Red	Green	Red
2	S18	Red	Red	Red
3	S19	Red	Red	Red
3	S20	Red	Green	Red
3	S21	Red	Green	Red
3	S22	Red	Red	Green
3	S23	Red	Red	Red
3	S24	Red	Red	Red
3	S25	Green	Green	Red
3	S26	Red	Red	Red
3	S27	Green	Green	Green
4	S28	Red	Red	Red
4	S29	Red	Red	Red
4	S30	Green	Green	Green

4	31			
4	32			
4	33			
4	34			
4	35			
5	36			
5	37			
5	38			
5	39			
5	40			
5	41			
5	42			
5	43			
5	44			
5	45			
6	46			
6	47			
7	48			
7	49			
7	50			
7	51			
7	52			
7	53			
7	54			
7	55			
7	56			
7	57			
8	58			
8	59			
8	60			
8	61			
8	62			
8	63			
8	64			
8	65			
8	66			
8	67			
8	68			
8	69			

7.4.3 Formation and confirmation of PA14 lysogens containing phages from each clade

Each of the phage lysates that could infect PA14 in table 7.3 were then used to form PA14 lysogens (using the lysogen formation method in section 2.5.6). DNA from these putative lysogens were then extracted (using a DNA extraction kit method detailed in section 2.6.2). Oligonucleotide primers were designed to differentiate between the phage clades. This lysogen gDNA was used as the input template for the discriminatory PCR. PA14 gDNA was used as the negative control to show that there was no presence of the target phage gene in the bacteria prior to infection of the phages. The gDNA from the original bacterial host was used as the positive control. Out of the 26 samples that showed to positively infect PA14 in the spot assays, 16 isolates from 4 clades displayed the presence of the marker genes, showing that the correct phages had integrated into the PA14 genome; clade 1 (n=5), clade 2 (n=4), clade 5 (n=4) and clade 8 (n=3) highlighted in table 7.2. An example of the results from the PCR products run on an electrophoresis gel (using method in section 2.7.3) are shown in figure 7.1, which shows the gene marker for phages from clade 5 were present in the PA14. This confirms a clade 5 phage integrated and that clade 6 marker genes were not present in the PA14, therefore it was not a lysogen. The other PCRs, including controls and confirmation that the oligonucleotides are specific to the particular clade, are shown in a working PCR test in (Appendix G).

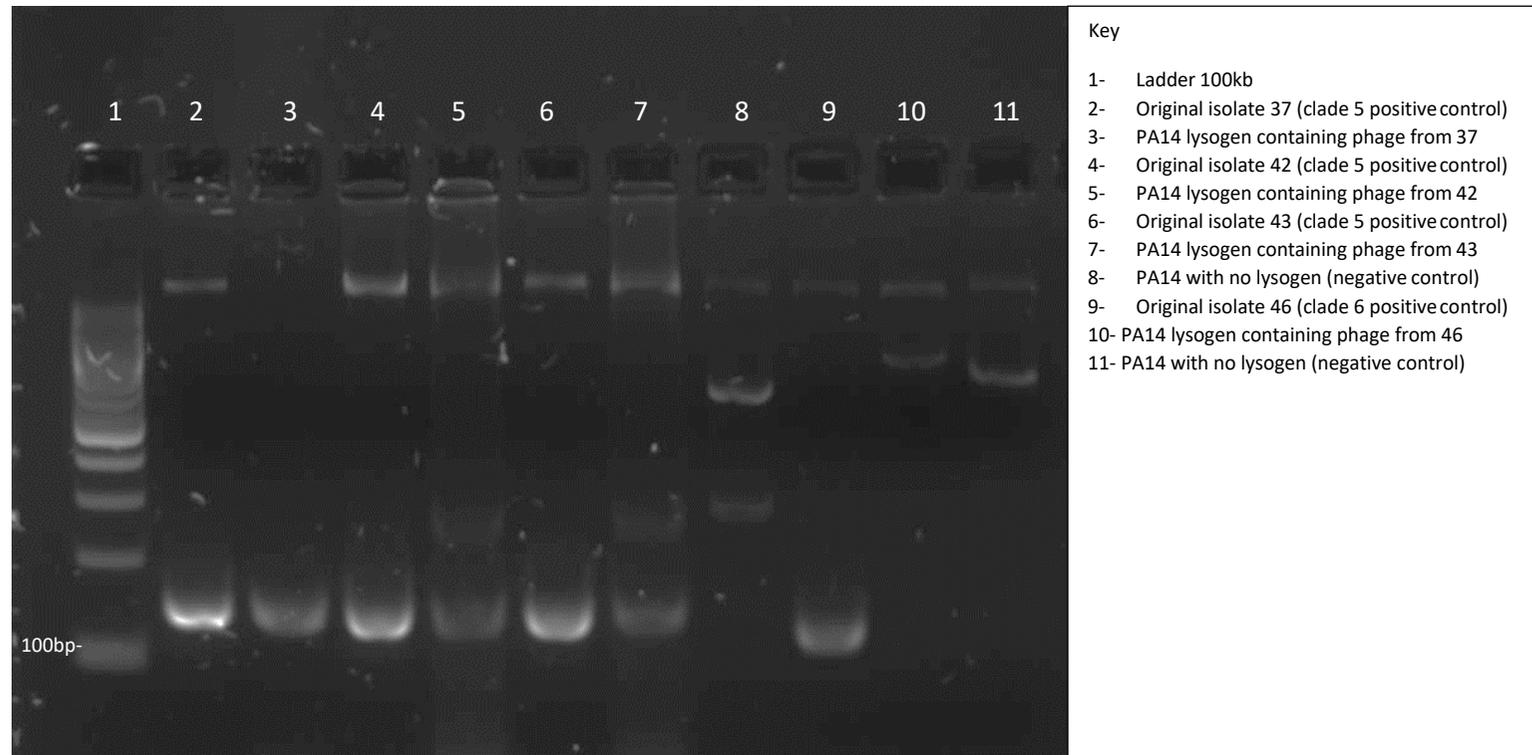


Figure 7.1 – Indicative electrophoresis gel showing the results of the PCR reactions to confirm lysogeny. This gel shows if the selected phage gene marker is present. This indicates that the phage integrated into the PA14 isolate, compared to the PA14 negative control without the phage integrated.

7.4.4 Metabolomic analysis and comparison of IPCD Phage-PA14 lysogens

The 16 confirmed lysogens that represent 4 of the clades were; clade 1 (n-4), clade 2 (n-3), clade 5 (n-4) and clade 8 (n-3). They were prepared for metabolomic analysis following the same protocol as used in chapter 5. They were grown in ASM (using method in section 2.1.3), with an n=10 for each lysogen, extracted and prepared for LC-MS processing (using method in section 2.4.1) and ran on a LC-MS (using methods in section 2.9.2). The untargeted data output gave 2667 metabolites. The metabolites that were identified to differentiate significantly between samples allowed reduction to 1647 metabolites. Due to high numbers of significant metabolites, they needed further reduction, to enable observation of the metabolomic shift between the control and the lysogen groups (using the same method described in section 5.4.2). Using this approach, it focused the data to 292 metabolites. The 292 metabolites were compared to the PAMDB (using method in 2.12) to give putative compound/metabolite names as done in chapter 5.

Comparisons of these metabolites were achieved using the online tool MetaboAnalyst, as previously mentioned and used in chapter 5. These metabolites were then ran through MetaboAnalyst statistical analysis software and the results of the multivariate data analysis comparing all the lysogen groups are shown on a PCoA plot (Principal Components Analysis) in figure 7.2. The PCoA shows that there is separation of the nodes, therefore difference between the metabolism of the control and lysogens in all clades. Lysogens of clade 1 and clade 8 cluster individually, which shows diversity of metabolisms to all other clades and the control. Whereas the other two clades (2 and 5) cluster together, which suggests that the phages from those clades cause a similar shift in host bacterial metabolism. However, a key point to be made is that all lysogens differentiate away from the PA14 control, showing that all phages from every clade have an impact on the cellular metabolism of their Pa host.

The PCoA plot has a corresponding heatmap and dendrogram, shown in figure 7.3, which shows the specific metabolites that are differentially regulated between lysogen clades. The dendrogram illustrates how the lysogens group based on their metabolism, showing the same results as the PCoA but also the check within-clade diversity, as there were multiple lysogens from each clade. This highlights any variation in how they affect the PA14's metabolism (seen in further detail in the dendrograms in figures 7.4-7.7).

To investigate metabolic shifts in further detail, each of the clades were compared separately against the control PA14 (Figures 7.4-7.7). The PCoAs in figures 7.4a, 7.5a, 7.6a and 7.7a confirm that there is a marked change in the metabolism between PA14 and each of the phage subverted lysogen clades. The dendrograms shown in figures 7.3b, 7.4b, 7.5b and 7.6b show again that the lysogens cluster separately from the PA14 control, showing a distinct difference between the control versus the lysogen metabolic profiles.

Comparing these clades separately against the control allows for direct comparison between PA14 and each of the lysogens within a clade, illustrating phage-specific alteration of metabolic function. This is shown in figures 7.3c, 7.4c, 7.5c and 7.6c (an image with greater pixel density and resolution of individual metabolites can be seen in appendix H figures A-C). The heatmaps illustrate the similarity in the phage subversion of bacterial host metabolism that is different from PA14. For example, in clade 2 all samples apart from S71, S72 and S73 (shown at the bottom of the heat map), which are technical replicates of PA14 lysogen containing phage 32, appear to regulate the metabolites very similarly (as shown in the heatmap in 7.5c) with other clade 2 type phages. (The list of which samples correspond to which lysogens can be seen in appendix H). Conversely, there are specific phages that alter other metabolic pathways that differ from these core patterns (S129-S134 in figure 7.7c). This shows that different temperate phages, even within the same clade, can subvert the Pa metabolism in an altered way. Due to the high number of metabolites, pathway analysis was required to narrow down and assess which pathways in the Pa metabolism is altered by conversion of these IPCD phages.

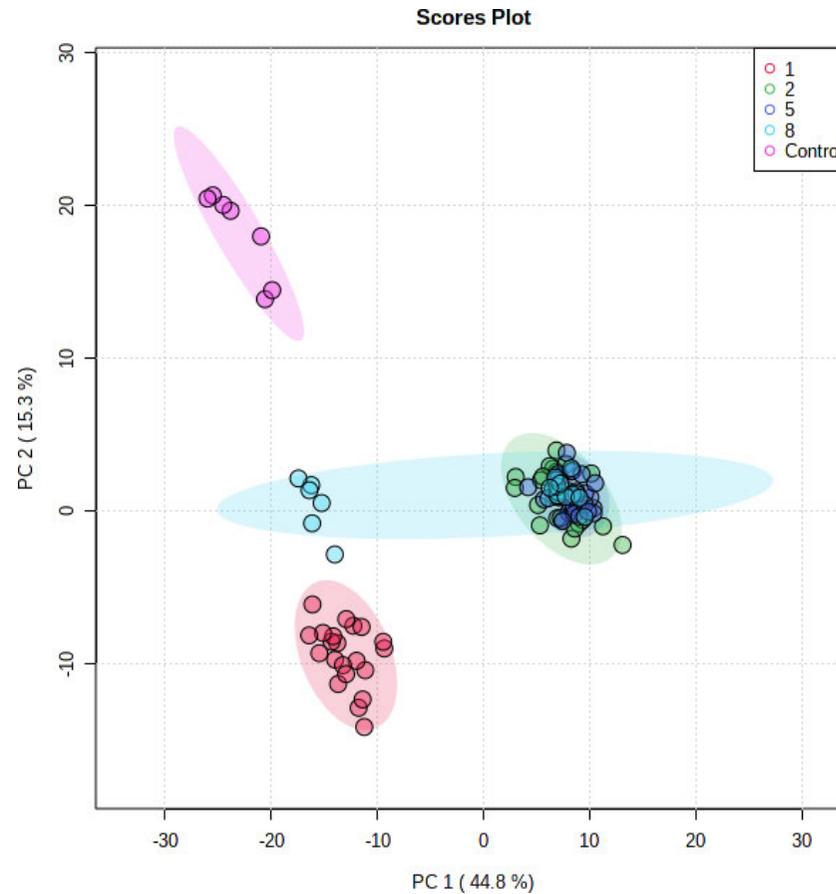


Figure 7.2 – PCoA of the metabolic profile comparison of lysogens from 4 different clade (phage types) grown in ASM. Comparison of all lysogens to the PA14 control. 2D PCoA plot with each node representing the MS results ran in negative mode (top 100 significant metabolites from each clade only), showing the dissimilarity between the metabolism of each set of lysogens.

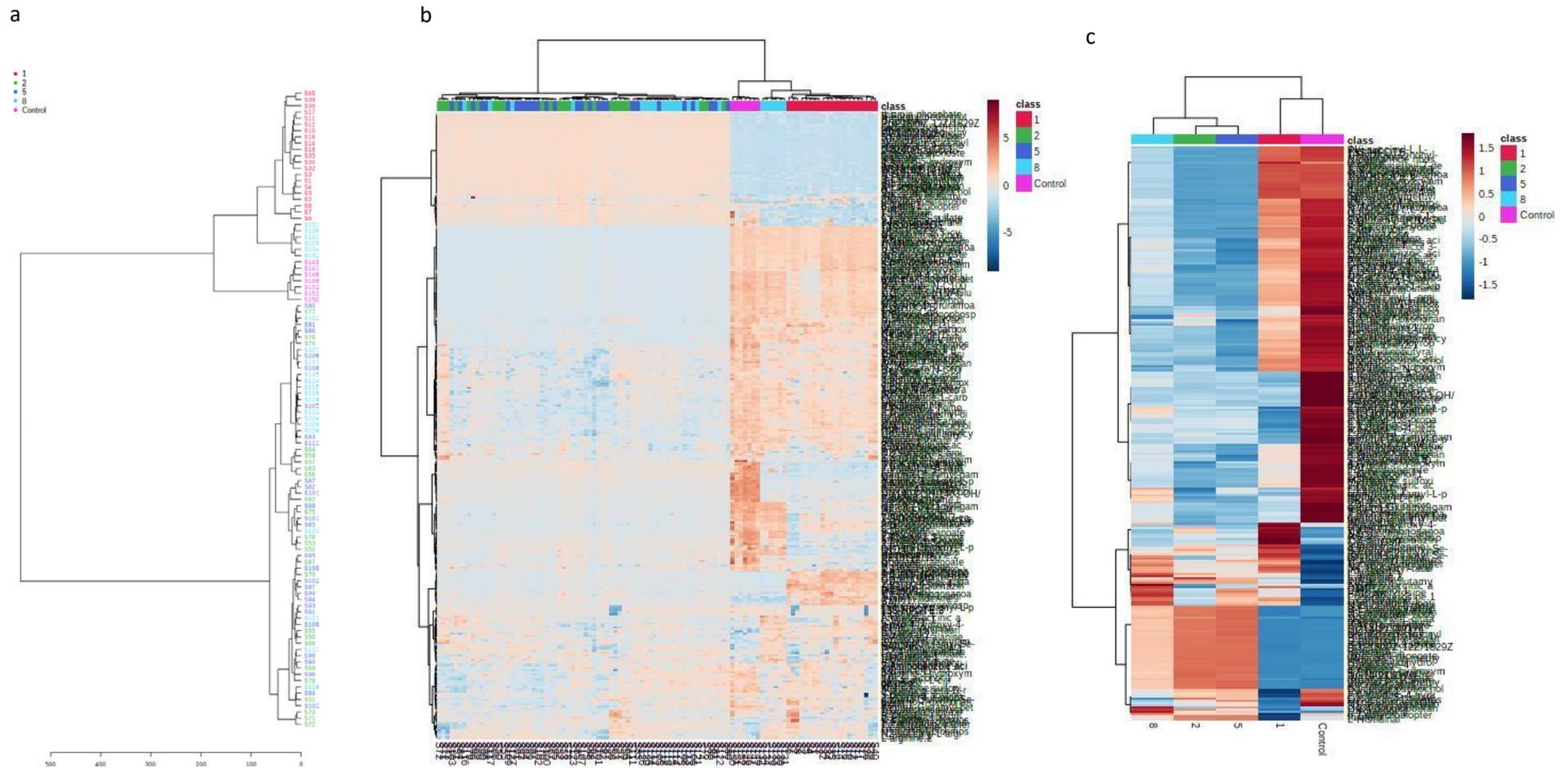


Figure 7.3 – Metabolic comparison of all lysogens metabolic profiles from all clades compared against one another and the PA14 control. (a) Hierarchical clustering result shown as a dendrogram, showing each clade’s lysogen samples and the control compared. (b) A heatmap representing the up and down regulation of each significant metabolite between each of the lysogen samples from each clade, compared against the other lysogens and the control. A clearer image with readable labels is shown in figure 1a in appendix H. (c) The average score of the samples from each clade was taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. The lysogen the sample/S numbers are associated to is detailed in appendix I.

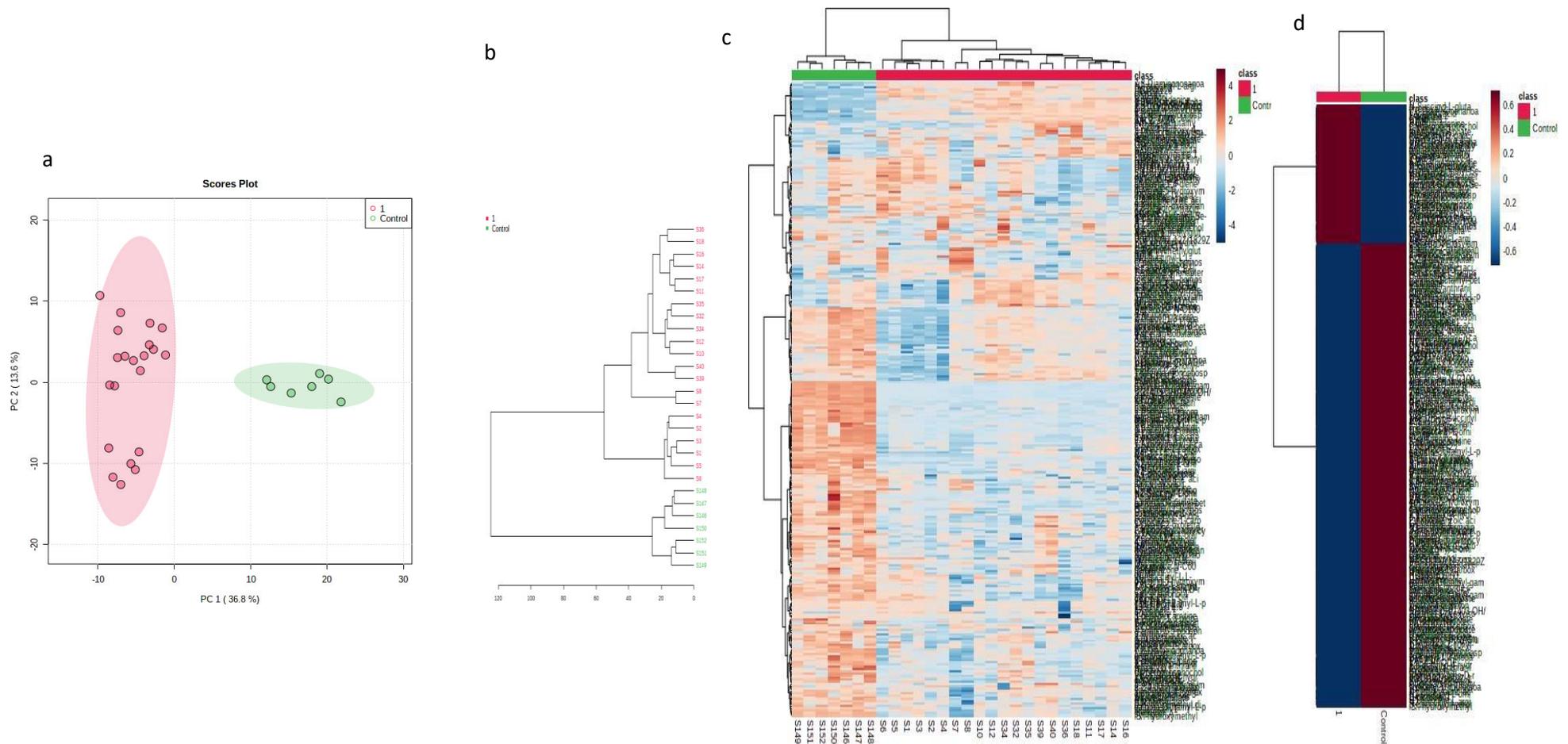


Figure 7.4 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 1 compared to PA14 control. (a) 2D PCoA plot with each node representing the MS results (significant metabolites only) of only the Clade1 lysogens and the PAO1, which shows the dissimilarity between the metabolism of the Clade1 lysogens and PAO1 without the prophage. (b) Hierarchical clustering result shown as a dendrogram, showing Clade1 lysogens and the control compared. (c) Clustering result shown as a heatmap, representing the up and down regulation of each significant metabolite between the Clade1 lysogens and the control. A clearer image with readable labels is shown in figure 1b in appendix H. (d) The average score of the samples from each clade was taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. The lysogen the sample/S numbers are associated to is detailed in appendix I.

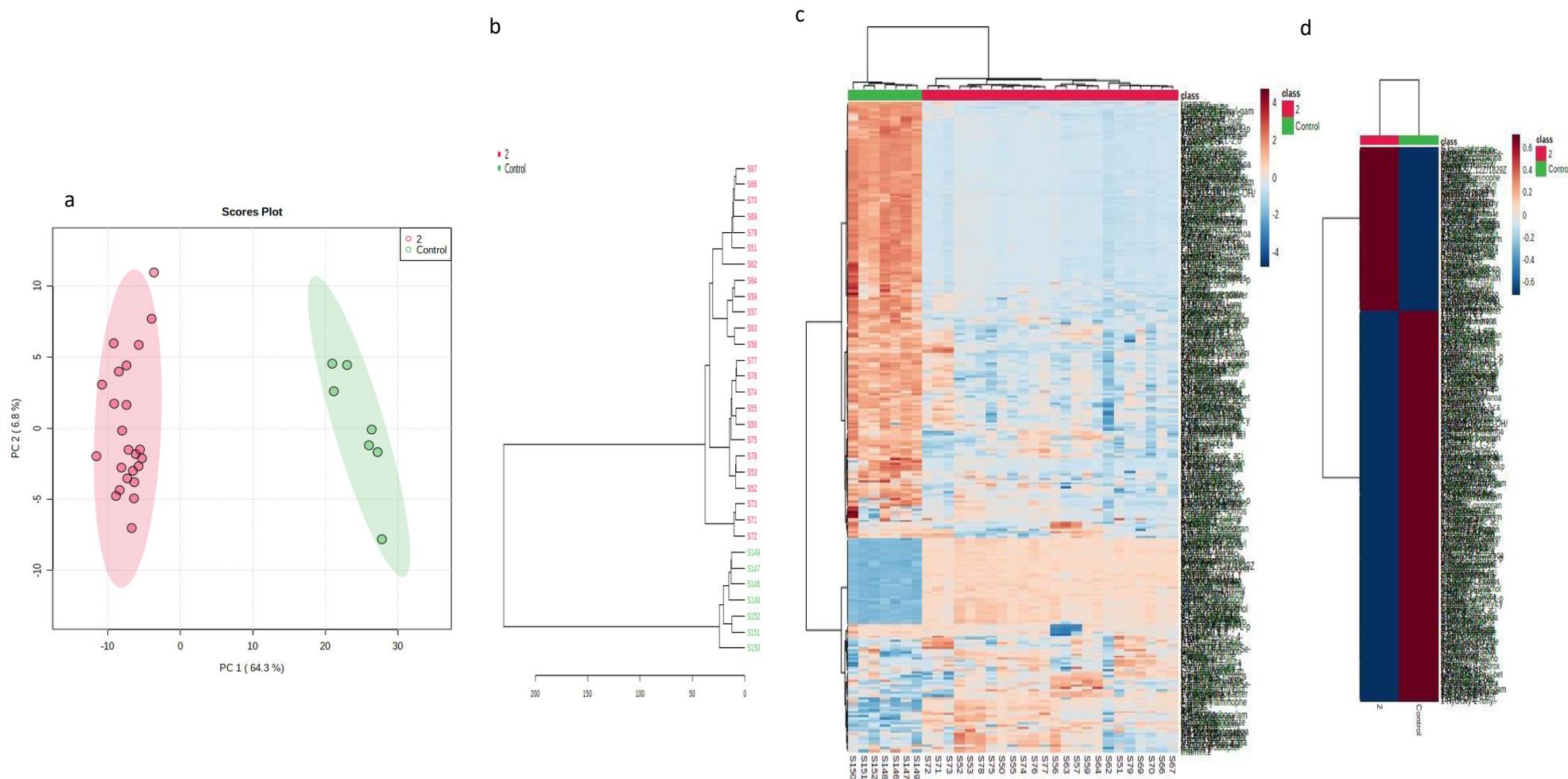


Figure 7.5 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 2 compared to PA14 control. (a) 2D PCoA plot each node representing the MS results (significant metabolites only) of only the Clade2 lysogens and the PAO1, showing the dissimilarity between the metabolism of the Clade2 lysogens and PAO1 without the prophage. (b) Hierarchical clustering result shown as a dendrogram, showing Clade 2 lysogens and the control compared. (c) Clustering result shown as a heatmap, representing the up and down regulation of each significant metabolite between the Clade 2 lysogens and the control. A clearer image with readable labels is shown in figure 1c in appendix H. (d) The average score of the samples from each clade was taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. The lysogen the sample/S numbers are associated to is detailed in appendix I.

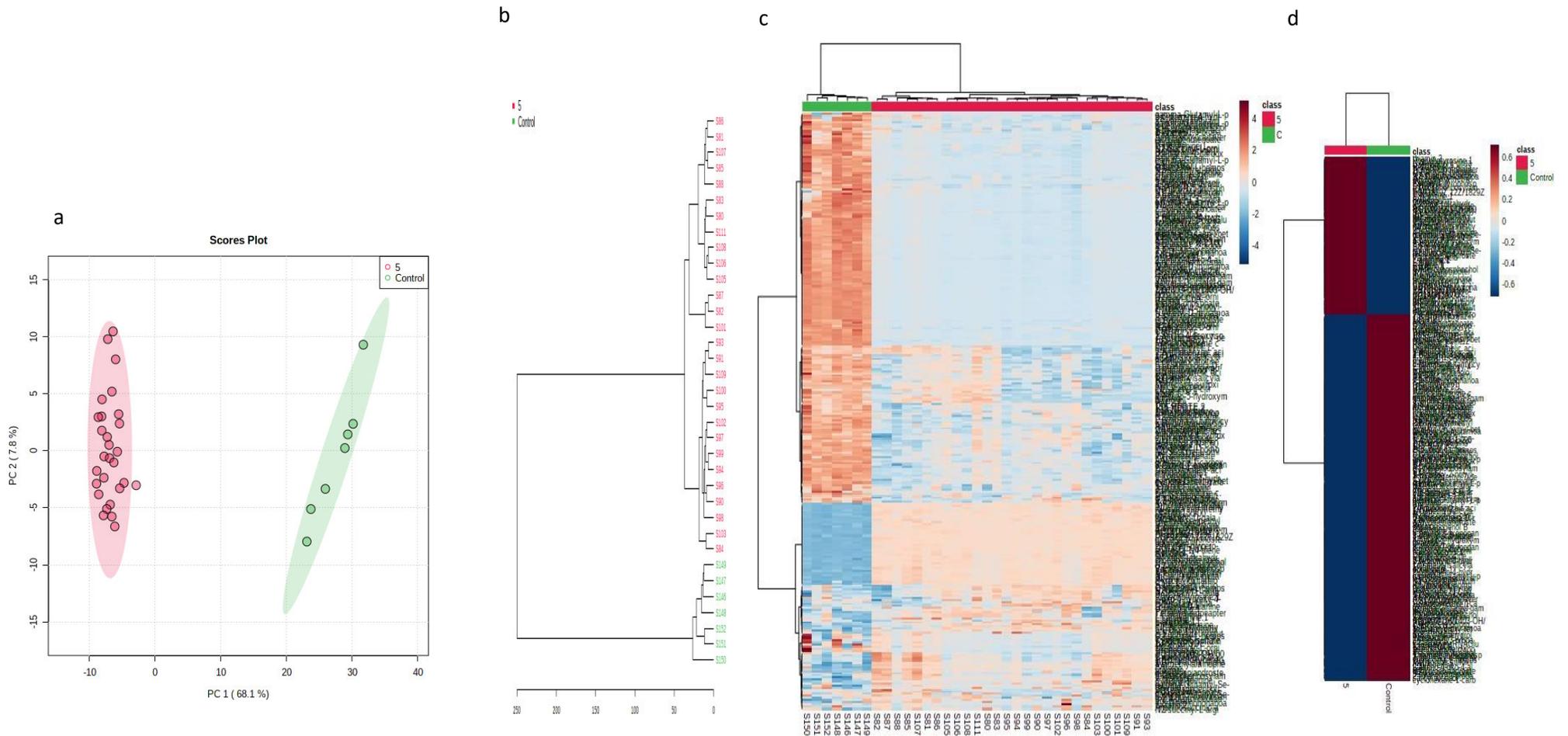


Figure 7.6 – PCoA, dendrogram and heatmaps of the metabolic profiles of the lysogens containing phages from Clade 5 compared to PA14 control. (a) 2D PCoA plot each node representing the MS results (significant metabolites only) of only the Clade 5 lysogens and the PAO1, showing the dissimilarity between the metabolism of the Clade 5 lysogens and PAO1 without the prophage. (b) Hierarchical clustering result shown as a dendrogram, showing Clade 5 lysogens and the control compared. (c) Clustering result shown as a heatmap, representing the up and down regulation of each significant metabolite between the Clade 5 lysogens and the control. A clearer image with readable labels is shown in figure 1d in appendix H. (d) The average score of the samples from each clade was taken for each metabolite and re-clustered as a heatmap to show the differences between the lysogens and the control more clearly. The lysogen the sample/S numbers are associated to is detailed in appendix I.

7.4.5 Pathway analysis of metabolomics

The data generated from MetabolAnalyst (statistical analysis) in figure 7.3 was then used for pathway analysis, performed also using MetaboAnalyst (pathway analysis) as per chapter 5, using PAMDB and metabolites only with associated KEGG IDs. Therefore, 94 metabolites out of the 292 significant metabolites were shown in the metabolomics results. Each significant metabolite with a KEGG ID in each of the lysogens from the clades was placed in its pathway and the total impact on each pathway was compared to that of the PA14 control. For each of the comparisons, the impact score for each of the pathways was translated into a heatmap, as shown in figure 7.8. These results show the pathways that were changed significantly (p value <0.05) in comparison to the PA14 control in at least one of the clades. It shows 33 pathways were significantly changed by the addition of phages from the 4 clades. The heatmap in figure 7.8 shows that the level of impact that clades 1, 2, and 5 have were all the same as one another for all 33 pathways. However, clade 8 was shown to impact the pathways at different levels, notably having a much lower impact on thiamine and lysine metabolism but a higher impact on retinol metabolism. This concurs with the metabolomic statistical results showing that clade 8 has a different metabolic profile to the other clades. However, clade 1 also looked to have a different metabolic profile to clades 2 and 5 from figures 7.1 and 7.2, but this is not represented by the pathway analysis results.

The pathway results only show if the pathway has been impacted (metabolites within that pathway have significantly changed between the lysogen and the control) rather than if it is up or down regulated. Therefore, the lysogens may have differing effects on the pathways to one another but impact the same pathways.

By referring to the heatmaps and identifying the metabolites that are in the pathway, it could be possible to work out for some of the pathways if the impact is referring to up or down regulation. However, for many results, where the metabolites from each pathway had some

up regulation and some down regulation, further research would be needed to target the pathways to determine greater resolution of the regulation of specific metabolism pathways. The pathways that were impacted the most by phages in clades 1, 2 and 5 (top 6) were; 'One carbon pool by folate', 'Arginine biosynthesis', 'Glutathione metabolism', 'Thiamine metabolism', 'Biotin metabolism' and 'Arginine and proline metabolism'. As clade 8 showed a distinctly different metabolic pathway profile, the most impacted pathways (top 6) by the phages in clade 8 were as follows in order of impact; 'One carbon pool by folate', 'Retinol metabolism', 'Phenylalanine, tyrosine and tryptophan biosynthesis', 'Riboflavin metabolism', 'Arginine biosynthesis' and 'Glutathione metabolism'. There is some overlap between clade 8 and the other clades in terms of the top 6 most impacted pathways, showing their importance as possible core pathways that change in response to phage infection.

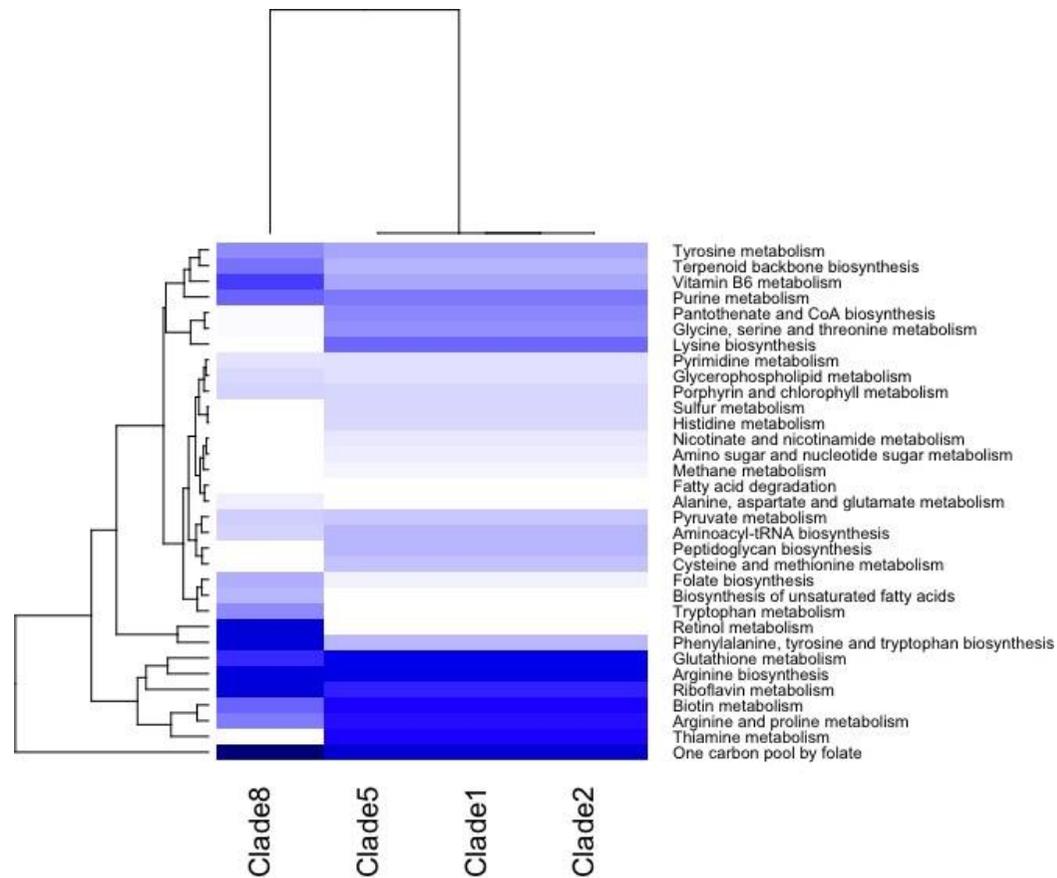


Figure 7.8 - Pathway analysis showing the ‘impact’ of the phages in each clade on the metabolic pathways compared to the PA14 control without phages. Shown as a heat map, darker blue represents more significant changes to the pathway i.e. more metabolites are changed within a pathway or a significant metabolite to that pathway has been affected compared to the control. This data was collated from running the MetaboAnalyst pathway analysis programme on each clade against the control and comparing the impact values against the control to enable comparison between phages and how they impact metabolic pathways differently.

7.5 DISCUSSION

The main aim of this study was to investigate the effects of temperate Pa phages diversity on the metabolism of their Pa host when converted to a lysogen. There were 8 clades identified from the prophages in the isolates from the IPCD in the previous chapter (chapter 6). These clades were based on the SaturnV results that clustered the prophage genomes at the protein level. As the phages in the same clade were the most similar at the protein level, we hypothesised that the lysogens generated from phages in the same clade would subvert their host in a similar way. Here we attempted to integrate phages identified from each clade into a common well-defined Pa strain to allow direct comparisons between the phages in each clade and how they alter the metabolism of a common host.

7.5.1 Lysogen formation

Firstly, lysogens containing phages from the 8 clades for metabolomic analysis needed to be formed. All phages that were selected for lysogen formation were from the IPCD from Pa isolates that were taken from lung infections, to keep in line with the overarching aim of this project and investigate the role of temperate Pa phage in chronic Pa lung infections.

The phages were induced from their original Pa isolate using NFLX as this has been shown to be effective at inducing Pa phages (Fothergill et al., 2011). However, Pa isolates are commonly resistant to fluoroquinolone antibiotics which in turn could prevent induction. The spot assay results displayed that several of the lysates showed no zones of clearance to any of the three Pa lab strains tested against, suggesting they either may not have been induced or were unable to infect any of the lab strains. Another option to induce prophage, if the isolate is resistant to NFLX, is Mitomycin C induction as it has a different mode of action to NFLX (Iyer and Szybalski, 1964) and is widely used as an inducing agent (Ramirez et al., 1999, Otsuji et al., 1959, Dziewit and Radlinska, 2016). The three lab strains were chosen as they have shown to be highly infective with the previously induced temperate phages, increasing the chances of lysogen formation. These Pa strains are also well studied, therefore the metabolites produced by them are more likely to provide matches on

the PAMDB.

The phages may behave differently in the lab strain to its original host. However, for comparative studies it is important to have the same metabolic background being affected by the integrated prophage. PA14 was able to be infected by the broadest range of phages (26 out of 69), therefore was used as the host and metabolic background for these phages to be compared in metabolomic analysis. PA14 has a degree of conservation compared to PAO1 (used in chapter 5 as the host Pa). However, the PA14 genome also contains two pathogenicity islands that are absent in PAO1 and was isolated originally from a chronic lung infection (He et al., 2004). This may make PA14 lysogens more clinically relevant to the chronic lung infection than PAO1 lysogens. As PA14 is a highly virulent strain opposed to PAO1, which is a moderately virulent strain, it would be of interest to determine whether the addition of prophages could also lower the virulence as seen for PAO1 lysogens in chapter 5.

The conversion of the 26 PA14 lysogens needed to be confirmed to ensure phages were present in the genome. In chapter 5, lysogeny was confirmed by genome sequencing, performed as part of previous studies (Tariq et al., 2019, James et al., 2012). However, the genome sequence of the target phages was already known through PHASTER identification, therefore confirmation of lysogeny could be performed by PCR, which has been used previously to detect prophages within bacteria (O'Sullivan et al., 2000, Zaburlin et al., 2017). ROARY was used to identify the most common genes carried by each clade to select the target gene to make clade-specific primers. Optimisation of the PCR and confirmatory amplification of the putative lysogens determined 16 individual phage infections in PA14 had been created from the 26 samples. These lysogens contained phages that covered 4 of the 8 clades. However, the 16 lysogens contained phages from four clades (1, 2, 5 and 8), which represents the largest number of different types of prophages in a common background to be compared by metabolomics. The lower-than-expected numbers of confirmed lysogens could be due to any phage resistance systems. For example, lack or modification of the host cell receptor required for adsorption. (High et al., 1993) or CRISPR

(Mojica et al., 2005). It could also be due to the original isolate containing an inducible phage that was not classed as intact by PHASTER. Therefore, the phage may not be from the expected clade and, instead, showed to be negative for phage infection from the PCR result, but showed a positive spot assay. With time, phages from the other clades could be infected into a common Pa to see if the phages from the other clades provoke the same or a different metabolic shift in their Pa host that is seen here.

7.5.2 Metabolic subversion of Pa by different types of prophage

All the phages in the lysogens in this study originated from Pa isolates from lung infections. and were grown in ASM media for metabolomic analysis, as in chapter 5, to more closely model the conditions in the infected lung that the lysogen originated from.

The results from the LC-MS identified 2667 features across the samples with 1647 of these being significant metabolites. This high number of significant metabolites could be due to there being 16 different variables (lysogens) were tested against one another. Therefore, there is larger number of phages that can affect the metabolites in different ways, compared to the LES lysogens where there were only 6 different variables/lysogens being compared and a lower number of significantly changed metabolites.

A metabolic shift was observed from metabolic analysis of the streamlined 292 metabolites when phages from the different clades were added compared to PA14 control. This supports the results seen in chapter 5 with the LES phages and the clinical CF and BR phages. There was also a marked difference in the metabolic profiles between some of the clades, such as clade 1 and 8. This demonstrates that temperate phages from the different clades can manipulate the Pa host metabolism in different ways. By looking at figure 7.2c, which shows a heat map of the up and down regulation of each of the metabolites, comparisons can be made between the clades. It also enables identification of metabolites of interest, which may be upregulated in one clade but not in the others, such as d-arabitol, which was highly upregulated in lysogens with clade 1 phages in but downregulated in the other 3 clades.

This work provides the basis for further investigation of specific target metabolites. This metabolic analysis of the addition of temperate phages has not been performed previously, therefore no comparisons can be made to the wider literature. However, it was expected that the prophages would change the metabolism of the Pa but it was not known to what extent, due to our experimental data showing the effects prophage addition has upon Pa. Phenotypic changes can occur, such as alterations to mucoid and over production of alginate, as well as and increased fitness for the host, all of which require an alteration of host metabolism and or gene expression to take place (Davies et al., 2016a, Miller and Rubero, 1984, Hentzer et al., 2001). To further this work, tandem MS would have to be performed to confirm the metabolites have been accurately identified. This involves a further fragmentation step making it possible to identify and separate ions that have very similar m/z-ratios in regular mass spectrometers.

The pathway analysis results showed from the metabolomics results there were 33 pathways that were significantly impacted by one clade of phages or more when compared to the control. The pathways that were shown to be the most impacted by phages in each clade were stated in section 7.4.5. Unexpectedly, 3 out of the 4 clades were seen to have the same level of impact on all the 33 pathways. Conversely, the metabolomics PCoA and heatmaps showed that clade 1 and 8 to cluster away from clade 2 and 5, suggesting they had shown a different metabolism. However, clade 8's pathway results did show that clade 8 phages impacted the PA14 pathways in different levels compared to the other clades. These results may be due to the lack of databases, as only metabolites with KEGG numbers could be assigned to a pathway, therefore, only taking 32% of the metabolites analysed previously into account. There are pathways that are impacted by all of the clades, possibly showing a core set of pathways that are impacted in some way by many temperate phage in Pa, similar to the pathway results of the clinical phages in chapter 5. In addition, different phages from the same clade affected the metabolism of the host in a similar way, suggesting the type of phage may have an impact on which metabolites are affected.

Some notable pathways that may benefit the phage and/or the Pa, which were highly

impacted by the phages in all clades, and therefore maybe seen as core pathways impacted within this cohort, were the one carbon pool by folate pathway, which supports multiple physiological processes. These include biosynthesis of purines and thymidine, amino acid homeostasis glycine, serine, and methionine. This pathway is impacted in all the lysogens, as purine and thymidine are needed by prophages for the construction of DNA and RNA, which is needed to make phage DNA as well as bacterial DNA. The biotin metabolism pathway again was impacted in all the lysogens, as seen in chapter 5. This pathway, in *E.coli*, is rate limiting to cell growth and also promotes broad-range antimicrobial tolerance by increasing cell wall lipids (Holt et al., 2017), which could also be the case in Pa. The arginine biosynthesis pathway and the arginine and proline metabolism pathway are all impacted by the addition of phages from all 4 clades. Arginine and proline metabolism is one of the central pathways for the biosynthesis of the amino-acids arginine and proline from glutamate (Isaac and Holloway, 1972). Similarly, the arginine biosynthesis pathway leads to the production of arginine, but is synthesised from citrulline in arginine and proline metabolism by the sequential action of the cytosolic enzymes arginosuccinate synthetase and arginosuccinate lyase (Isaac and Holloway, 1972). Arginine and proline are necessary amino-acids for the prophages to build their own proteins, but also necessary for the host. Glutathione metabolism has been shown to be important for the expression of Pa type III secretion system (T3SS) genes. It has been suggested that glutathione serves as an intracellular redox signal, sensed by Vfr protein, to upregulate T3SS expression in Pa (Zhang et al., 2019). T3SS expression is associated with increased virulence of Pa in humans and in animal models, suggesting the upregulation of glutathione metabolism may increase the T3SS and therefore increase virulence of the Pa (Zhang et al., 2019). Glutathione was highly impacted by all lysogen groups suggesting that carriage of these prophages may increase virulence of the Pa compared to the PA14 control.

Thiamine metabolism plays important roles as (TPP) is a cofactor in various essential cellular processes, including carbohydrate, lipid, and amino acid metabolism. Thiamine monophosphate kinase (ThiL) catalyses the final step of the pathway by phosphorylating TMP to TPP, which is the biologically active form of the cofactor. It has also been shown in Pa that the pathogenesis of the Δ thiL mutants were markedly attenuated compared to wild-type bacteria, suggesting that the thiamine metabolic pathway plays a role in Pa pathogenesis (Kim et al., 2020). As this pathway was impacted by the addition of phages in the majority of the clades, this could mean that the phages are impacting this to affect the pathogenicity of the Pa.

There were drawbacks to this pathway analysis method (mentioned in chapter 5), as comparisons could only be made to pathways of metabolisms that are within its database. Therefore, the *Pseudomonas putida* metabolism had to be used as the closest to the Pa metabolism. Also, only metabolites that have an associated KEGG number were taken forward for the pathway analysis (32% of metabolites), as those without a KEGG number were not recognised. This may lead to missing the intricacy of the analysis due to not being able to offer a pathway to a large proportion of the metabolites determined. This could be the reason that the heatmap showing the impact each clade has on each metabolic pathway (Figure 7.8) shows that 3 out of 4 clades have the same impact on all 33 of the pathways. These results were still insightful and enabled inferences to be made about how the different clades of phages may differ from one another in terms of how they affect their host Pa, as well as pinpointing metabolic targets of interest that can be verified experimentally with functional assays, such as biotinylation, changes in fatty acids synthesis and alteration of fatty acids at the cell surface that can be analysed by lipid profiling (Han et al., 2018).

Although these metabolomic results are based on a single MS, they are the preliminary results and untargeted. Future studies are now able to focus now on these individual metabolic pathways at a greater resolution. As there are few studies on the metabolic effects of temperate phages (Hargreaves et al., 2014, Feiner et al., 2015) and no other published

work looking specifically at Pa, nor temperate Pa phages from the chronically infected lung, this illustrates the importance of this research. Further study into the metabolites altered by the addition of phages will allow us to identify if any of these changes to the host's metabolism gives rise to enhanced fitness (Davies et al., 2016a), antibiotic tolerance or resistance (Burgener et al., 2019) or change in virulence of the Pa (shown in chapter 5), which could then enable the Pa lung infection to become chronic. If these characteristics that benefit the Pa can be associated with a particular phage type or phage gene, it could be used as a marker for severe infection, as done with Pf phages.

7.6 SUMMARY

This study is the first to show that different types of Pa temperate phages can subvert the same host in different ways, therefore, phage diversity can have an impact on how they change their Pa host's metabolism. Also, there may be core changes in the metabolism in PA14 when prophages are added to the genome, further supported by the analysis of metabolomics data from the clinical CF and BR lysogens in chapter 5.

8 COMPARISON OF *PSEUDOMONAS AERUGINOSA* TEMPERATE BACTERIOPHAGES ISOLATED FROM CHRONICALLY INFECTED NON-CYSTIC FIBROSIS BRONCHIECTASIS PATIENT ISOLATES, OVER A PERIOD OF 10 YEARS

8.1 INTRODUCTION

The previous chapter illustrates the diversity of temperate phages from a large cross-sectional repository of *Pseudomonas aeruginosa* (Pa). It illustrates the diversity across Pa from multiple environments that begins to illustrate the differences that relate to environmental adaptation, selection and evolution. To focus on how phages change in the lung over time, this longitudinal study focuses on temperate bacteriophages in the lung environment of patients diagnosed with non-CF Bronchiectasis (BR). Over a period of 10 years sputum samples were collected from BR patients at the Freeman hospital, Newcastle, at a frequency that linked to patients presenting at clinic with exacerbation of symptoms. These samples were then screened for the presence of bacteria and if Pa was identified, isolates were stored at -80°C. This collection is ongoing with new patients being added regularly and supplementing each patient panel. Through collaboration with Dr Anthony de Soyza (who these patients are under the care of) and Audrey Perry and John Perry (who manage the collection at the Freeman hospital), access to these Pa samples was permitted for this study. At the time of the study there were 27 patients that had more than one sample and a total of 220 Pa isolates in the collection. By looking at the Pa isolates longitudinally in patients with chronic Pa lung infections and with 2 or more samples, it may be possible to illustrate temperate phage genomic change that may be linked to adaptation and evolution of Pa in the lung.

8.1.1 Longitudinal studies of Pa from chronically infected BR patient lungs

There have been multiple longitudinal studies on the evolution and patho-adaptation of Pa infections in CF patients (Winstanley et al., 2016, Burns et al., 2001, Bartell et al., 2019),

whereas there are only a small number focusing on Pa infections in BR patients (Hilliam et al., 2017, Woo et al., 2019). Much of the current understanding of Pa lung infections in BR has been inferred from data derived from CF patients' chronic respiratory infections. Observations have been made regarding Pa infections in CF and also inferred to BR, until recently, as they have a similar pathophysiology of infection. Some of the observations that were made in CF and assumed to be the same in BR include, the strains seen in adult chronic Pa airway infections are generally stable, with the exception of rare events of acquisition and eventual replacement or co-infection of epidemic Pa strains (Aaron et al., 2004, Williams et al., 2015). Also, transmission of Pa infections is possible and may be associated with adverse clinical consequences (Fothergill et al., 2012). Again, this is only seen in CF and has yet to be investigated in BR. The few studies on Pa from BR have focused on investigating these trends, such as the rate of co-infection and change in strain/replacement on Pa strain in BR patients (Hilliam et al., 2017, Mitchelmore et al., 2017), as well as the evolution of Pa from BR patients. A recent study by Woo et al (2018) was the first extensive longitudinal study to look at Pa from BR compared to CF samples, to compare between findings (Woo et al., 2018). They observed several distinctions in the epidemiology of Pa infections in BR compared to CF, one being Pa infections in CF adults are usually clonal and strain displacement is not seen, whilst unique strains were observed more commonly in BR. Where displacement was observed it was not associated with pulmonary exacerbation events, as previously observed in CF through acquisition of a new strain. They also observed that there was no evidence of cross-infection amongst BR patients, whereas cross-infection of Pa has been observed in CF and measures put in place to prevent it (Woo et al., 2018).

In this study, the aim is to not only investigate displacement of Pa strains from BR patients longitudinally, but also consider the possibility of displacement of prophages in the genome. In addition, the transfer of prophages between Pa strains, if there is bacterial displacement occurring, will also be investigated.

8.1.2 Evolution of temperate phages in *Pseudomonas aeruginosa* over time

The study of temperate phages compared to lytic bacteriophages is less common but some work has focussed on prophages and how they lead to the evolution of the bacterial Pa host (O'Brien et al., 2019, Davies et al., 2016b). However, there have been fewer studies investigating the evolution of temperate phages within a Pa bacterial background (James et al., 2015). As a result of phage evolution being complex and their genomes being composed of genes with distinct and varied evolutionary histories (Hendrix et al., 1999), it makes the comparison of phages and determining their origin difficult. Phage evolution happens at a high rate, therefore changes can occur even over a short period of time, meaning phages can evolve to overcome bacterial anti-phage mechanisms, which allows phages to infect bacteria more effectively (Murphy et al., 2013). In addition, they may also acquire genes from HGT or recombination, which may be advantageous to the bacteria and prolong the prophages survival in the genome. Here we look at the changes of the prophages in Pa over a period of chronic infection and disease progression to observe if there are any changes to the phage genome that may show evolution in this way.

8.2 AIMS

- Use cross-infection results of induced phages from the longitudinal Pa isolates panel and infect against the same longitudinal Pa isolates from BR patients (n=220 from 27 patients). The aim, if adaptation in the host range can be determined, would be to track the evolution of the bacteria in the lung environment over time.
- Use cross-infection results to select patient panels and isolates that may have temperate phages of interest to study further.
- Use twitching motility to show different phenotypes and possible different strains of Pa.

- Compare bacterial genomes sequenced between the isolates taken from each patient over time to see if they are infected by the same or a different strain of Pa.
- Identify prophages within each of the Pa genomes that are sequenced and compare the diversity of all the prophages identified.
- Compare the phages between isolates from the same patient longitudinally.
- Sequence phages to ascertain which phages identified are inducible.

8.3 OBJECTIVES

- Cross infect the induced phages from the Pa isolates against the other isolates from the same patient to ascertain if phages isolated at a later stage of disease progression can more readily infect the isolates from an early stage of the disease.
- Select patient isolate panels and isolates within them that may have temperate phages of interest to study further by using cross-infection data and resistance data supplied by the hospital and subsequently sequencing them.
- Use pairwise comparison to compare the phage genomes between isolates from the same patient longitudinally.

8.4 RESULTS

8.4.1 Selection of longitudinal Pa isolates to sequence

Our collaboration with Professor Levesque at the University of Laval, with the Microbiology Society funding a visit for 3 months to use their informatic tools and receive training, offered the ability to sequence and analyse 60 Pa isolates from the longitudinal collection. From the 220 isolates, we streamlined to 60 isolates from 6 BR patients, with each patient presenting at clinic on multiple occasions (a minimum of 6 times). These isolates were selected to maximise the longitudinal carriage of phages. Due to the limitation on the number of samples, patients were selected with ≥ 6 samples spanning greater than 3 years. The clinical metadata for each patient was also taken into account, including antibiotics used at

the time the sample was isolated (data provided by the hospital), detailed in appendix J. Importantly, all samples that were selected for each patient had samples taken before, during and after the administration of different antibiotics. The selected isolates span the disease progression of the patient, illustrated in figure 8.1. The figure shows that patient B has had samples collected for the longest time period, spanning from 2008 to 2017, and patient F the shortest selected. During occasions where multiple colony morphologies of Pa from the same time point were seen, samples were selected for the panel and are shown as overlapping points (Figure 8.1). The prophage carriage of patients harbouring Pa of multiple colony morphologies at a single time point was investigated initially in chapter 4. Therefore, if these non-clonal isolates display similar prophage carriage findings, this would complement the previous work in chapter 4 on carriage of temperate bacteriophages in non-clonal Pa within the lung.

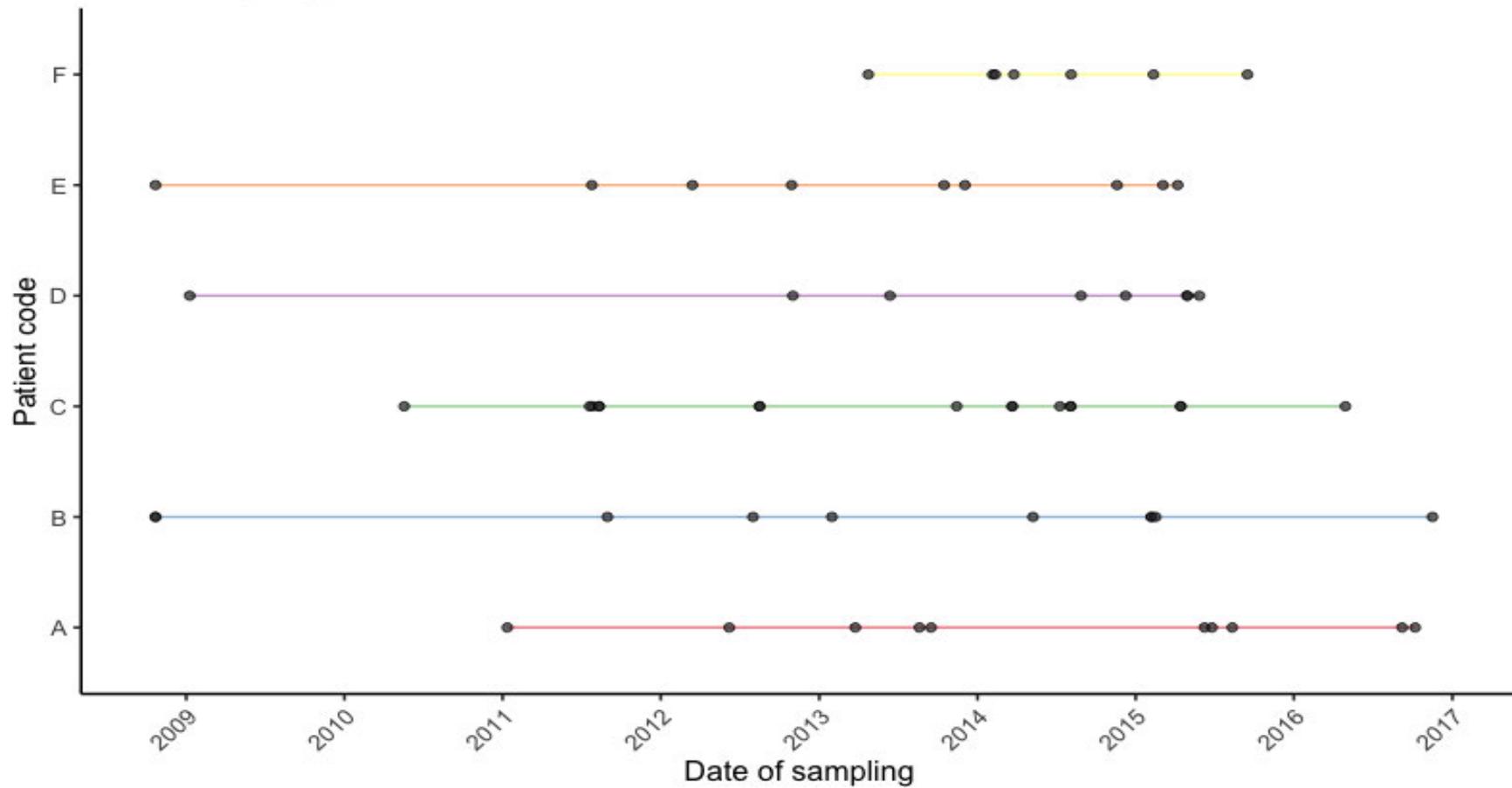


Figure 8.1 – Disease progression and sample collection of 6 BR patients. Points show when the samples (that were taken forward for sequencing) were taken at points of exacerbation. A complete list of samples that were taken for these patients is shown in appendix J. Where multiple samples were taken at the same time point, 2 overlapping points are seen.

8.4.1.1 Antimicrobial resistance testing of the Pa chosen for sequencing

The antibiotic resistance profiles for the 60 isolates against a range of antibiotics were obtained from the Freeman hospital and are shown in figure 8.2. They showed that the Pa isolates taken from the same patients longitudinally have differing resistance profiles, suggesting that the Pa isolated may be different at each time point, despite being chronically infected. Through selective pressure of multiple antibiotic challenges, it is possible that the Pa has adapted and evolved over time to exhibit or acquire tolerance or resistance to an antimicrobial. However, some patients' samples have shown the opposite results, showing an isolate having resistance to a particular antibiotic and a later isolate being sensitive to the same antibiotic. If the isolate is the same as the previous sample, it is possible if the selection pressure is taken away, i.e. it is no longer treated with that antibiotic (due to it being resistant), it becomes sensitive to the antibiotic again through evolution.

This microbial data, coupled with the patients' clinical data, can be used to see if any of the antibiotics that the bacteria acquired resistance to are from those previously administered to the patient.

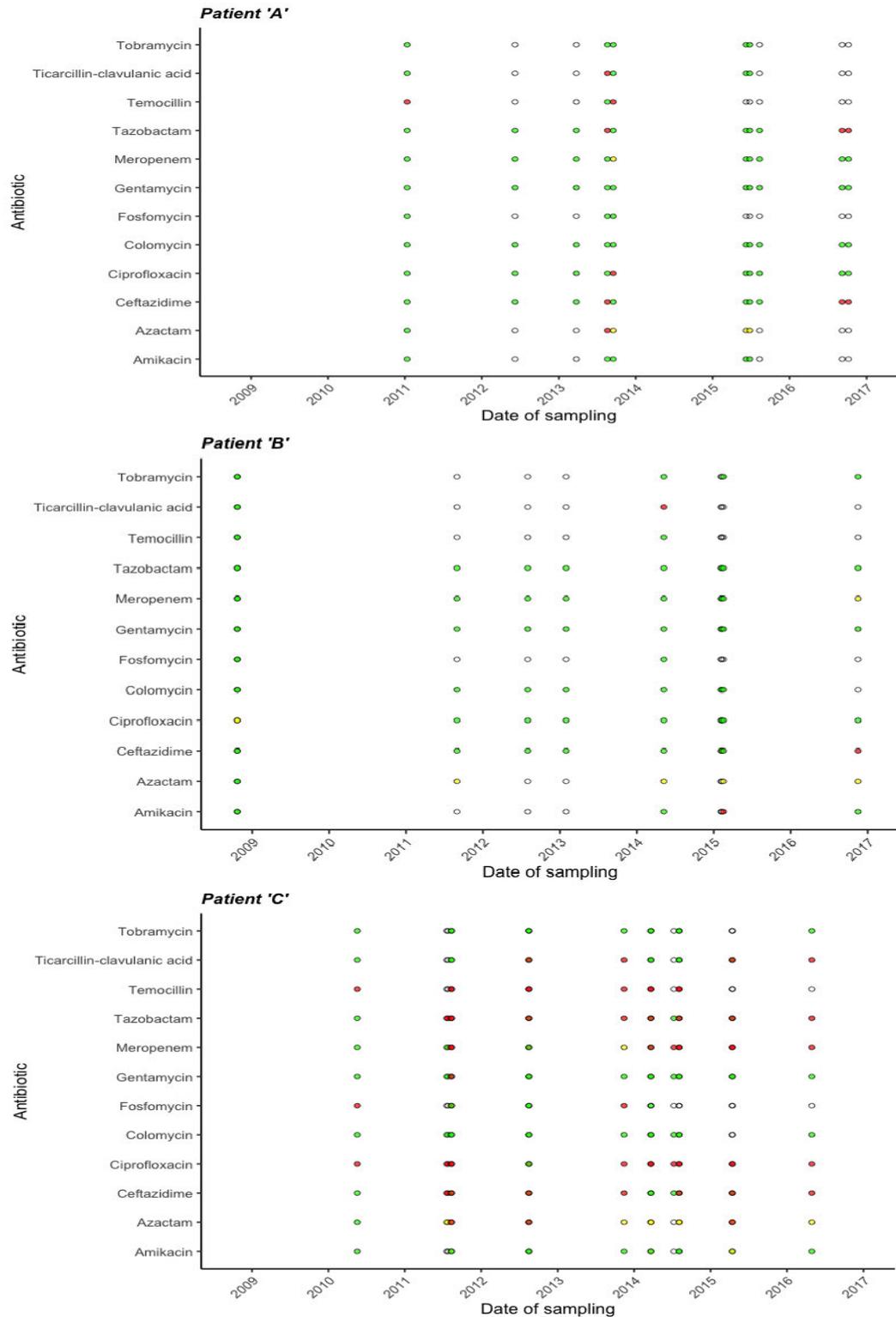


Figure 8.2a-c Antibiotic resistance results of the longitudinal Pa isolates from BR patients A-C that were taken forward for sequencing. The points match those in figure

8.1 to show when the isolate was taken and the colour shows if the isolate was resistant (red) or sensitive (green). Points which are not coloured indicate that the resistance was not tested.

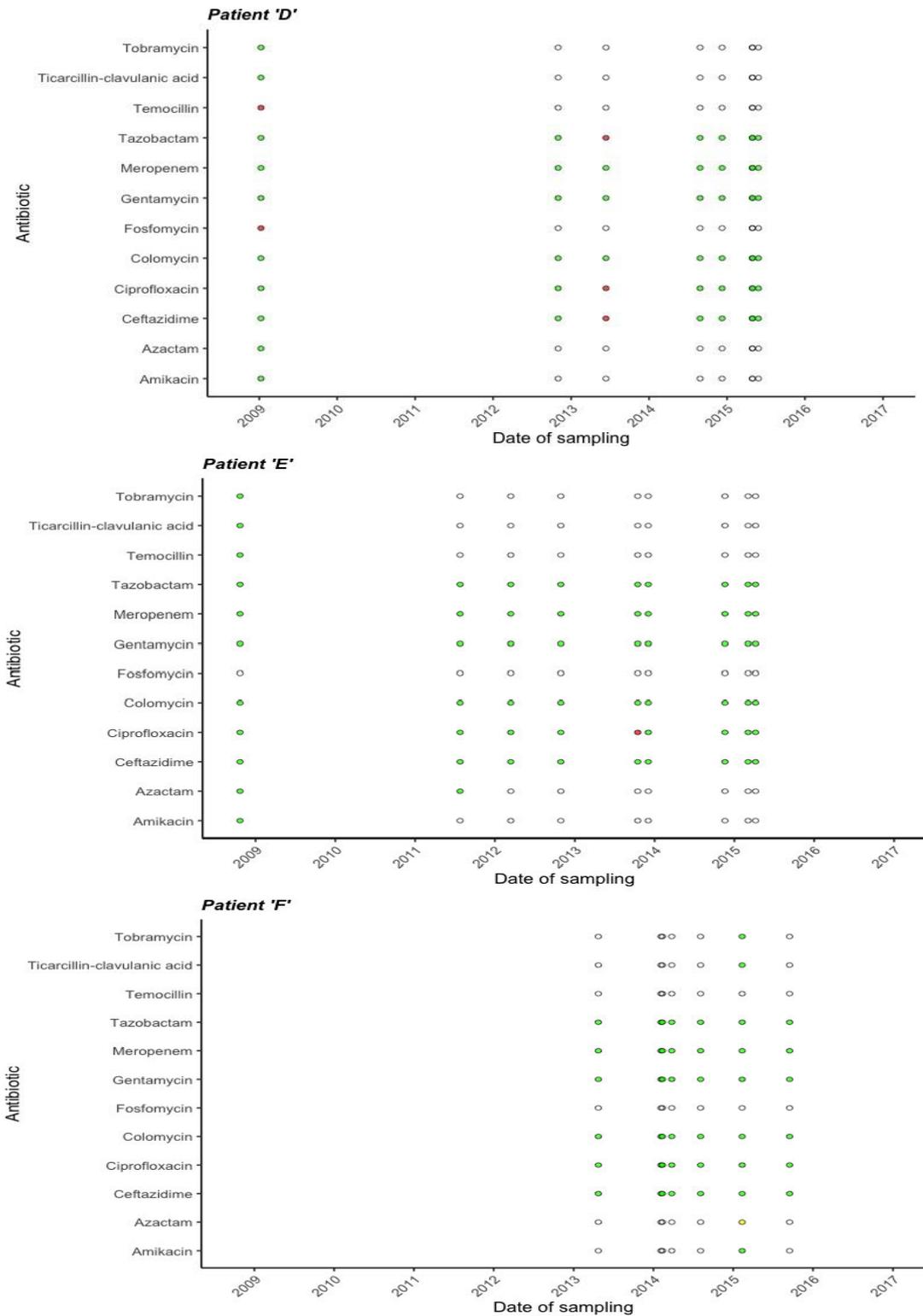


Figure 8.2d-f Antibiotic resistance results of the longitudinal Pa isolates from BR patients D-F that were taken forward for sequencing. The points match those in figure 8.1 to show when the isolate was taken and the colour shows if the isolate was resistant (red) or sensitive (green) and the points not coloured indicate that the resistance was not tested.

8.4.1.2 Twitching motility assays

The 60 isolates were tested for twitching motility to see if there were differences in motility of the isolates taken at each of the time points. If differences were seen it would suggest they were different strains or that the strain had adapted in the lung over time in some way that affected the motility of the Pa. Twitching motility assays are displayed in table 8.1. Images for the twitching motility plates for patients F, the positive control PAO1 and negative control LESB58 are shown in appendix K. These data show a gain or loss in motility and whether gain of motility is associated with displacement of Pa. There was change in the twitching motility between the longitudinal isolates in all patients except patient F. There were slightly more isolates that were positive for twitching motility with 32 being positive and 28 negative. However, there was not a trend for a patients isolate panel to turn either positive or negative over time, and the results show sporadic changes over time. Patient C's panel did show more motile strains towards the later stages of testing, with all 8 isolates collected in the final 2 years showing motility, where it had fluctuated prior to that, suggesting possible stabilisation in that phenotype.

Table 8.1 – Twitching motility of longitudinal Pa isolates from BR patients

Isolate ID	Patient	Isolation date	Twitching motility
141	A	12/01/2011	Negative
550	A	07/06/2012	Positive
915	A	25/03/2013	Positive
1082	A	20/08/2013	Positive
1192	A	16/09/2013	Negative
2101	A	09/06/2015	Positive
2107	A	26/06/2015	Positive
2165	A	12/08/2015	Positive
2437	A	07/09/2016	Positive
2459	A	07/10/2016	Negative
83a	B	22/10/2008	Negative
89b	B	22/10/2008	Negative
253	B	31/08/2011	Negative
604	B	01/08/2012	Negative
844	B	30/01/2013	Positive

1544	B	09/05/2014	Negative
1956a	B	06/02/2015	Positive
1957b	B	06/02/2015	Negative
1970	B	16/02/2015	Negative
2514	B	16/11/2016	Negative
125	C	19/05/2010	Negative
209	C	27/07/2011	Positive
213	C	21/07/2011	Positive
240a	C	12/08/2011	Negative
241b	C	12/08/2011	Positive
664a	C	16/08/2012	Positive
665b	C	16/08/2012	Negative
1297	C	14/11/2013	Negative
1487a	C	22/03/2014	Positive
1501b	C	22/03/2014	Positive
1592	C	10/07/2014	Positive
1623a	C	04/08/2014	Positive
1624	C	04/08/2014	Positive
2009a	C	15/04/2015	Positive
2010b	C	15/04/2015	Positive
2334	C	29/04/2016	Positive
102	D	09/01/2009	Positive
727	D	01/11/2012	Positive
997	D	13/06/2013	Negative
1669	D	28/08/2014	Negative
1911	D	09/12/2014	Negative
2034a	D	30/04/2015	Positive
2035b	D	30/04/2015	Positive
2088	D	28/05/2015	Positive
79	E	22/10/2008	Positive
208	E	26/07/2011	Negative
464	E	14/03/2012	Positive
722	E	29/10/2012	Positive
1243	E	16/10/2013	Positive
1348	E	03/12/2013	Negative
1846	E	19/11/2014	Positive
1986	E	05/03/2015	Positive
2015	E	08/04/2015	Negative
945	F	24/04/2013	Negative
1427	F	05/02/2014	Negative
1434	F	11/02/2014	Negative

1481	F	26/03/2014	Negative
1620	F	05/08/2014	Negative
1954	F	11/02/2015	Negative
2207	F	16/09/2015	Negative

8.4.2 Genomic Diversity of longitudinal Pa isolates from BR patients

Of the 60 Pa isolates that were extracted for sequencing, 3 had degraded DNA with insufficient libraries and could not be sequenced. Table 1 of appendix L shows the numbers of sequences with >Q40 quality used and also illustrates the numbers of contigs generated for each assembly.

The 57 assemblies and genomes were first annotated using Prokka, and were then ran through ROARY, which takes annotated assemblies and calculates the pan-genome at the protein level. Using BLASTp, ROARY compares the gene products of each genome in an all-against-all manner, using the default cut-off of 95% sequence identity. This allows comparisons to be made between isolates from the same patient to see if the same strain was isolated each time or if there was displacement taking place. Also, observations were made by comparing all of the bacteria genomes against one another using BLASTn, to see if there were any strains seen in multiple patients, which may suggest cross infection. These data were utilised to create the dendrogram displayed in figure 8.3. The distance between the nodes denotes the Pa bacterial isolate's genomes once translated into proteins. This is beneficial as it negates silent point mutations that may be picked up in the genome over time which would be viewed as differences if compared at the nucleotide level. These results show that most of the isolates from the same patient are very similar. Each patient had a definitive clade of isolates that have similar gene carriage, which differentiates them from the other patients, implying they are the same strain with some mutations. However, there are some isolates from the patient panels that do not cluster together, such as B_Pa3, which shows this isolate is a different strain. This suggests that co-infection or displacement by another strain occurred and this is seen in at least one isolate from patients A, B, D and E. There is also evidence of cross infection as A_Pa1 clusters with, and is therefore most

similar to, isolates from patient E. Also, B_Pa1 clusters with, and is therefore most similar to, isolates from patient D. Interestingly, these isolates from samples taken later than the first isolates that were collected, proposing that these isolates may have been displaced by a new Pa strain in the patient lung.

A dendrogram was generated using the pan-genome tool ROARY, as mentioned previously, for each patient's isolates individually to see how different the isolates from each patient are in more detail and these are shown in appendix M figures 1a-f. This data shows which isolates are the most similar and we could hypothesise that diversity within the clade of bacteria should branch and relate to the longevity of colonisation, as more mutations are accumulated over time due to evolution.

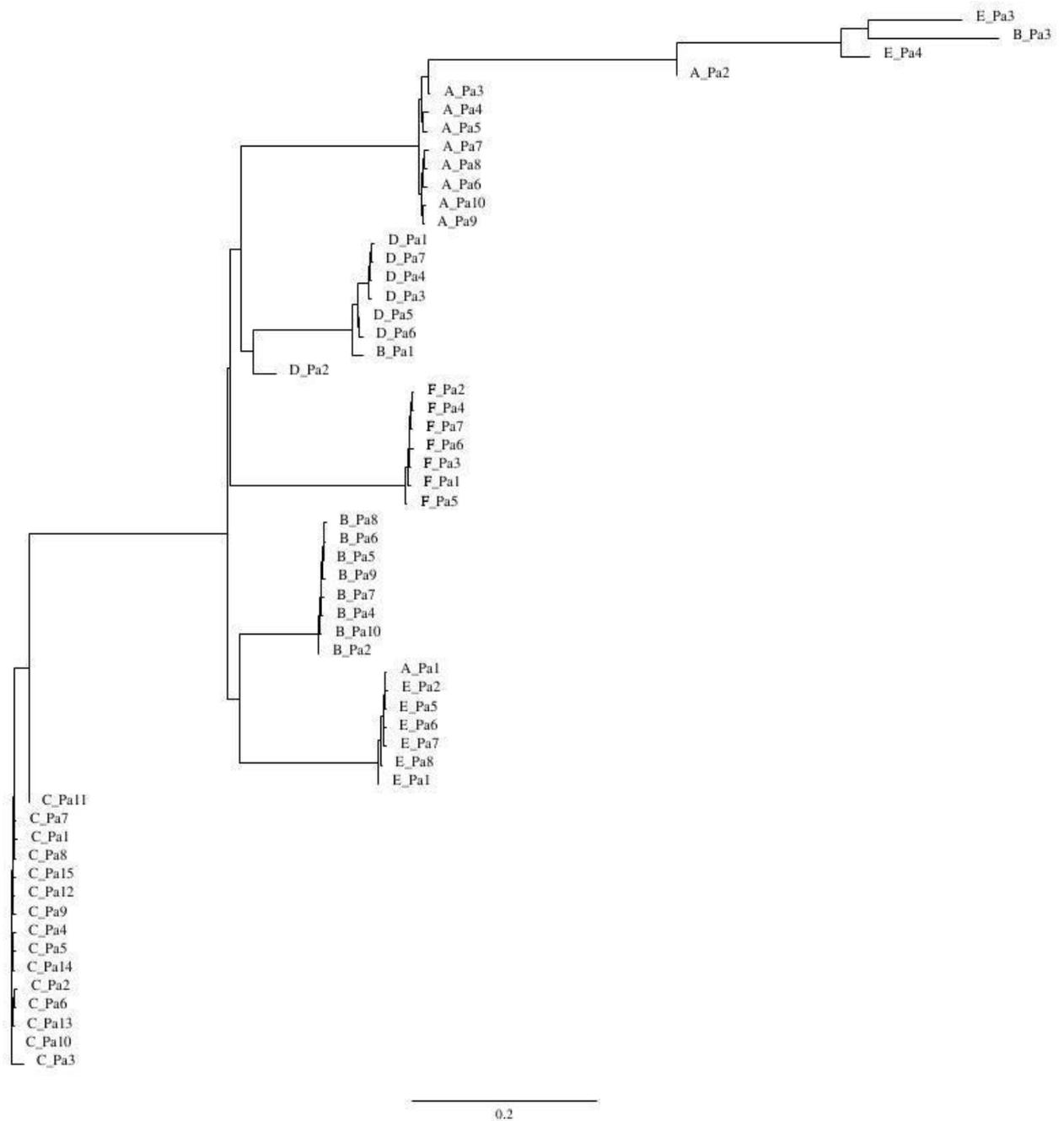


Figure 8.3 – Bacterial diversity of longitudinal Pa isolates from BR patients using ROARY pangenome analysis. ROARY generated bacterial genome comparison of Pa isolates from 6 patients. The labelling, such as A_Pa1, shows that the genome is from patient A and is isolate Pa1, which was the first isolate to be collected from that patient. Subsequent isolates are labelled Pa2, Pa3 etc. The length and subdivisions of the branches shows the relatedness/similarity of the genomes, with isolates with short branches being most similar. The scale bar represents the dissimilarity score.

8.4.3 Prophage identification from the longitudinal Pa isolates from each BR patient

From the 57 genomes sequenced, analysis with PHASTER predicted 92 intact, 101 questionable and 226 incomplete prophages. Firstly, there is a difference in the total number of prophages within longitudinal isolates from the same patient. This could illustrate accrual, displacement or a marker of different Pa being present. There were relatively low numbers of intact phages within this overall cohort of Pa isolates, ranging from 0-4 per isolate with higher numbers of incomplete phages from 1-8. There is not a trend of an increase or decrease in numbers of intact prophages over time, rather they are a set of prophages that are seen to appear and disappear from the bacterial genomes over time (Figure 8.4).

Figure 8.4 shows the number of intact phage regions in each isolate over time and which clade the phage is in, as determined by PHASTER (the identification of these clades for these phages are shown in the following section 8.4.4). This enabled visualisation of the differences in prophage carriage in the different Pa isolates over time. For example, in the phages from figure 8.4a, the same clade is seen throughout the longitudinal isolate series, which is shown to be an F10-like phage and is seen in the 2nd to the 10th isolate. Furthermore, it shows the addition of a *Burkholderia*-like phage in isolate 8 but is not seen in the subsequent isolates, indicating it was possibly gained but then lost before the next sample was taken.

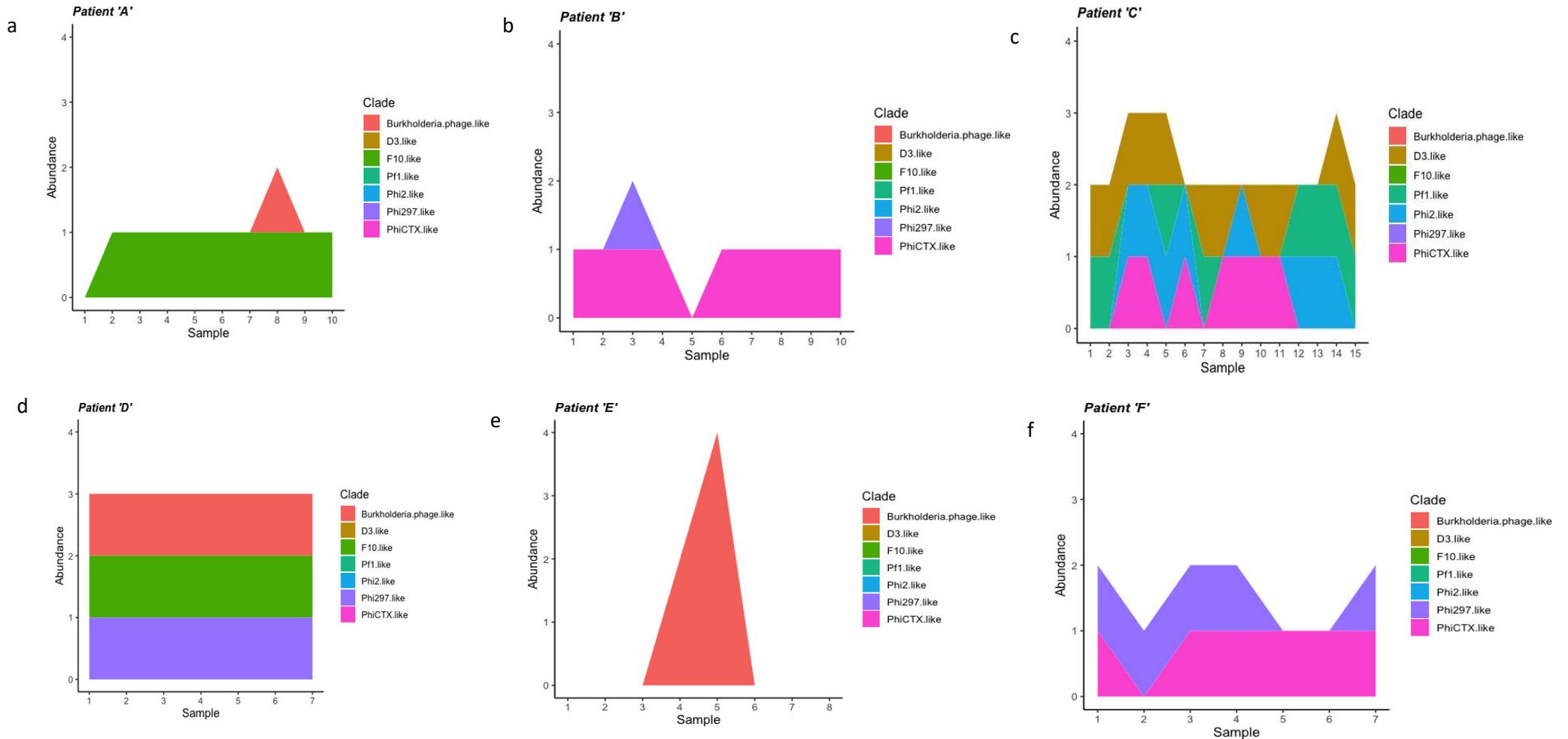


Figure 8.4 - Carriage of prophages by Pa isolates collected longitudinally from chronically infected BR patients. Sample denotes the sample number collected longitudinally, with 1 being the earliest, and abundance shows the number of different types/clades of phage in each of the isolates. a) Patient A's Pa isolates carriage of prophages of differing types. b) Patient B's Pa isolates carriage of prophages of differing types. c) Patient C's Pa isolates carriage of prophages of differing types. d) Patient D's Pa isolates carriage of prophages of differing types. e) Patient E's Pa isolates carriage of prophages of differing types. f) Patient F's Pa isolates carriage of prophages of differing types.

8.4.3.1 Prophage diversity within longitudinal Pa isolates from the same patient over time at the nucleotide and protein level

In the previous section we compared between the longitudinal bacterial isolates at a genomic level, focusing on the patient longitudinally. Here we use SaturnV to translate these putative prophage genomes to amino acids and compare the results of each phage region in a pairwise manner against all translated putative prophage coding regions. This is useful as it takes into account heterogeneity within a gene that does not offer a coding change. This was used to compare all the prophages identified in the 57 Pa isolates against each other (as carried out in chapter 6 for the prophages in the IPCD). A SaturnV generated cladogram is displayed using figtree in figure 8.5, which focuses on the diversity of all prophages identified in the bacterial genomes sequenced. The cladogram illustrates the similarity and level of diversity between the prophages identified. The patients are colour coded at the node level and each clade has been given the closest genome comparator from GenBank using BLASTn as its phage type (as carried out in chapter 6). The prophages mostly cluster in groups of the same colour (same patient), showing that prophages that are from different isolates from the same patient are similar. However, the spread of the cluster denotes that the prophage genomes are not identical and illustrates genome differences at an amino acid level. There are several phages illustrated in figure 8.5 that do not cluster with other phages from the same patient (same colour), which shows that it is not present in any other isolate from that patient. For example; patient B (blue nodes) has 1 phage (node) that cluster with patient A's phages (nodes) and 2 phages that clusters with patient D's. This suggests that the phage is a new phage that has infected the Pa isolate at that time-point or it is from a different strain of Pa containing new phages. Prophages are also lost as well as gained from one time point to the next. This suggests that the phage has been induced. The specific phages that were lost, and at which time-point, can be seen in figure 8.4. From these results, the phages that are present but are then not (disappear/induce) in the Pa isolate at the next time-point are frequently seen again in isolates further down the timeline. For example, in patient C the phage in clade 7 is present

in isolates 1 and 2, then not seen in isolates 3 and 4 and then present again in 5. This suggests that the phages are inducing and re-infecting, possibly due to the administration of antibiotics that may cause induction of phages.

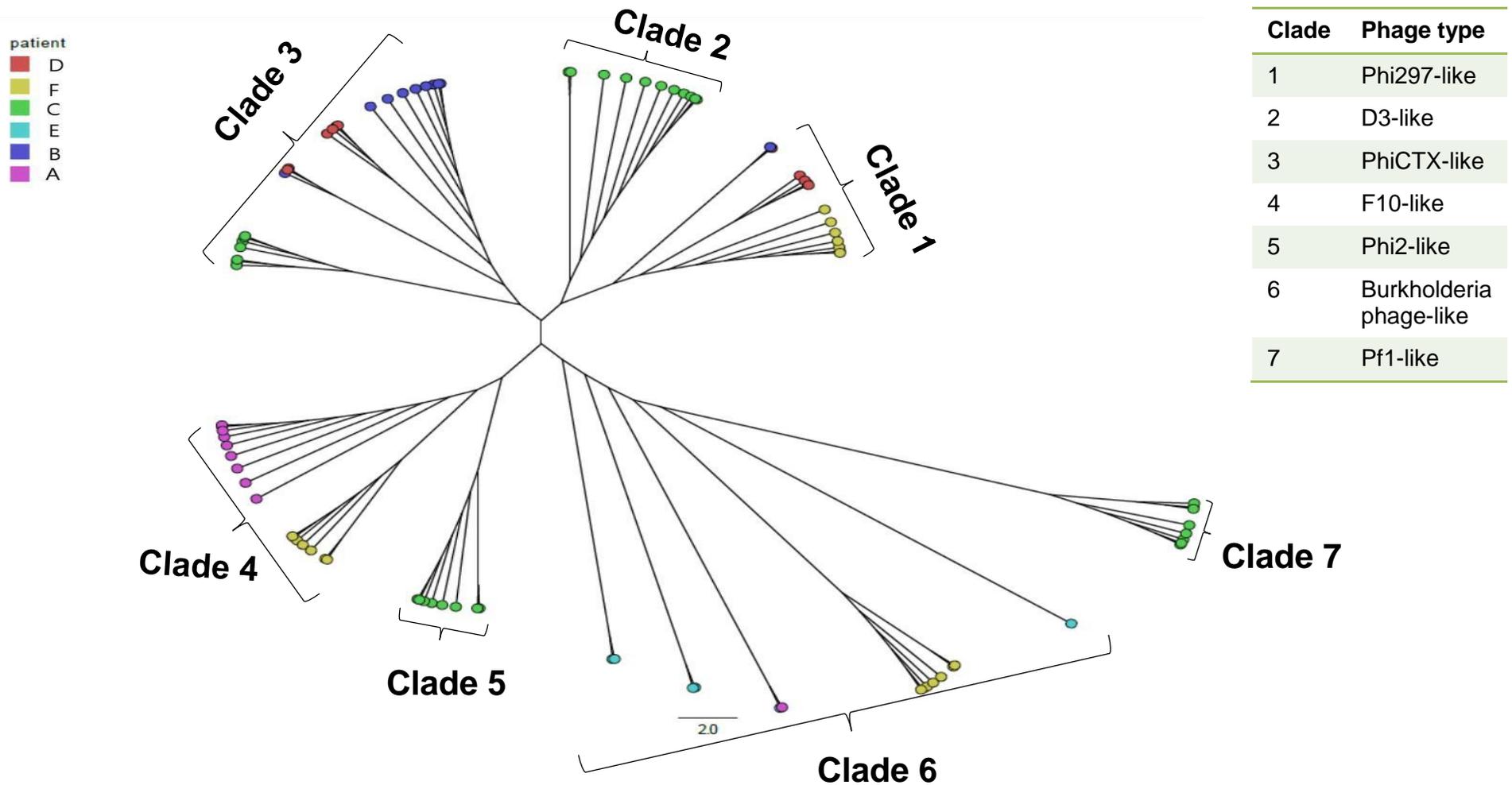


Figure 8.5 – Diversity of prophages identified in longitudinal Pa isolates from BR patient. Cladogram produced from SaturnV results showing the protein-wise comparison between prophages identified in the longitudinally collected Pa genomes of BR patients. The nodes are colour coded depending on the patient that the Pa the prophage was identified in. The clades show what type of phage the group are most similar to from Genbank and were compared using blast.

8.4.3.2 Comparison of the prophage genomes carried by the *Pa* longitudinal isolates from each patient

The previous sections in this chapter have illustrated that there is a shift in both bacterial and prophage genomes that links to the longevity of bacterial infection of the lung. Pairwise comparisons of the phages that are present longitudinally may uncover additions or deletions from their genomes that are important to support the phages or bacterial longevity in the lung. The comparison tool Nucmer was used to complete this pairwise analysis of the query phage genome to all phages identified within that patient's isolates. The output from Nucmer was visualised using Circos as it allows imaging of one phage versus all in the comparison panel.

Patients A, B and E all have isolates that only carry 2 or fewer intact phages. Due to this low number of phages, it was possible to illustrate this on a single Circos image per patient (Figure 8.6a-c), comparing each phage in succession over the time of collection, therefore comparing over time to see any changes. In Figure 8.6a, phage A2_1 (from the 2nd isolate collected from patient A) is smaller than A3_1. However, the majority of the genome (65% of A3_1) is almost identical (99.93% similarity) to A2_1 shown by the red ribbon. The areas circled in yellow represents a prophage genome expansion (a 35% expansion, looked at in further detail in section 8.4.3.2.1). The phage A3_1 is then shown to be the same as A4_1 (100% similarity) throughout all the following isolates, shown by the red ribbon from A3_1 to a phage in every other isolate in the series, meaning it is conserved longitudinally. Phage A8_1 does not have any similar phages in any of the other isolates and A8 is the only isolate with 2 phages, suggesting it was gained and lost again over time. Figure 8.6b shows the comparison of the phages identified in the *Pa* isolates from patient B. These results show that phage B1_1 and B2_1 show regions of similarity accounting for 70% coverage at 98% identity, interspersed with regions with no sequence similarity, suggesting expansions throughout the genome rather than only at the beginning or the end. These expansions are looked at in further detail in section 8.4.3.2.1. The following 5 isolates contain a phage that is

identical to B2_1, suggesting it stabilised with the expansions. Isolate B3 had 2 phages, unlike the other isolates, and the second B3_2 shares no similarity to any phages from the other isolates from the patient. This implies that it is a new phage that infected the bacteria or it may be a different strain of Pa that infected/co-infected the patient. By looking back at figure 8.3, showing the bacterial genome diversity, B3 does not cluster with the other isolates, suggesting it is a different strain and may provide a reason why it would have a new phage present in its genome. Figure 8.6c shows the phage comparisons for patient E. Intact phages were identified in only two of eight isolates, . Sequence comparison reveals that the two phages in E3 are highly similar to two of the phages in E4. However, E4_4 has an expansion of 4% compared to E3_2, with the rest of the genome having 100% similarity, which suggests the addition of genes on the ends of the genome (discussed further in section 8.4.3.2.1). The other 2 phages in E4 are new and have no similarity to phages in E3. These 2 isolates are shown to cluster together but away from the other isolates for patient E (Figure 8.3), suggesting they are the same strain but due to differences in the accessory genome. However, E4 was infected with a further 2 phages as well as the phage genome expansion.

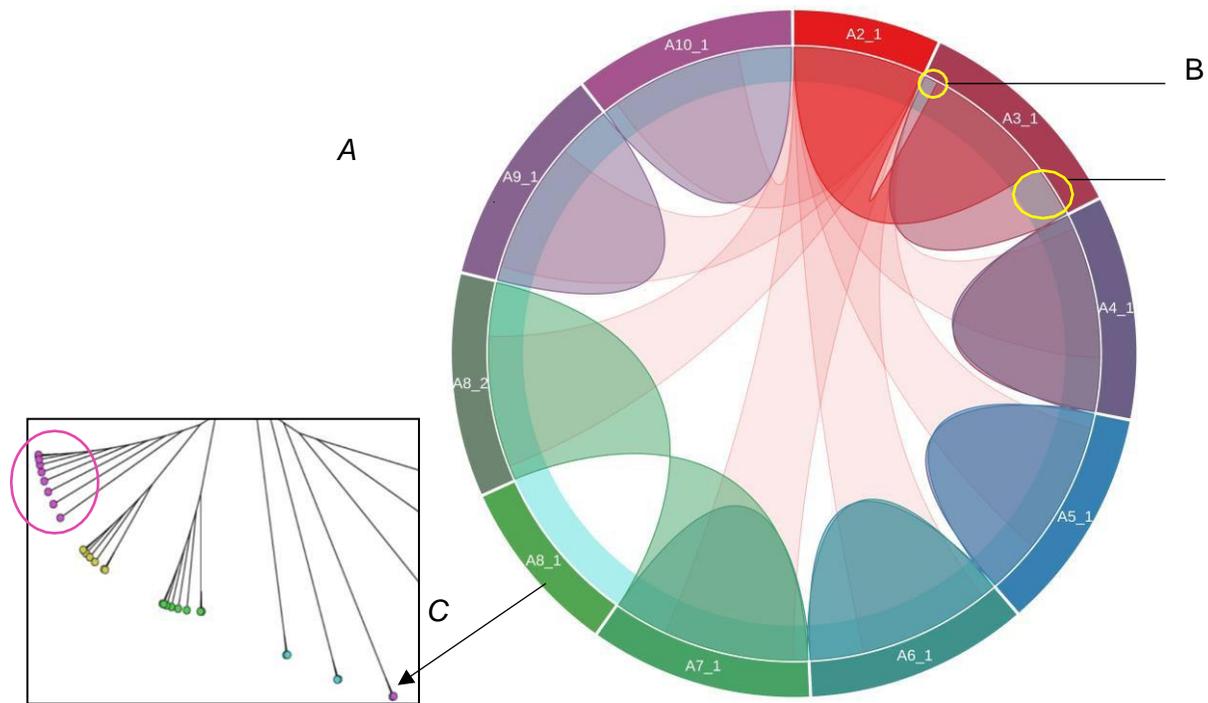


Figure 8.6a – Prophage comparison from the prophages identified in Pa isolates from patient A. The phage genomes were compared using pair-wise comparison from each isolate’s phage/s compared to the next isolate’s phage/s in the order of earliest to the latest collected. The labels denote patient_isolate number_phage number in that isolate e.g. A2_1. A) The outer track shows the phage genomes present in each of the patient’s isolates (A8_2 is the second phage in isolate A8). B) The parts of A3_1 circled in yellow are where the phage has expanded and gained more genetic information, compared to the phage found in the earlier isolate A2_1 (this is looked at in more detail in section 8.4.3.2.1). C) Phage A8_1 is seen as the lone pink node in clade 6, while the other pink nodes that are clustered together in clade 4 are all the other phages that share a high similarity (referring to figure 8.5 enlarged section).

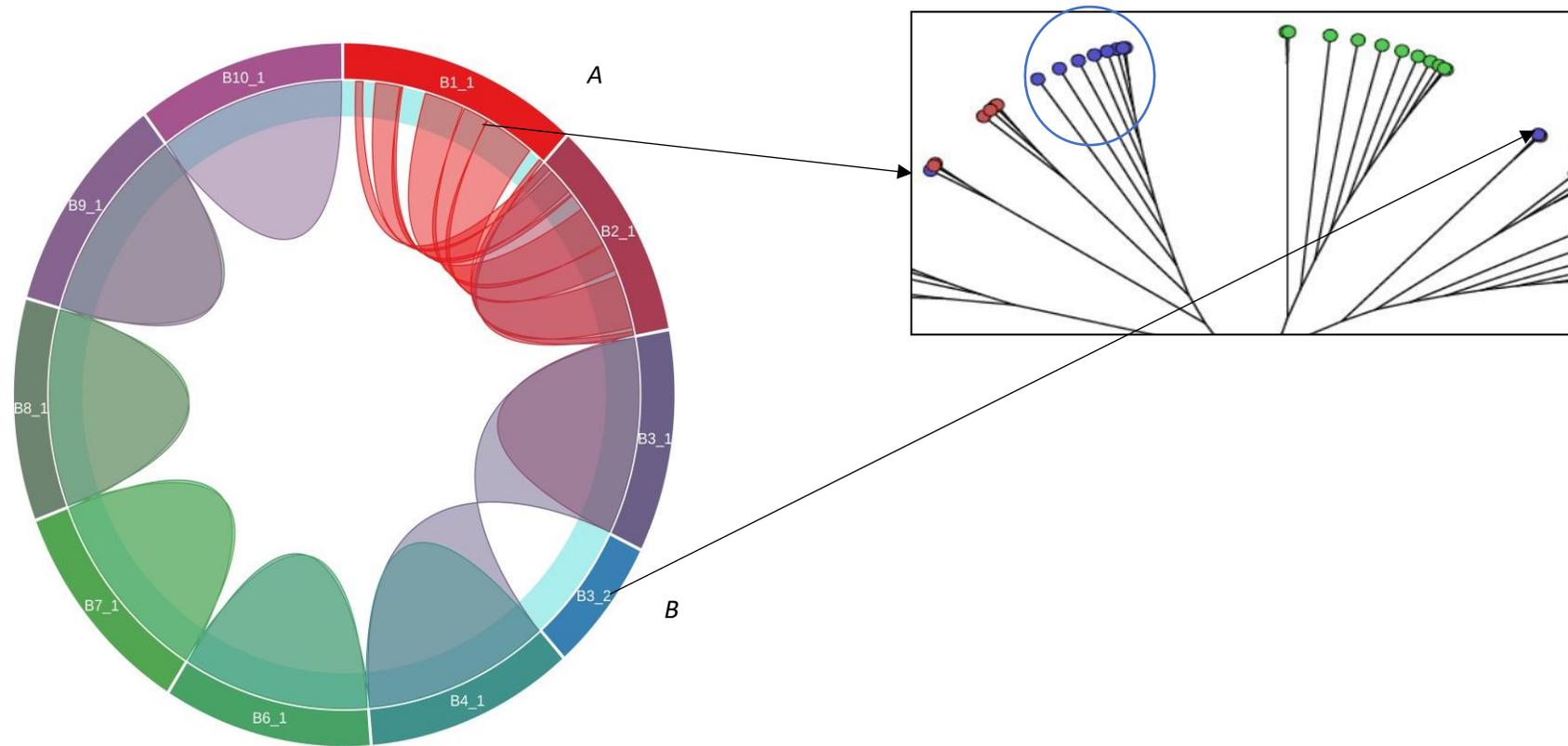


Figure 8.6b – Prophage comparison from the prophages identified in Pa isolates from patient B. The phage genomes were compared using pair-wise comparison from each isolate’s phage/s compare to the next isolate’s phage/s in the order of earliest to the latest collected. The labels denote patient_isolate number_phage number in that isolate e.g. B1_1. The arrows show where particular phages are in the cladogram compared to others. A) B1_1 falls in the same clade as the other phages due to it being highly similar but there is distinct space from it and the phages in the blue circle with B) B3_2 falling in clade one and showing no similarity (referring to figure 8.5 enlarged section).

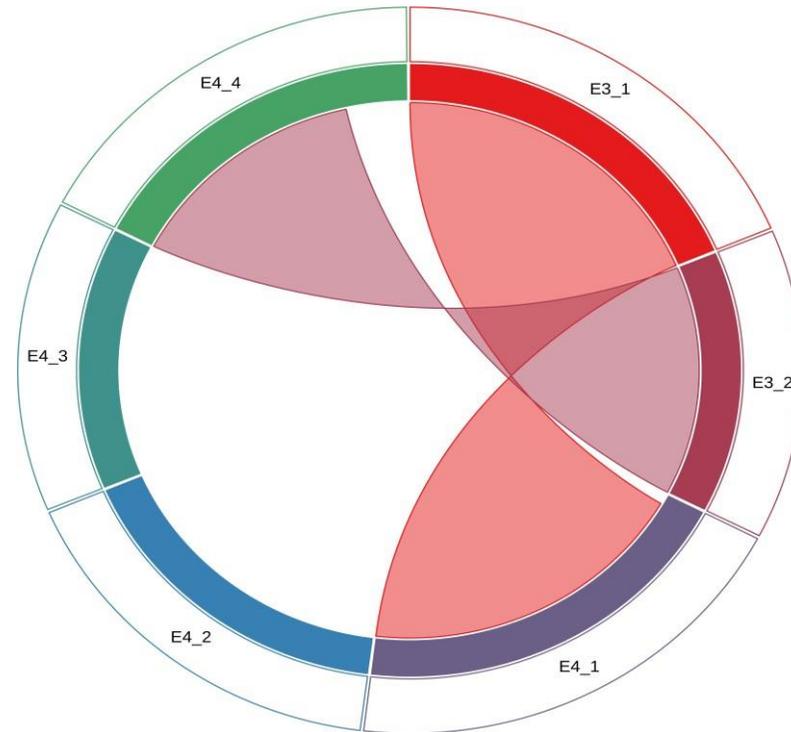


Figure 8.6c – Prophage comparison from the prophages identified in Pa isolates from patient E. The phage genomes were compared using pair-wise comparison from each isolate’s phage/s compare to the next isolate’s phage/s in the order of earliest to the latest collected. The labels denote patient_isolate number_phage number in that isolate e.g. E3_1. Ribbons from one genome to another show exact similarity. Patient E’s other 6 isolates did not have any intact phages identified, hence only 2 isolate’s prophages are being compared.

Patients C, D and F all carry >2 phages and they were compared to see how these phages compare across the panel of phages within the patient longitudinal isolates. Therefore, separate Circos diagrams were constructed that compared the phages per isolate. Each phage cohort within each isolate was then compared against all phages from all the other isolates from the same patient. Within each patient isolate there are phages that, when acquired, are prevalent across the isolates from that patient longitudinally where others are identified and then disappear over time. In figure 8.7 there are 14 circos images, each showing one isolate's prophages compared to all the other phages identified in the Pa panel. The phages present in Pa isolate C1 (the earliest sample from patient C) have a high similarity (100% identity) to all or over 65% coverage of phage genomes in isolates C2, C4, C11 and C14 (figure 8.7a). When comparing the second isolate (C2) phages (shown in figure 8.7b), C2_2 is similar in whole or in part to one phage in all other isolates, apart from C3, C6, C9 and C12, which shows that the same phage may be transient over time. As seen in the previous patients we can also see noticeable expansions of the phage genomes in the comparison between C2_2 and C5_3, where there is 56% coverage at 100% identity, which means there is a 44% expansion (looked at further in section 8.4.3.2.1). This also shows that there are a number of phages that are not similar to any other phages showing infection, and then loss of the phage in a short space of time.

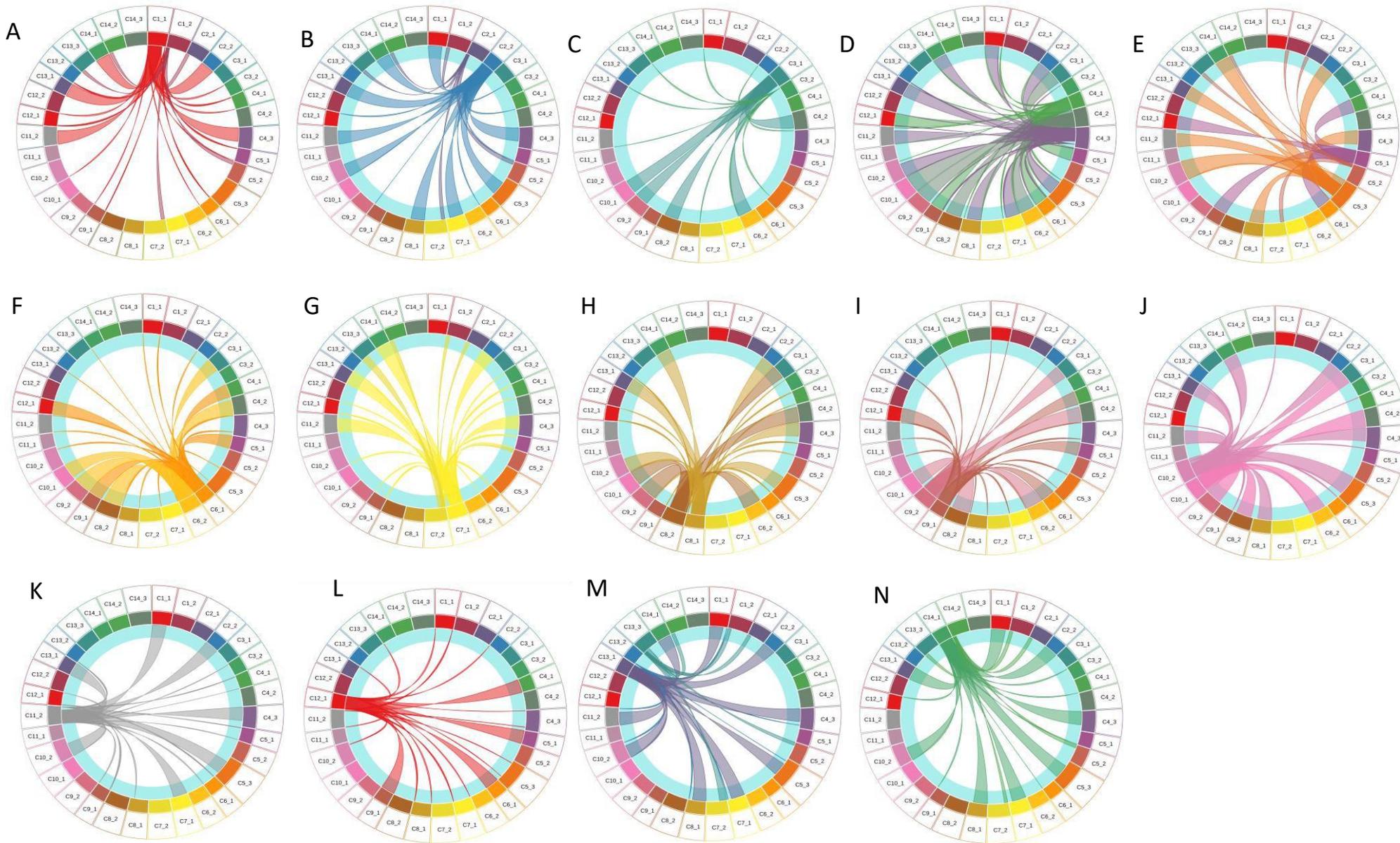


Figure 8.7 Comparison of the prophage genomes identified in Pa longitudinal isolates from patient C. The prophages identified in each isolate are compared against all other prophage genomes in the other 13 isolates. The labelling is as follows; isolate number _phage number e.g. C1_1. A) Isolate C1's Φ vs all other isolates Φ . B) Isolate C2's Φ vs all other isolates Φ . C) Isolate C3's Φ vs all other isolates Φ . D) Isolate C4's Φ vs all other isolates Φ . E) Isolate C5's Φ vs all other isolates Φ . F) Isolate C6's Φ vs all other isolates Φ . G) Isolate C7's Φ vs all other isolates Φ . H) Isolate C8's Φ vs all other isolates Φ . I) Isolate C9's Φ vs all other isolates Φ . J) Isolate C10's Φ vs all other isolates Φ . K) Isolate C11's Φ vs all other isolates Φ . L) Isolate C12's Φ vs all other isolates Φ . M) Isolate C13's Φ vs all other isolates Φ . N) Isolate C14's Φ vs all other isolates Φ .

The phages identified in Patient D's isolates are compared in figure 8.8. The results show that there are phages that have 100% similarity. These are D1_1, D3_1 and D7_1, as shown by the ribbons from D1_1 in figure 8.8a. There is also evidence of possible mosaicism shown in these results as a region of D1_1 and the whole genome of D1_2 are shown to be present in D5_1 and D6_1. However, from figures 8.7d and 8.7e it shows ribbons from D5_1 and D6_1 are showing that the mosaicism that occurred in D5 was conserved into the following isolate (D6), though the original phages were then seen in the isolates thereafter. Figure 8.8 shows the comparison of the prophages from the isolates from patient F. These results show that each isolate has 2 phages that are highly similar (100% similarity) and did not show any change over time. The three phages seen in F1 seem to be stable within the Pa strain and there is no change in 2 of them. However, in isolates F5 and F6, where the smallest of the 3 phages seem to either be replaced with a new phage or it changes over time as the 'new' phage, F5_1 and F6_1 still have some similarity but a coverage of 28% to the original, which indicates that it is a change in the phage rather than a phage displacement (The expansion seen here is looked at in more detail in 8.4.3.2.1). The phages F5_1 and F6_1 have 100% similarity, however, this phage is only present for these 2 consecutive isolates. In the later isolates, a phage carriage reverts back to a phage seen in earlier isolates with a high similarity to F1_2 or possibly a different bacterium. This shows that prophages can change and be lost and gained over time during a chronic lung infection. In addition, it is possible that they swap and change, depending on environmental changes, into phages that may give them benefits at a given moment.

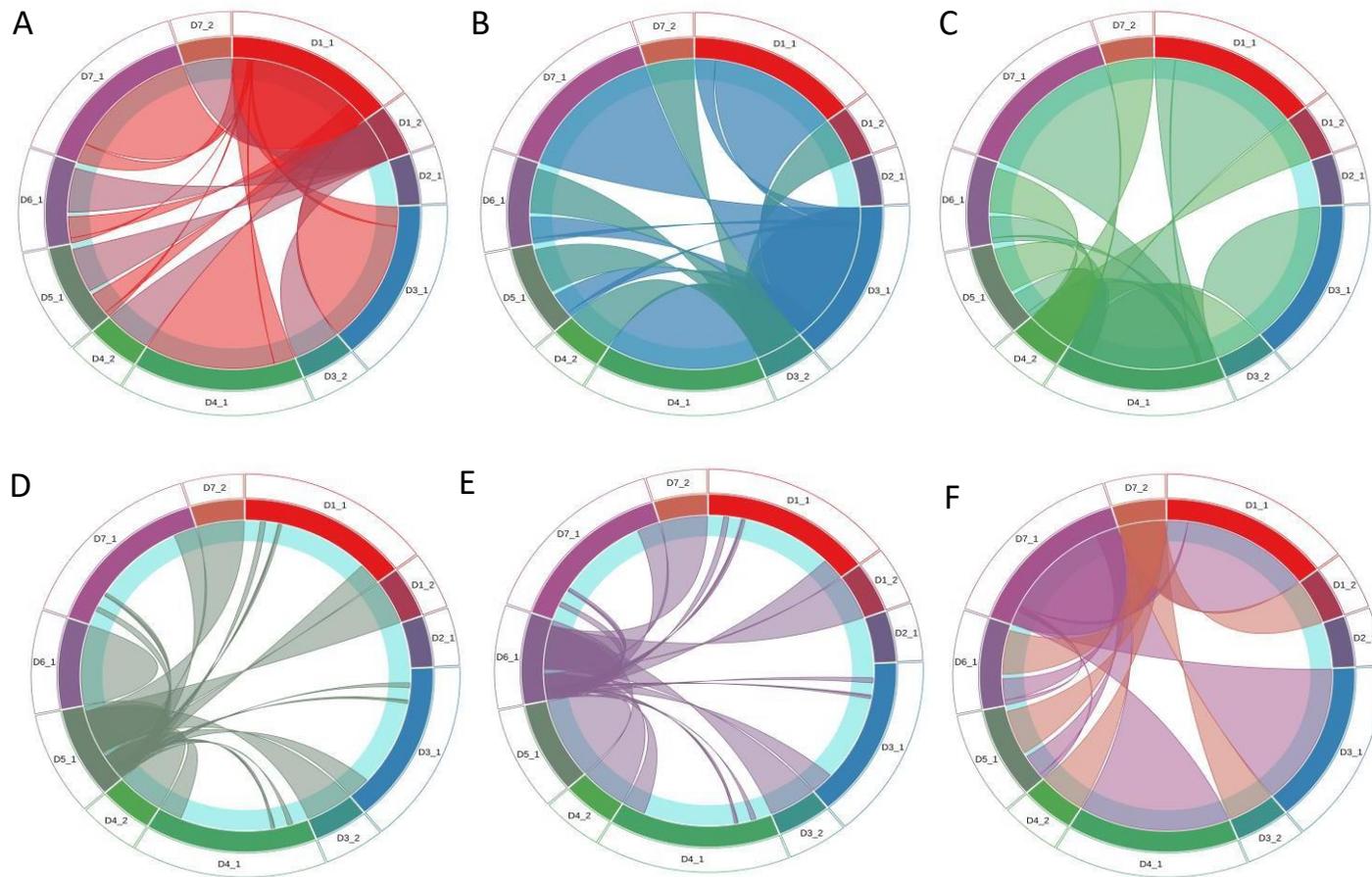


Figure 8.8 Comparison of the prophage genomes identified in *Pa* longitudinal isolates from patient D. The prophages identified in each isolate are compared against all other prophage genomes in the other 5 isolates. The labelling is as follows; isolate number _phage number e.g. D1_1. A) Isolate D1's Φ vs all other isolates Φ . B) Isolate D2's Φ vs all other isolates Φ . C) Isolate D3's Φ vs all other isolates Φ . D) Isolate D4's Φ vs all other isolates Φ . E) Isolate D5's Φ vs all other isolates Φ . F) Isolate D6's Φ vs all other isolates Φ .

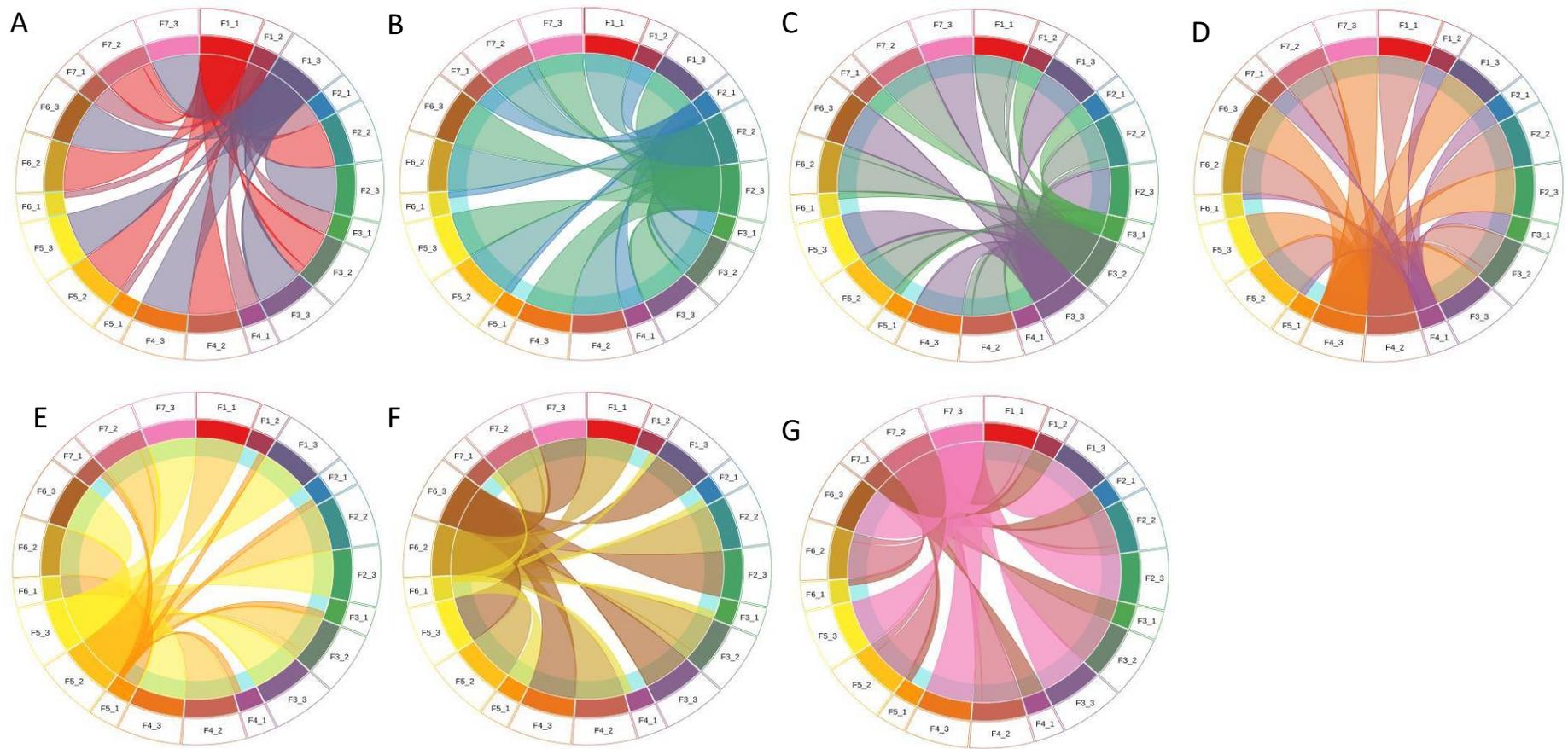


Figure 8.9 Comparison of the prophage genomes identified in *Pa* longitudinal isolates from patient F. The prophages identified in each isolate are compared against all other prophage genomes in the other 6 isolates. The labelling is as follows isolate number _phage number e.g. F1_1. A) Isolate F1's Φ vs all other isolates Φ . B) Isolate F2's Φ vs all other isolates Φ . C) Isolate F3's Φ vs all other isolates Φ . D) Isolate F4's Φ vs all other isolates Φ . E) Isolate F5's Φ vs all other isolates Φ . F) Isolate F6's Φ vs all other isolates Φ . G) Isolate F7's Φ vs all other isolates Φ .

8.4.3.2.1 Evolutionary expansions in prophage genomes between longitudinal Pa isolates

From the previous section, it was shown that prophages within longitudinally collected Pa samples displayed expansion of their genomes over time when comparing prophage genomes from one Pa genome to the next chronologically. These expansions in the genome over time suggest that they may have an evolutionary benefit to the phage and/or the Pa isolate. Here, the regions of expansion are looked at in closer detail to determine the genes that are added and if they could be advantageous. Where there were expansions seen between the phages across the longitudinal isolates from the same patient, the regions were annotated by Prokka to predict the function, as many of the genes were shown to code for hypothetical proteins. The hypothetical genes were also run queried against the *Pseudomonas* genome database in an attempt to find the function of the protein in Pa. Table 8.2 shows the expansion shown between phages A2_1 and A3_1 (shown in figure 8.6a) with the expansion regions circled in yellow. These regions were at the start and end of the phage genome rather than midway through, lengthening the genome by 35%.

Where expansions of phage genomes were seen between the phages from the longitudinal isolates from the same patient in the following patients; B, C, D, E, and F (Figures 8.6b-f), the proteins in the expanded regions were extracted and annotated (as carried out for patient A in Table 8.2). These results are shown in appendix N tables 1-12 and display the proteins carried in these expanded regions of prophage within the Pa isolates from each patient. Patient B's earliest isolate (B1) contained a phage that was seen to be expanded in the second Pa isolate (B2), as shown in figure 8.6b. It showed, unlike the expanded regions in patient A that appears at the start and end of the phage genome, the additional proteins here are seen added throughout the prophage genome.

Proteins present in the expansions in the prophages in isolates from patient A that may give the phage or Pa an advantage (Table 8.2) included 'TetR family transcriptional regulator', which are proteins that play a role in DNA binding which regulate gene expression. Many of

these act as global multi-target regulators, which could be beneficial to the Pa host. Also, 'Rhs element Vgr protein', a Rhs protein found on a plasmid in Pa, has been implicated in bacteriocin production in *Pseudomonas savastanoi*. The 'ice nucleation protein' that was seen in the expansion enables Gram-negative bacteria to promote nucleation of ice at relatively high temperatures. The benefit that these additional proteins may give to their host might be why this expansion was seen in prophages sequentially in the longitudinal Pa isolates from patient A. Other notable proteins found in the prophage genome expansions with the Pa isolates from the other patients (shown in appendix N) are the 'Bax inhibitor/YccA family' and 'DNA cytosine methyltransferases'. Both of which are from a prophage expansion seen in patient B's isolate, with 'methyltransferases' also being seen in multiple expanded phage regions when comparing the prophages in patient C's Pa isolates. The YccA in *E.coli* has been shown to interact with ATP-dependent protease FtsH, which degrades abnormal membrane proteins as part of a quality control mechanism to maintain the integrity of biological membranes and it is possible it plays a similar role in Pa . Genomes of lytic and lysogenic phages have been shown to encode multi- and mono-specific methyltransferases that have the ability to confer protection from restriction endonucleases of their bacterial host, enabling the phages to stay within the Pa host (Murphy et al., 2013). Therefore, the addition and carriage of this protein would be beneficial for maintenance of the phage genome. 'ImmA/IrrE family metallo-endopeptidase' was also identified in multiple expanded phage regions when comparing phages of isolates from patient's C and D. ImmA is encoded on a conjugative transposon, which could mean the expansions are the addition of transposons and conjugating bacteria are able to transfer conjugative transposons to other bacteria or possibly other phages.

These expansions are of interest. However, without further experimentation and looking at these genes in the context of the whole phage and the whole Pa genome, it is hard to know their true function. This could be further investigated with metabolomics or RNA sequencing to see if these proteins are transcribed and translated, and therefore are more likely to affect the Pa host. It would also be valuable to confirm whether these genome expansions are excised within the phage genome or if they are an addition to the Pa genome, such as a transposon, which has inserted near to a phage and has been classed as part of the phage genome incorrectly.

Table 8.2 - Gene annotation results from the prophage expansion region in Pa isolate 3 from patient A.

Protein name	Prokka identification	<i>Pseudomonas</i> database identification
NKDKGOEBB_00056	hypothetical protein	MULTISPECIES: hypothetical protein
NKDKGOEBB_00057	hypothetical protein	Transposase
NKDKGOEBB_00058	Ice nucleation protein	ice nucleation protein
NKDKGOEBB_00059	hypothetical protein	arc-like DNA binding domain protein
NKDKGOEBB_00060	hypothetical protein	hypothetical protein
NKDKGOEBB_00061	hypothetical protein	hypothetical protein
NKDKGOEBB_00062	hypothetical protein	putative membrane protein
NKDKGOEBB_00063	hypothetical protein	hypothetical protein PPF10_gp031
NKDKGOEBB_00064	hypothetical protein	hypothetical protein
NKDKGOEBB_00065	hypothetical protein	hypothetical protein

NKDKGOEBB_00066	Putative defective protein IntQ	site-specific integrase
NKDKGOEBB_00067	Antitoxin SocA	DUF4065 domain-containing protein
NKDKGOEBB_00068	hypothetical protein	hypothetical protein
NKDKGOEBB_00069	hypothetical protein	hypothetical protein
NKDKGOEBB_00070	hypothetical protein	hypothetical protein
NKDKGOEBB_00071	hypothetical protein	MULTISPECIES: TM2 domain-containing protein
NKDKGOEBB_00072	hypothetical protein	Rhs element Vgr protein
NKDKGOEBB_00073	hypothetical protein	MULTISPECIES: VRR-NUC domain-containing protein
NKDKGOEBB_00074	hypothetical protein	DUF3396 domain-containing protein
NKDKGOEBB_00001	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase	alpha/beta fold hydrolase
NKDKGOEBB_00002	hypothetical protein	TetR family transcriptional regulator
NKDKGOEBB_00003	hypothetical protein	hypothetical protein CLM86_28910
NKDKGOEBB_00052	hypothetical protein	ADP-ribosyl-(dinitrogen reductase) hydrolase
NKDKGOEBB_00053	hypothetical protein	hypothetical protein T223_06585
NKDKGOEBB_00054	hypothetical protein	helix-turn-helix transcriptional regulator
NKDKGOEBB_00055	hypothetical protein	transcriptional regulator, LuxR family

8.4.4 Phage Induction of longitudinal isolates and cross-infection and host range analysis.

Each of the Pa isolates were induced using norfloxacin (methods section 2.5.1) and bacteriophage lysates obtained were used for cross-infection against the other isolates from the same patient. The results for these cross infections for patients A-F are shown in tables 8.2a-f and the isolates were numbered from the earliest isolated (0 showing no infection and 1 showing infection of the bacteria by the phage lysate). These results show that there was at least one isolate per patient set that was able to infect at least one of the other isolates. Out of the 57 isolates induced, there were 26 that were not able to infect any of the other isolates from the same patient, and the isolates that showed to have active phages showed differing infection profiles. However, there are no correlations that can be drawn from how they infect. It was hypothesised that longevity may determine phages with broader infectivity across earlier isolates as they may evolve to overcome bacterial phage restriction functions over time. The phages that were the most infectious were induced from isolates 2 and 6 from patient C, where they were able to infect all other isolates from patient C.

Table 8.3a – Cross-infection of Pa isolates from patient A against phages induced from them

		Phage lysate										
		Patient A	1	2	3	4	5	6	7	8	9	10
Bacteria	1	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	1	1	1	0	0	0	0	0	0
	3	0	0	1	1	0	0	0	0	0	0	0
	4	0	0	1	1	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	1	1	1	0	0	0	0	0	0
	7	0	0	1	1	1	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0

Table 8.3b - Cross-infection of Pa isolates from patient B against phages induced from them

		Phage lysate										
		Patient B	1	2	3	4	5	6	7	8	9	10
Bacteria	1	1	0	1	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0
	4	1	0	0	0	0	0	0	0	0	0	0
	5	1	1	0	0	0	0	0	1	0	0	1
	6	1	0	0	0	0	0	0	0	0	0	1
	7	1	0	0	0	0	0	0	0	0	0	0
	8	1	0	0	0	0	0	0	0	0	0	0
	9	1	0	0	0	0	1	0	0	0	0	0
	10	1	0	0	0	0	0	0	0	0	0	0

Table 8.3c - Cross-infection of Pa isolates from patient C against phages induced from them

		Phage lysate													
		Patient C	1	2	3	4	5	6	7	8	9	10	11	12	13
Bacteria	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0
	3	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	4	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	5	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	6	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	7	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	8	0	1	0	0	0	1	0	0	0	0	0	0	0	0
	9	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	10	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	11	0	1	0	0	0	1	0	0	0	0	0	0	0	0
	12	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	13	0	1	0	0	0	1	0	0	0	0	1	0	0	0
	14	0	1	0	0	0	1	0	0	0	0	1	0	0	0

Table 8.3d – Cross-infection of Pa isolates from patient D against phages induced from them

		Phage lysate						
		Patient D	1	2	3	4	5	6
Bacteria	1	1	1	1	1	1	1	1
	2	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0
	6	0	1	0	1	1	1	1
	7	0	1	0	1	1	0	0

Table 8.3e – Cross-infection of Pa isolates from patient E against phages induced from them

		Phage lysate							
Patient E		1	2	3	4	5	6	7	8
Bacteria	1	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0
	4	0	0	1	1	1	1	0	0
	5	0	0	0	0	0	0	0	0
	6	0	0	1	0	0	0	0	0
	7	0	0	1	0	1	0	0	0
	8	0	0	1	1	1	0	0	0

Table 8.3f – Cross-infection of Pa isolates from patient F against phages induced from them

		Phage lysate						
Patient F		1	2	3	4	5	6	7
Bacteria	1	1	1	1	0	1	0	0
	2	1	1	1	0	1	0	0
	3	1	1	1	0	1	0	0
	4	1	1	1	0	0	0	0
	5	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0

8.5 DISCUSSION

This chapter illustrates the first longitudinal study in determining the temperate phage genomes of Pa isolates from BR patients. These results show that phages may be a way of showing that bacterial isolates from the same chronically infected lungs are not necessarily clonal. This was determined by analysis of genome sequencing that showed multiple strains of Pa were observed over time based on analysis of gene content, rather than one strain being present throughout a chronic infection. This change in specific isolates of Pa is probably caused by displacement or co-infection, supporting research that suggests displacement occurs at a higher rate in BR rather than in CF patients (Woo et al., 2018). Patients A and B both show possible displacement as their earliest isolate is a different strain to the other isolates collected later from the patient, which implies that the new Pa strain displaced the isolate seen in the first sample. There could also possibly be a mixed community of Pa present. Patient B also shows a different strain again in the 3rd sample that was taken. This was possibly a co-infection as the samples thereafter were the same strain as had been seen prior to this sample. From the comparison of the bacterial genomes from the same patient, where there is a single strain seen over time, changes are seen in the genomes probably due to continuing adaption and evolution of the bacteria within the lung, which has been reported previously in Pa isolates from BR and CF patients (Hilliam et al., 2017, Winstanley et al., 2016).

8.5.1 Changes to antibiotic resistance of Pa from longitudinally during chronic infection

These genetic changes in the Pa genomes are important as they may cause antibiotic resistance, as shown in figure 8.2. For example, patient C results show that all 14 of the isolates were highly similar/ the same strain. However, there were changes over time that changed the resistance profile with patient C's isolates developing resistance to Ceftazidime, Temocillin, and Ticarcillin that the earliest Pa isolate was susceptible to. However, over the 14 time points for patient C there were also isolates that were shown to

be resistant to an antibiotic but were then sensitive at a later time point, such as Fosfomycin where the earliest isolate is resistant and then the 2nd, 3rd and 4th isolates are sensitive to it again. This was seen in 6 out of 12 antibiotics tested in patient C's isolates. This could be due to the selection pressure being taken away (an antibiotic), meaning the resistance is not necessary and is lost through evolution. From analysing the IV antibiotics administered to the patient this appears to be the case for Fosfomycin in patient C, as Fosfomycin is administered when the antibiotic testing shows the Pa to be sensitive, the next test shows it is resistant and Fosfomycin stops being administered, then during the following AB testing it is shown to have regained sensitivity. There are other reasons for a regain of sensitivity to an antibiotic, such as loss of a resistance plasmid (Igumbor et al., 2000), or possibly the induction of a prophage that confers resistance, which has been reported to occur in *Salmonella* (Zhang and LeJeune, 2008). There have been many phage-associated AMR genes in Pa from CF patient lungs (Rolain et al., 2011, Fancello et al., 2011, Modi et al., 2013), indicating it is a possible explanation for AMR changes in Pa within the lung.

These results illustrate that there is broader resistance that relates to a later stage of infection and disease progression, providing evidence that if these are the same strains, resistance was gained over time.

8.5.2 Motility alteration to Pa from over the course of chronic infection

This evolution of the Pa strains over time also led to changes in motility of the isolates. As there was no displacement, the motility changes were also due to adaption. For example, the earliest isolates were negative, the following 3 isolates collected fluctuate between negative and positive motility and the last 10 isolates all were positively motile. This suggests an adaption by the Pa to become motile, which is in contrast to the adaptations that have been seen previously from Pa isolates in the infected lung environment that have a reduction in motility, which has been associated with increased bacterial burden and increased disease severity (Faure et al., 2018, Veses-Garcia et al., 2015). Phages could be considered a driving force of motility being re-established to a strain if it coincided with a new prophage being present in the genome. The final point that the bacterial comparisons

show is that the bacterial genomes of isolates from patients were very similar to those of other patients. For example, patient A isolate A1 is highly similar to patient E's isolates, and patient B's isolate B1 clusters with patient D's isolates, suggesting there may have been cross-infection between patients, which is thought to be not as common in BR as in CF, where there are prevention measures in place (Woo et al., 2018).

8.5.3 Analysis of prophage genomes from longitudinal Pa isolates to determine evolutionary adaptation over time

Identification of prophages from the Pa genomes revealed different temperate phage profiles are seen in the isolates longitudinally collected from each patient. There is evidence of either displacement of Pa strains or co-colonisation, which may be a reason for new phages being gained or lost, rather than the same strain being infected by a new phage. Phage numbers were also seen to change where no displacement was seen (in chapter 4), illustrating that there is gain and loss of phages over time as well as expansion of the phage genomes. This suggests selective pressures lead to evolution of phages over the progression of the chronic lung infection. Pressures, such as antibiotic administration, may cause induction of temperate phages and alterations to their genomes. For phages, this inclusion of other bacteria at low frequency offers an extended gene pool and possible other phages that may allow a reservoir for recombination and allow positive selection for mobilisable phages.

The prophage diversity results illustrated that the phages from isolates from the same patient share a higher similarity between each other than those from other patients, as they are probably the same phage within the same Pa isolate. This shows that phage diversity is relatively stable with few changes in phages over time if it is within the same isolate. Yet in figure 8.5, there is a slight dispersion in the clusters from the same patient, showing there is some diversity in the phages when compared at the protein level. They are not all identical, suggesting genetic drift has occurred over time, which may play a role in the bacterial evolution.

To focus in more detail on the diversity over time, the phages in each of the patient's isolates

were compared to the entire panel from this part of the study. This also shows which phages were integrated or lost over time. The results show there are regions of genome expansion and deletion over time, suggesting there may be domestication of phage or recombination and accrual of genes that could offer positive evolutionary selection for the phage. However, there is some inaccuracy in PHASTER determining the start and end of phages, although it has determined these ends in the previous search when comparing a limited study of the LES prophages. For subsequent publication, manual curation of the ends of the phage will be needed to unequivocally verify their start and end points. Referencing back to the bacterial genome to see if the deleted/added region is present would also confirm whether this is the case. Also verification of bacterial genome assembly. Importantly, this research provides evidence of transfer of phages between strains within the lungs shown in patient B, as B3 is a different strain to B4 (shown in figure 8.3) but both have a phage that is highly similar, suggesting there was transfer within the lung between these bacterial strains. The findings presented here confirm that phages adapt quickly over time and are able to alter the genomes of their Pa host over the course of a chronic infection, with the possibility that they contribute to the virulence and persistence of the infection. Further work using transcriptomics would shed more light on whether they are a contributing factor as it would be possible to compare the expression between Pa isolates with and without the prophage, but also compare what changes that a phage genome expansion may offer the Pa host and why the expansion may be conserved over time

From comparing the prophages within Pa isolates longitudinally, unsurprisingly, the same phages appeared. Most importantly however, evolution of phages is difficult to map as previously discussed. This study is able to demonstrate this evolutionary progression as in some cases these phages had expansions in their genomes at the start, end or middle of the genome, identified by PHASTER. In table 8.2 the genes in one expansion, comparing a prophage in isolates A2 and A3, (from patient A shown circled in yellow in figure 8.6a) are annotated to determine whether these additional genes give any advantage to the prophage or the Pa host. From these results it suggests that there are a number of genes that have a

possible benefit and can drive evolution in phages, which has been shown previously. However, the majority of the genes in the expansion, like those in phage genomes, have no known function. As most of the phage expansions seen between the longitudinal Pa isolates are shown to be retained over time, it would suggest that they give a selective advantage and therefore are not lost over time. The genes noted in section 8.4.3.2.1 give advantages to the Pa in a number of different ways, from conferring antibiotic resistance to promoting the nucleation of ice. The expansion of bacterial genomes, due to the addition of phages, plasmids or transposons, has been well documented (Gao et al., 2019). However, the expansion of prophages within longitudinal Pa and the impact that it has on the prophage and the host has not been studied, even though phages are known to be diverse and changeable entities that can drive evolution (Rodriguez-Valera et al., 2016). As this is one of the first studies looking into the temporal changes within prophages within Pa from chronically infected lungs, it could give rise to more studies in this area as there is some evidence that temperate phages change over time and possibly gain genes in expansions that benefit the host Pa, which the phage could then transfer to other Pa strains. Through this process, traits beneficial to host such as antibiotic resistance, could be transferred resulting in a Pa isolate that is able to persist in the lungs for longer.

The results from the induction of these Pa isolates and the re-infection of them into isolates from the same patient, shows that they contain inducible phages that are able to infect isolates from the same patient and also the originating bacterial host. This shows that superinfection is possible as many of the isolates are of the same strain (Tariq et al., 2019). Tariq and colleagues showed that phages induced from Pa isolates from patients who were in the later stages of chronic Pa infection could infect Pa isolates from patients in the early stages more readily than in later stages for chronic infection. This finding was not seen in these results, possibly as this was tested against isolates from the same patient.

Isolates from the same patient, which have been shown to be different strains, showed that induced phages from one strain were able to infect the other strain from the same patient,

such as in patient E, illustrating that transfer of phage between strains within the lung environment is possible. This supports our study on the transfer of phages within the chronically infected lungs of BR patients in chapter 3. These results also display that some isolates such as E5 and E6, which were predicted by PHASTER to contain no intact phages, harbour active phages. This shows that PHASTER is not to be fully relied on and that experimental methods are necessary to confirm the presence of inducible phages.

8.6 SUMMARY

This is the first longitudinal study on prophage carriage in Pa isolates from chronically infected BR patients, showing loss, gain and transfer over time, which could be due to the induction and reinfection caused by exposure to antibiotics. By experimentally inducing the phages and showing re-infection was possible, it confirms transfer can take place in the lungs. The antibiotic resistance profiles show a general trend of resistance over time, but also that the bacteria can regain sensitivity to antibiotics over time, which is possibly impacted by prophages as gain and loss of prophages are also seen within these isolates. Bacteria/Pa could also be affected by expansions to the prophage genome, suggesting HGT within the phage genome, which has not been reported in Pa in the lung previously. This study sheds light on the complex bacterial-phage-antibiotic interactions that take place during the course of a chronic lung infection.

9 GENERAL DISCUSSION

9.1 THE CONTEXT OF THIS THESIS

Pseudomonas aeruginosa is an important pathogen that causes disease across the world. In particular, lung infections by Pa in CF and BR patients are of major concern as chronic infection leads to poor clinical outcomes with a high associated mortality. Many Pa lung infections become chronic with the Pa becoming increasingly multi-drug resistant (Obritsch et al., 2004). One of the possible reasons for this is the genetic pliability is due to it being open to horizontal genetic transfer, mediating an enhanced propensity to cause chronic infection. This genetic diversity is highlighted, in part, by temperate phages that stud the bacterial chromosome. Their role in the biology of their bacterial host is understudied, but with evidence of temperate phages of Pa promoting a selective fitness advantage (Davies et al., 2016a) and harbouring antibiotic resistance genes, it implies they play a role in the chronicity of infection (Kondo et al., 2020). With temperate phages able to induce and mobilise, and that the antibiotics used to treat their bacterial host are capable of causing induction and possible transfer, it makes them a dangerous entity, especially within the chronically infected lung, which offers a positive evolutionary selection for the bacteria that could lead to poorer clinical outcomes for patients. This study set out to investigate further how temperate phages contribute to the adaption and evolution of the Pa and whether their ability to modulate host metabolism plays a part in the chronic phenotype. A better understanding of why infection with Pa becomes chronic may aid development of new diagnostic approaches and treatment methods for patients with chronic Pa infections.

9.1.1 Temperate phage transfer in the chronic lung environment

The nature of temperate phages, being transducible, and the reports of free living temperate phages being isolated from the sputa of CF patients (James et al., 2015), led to understandings that temperate phages were able to be induced and re-infect while in the chronically infected lung. In doing so these phages are capable of transferring their genetic information to other strains, or within non-clonal populations under conditions of lung co-colonisation. The ability of these dsDNA phages, lambdoid in genome organisation, to

overcome phage immunity is seen as common in these phages and has been shown in other phages in *E.coli*, but not at the rates seen in Pa colonised from the lung (Tariq et al., 2015). It has been presumed that phages must transfer between strains of Pa within the lung, as temperate phages have been detected free floating however, to date, no study to the best of our knowledge has focussed on the carriage and transfer of prophages within lung populations of Pa at a specific point in time, or how this develops longitudinally and their role in chronicity and pathogenicity of infection.

The chronically infected lung of a CF or BR patient can be colonised by multiple phenotypes and genotypes of Pa (Jelsbak et al., 2007). A study looking at explanted lungs showed that the Pa isolates occupying different regions of the lung physiology had evolved independently and they differed in their comparative phenotypic characteristics (Jorth et al., 2015). These different phenotypes or co-colonisation of multiple strains in the same lung environment poses the questions; can temperate phages transfer between them, adding an extra compartment of genetic diversity and evolutionary potential? Or, do temperate phages and alteration of the cell metabolism offer these strains a selective advantage in colonising different compartments of the lung that offer the lysogen positive evolutionary selection?

In chapter 4, the genomes of Pa isolates from 6 BR patients, where each patient had multiple phenotypes of Pa identified in one sputum sample, were analysed for prophage carriage and compared between isolates from the same patient. The Pa genomes of the different phenotypes from the same patient had different MLST types within a sample shown in the paper by Hilliam et al 2016, suggesting co-colonisation rather than evolutionary adaptation. A different complement of prophage/s were seen in each isolate from the same patient from genome sequencing data, with at least one phage being identified in every genome, with an upper threshold of 10 phages of varying levels of genome completeness identified. From comparing the intact prophage genomes carried between isolates from the same patient, it was reported that there were some prophages with high percentage homogeneity to prophages from the comparable isolates that were of a different MLST,

which suggests possible transfer of temperate phages. Attempts were made to induce these phages and sequence them for confirmation that phages thought to be transferred were mobilisable. It was confirmed that two out of six patients carried non-clonal Pa isolates (n-2 and n-3) that contained phages that were inducible with 99-100% similarity (with only a few base pair changes). This strongly suggests that these phages were transferred while in the lung environment. This movement of phages between mixed populations of Pa has not been reported previously and may offer an advantage for evolution of the bacteriophage, offering an increased genetic reservoir for the phage for recombination and evolutionary events. The induction of temperate phages and reinfection as seen here is likely, especially in a patient that is exposed to antibiotics on a long-term basis (Tariq et al., 2019). For example, BR patients with chronic infections, where lytic activity of temperate phages has been reported previously in CF (James et al., 2015) and induction by antibiotics that are regularly used to treat Pa infection, such as Ciprofloxacin (Fothergill et al., 2011). Inflammation seen in a chronic lung infection can also induce phages through the inflammatory response and the release of cytokines, has been observed in *Salmonella* in the gut, increasing the transfer of phages (Diard et al., 2017). The hypothesis that antibiotic therapy may cause the induction of temperate phages and lead to transfer into other (naive) strains within the chronic environment would work in a similar way. This could possibly cause the transfer of genes (specialised transduction), leading to adaptation and increased bacterial fitness in the lung environment, hence, playing a role in Pa chronicity within the lung. However, transduction has been shown to happen in Pa outside the lung (Monson et al., 2011), and in different hosts and environments (Goh et al., 2013), suggesting temperate phages from environmental Pa may be able to infect lung Pa strains, possibly giving them an advantage.

By identifying and studying the integration sites of these prophages in the Pa bacterial genomes, it was established that 30% of the prophages integrated into a bacterial gene, which has the potential to lead to phenotypic changes due to the interruption of gene function. This suggests temperate phages can also drive bacterial diversity by introducing

altered function with each integration event. An example of disruptive lysogenic conversion is in *Staphylococcal* phage L54a, which has been shown to integrate into the lipase-encoding gene (*geh*) resulting in a loss of phenotype (Lee and Landolo, 1986), and certain transposable phages like Mu- phages are known to insert randomly and disrupt bacterial genes as part of their replication strategy. The true extent of the impact of phage-mediated gene disruption on bacterial evolution is not well understood. However, this study highlights that mobilizable prophage genomes are present within Pa isolates in the chronically infected lung and are capable of transferring within the lung between isolates, this informs further studies to look in more detail at these genomes to ascertain why these particular prophages are kept and passed between isolates in the genomes as they possibly harbour genes that benefit the Pa.

9.1.2 Metabolomics approach to identify bacterial physiological changes caused by temperate phages

Metabolomics using sensitive mass spectrometry is a relatively new omics-based approach in bacterial populations to investigate the effects certain conditions have on a bacterium's metabolism (Garcia et al., 2008). This study is the first time the addition of prophages to the genome of a bacteria has been investigated using metabolomics to reveal how they influence the bacterial metabolism. The effects of lytic phage infection on bacteria including Pa has been looked at using metabolomics (De Smet et al., 2016, Chevallereau et al., 2016, Zhao et al., 2017). However, as lytic and lysogenic phages interact with their host in very distinct ways, it was expected that the results would differ significantly.

Untargeted metabolomic analysis was used to discover the impact the three sets of phages (all relating to chronic respiratory infection) had on the metabolism of a common well-defined Pa host, when integrated as a prophage. They were all cultured in ASM to mimic the sputum in the lung environment and to aid biofilm production. As an *in vitro* model, it has the same constituents as human sputum, making it an informative approach in comparing the Pa metabolism under biofilm conditions. However, unlike the chronic lung

where these components can fluctuate, such as levels of mucin, with levels being patient dependent, as shown by a study looking at the metabolome of CF sputum (Quinn et al., 2016). The immune cells in CF sputum may play a part in how Pa behaves, which is a limitation of this study using ASM. All three sets of phages that were studied show that phage conversion is associated with a metabolomic shift in their Pa host, which is the first time this has been shown and using metabolomics, advising further studies in this field.

9.1.2.1 Subversion of the PAO1 metabolism caused by LES prophages

From analysing the metabolomics of lysogens containing the LES phages, the results clearly show the LES phages have different effects on the metabolism of the Pa, both on their own and in combination with the other LES phages, with LES-2 evoking a very different metabolic profile to the other two LES phages. These results are supported by previous studies investigating the LES phages, which show how they have different infection properties such as, that LES-2 is more sensitive to induction into the lytic cycle or has a more efficient replicative cycle than the other LES phages (James et al., 2012). This indicates that the prophages affect their host in different ways. This data is also supported by the work of Davies et al (2016), who demonstrates that the LES phages increase competitive fitness during lung infection at different rates, which could be due to the shift/drive in the Pa metabolism that increases competitive fitness. (Davies et al., 2016).

9.1.2.2 Effects of temperate phage communities from Pa isolated from different aetiologies

After analysis of well-defined Pa phages, the effects of different clinically relevant prophages from BR and well as CF isolates was analysed. The temperate phages were induced from clinical Pa isolates from patients in four clinical groups or aetiologies (early and late stages of disease progression from CF and BR patients). Also, the full complement of phages that were induced were integrated into PAO1, rather than a single phage. Polylysogeny is common in Pa (Winstanley et al., 2009) and multiple prophages can exert combinatorial effects conferring phenotypic and metabolic change (James et al., 2012, Matos et al., 2013). The metabolomics results showed that there was a significant difference in how all the sets

of phages subverted the PAO1's metabolism. However, there was no significant difference between how the phage from the differing aetiologies subverted the 142 metabolites that were analysed. This may be due to the metabolites that were different between aetiologies only provoking a small effect and therefore would not be classed as significant. Nevertheless, these results do demonstrate that there are some core significant metabolic changes that many temperate Pa phages evoke on their host, as 20 lysogens, carrying 20 different sets of phages, from 4 clinical groups, all subverted their host Pa in a similar way. This suggests that these conserved metabolic changes are important for a Pa to retain in the lung environment, as all the phages evolved in the chronically infected lung and share the same metabolic subversion to their host. This suggests that there may be phage-mediated metabolites that are commonly upregulated and downregulated in the lung environment, which may confer an advantage to the host. It would be possible to use these significantly altered metabolites as a biological marker of infection by an isolate that may promote more severe infection, as done using a genetic marker for Pf phages whose presence is associated with more severe infections in CF patients (Burgener et al., 2019).

9.1.2.3 Metabolic subversion of Pa by different types of prophage

In chapter 7, the effects of temperate Pa phage diversity on the metabolism was investigated by integrating phages from diverse genetic groups. The metabolomic results showed out of 4 phage types, 2 subverted the metabolism of the host in a distinct way compared to the other 2 that had more similar metabolic profiles. This suggests that different phage types show independent mechanisms of altering the bacterial host metabolism. These data illustrate metabolomic approaches are necessary to supplement the information obtained from gene annotation, comparative genomics and transcriptomics to better understand the role of prophage within the host bacteria, as only limited numbers of genes in phages offer predicted function (Klumpp et al., 2013). A study in *E.coli* reported that around 65% of the genes carried by shiga-toxin encoding prophage 24B, another temperate phage, could not be related to a known cellular function (Smith et al., 2012). This is similar to the levels seen

anecdotally in Pa temperate phages. Metabolomics can help resolve questions previously unanswered by using other approaches (Patti et al., 2012) and elucidate new metabolic pathways involved in lysogenic infection. Also, metabolomics offers a more cost-effective approach to comparing the impact of phage conversion, without the need to perform RNA-sequencing, which has shown to be highly effective (Owen et al., 2020) but expensive, comparing tens of pounds to hundreds of pounds, respectively, per test. This allows for a cost-effective way to analyse the impact of a number of phages on a common host in a single run and give an overview of the metabolites and the metabolic pathways that the phages alter. The pathways or metabolites that are significantly affected by the addition of a prophage and are of interest can then be examined further using experimental and/or functional assays to confirm the metabolic result. Further studies including a second MS stage could also be used to compare to metabolite standards or fragmentation analysis to determine if each metabolite is validated and correct. Moreover, a key question remains: which phage gene products play a role or mediate these metabolic changes? Which future transcriptomics would be able to reveal.

The metabolic pathway analysis carried out on the metabolomics results on the three sets of lysogens, displayed that there were a set of common pathways that were impacted by all the groups of phages. This illustrates a core shift in metabolism caused by the addition of prophages, suggesting these pathways are key to the subversion of Pa by prophages. Many of the core pathways that were significantly impacted by all phage groups are involved in fatty acid metabolism, which is associated to the cell membrane. These pathways included but were not limited to peptidoglycan, pantothenate and CoA biosynthesis, glycerophospholipid metabolism, and biotin. A previous study showed that prophages upregulate the biotin pathway in *E.coli*, which is rate limiting to cell growth, and promotes broad-range antimicrobial tolerance by increasing cell wall lipids (Holt et al., 2017). This suggests that the cell membrane could be an area targeted by the Pa temperate phages. Interestingly, it seems that temperate phages of two different Gram-negative species, Pa and *E.coli*, both utilise biotin as part of conversion play some role in cell physiology that

promotes lysogeny. It may be that this work is beginning to determine a core approach of temperate phages to subvert cell function in Gram negative bacteria. Further work using commercially available kits would allow determination of biotin concentration unequivocally. This provides valuable information about metabolic pathways leading to an understanding of the phage-mediated adaptations, which may also be used in the development of new strategies regarding the prevention and treatment of infections (Xu et al., 2014). This also informs future studies off the following preliminary pathway results which can be analysed in further detail with subsequent studies.

9.1.3. Ability of prophages to change the virulence of their Pa host

It has been shown that temperate phages have the potential to impact their host's virulence, such as the Stx phage 24B in *E.coli*, which increases the virulence due to the production of the Shiga-toxin when the prophage is induced (Wagner et al., 2001b). Pa filamentous phages (Pf) have been associated with an increase in the virulence of Pa (Secor et al., 2015), as they increase biofilm formation by spontaneously assembling into a liquid crystalline structure (Burgener et al., 2019). This also insulates against antibiotics, driving antibiotic tolerance against the most common anti-pseudomonas antibiotics, with data showing evidence that Pf phages may contribute to the clinical outcomes in Pa infection in CF.

The results from the *Galleria* virulence model (section 5.4.3) comparing the virulence of the CF and BR PAO1 lysogens (in the 4 clinical/aetiology groups) compared to wild-type PAO1 showed that temperate phage conversion lowers the virulence of the PAO1. A reduction of virulence has been reported in clinical Pa isolates over time (Lelong et al., 2011) and this data indicates that lysogenic phage infection may play a part in this phenotype. These results are important as this change was seen at differing levels across all 20 PAO1 lysogens converted with CF and BR Pa phages. This also means that prophages are very possibly one of the reasons for Pa becoming chronic in the lung. The core shift in the bacterial cell metabolism, illustrated here with conversion, may be one of the reasons that

this occurs. Alteration of fatty acid or peptidoglycan synthesis may stimulate a change at the cell surface (Rowlett et al., 2017) that masks its recognition by the larval immune response and leads to persistence. There is little understanding as to why chronically infected patients experience unpredictable exacerbations. The lowering of virulence of Pa in the lung could be part of the answer, as it would promote host immune evasion and may explain how Pa remain in the lung between periods of exacerbation without stimulating inflammation, alongside other factors, such as biofilm formation. It has been suggested that an increase in bacterial density may be a trigger, although it has been shown that this is not the case in CF patients (Stressmann et al., 2011). One hypothesis is that induction of phages could be a trigger for exacerbation. These inducible phages are less virulent when carrying a prophage, therefore loss through induction would cause the Pa to return to a higher virulent state and may cause exacerbation. This would need further study, however, there are numerous causes of induction that can be seen in lungs, such as exposure to reactive oxygen species or certain antibiotics, such as ciprofloxacin (Przerwa et al., 2006, Fothergill et al., 2011). These results were not in a model comparable to the CF lung and strains can exhibit differences in virulence between different animal models (Fothergill et al., 2012). Therefore, it would be beneficial to determine if a similar trend could be seen in a more clinically relevant virulence model, such as the mouse CF lung model (Wilke et al., 2011), the porcine lung *ex vivo* model (Harrison et al., 2014) or the ferret model (Sun et al., 2014).

Pf phages of Pa have been shown to change virulence of Pa by the production alginate and rhamnolipid (Zulianello et al., 2006, Tsao et al., 2018). They also suppress phagocytosis against bacterial infection, in turn, lowering the inflammation and associated virulence. Pf phages do this by being internalised by the leukocytes, which results in ssDNA phage producing phage RNA, which triggers Toll-like receptor 3 (TLR3) and TIR domain-containing adapter-inducing interferon- β (TRIF)-dependent type I interferon production, inhibition of tumor necrosis factor (TNF), and the suppression of phagocytosis (Sweere et al., 2019), therefore, allowing the bacteria to continue to colonise and possibly becoming

chronic. From the virulence results, together with the metabolomics from the same lysogens, it would be possible with further experimentation to identify which pathways or metabolites are responsible for lowering the virulence of PAO1. For example, looking at known virulence reducing mechanisms, such as in the Pf phages (Tsao et al., 2018) may determine the metabolic changes that could happen, and these may also be seen in other Pa phage and could be identified. This would shed some light onto how Pa avoids the human immune system and causes Pa infections to become chronic.

9.1.4. *Pseudomonas aeruginosa* prophage genetic diversity

The genetic diversity of temperate phages is frequently overlooked in comparison to lytic phages, as study of lytic phages and their genomes is more common. There's a total of 8,437 complete phage genomes divided into 12 families (based on the ICTV classification) as of September 2019. Differentiating between lytic and lysogenic phages is difficult as temperate phage genomes within bacterial genomes that can induce and act as mobilisable viral elements, and therefore be isolated as a lytic phage, meaning that genomic analysis is needed to see if the phage has lysogeny genes. Prophages can also be overlooked especially if they are not inducible, offering most repositories that are biased towards lytic phages. This makes it difficult to acquire data to analyse the diversity of prophages as individual phage genomes to compare. The research into temperate phages in Pa has focused upon single or small numbers of prophages of interest in detail, such as the work on the prophages of the Liverpool epidemic strain (LES) of Pa (Winstanley et al., 2009, Fothergill et al., 2011, James et al., 2015, Davies et al., 2016a, Davies et al., 2016c).

9.1.5 Investigation of prophage diversity within Pa from the International *Pseudomonas* Consortium Database (IPCD)

Whilst investigating temperate phages from Pa, access was obtained to the IPCD (the largest collection of internationally sourced Pa genomes), which was valuable to look at the genetic diversity of prophages in the Pa species. This could determine the distribution of different types of Pa temperate phages and genes present in differing environments, other than the lung. Also, it confirms whether any phage types or genes are more prevalent in Pa

from lungs. There have been a small number of studies into lytic phages that infect *Pa* (Sepulveda-Robles et al., 2012, Essoh et al., 2015, Ha and Denver, 2018) and even fewer looking at the diversity of temperate phages (prophages) in *Pa*, which have only looked at small numbers of phages (Tariq et al., 2015).

In chapter 6 we show the largest study to date, comparing carriage and genetic diversity of temperate phages as prophages within *Pa* isolates from the IPCD (1030 genomes). The findings revealed that there are a high number of prophages carried by *Pa* isolates. A total of 6499 phages were identified in the 1030 genomes using PHASTER, with higher numbers of incomplete prophages in the *Pa* genomes than intact or questionable prophages. This level of carriage has been reported in other bacterial backgrounds. For example, in *E.coli* O157:H7 strain EDL933, 17 remnant or cryptic prophages are reported alongside the single inducible bacteriophage 933W (Plunkett et al, 1999) and therefore carriage in some Gram-negative bacteria is high. Cryptic phages have been associated with increase adherence in epithelial cells and upregulation of type III secretion systems in *E.coli* (Flockhart et al., 2012). This underlines that prophages are an important area of bacterial accessory genome that may carry genes beneficial to the host. These genomic regions of remnant prophages that are conserved across bacteria should be a future area of study. This thesis illustrates that the complexity in the numbers of prophage carriage is shared between both clinical and environmental isolates. Only one isolate out of 1030 isolates showed no prophage carriage and would be a bacterium to study downstream of this project to see how this exclusion is occurring. It is clear that these phages significantly contribute to the genomic diversity of *Pa* across the IPCD database.

The diversity of the intact temperate phages in the IPCD is shown in figure 6.4, which demonstrates 8 groups of prophages clustered based on similarities at the protein level, however as stated in the results these 8 groups are likely to be underrepresenting the actual number of groups there are here as many smaller groups were classed as one larger one, meaning that the diversity of the prophages is likely to be even greater than shown here

which will be shown when publishing this data. The 8 clades that were shown demonstrate different types of phages, which if looked at in more detail and a lower cut off could show even more different types of phages within these 8 groups, showing there is a high level of diversity seen in the temperate phages of Pa. These results also confirm other smaller studies comparing temperate and lytic phages of Pa are as diverse (Sepulveda- Robles et al., 2012). Tariq et al (2019) also found 8 clades of phages when comparing temperate phages of Pa, although this was from a smaller panel of only 105 phages, suggesting that even with a much larger number, the number of phage groups associated as prophage are similar, although there are differences between both this and the previous study. Of the eight clades identified by Tariq et al. and colleagues (2019) only three showed similarity to those identified in this study, namely F10-like, PhiCTX- like and Phi297-like. The most prevalent groups in this study was the YMC11-like phages, however, this was not seen in Tariq et al (2019), showing there may be differing phage types found because of the larger panel of Pa genome studied here. The smallest clade here was Phi3-like, which was highly prevalent in their panel. Due to the IPCD being highly biased towards Pa from CF and lung infection, this may be why there are comparisons with Tariq (2019), as these phages were all from Pa from chronic lung infection. It would be useful to compare to a study of only environmental Pa prophages to see if there were any that were missed here, as there were a few outliers that did not cluster, which may be of interest for further study.

By comparing whole phage genomes, the clustering is based on the numbers of shared proteins of the phage, which is why they cluster into types of phage, as most viruses have a conserved genome order linked to their integration and replication strategies. Therefore, it is a challenge to focus and compare the accessory genomes of prophages, and to determine the function of these phage accessory genes that are ones associated with the benefits to the host and can be carried between bacteria. There further complexities to determining the gene products responsible as recent work has illustrated that the CII protein (a core phage gene) expressed from one of the most conserved genes in lambdoid like temperate phages (24B) offers the bacteria cell bile acid tolerance (Veses-Garcia et al.,

2015, Smith et al., 2012). Therefore, some core genes may also have hidden functions outside of prophage replication.

Thus, the future focus should be on the accessory genomes of prophages to determine their role in the biology of the bacterial pathogen. Therefore, the next step in this research would be to look at the phages within each of the clades in more detail and disregard their core genome similarities, concentrating on the similarities in their accessory genomes. It would be of particular interest to look at those that have high conservation and are present in isolates associated with chronic lung infection, which may link to the importance of that gene for the biology of the Pa or phage within the lung environment. Therefore functional studies would also be necessary to show the effects of these genes on their host Pa.

The panel of 6499 prophages identified from the IPCD were run against the VFDB to establish how prevalent virulence genes were within prophages of Pa, as it has been previously reported that prophages can carry functional virulence genes that may alter the pathogenic profile of the infected bacterium. Examples include, *stx*, *lom* and *bor* (Wagner et al., 2001, Vica Pacheco et al., 1997, Plunkett et al., 1999). The results indicated that 2100 prophages carry a VF gene with 23-100% sequence identity of which 524 intact (putatively mobile) phages. A total of 23 intact phages carried a gene that had 100% identity to the genes relating to virulence genes within the VFDB, suggesting these are more likely to be functional genes carried by the phages. These results show that over 30% of the temperate phages carry a gene or part of a gene with similarity to known virulence factors. Further work here could focus on the VF genes carried by prophages associated to lung infection and examine whether they affect the virulence of the Pa host or give it an advantage within the lung environment, such as genes associated with biofilm formation.

9.1.6 Evolution of prophages within Pa isolates over time in the chronically infected lung

This is the first study, to the best of our knowledge, which examines the changes within prophages of Pa over the course of chronic infection of the lung. It has shown evidence that

temperate phages gain genes in expansions that could benefit the host Pa, resulting in a Pa isolate able to persist in the lungs for longer, hence, facilitating chronic infection. There have been few studies investigating the evolution of temperate phages within a Pa bacterial background (James et al., 2015), especially from longitudinal isolates from patients with a chronic infection, as experimental evolution is an easier, faster and more commonly used method (O'Brien et al., 2019). However, the conditions in the lung during a chronic infection cannot be fully modelled *in vitro*. Phages are genetically diverse with their genome architectures being mosaic and evolution driven by horizontal gene transfer (HGT) with other phages and host genomes (Pedulla et al., 2003). Lambdoid-like dsDNA phages usually carry an artillery of multiple ssDNA recombinases, which increases the likelihood/frequency of recombination to occur. Therefore, with prophages integrated, carrying these may increase the rates of gene exchange between remnant and mobilisable prophages. This may also mean that changes can occur even over a short period of time. This high rate of evolution would benefit phages and the bacterial host as we now know that infection in the lung is not clonal and therefore there is a genetic reservoir that can be transduced by phages between bacteria in the same environment. We illustrate this movement in the longitudinal samples in chapter 8. These genomic changes to the phage and bacteria are then subject to evolutionary selection in the environment they reside.

The research in chapter 8 focussed on the prophages from a longitudinal collection of Pa isolates from BR patients over a period of several years, looking closely at the carriage, loss, and genetic change of temperate phages. The results revealed that different temperate phage profiles are seen in the isolates longitudinally collected from each patient. When the same strain of Pa is present over a number of samples, genetic changes can be seen within the prophages identified, such as expansions to the prophage genomes, which suggests evolution of the phage. A change in prophage number was also seen, showing infection by a new phage or loss of a temperate phage by induction, illustrating that there is gain and loss of phages over time. This gain, loss and evolution is assumed to happen due to their life cycle. However, the literature reporting the activity and evolution of temperate phages

in areas of chronic infection such as the lung is very limited. Selective pressures may lead to evolution of phages over the progression of the chronic lung infection (O'Brien et al., 2019), such as antibiotic administration, which may cause induction of temperate phages and alterations to their genomes (Fothergill et al., 2011) and may be advantageous for the Pa host over time. We hypothesise that prophages benefit the bacterial fitness by subverting the Pa host in a way that enables better biofilm production in the lung or more tolerate antibiotics or other chemical treatments that are administered to the lungs which enables the Pa to remain within the lungs chronically.

Most of the phage expansions, such as phage A3_1 in patient A (Figure 8.6a), shows an expanded region compared to a similar phage seen in the isolate from the previous time point. These expansions were seen to be kept longitudinally (in the case of A3_1 it is kept throughout the remaining 7 isolates over a period of 3 years) with high conservation in sequence identity, which would suggest that they are important or neutral to the host, as they are conserved and not lost. From annotating the expanded regions, there were genes that could be beneficial to the host Pa, including TetR family transcriptional regulator, which play a role in DNA binding that regulate gene expression, which could be beneficial to the Pa host. The impact that these expanded prophage genomes may have on the Pa host have not yet been studied, even though prophages are known to be diverse and changeable entities that can drive evolution (Rodriguez-Valera et al., 2016). Therefore, this study opens up an area that could be of great importance in the field of Pa, and that should be examined in more detail in future studies.

9.2 FURTHER STUDY

In order to elucidate the core pathways and metabolites that are impacted by the addition of phages, a secondary MS step would need to be completed or experimental confirmation of the upregulation or downregulation of the pathway or metabolite would need to be carried out. It would also be beneficial to integrate some of the phages used in the metabolomics into a different host to see if the same core set of pathways are seen to be impacted. If a

core set of metabolites are confirmed, they can be tested against other lysogens containing other Pa prophages originating from the lung environment to see how wide spread these core changes are. Also, testing against lysogens containing environmental Pa prophages may elucidate pathways that are only impacted by prophages from the lungs, further pinpointing the pathways that are involved in adaption of the Pa to the lung environment.

It would be interesting to see if the phages integrating into PA14 from the IPCD have the same effects on the virulence, as seen in the clinical Pa phages. As PA14 is a highly virulent strain opposed to PAO1, if a significant reduction was seen it would add evidence to the effect of prophages on the virulence of the Pa.

As only the intact prophages that were identified from the IPCD were investigated here, but both questionable and the incomplete phages were also identified, this is something that should be examined in further detail. This would allow us to see if there are commonalities between the cryptic (incomplete) phages from Pa originating from lung infection, as there were virulence gene hits against the VFDB for some of the incomplete phages, suggesting they may play a part in the virulence of the Pa also, this could be explored further by using knockouts to see how the absence of this VF gene effects the host, in terms of fitness and virulence in a virulence model such as the galleria model used.

The prophages that were seen to change over time in the same strain from the longitudinal isolates could be analysed by metabolomics as they already have a common background (if there are no other prophage changes). This could be done to see what effects phage evolution has compared to the original isolate prior to the expansion of the phage genome. This would allow us to examine the metabolites that change during phage evolution from the same patients, illustrating exactly what metabolic effects the conserved expansion has on the host Pa, which may convey a benefit to the host.

This work has shown that prophages of Pa significantly change the metabolism of their Pa host and are able to reduce the virulence in *Galleria* larvae model. Temperate phages are also able to evolve while in the chronic lung and transfer between different strains of Pa

within the lung. This provides solid evidence that further work is vital to understand the impact of temperate phages on Pa and their role in allowing the host to persist and cause chronic infection.

10 REFERENCES

- AARON, S. D., RAMOTAR, K., FERRIS, W., VANDEMHEEN, K., SAGINUR, R., TULLIS, E., HAASE, D., KOTTACHCHI, D., ST DENIS, M. & CHAN, F. 2004. Adult cystic fibrosis exacerbations and new strains of *Pseudomonas aeruginosa*. *Am J Respir Crit Care Med*, 169, 811-5.
- ABEDON, S. 2011. Phage therapy pharmacology: calculating phage dosing. *Adv Appl Microbiol*, 77, 1-40.
- ACKERMANN, H. W. 2009. Phage classification and characterization. *Methods Mol Biol*, 501, 127-40.
- ADRIAENSSENS, E. M., SULLIVAN, M. B., KNEZEVIC, P., VAN ZYL, L. J., SARKAR, B. L., DUTILH, B. E., ALFENAS-ZERBINI, P., ŁOBOCKA, M., TONG, Y., BRISTER, J. R., MORENO SWITT, A. I., KLUMPP, J., AZIZ, R. K., BARYLSKI, J., UCHIYAMA, J., EDWARDS, R. A., KROPINSKI, A. M., PETTY, N. K., CLOKIE, M. R. J., KUSHKINA, A. I., MOROZOVA, V. V., DUFFY, S., GILLIS, A., RUMNIEKS, J., KURTBOKE, I., CHANISHVILI, N., GOODRIDGE, L., WITTMANN, J., LAVIGNE, R., JANG, H. B., PRANGISHVILI, D., ENAULT, F., TURNER, D., PORANEN, M. M., OKSANEN, H. M. & KRUPOVIC, M. 2020. Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch Virol*, 165, 1253-1260.
- AKHTER, S., AZIZ, R. K. & EDWARDS, R. A. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res*, 40, e126.
- AKSYUK, A. A. & ROSSMANN, M. G. 2011. Bacteriophage assembly. *Viruses*, 3, 172-203.
- AL-ALOUL, M., CRAWLEY, J., WINSTANLEY, C., HART, C. A., LEDSON, M. J. & WALSHAW, M. J. 2004. Increased morbidity associated with chronic infection by an epidemic *Pseudomonas aeruginosa* strain in CF patients. *Thorax*, 59, 334-6.
- AL-HASAN, M. N., WILSON, J. W., LAHR, B. D., ECKEL-PASSOW, J. E. & BADDOUR, L. M. 2008. Incidence of *Pseudomonas aeruginosa* bacteremia: a population-based study. *Am J Med*, 121, 702-8.
- ALEMAYEHU, D., CASEY, P. G., MCAULIFFE, O., GUINANE, C. M., MARTIN, J. G., SHANAHAN, F., COFFEY, A., ROSS, R. P. & HILL, C. 2012. Bacteriophages phiMR299-2 and phiNH-4 can eliminate *Pseudomonas aeruginosa* in the murine lung and on cystic fibrosis lung airway cells. *MBio*, 3, e00029-12.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol*, 215, 403-10.
- ALTUVIA, S., LOCKER-GILADI, H., KOPY, S., BEN-NUN, O. & OPPENHEIM, A. B. 1987. RNase III stimulates the translation of the cIII gene of bacteriophage lambda. *Proc Natl Acad Sci U S A*, 84, 6511-5.
- ANGRILL, J., AGUSTI, C., DE CELIS, R., RANO, A., GONZALEZ, J., SOLE, T., XAUBET, A., RODRIGUEZ-ROISIN, R. & TORRES, A. 2002. Bacterial colonisation in patients with bronchiectasis: microbiological pattern and risk factors. *Thorax*, 57, 15-9.
- ANKRAH, N. Y., MAY, A. L., MIDDLETON, J. L., JONES, D. R., HADDEN, M. K., GOODING, J. R., LECLEIR, G. R., WILHELM, S. W., CAMPAGNA, S. R. & BUCHAN, A. 2014. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *Isme j*, 8, 1089-100.
- ARMSTRONG, D. S., GRIMWOOD, K., CARLIN, J. B., CARZINO, R., GUTIERREZ, J. P., HULL, J., OLINSKY, A., PHELAN, E. M., ROBERTSON, C. F. & PHELAN, P. D. 1997. Lower airway inflammation in infants and young children with cystic fibrosis. *Am J Respir Crit Care Med*, 156, 1197-204.
- ARNDT, D., GRANT, J. R., MARCU, A., SAJED, T., PON, A., LIANG, Y. & WISHART, D. S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*, 44, W16-21.

- ASADULGHANI, M., OGURA, Y., OOKA, T., ITOH, T., SAWAGUCHI, A., IGUCHI, A., NAKAYAMA, K. & HAYASHI, T. 2009. The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog*, 5, e1000408.
- ASHISH, A., SHAW, M., WINSTANLEY, C., LEDSON, M. J. & WALSHAW, M. J. 2012. Increasing resistance of the Liverpool Epidemic Strain (LES) of *Pseudomonas aeruginosa* (Psa) to antibiotics in cystic fibrosis (CF)--a cause for concern? *J Cyst Fibros*, 11, 173-9.
- ATHANAZIO, R. 2012. Airway disease: similarities and differences between asthma, COPD and bronchiectasis. *Clinics (Sao Paulo)*, 67, 1335-43.
- BAHAROGLU, Z. & MAZEL, D. 2014. SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol Rev*, 38, 1126-45.
- BALDING, C., BROMLEY, S. A., PICKUP, R. W. & SAUNDERS, J. R. 2005. Diversity of phage integrases in Enterobacteriaceae: development of markers for environmental analysis of temperate phages. *Environ Microbiol*, 7, 1558-67.
- BALTIMORE, R. S. & MITCHELL, M. 1980. Immunologic investigations of mucoid strains of *Pseudomonas aeruginosa*: comparison of susceptibility to opsonic antibody in mucoid and nonmucoid strains. *J Infect Dis*, 141, 238-47.
- BARONDESS, J. J. & BECKWITH, J. 1990. A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature*, 346, 871-4.
- BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315, 1709-12.
- BARTELL, J. A., SOMMER, L. M., HAAGENSEN, J. A. J., LOCH, A., ESPINOSA, R., MOLIN, S. & JOHANSEN, H. K. 2019. Evolutionary highways to persistent bacterial infection. *Nat Commun*, 10, 629.
- BEAR, C. E., LI, C. H., KARTNER, N., BRIDGES, R. J., JENSEN, T. J., RAMJEESINGH, M. & RIORDAN, J. R. 1992. Purification and functional reconstitution of the cystic fibrosis transmembrane conductance regulator (CFTR). *Cell*, 68, 809-18.
- BELLELLI, G., MORANDI, A., DI SANTO, S. G., MAZZONE, A., CHERUBINI, A., MOSSELLO, E., BO, M., BIANCHETTI, A., ROZZINI, R., ZANETTI, E., MUSICCO, M., FERRARI, A., FERRARA, N. & TRABUCCHI, M. 2016. "Delirium Day": a nationwide point prevalence study of delirium in older hospitalized patients using an easy standardized diagnostic tool. *BMC Med*, 14, 106.
- BEN MOHAMED, F., GARCIA-VERDUGO, I., MEDINA, M., BALLOY, V., CHIGNARD, M., RAMPHAL, R. & TOUQUI, L. 2012. A crucial role of Flagellin in the induction of airway mucus production by *Pseudomonas aeruginosa*. *PLoS One*, 7, e39888.
- BENSING, B. A., RUBENS, C. E. & SULLAM, P. M. 2001. Genetic loci of *Streptococcus mitis* that mediate binding to human platelets. *Infect Immun*, 69, 1373-80.
- BERLANGA, M. & GUERRERO, R. 2016. Living together in biofilms: the microbial cell factory and its biotechnological implications. *Microb Cell Fact*, 15, 165.
- BERNARD, C. S., BORDI, C., TERMINE, E., FILLOUX, A. & DE BENTZMANN, S. 2009. Organization and PprB-dependent control of the *Pseudomonas aeruginosa* tad Locus, involved in Flp pilus biology. *J Bacteriol*, 191, 1961-73.
- BERTANI, L. E. 1971. Stabilization of P2 tandem double lysogens by int mutations in the prophage. *Virology*, 46, 426-36.
- BJARNSHOLT, T., JENSEN, P. O., FIANDACA, M. J., PEDERSEN, J., HANSEN, C. R., ANDERSEN, C. B., PRESSLER, T., GIVSKOV, M. & HOIBY, N. 2009. *Pseudomonas aeruginosa* biofilms in the respiratory tract of cystic fibrosis patients. *Pediatr Pulmonol*, 44, 547-58.
- BLACK, L. W. 1989. DNA packaging in dsDNA bacteriophages. *Annu Rev Microbiol*, 43, 267-92.
- BLAHOVA, J., HUPKOVA, M. & KRČMERY, V., SR. 1994. Phage F-116 transduction of antibiotic resistance from a clinical isolate of *Pseudomonas aeruginosa*. *J Chemother*, 6, 184-8.
- BLAIR, D. F. 2003. Flagellar movement driven by proton translocation. *FEBS Lett*, 545, 86-95.

- BOBAY, L. M., ROCHA, E. P. & TOUCHON, M. 2013. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol*, 30, 737-51.
- BOJANOVIČ, K., D'ARRIGO, I. & LONG, K. S. 2017. Global Transcriptional Responses to Osmotic, Oxidative, and Imipenem Stress Conditions in *Pseudomonas putida*. *Appl Environ Microbiol*, 83.
- BONADIA, L. C., DE LIMA MARSON, F. A., RIBEIRO, J. D., PASCHOAL, I. A., PEREIRA, M. C., RIBEIRO, A. F. & BERTUZZO, C. S. 2014. CFTR genotype and clinical outcomes of adult patients carried as cystic fibrosis disease. *Gene*, 540, 183-90.
- BONDY-DENOMY, J., PAWLUK, A., MAXWELL, K. L. & DAVIDSON, A. R. 2013. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493, 429-32.
- BOPAKA, R. G., EL KHATTABI, W., JANAH, H., JABRI, H. & AFIF, H. 2015. Bronchiectasis: a bacteriological profile. *Pan Afr Med J*, 22, 378.
- BOSE, M. & BARBER, R. D. 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol*, 6, 223-7.
- BOUCHER, R. C. 2007. Airway surface dehydration in cystic fibrosis: pathogenesis and therapy. *Annu Rev Med*, 58, 157-70.
- BRADLEY, D. E. 1980. A function of *Pseudomonas aeruginosa* PAO polar pili: twitching motility. *Can J Microbiol*, 26, 146-54.
- BRUEGGEMANN, A. B., HARROLD, C. L., REZAEI JAVAN, R., VAN TONDER, A. J., MCDONNELL, A. J. & EDWARDS, B. A. 2017. Pneumococcal prophages are diverse, but not without structure or history. *Sci Rep*, 7, 42976.
- BRUSSOW, H., CANCHAYA, C. & HARDT, W. D. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev*, 68, 560-602, table of contents.
- BUCIOR, I., PIELAGE, J. F. & ENGEL, J. N. 2012. *Pseudomonas aeruginosa* pili and flagella mediate distinct binding and signaling events at the apical and basolateral surface of airway epithelium. *PLoS Pathog*, 8, e1002616.
- BURGENER, E. B., SWEERE, J. M., BACH, M. S., SECOR, P. R., HADDOCK, N., JENNINGS, L. K., MARVIG, R. L., JOHANSEN, H. K., ROSSI, E., CAO, X., TIAN, L., NEDELEC, L., MOLIN, S., BOLLYKY, P. L. & MILLA, C. E. 2019. Filamentous bacteriophages are associated with chronic *Pseudomonas* lung infections and antibiotic resistance in cystic fibrosis. *Sci Transl Med*, 11.
- BURKAL'TSEVA, M. V., KRYLOV, S. V., KROPINSKIĬ, A. M., PLETNEVA, E. A., SHABUROVA, O. V. & KRYLOV, V. N. 2011. [Bacteriophage phi297--the new species of temperate phages *Pseudomonas aeruginosa* with a mosaic genome: potential use in phagotherapy]. *Genetika*, 47, 900-4.
- BURNS, J. L., GIBSON, R. L., MCNAMARA, S., YIM, D., EMERSON, J., ROSENFELD, M., HIATT, P., MCCOY, K., CASTILE, R., SMITH, A. L. & RAMSEY, B. W. 2001. Longitudinal assessment of *Pseudomonas aeruginosa* in young children with cystic fibrosis. *J Infect Dis*, 183, 444-52.
- BURNS, N., JAMES, C. E. & HARRISON, E. 2015. Polylysogeny magnifies competitiveness of a bacterial pathogen in vivo. *Evol Appl*, 8, 346-51.
- BUSBY, B., KRISTENSEN, D. M. & KOONIN, E. V. 2013. Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. *Environ Microbiol*, 15, 307-12.
- CADY, K. C., WHITE, A. S., HAMMOND, J. H., ABENDROTH, M. D., KARTHIKEYAN, R. S., LALITHA, P., ZEGANS, M. E. & O'TOOLE, G. A. 2011. Prevalence, conservation and functional analysis of *Yersinia* and *Escherichia* CRISPR regions in clinical *Pseudomonas aeruginosa* isolates. *Microbiology*, 157, 430-437.
- CALLAGHAN, M. & MCCLEAN, S. 2012. Bacterial host interactions in cystic fibrosis. *Curr Opin Microbiol*, 15, 71-7.
- CAMPBELL, A. M. 1992. Chromosomal insertion sites for phages and plasmids. *J Bacteriol*, 174, 7495-9.

- CANCHAYA, C., DESIERE, F., MCSHAN, W. M., FERRETTI, J. J., PARKHILL, J. & BRÜSSOW, H. 2002. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. *Virology*, 302, 245-58.
- CANCHAYA, C., FOURNOUS, G. & BRUSSOW, H. 2004. The impact of prophages on bacterial chromosomes. *Mol Microbiol*, 53, 9-18.
- CASJENS, S. 2003. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol*, 49, 277-300.
- CASJENS, S., ADAMS, M. B., HALL, C. & KING, J. 1985. Assembly-controlled autogenous modulation of bacteriophage P22 scaffolding protein gene expression. *J Virol*, 53, 174-9.
- CASJENS, S. & HENDRIX, R. 1974. Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. *J Mol Biol*, 90, 20-5.
- CASPAR, D. L. & KLUG, A. 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol*, 27, 1-24.
- CEYSSENS, P. J., NOBEN, J. P., ACKERMANN, H. W., VERHAEGEN, J., DE VOS, D., PIRNAY, J. P., MERABISHVILI, M., VANECHOUTTE, M., CHIBEU, A., VOLCKAERT, G. & LAVIGNE, R. 2009. Survey of *Pseudomonas aeruginosa* and its phages: de novo peptide sequencing as a novel tool to assess the diversity of worldwide collected viruses. *Environ Microbiol*, 11, 1303-13.
- CF_TRUST. 2018a. *UK Cystic Fibrosis Registry Annual Data Report 2016* [Online]. Available: <https://www.cysticfibrosis.org.uk/~media/documents/the-work-we-do/uk-cf-registry/2016-registry-annual-data-report>. [Accessed 2019].
- CF_TRUST. 2018b. *What is cystic fibrosis?* [Online]. UK. Available: <https://www.cysticfibrosis.org.uk/what-is-cystic-fibrosis> [Accessed 10/12/18 2018].
- CHAO, L. & COX, E. C. 1983. COMPETITION BETWEEN HIGH AND LOW MUTATING STRAINS OF ESCHERICHIA COLI. *Evolution*, 37, 125-134.
- CHEN, Y., GOLDING, I., SAWAI, S., GUO, L. & COX, E. C. 2005. Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biol*, 3, e229.
- CHENG, K., SMYTH, R. L., GOVAN, J. R., DOHERTY, C., WINSTANLEY, C., DENNING, N., HEAF, D. P., VAN SAENE, H. & HART, C. A. 1996. Spread of beta-lactam-resistant *Pseudomonas aeruginosa* in a cystic fibrosis clinic. *Lancet*, 348, 639-42.
- CHEVALLEREAU, A., BLASDEL, B. G., DE SMET, J., MONOT, M., ZIMMERMANN, M., KOGADEEVA, M., SAUER, U., JORTH, P., WHITELEY, M., DEBARBIEUX, L. & LAVIGNE, R. 2016. Next-Generation "-omics" Approaches Reveal a Massive Alteration of Host RNA Metabolism during Bacteriophage Infection of *Pseudomonas aeruginosa*. *PLoS Genet*, 12, e1006134.
- CHEVALLEREAU, A., MEADEN, S., FRADET, O., LANDSBERGER, M., MAESTRI, A., BISWAS, A., GANDON, S., VAN HOUTE, S. & WESTRA, E. R. 2020. Exploitation of the Cooperative Behaviors of Anti-CRISPR Phages. *Cell Host Microbe*, 27, 189-198.e6.
- CHIANG, P. & BURROWS, L. L. 2003. Biofilm formation by hyperpiliated mutants of *Pseudomonas aeruginosa*. *J Bacteriol*, 185, 2374-8.
- CHOKKATHUKALAM, A., KIM, D. H., BARRETT, M. P., BREITLING, R. & CREEK, D. J. 2014. Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis*, 6, 511-24.
- CHOPIN, M. C., CHOPIN, A. & BIDNENKO, E. 2005. Phage abortive infection in lactococci: variations on a theme. *Curr Opin Microbiol*, 8, 473-9.
- COLVIN, K. M., GORDON, V. D., MURAKAMI, K., BORLEE, B. R., WOZNAK, D. J., WONG, G. C. & PARSEK, M. R. 2011. The pel polysaccharide can serve a structural and protective role in the biofilm matrix of *Pseudomonas aeruginosa*. *PLoS Pathog*, 7, e1001264.
- COLVIN, K. M., IRIE, Y., TART, C. S., URBANO, R., WHITNEY, J. C., RYDER, C., HOWELL, P. L., WOZNAK, D. J. & PARSEK, M. R. 2012. The Pel and Psl polysaccharides provide *Pseudomonas aeruginosa* structural redundancy within the biofilm matrix. *Environ Microbiol*, 14, 1913-28.

- COSTELLO, A., HERBERT, G., FABUNMI, L., SCHAFFER, K., KAVANAGH, K. A., CARAHER, E. M., CALLAGHAN, M. & MCCLEAN, S. 2011. Virulence of an emerging respiratory pathogen, genus *Pandoraea*, in vivo and its interactions with lung epithelial cells. *J Med Microbiol*, 60, 289-99.
- COSTERTON, J. W., CHENG, K. J., GEESSEY, G. G., LADD, T. I., NICKEL, J. C., DASGUPTA, M. & MARRIE, T. J. 1987. Bacterial biofilms in nature and disease. *Annu Rev Microbiol*, 41, 435-64.
- COX, M. J., ALLGAIER, M., TAYLOR, B., BAEK, M. S., HUANG, Y. J., DALY, R. A., KARAOZ, U., ANDERSEN, G. L., BROWN, R., FUJIMURA, K. E., WU, B., TRAN, D., KOFF, J., KLEINHENZ, M. E., NIELSON, D., BRODIE, E. L. & LYNCH, S. V. 2010. Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One*, 5, e11044.
- CUNNICK, W. R., CROMIE, J. B., CORTELL, R., WRIGHT, B., BEACH, E., SELTZER, F. & MILLER, S. 1972. Value of biochemical profiling in a periodic health examination program: analysis of 1,000 cases. *Bull N Y Acad Med*, 48, 5-22.
- D'HERELLE, F. 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad. Sci. Paris*, 373-5.
- DARLING, A. C., MAU, B., BLATTNER, F. R. & PERNA, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, 14, 1394-403.
- DASGUPTA, N., WOLFGANG, M. C., GOODMAN, A. L., ARORA, S. K., JYOT, J., LORY, S. & RAMPHAL, R. 2003. A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol Microbiol*, 50, 809-24.
- DAVIES, E. V., JAMES, C. E., BROCKHURST, M. A. & WINSTANLEY, C. 2017. Evolutionary diversification of *Pseudomonas aeruginosa* in an artificial sputum model. *BMC Microbiol*, 17, 3.
- DAVIES, E. V., JAMES, C. E., KUKAVICA-IBRULJ, I., LEVESQUE, R. C., BROCKHURST, M. A. & WINSTANLEY, C. 2016a. Temperate phages enhance pathogen fitness in chronic lung infection. *Isme j*, 10, 2553-5.
- DAVIES, E. V., JAMES, C. E., WILLIAMS, D., O'BRIEN, S., FOTHERGILL, J. L., HALDENBY, S., PATERSON, S., WINSTANLEY, C. & BROCKHURST, M. A. 2016b. Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proc Natl Acad Sci U S A*, 113, 8266-71.
- DAVIES, E. V., WINSTANLEY, C., FOTHERGILL, J. L. & JAMES, C. E. 2016c. The role of temperate bacteriophages in bacterial infection. *FEMS Microbiol Lett*, 363, fnw015.
- DE SMET, J., ZIMMERMANN, M., KOGADEEVA, M., CEYSSENS, P. J., VERMAELEN, W., BLASDEL, B., BIN JANG, H., SAUER, U. & LAVIGNE, R. 2016. High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *Isme j*, 10, 1823-35.
- DE SOYZA, A., PERRY, A., HALL, A. J., SUNNY, S. S., WALTON, K. E., MUSTAFA, N., TURTON, J., KENNA, D. T. & WINSTANLEY, C. 2014. Molecular epidemiological analysis suggests cross-infection with *Pseudomonas aeruginosa* is rare in non-cystic fibrosis bronchiectasis. *Eur Respir J*, 43, 900-3.
- DE WYNGAERT, M. & HINKLE, D. C. 1979. Involvement of DNA gyrase in replication and transcription of bacteriophage T7 DNA. *J Virol*, 29, 529-35.
- DEBARBIEUX, L., LEDUC, D., MAURA, D., MORELLO, E., CRISCUOLO, A., GROSSI, O., BALLOY, V. & TOUQUI, L. 2010. Bacteriophages can treat and prevent *Pseudomonas aeruginosa* lung infections. *J Infect Dis*, 201, 1096-104.
- DEPKE, T., THÖMING, J. G., KORDES, A., HÄUSSLER, S. & BRÖNSTRUP, M. 2020. Untargeted LC-MS Metabolomics Differentiates Between Virulent and Avirulent Clinical Strains of *Pseudomonas aeruginosa*. *Biomolecules*, 10.
- DETTMAN, J. R., RODRIGUE, N., AARON, S. D. & KASSEN, R. 2013. Evolutionary genomics of epidemic and nonepidemic strains of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*, 110, 21065-70.

- DEZIEL, E., COMEAU, Y. & VILLEMUR, R. 2001. Initiation of biofilm formation by *Pseudomonas aeruginosa* 57RP correlates with emergence of hyperpiliated and highly adherent phenotypic variants deficient in swimming, swarming, and twitching motilities. *J Bacteriol*, 183, 1195-204.
- DIARD, M., BAKKEREN, E., CORNUAULT, J. K., MOOR, K., HAUSMANN, A., SELLIN, M. E., LOVERDO, C., AERTSEN, A., ACKERMANN, M., DE PAEPE, M., SLACK, E. & HARDT, W. D. 2017. Inflammation boosts bacteriophage transfer between *Salmonella* spp. *Science*, 355, 1211-1215.
- DODD, I. B., PERKINS, A. J., TSEMITSIDIS, D. & EGAN, J. B. 2001. Octamerization of lambda CI repressor is needed for effective repression of P(RM) and efficient switching from lysogeny. *Genes Dev*, 15, 3013-22.
- DOIG, P., TODD, T., SASTRY, P. A., LEE, K. K., HODGES, R. S., PARANCHYCH, W. & IRVIN, R. T. 1988. Role of pili in adhesion of *Pseudomonas aeruginosa* to human respiratory epithelial cells. *Infect Immun*, 56, 1641-6.
- DONNELLY, A., YATA, T., BENTAYEBI, K., SUWAN, K. & HAJITOU, A. 2015. Bacteriophage Mediates Efficient Gene Transfer in Combination with Conventional Transfection Reagents. *Viruses*, 7, 6476-89.
- DUDA, R. L., MARTINCIC, K. & HENDRIX, R. W. 1995. Genetic basis of bacteriophage HK97 prohead assembly. *J Mol Biol*, 247, 636-47.
- DUFF, R. M., SIMMONDS, N. J., DAVIES, J. C., WILSON, R., ALTON, E. W., PANTELIDIS, P., COX, M. J., COOKSON, W. O., BILTON, D. & MOFFATT, M. F. 2013. A molecular comparison of microbial communities in bronchiectasis and cystic fibrosis. *Eur Respir J*, 41, 991-3.
- DY, R. L., PRZYBILSKI, R., SEMEIJN, K., SALMOND, G. P. & FINERAN, P. C. 2014. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res*, 42, 4590-605.
- DZIEWIT, L. & RADLINSKA, M. 2016. Two Inducible Prophages of an Antarctic *Pseudomonas* sp. ANT_H14 Use the Same Capsid for Packaging Their Genomes - Characterization of a Novel Phage Helper-Satellite System. *PLoS One*, 11, e0158889.
- EDDY, S. R. 1998. Profile hidden Markov models. *Bioinformatics*, 14, 755-63.
- ELAMIN, A. A., STEINICKE, S., OEHLMANN, W., BRAUN, Y., WANAS, H., SHURALEV, E. A., HUCK, C., MARINGER, M., ROHDE, M. & SINGH, M. 2017. Novel drug targets in cell wall biosynthesis exploited by gene disruption in *Pseudomonas aeruginosa*. *PLoS One*, 12, e0186801.
- ESSOH, C., LATINO, L., MIDOUX, C., BLOUIN, Y., LOUKOU, G., NGUETTA, S. P., LATHRO, S., CABLANMIAN, A., KOUASSI, A. K., VERGNAUD, G. & POURCEL, C. 2015. Investigation of a Large Collection of *Pseudomonas aeruginosa* Bacteriophages Collected from a Single Environmental Source in Abidjan, Côte d'Ivoire. *PLoS One*, 10, e0130548.
- ESTAHBANATI, H. K., KASHANI, P. P. & GHANAATPISHEH, F. 2002. Frequency of *Pseudomonas aeruginosa* serotypes in burn wound infections and their resistance to antibiotics. *Burns*, 28, 340-8.
- FALKOWSKI, P. G., FENCHEL, T. & DELONG, E. F. 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320, 1034-9.
- FANCELLO, L., DESNUES, C., RAOULT, D. & ROLAIN, J. M. 2011. Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota. *J Antimicrob Chemother*, 66, 2448-54.
- FAURE, E., KWONG, K. & NGUYEN, D. 2018. *Pseudomonas aeruginosa* in Chronic Lung Infections: How to Adapt Within the Host? *Front Immunol*, 9, 2416.
- FEINER, R., ARGOV, T., RABINOVICH, L., SIGAL, N., BOROVOK, I. & HERSKOVITS, A. A. 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol*, 13, 641-50.
- FELDMAN, M., BRYAN, R., RAJAN, S., SCHEFFLER, L., BRUNNERT, S., TANG, H. & PRINCE, A. 1998. Role of flagella in pathogenesis of *Pseudomonas aeruginosa* pulmonary infection. *Infect Immun*, 66, 43-51.

- FERNANDES, S. & SÃO-JOSÉ, C. 2018. Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. *Viruses*, 10.
- FILIPPOV, A. A., SERGUEEV, K. V., HE, Y., HUANG, X. Z., GNADE, B. T., MUELLER, A. J., FERNANDEZ-PRADA, C. M. & NIKOLICH, M. P. 2011. Bacteriophage-resistant mutants in *Yersinia pestis*: identification of phage receptors and attenuation for mice. *PLoS One*, 6, e25486.
- FINERAN, P. C., BLOWER, T. R., FOULDS, I. J., HUMPHREYS, D. P., LILLEY, K. S. & SALMOND, G. P. 2009. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A*, 106, 894-9.
- FLOCKHART, A. F., TREE, J. J., XU, X., KARPIYEVICH, M., MCATEER, S. P., ROSENBLUM, R., SHAW, D. J., LOW, C. J., BEST, A., GANNON, V., LAING, C., MURPHY, K. C., LEONG, J. M., SCHNEIDERS, T., LA RAGIONE, R. & GALLY, D. L. 2012. Identification of a novel prophage regulator in *Escherichia coli* controlling the expression of type III secretion. *Mol Microbiol*, 83, 208-23.
- FOTHERGILL, J. L., MOWAT, E., WALSHAW, M. J., LEDSON, M. J., JAMES, C. E. & WINSTANLEY, C. 2011. Effect of antibiotic treatment on bacteriophage production by a cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother*, 55, 426-8.
- FOTHERGILL, J. L., NEILL, D. R., LOMAN, N., WINSTANLEY, C. & KADIOGLU, A. 2014. *Pseudomonas aeruginosa* adaptation in the nasopharyngeal reservoir leads to migration and persistence in the lungs. *Nat Commun*, 5, 4780.
- FOTHERGILL, J. L., WALSHAW, M. J. & WINSTANLEY, C. 2012. Transmissible strains of *Pseudomonas aeruginosa* in cystic fibrosis lung infections. *Eur Respir J*, 40, 227-38.
- FOUTS, D., FRASER, C., READ, T. & NELSON, K. 2004. *Bacteriophage bioinformatics*, Humana Press Inc.
- FOUTS, D. E. 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res*, 34, 5839-51.
- FRANKLIN, M. J., NIVENS, D. E., WEADGE, J. T. & HOWELL, P. L. 2011. Biosynthesis of the *Pseudomonas aeruginosa* Extracellular Polysaccharides, Alginate, Pel, and Psl. *Front Microbiol*, 2, 167.
- FRESCHI, L., JEUKENS, J., KUKAVICA-IBRULJ, I., BOYLE, B., DUPONT, M. J., LAROCHE, J., LAROSE, S., MAAROUFI, H., FOTHERGILL, J. L., MOORE, M., WINSOR, G. L., AARON, S. D., BARBEAU, J., BELL, S. C., BURNS, J. L., CAMARA, M., CANTIN, A., CHARETTE, S. J., DEWAR, K., DEZIEL, E., GRIMWOOD, K., HANCOCK, R. E., HARRISON, J. J., HEEB, S., JELSBK, L., JIA, B., KENNA, D. T., KIDD, T. J., KLOCKGETHER, J., LAM, J. S., LAMONT, I. L., LEWENZA, S., LOMAN, N., MALOUIN, F., MANOS, J., MCARTHUR, A. G., MCKEOWN, J., MILOT, J., NAGHRA, H., NGUYEN, D., PEREIRA, S. K., PERRON, G. G., PIRNAY, J. P., RAINEY, P. B., ROUSSEAU, S., SANTOS, P. M., STEPHENSON, A., TAYLOR, V., TURTON, J. F., WAGLECHNER, N., WILLIAMS, P., THRANE, S. W., WRIGHT, G. D., BRINKMAN, F. S., TUCKER, N. P., TUMMLER, B., WINSTANLEY, C. & LEVESQUE, R. C. 2015a. Clinical utilization of genomics data produced by the international *Pseudomonas aeruginosa* consortium. *Front Microbiol*, 6, 1036.
- FRESCHI, L., JEUKENS, J., KUKAVICA-IBRULJ, I., BOYLE, B., LAROCHE, J., DUPONT, M.-J., LAROSE, S., MAAROUFI, H. & LEVESQUE, R. 2015b. *International Pseudomonas Consortium: the 1000 plus genome project*.
- FRESCHI, L., VINCENT, A. T., JEUKENS, J., EMOND-RHEAULT, J. G., KUKAVICA-IBRULJ, I., DUPONT, M. J., CHARETTE, S. J., BOYLE, B. & LEVESQUE, R. C. 2019. The *Pseudomonas aeruginosa* Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. *Genome Biol Evol*, 11, 109-120.
- FRIEDMAN, S. & GOTS, J. S. 1953. The purine and pyrimidine metabolism of normal and phage-infected *Escherichia coli*. *J Biol Chem*, 201, 125-35.
- FRIMMERSDORF, E., HORATZEK, S., PELNIKEVICH, A., WIEHLMANN, L. & SCHOMBURG, D. 2010. How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach. *Environ Microbiol*, 12, 1734-47.
- FROBISHER, M. B., M. 1927. *Transmissible toxicogenicity of streptococci*, Bull. .

- GAO, N. L., CHEN, J., WANG, T., LERCHER, M. J. & CHEN, W. H. 2019. Prokaryotic Genome Expansion Is Facilitated by Phages and Plasmids but Impaired by CRISPR. *Front Microbiol*, 10, 2254.
- GARCIA, D. E., BAIDOO, E. E., BENKE, P. I., PINGITORE, F., TANG, Y. J., VILLA, S. & KEASLING, J. D. 2008. Separation and mass spectrometry in microbial metabolomics. *Curr Opin Microbiol*, 11, 233-9.
- GARVEY, P., FITZGERALD, G. F. & HILL, C. 1995. *Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid* Applied and Environmental Microbiology.
- GERMAN, G. J. & MISRA, R. 2001. The TolC protein of Escherichia coli serves as a cell-surface receptor for the newly characterized TLS bacteriophage. *J Mol Biol*, 308, 579-85.
- GERTH, M. L., NIGON, L. V. & PATRICK, W. M. 2012. Characterization of the indole-3-glycerol phosphate synthase from *Pseudomonas aeruginosa* PAO1. *Protein J*, 31, 359-65.
- GESSARD 1984. Classics in infectious diseases. On the blue and green coloration that appears on bandages. By Carle Gessard (1850-1925). *Rev Infect Dis*, 6 Suppl 3, S775-6.
- GHAFOOR, A., HAY, I. D. & REHM, B. H. 2011. Role of exopolysaccharides in *Pseudomonas aeruginosa* biofilm formation and architecture. *Appl Environ Microbiol*, 77, 5238-46.
- GIARDINE, B., RIEMER, C., HARDISON, R. C., BURHANS, R., ELNITSKI, L., SHAH, P., ZHANG, Y., BLANKENBERG, D., ALBERT, I., TAYLOR, J., MILLER, W., KENT, W. J. & NEKRUTENKO, A. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res*, 15, 1451-5.
- GILTNER, C. L., VAN SCHAİK, E. J., AUDETTE, G. F., KAO, D., HODGES, R. S., HASSETT, D. J. & IRVIN, R. T. 2006. The *Pseudomonas aeruginosa* type IV pilin receptor binding domain functions as an adhesin for both biotic and abiotic surfaces. *Mol Microbiol*, 59, 1083-96.
- GINGERAS, T. R. & BROOKS, J. E. 1983. Cloned restriction/modification system from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*, 80, 402-6.
- GJERSING, E. L., HERBERG, J. L., HORN, J., SCHALDACH, C. M. & MAXWELL, R. S. 2007. NMR metabolomics of planktonic and biofilm modes of growth in *Pseudomonas aeruginosa*. *Anal Chem*, 79, 8037-45.
- GOH, S., HUSSAIN, H., CHANG, B. J., EMMETT, W., RILEY, T. V. & MULLANY, P. 2013. Phage ϕ C2 mediates transduction of Tn6215, encoding erythromycin resistance, between *Clostridium difficile* strains. *mBio*, 4, e00840-13.
- GUERRERO-FERREIRA, R. C., VIOLLIER, P. H., ELY, B., POINDEXTER, J. S., GEORGIEVA, M., JENSEN, G. J. & WRIGHT, E. R. 2011. Alternative mechanism for bacteriophage adsorption to the motile bacterium *Caulobacter crescentus*. *Proc Natl Acad Sci U S A*, 108, 9963-8.
- GUSS, A. M., ROESELERS, G., NEWTON, I. L., YOUNG, C. R., KLEPAC-CERAJ, V., LORY, S. & CAVANAUGH, C. M. 2011. Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *Isme j*, 5, 20-9.
- HA, A. D. & DENVER, D. R. 2018. Comparative Genomic Analysis of 130 Bacteriophages Infecting Bacteria in the Genus *Pseudomonas*. *Front Microbiol*, 9, 1456.
- HACKER, J. & CARNIEL, E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep*, 2, 376-81.
- HAGGÅRD-LJUNGQUIST, E., HALLING, C. & CALENDAR, R. 1992. DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *J Bacteriol*, 174, 1462-77.
- HAIKO, J. & WESTERLUND-WIKSTROM, B. 2013. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel)*, 2, 1242-67.
- HALEY, C. L., COLMER-HAMOOD, J. A. & HAMOOD, A. N. 2012. Characterization of biofilm-like structures formed by *Pseudomonas aeruginosa* in a synthetic mucus medium. *BMC Microbiol*, 12, 181.

- HAN, M. L., ZHU, Y., CREEK, D. J., LIN, Y. W., ANDERSON, D., SHEN, H. H., TSUJI, B., GUTU, A. D., MOSKOWITZ, S. M., VELKOV, T. & LI, J. 2018. Alterations of Metabolic and Lipid Profiles in Polymyxin-Resistant *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother*, 62.
- HARGREAVES, K. R., KROPINSKI, A. M. & CLOKIE, M. R. 2014. Bacteriophage behavioral ecology: How phages alter their bacterial host's habits. *Bacteriophage*, 4, e29866.
- HARPER, D. R., PARRACHO, H. M. R. T., WALKER, J., SHARP, R., HUGHES, G., WERTHÉN, M., LEHMAN, S. & MORALES, S. 2014. Bacteriophages and Biofilms. *Antibiotics*, 3, 270-284.
- HARRISON, F., MURULI, A., HIGGINS, S. & DIGGLE, S. P. 2014. Development of an *Ex Vivo* Porcine Lung Model for Studying Growth, Virulence, and Signaling of *Pseudomonas aeruginosa*. *Infection and Immunity*, 82, 3312-3323.
- HART, C. A. & WINSTANLEY, C. 2002. Persistent and aggressive bacteria in the lungs of cystic fibrosis children. *Br Med Bull*, 61, 81-96.
- HAUSER, N. & ORSINI, J. 2015. Cepacia Syndrome in a Non-Cystic Fibrosis Patient. *Case Rep Infect Dis*, 2015, 537627.
- HAYASHI, F., SMITH, K. D., OZINSKY, A., HAWN, T. R., YI, E. C., GOODLETT, D. R., ENG, J. K., AKIRA, S., UNDERHILL, D. M. & ADEREM, A. 2001a. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, 410, 1099-103.
- HAYASHI, T., BABA, T., MATSUMOTO, H. & TERAWAKI, Y. 1990. Phage-conversion of cytotoxin production in *Pseudomonas aeruginosa*. *Mol Microbiol*, 4, 1703-9.
- HAYASHI, T., MAKINO, K., OHNISHI, M., KUROKAWA, K., ISHII, K., YOKOYAMA, K., HAN, C. G., OHTSUBO, E., NAKAYAMA, K., MURATA, T., TANAKA, M., TOBE, T., IIDA, T., TAKAMI, H., HONDA, T., SASAKAWA, C., OGASAWARA, N., YASUNAGA, T., KUHARA, S., SHIBA, T., HATTORI, M. & SHINAGAWA, H. 2001b. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res*, 8, 11-22.
- HE, J., BALDINI, R. L., DÉZIEL, E., SAUCIER, M., ZHANG, Q., LIBERATI, N. T., LEE, D., URBACH, J., GOODMAN, H. M. & RAHME, L. G. 2004. The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proc Natl Acad Sci U S A*, 101, 2530-5.
- HENDRIX, R. W. & DUDA, R. L. 1998. Bacteriophage HK97 head assembly: a protein ballet. *Adv Virus Res*, 50, 235-88.
- HENDRIX, R. W., SMITH, M. C., BURNS, R. N., FORD, M. E. & HATFULL, G. F. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci U S A*, 96, 2192-7.
- HENRY, M., LAVIGNE, R. & DEBARBIEUX, L. 2013. Predicting in vivo efficacy of therapeutic bacteriophages used to treat pulmonary infections. *Antimicrob Agents Chemother*, 57, 5961-8.
- HENTZER, M., TEITZEL, G. M., BALZER, G. J., HEYDORN, A., MOLIN, S., GIVSKOV, M. & PARSEK, M. R. 2001. Alginate overproduction affects *Pseudomonas aeruginosa* biofilm structure and function. *J Bacteriol*, 183, 5395-401.
- HIGH, N. J., DEADMAN, M. E. & 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. *Mol Microbiol*, 9, 1275-82.
- HILL, D., ROSE, B., PAJKOS, A., ROBINSON, M., BYE, P., BELL, S., ELKINS, M., THOMPSON, B., MACLEOD, C., AARON, S. D. & HARBOUR, C. 2005. Antibiotic susceptibilities of *Pseudomonas aeruginosa* isolates derived from patients with cystic fibrosis under aerobic, anaerobic, and biofilm conditions. *J Clin Microbiol*, 43, 5085-90.
- HILL, D. B., LONG, R. F., KISSNER, W. J., ATIEH, E., GARBARINE, I. C., MARKOVETZ, M. R., FONTANA, N. C., CHRISTY, M., HABIBPOUR, M., TARRAN, R., FOREST, M. G., BOUCHER, R. C. & BUTTON, B. 2018. Pathological mucus and impaired mucus clearance in cystic fibrosis patients result from increased concentration, not altered pH. *Eur Respir J*, 52.

- HILLIAM, Y., MOORE, M. P., LAMONT, I. L., BILTON, D., HAWORTH, C. S., FOWERAKER, J., WALSHAW, M. J., WILLIAMS, D., FOTHERGILL, J. L., DE SOYZA, A. & WINSTANLEY, C. 2017. *Pseudomonas aeruginosa* adaptation and diversification in the non-cystic fibrosis bronchiectasis lung. *Eur Respir J*, 49.
- HOIBY, N., BJARNSHOLT, T., GIVSKOV, M., MOLIN, S. & CIOFU, O. 2010a. Antibiotic resistance of bacterial biofilms. *Int J Antimicrob Agents*, 35, 322-32.
- HOIBY, N., CIOFU, O. & BJARNSHOLT, T. 2010b. *Pseudomonas aeruginosa* biofilms in cystic fibrosis. *Future Microbiol*, 5, 1663-74.
- HOLLOWAY, B. W., EGAN, J. B. & MONK, M. 1960. Lysogeny in *Pseudomonas aeruginosa*. *Aust J Exp Biol Med Sci*, 38, 321-9.
- HOLT, G. S., LODGE, J. K., MCCARTHY, A. J., GRAHAM, A. K., YOUNG, G., BRIDGE, S. H., BROWN, A. K., VESES-GARCIA, M., LANYON, C. V., SAILS, A., ALLISON, H. E. & SMITH, D. L. 2017. Shigatoxin encoding Bacteriophage varphi24B modulates bacterial metabolism to raise antimicrobial tolerance. *Sci Rep*, 7, 40424.
- HOPF, A., SCHREIBER, R., MALL, M., GREGER, R. & KUNZELMANN, K. 1999. Cystic fibrosis transmembrane conductance regulator inhibits epithelial Na⁺ channels carrying Liddle's syndrome mutations. *J Biol Chem*, 274, 13894-9.
- HULL, S. C. & KASS, N. E. 2000. Adults with cystic fibrosis and (in)fertility: how has the health care system responded? *J Androl*, 21, 809-13.
- HUMPHREY, S. B., STANTON, T. B. & JENSEN, N. S. 1995. Mitomycin C induction of bacteriophages from *Serpulina hyodysenteriae* and *Serpulina innocens*. *FEMS Microbiol Lett*, 134, 97-101.
- HYATT, D., CHEN, G. L., LOCASCIO, P. F., LAND, M. L., LARIMER, F. W. & HAUSER, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119.
- IGUMBOR, E., GWANZURA, L., CHIRARA, M., OBI, C. & MUZA, D. 2000. Antibiotic sensitivity and plasmid profiles of *Pseudomonas aeruginosa*. *Cent Afr J Med*, 46, 296-300.
- IRIE, Y., BORLEE, B. R., O'CONNOR, J. R., HILL, P. J., HARWOOD, C. S., WOZNIAK, D. J. & PARSEK, M. R. 2012. Self-produced exopolysaccharide is a signal that stimulates biofilm formation in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*, 109, 20632-6.
- IRVIN, R. T. 1993. *Attachment and Colonization of Pseudomonas aeruginosa: Role of the Surface Structures*. In: *Pseudomonas aeruginosa as an Opportunistic Pathogen*, Boston, MA, Springer.
- ISAAC, J. H. & HOLLOWAY, B. W. 1972. Control of arginine biosynthesis in *Pseudomonas aeruginosa*. *J Gen Microbiol*, 73, 427-38.
- IYER, V. N. & SZYBALSKI, W. 1964. MITOMYCINS AND PORFIROMYCIN: CHEMICAL MECHANISM OF ACTIVATION AND CROSS-LINKING OF DNA. *Science*, 145, 55-8.
- JAMES, C. E., DAVIES, E. V., FOTHERGILL, J. L., WALSHAW, M. J., BEALE, C. M., BROCKHURST, M. A. & WINSTANLEY, C. 2015. Lytic activity by temperate phages of *Pseudomonas aeruginosa* in long-term cystic fibrosis chronic lung infections. *Isme j*, 9, 1391-8.
- JAMES, C. E., FOTHERGILL, J. L., KALWIJ, H., HALL, A. J., COTTELL, J., BROCKHURST, M. A. & WINSTANLEY, C. 2012. Differential infection properties of three inducible prophages from an epidemic strain of *Pseudomonas aeruginosa*. *BMC Microbiol*, 12, 216.
- JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOOLS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, 43, 1565-75.
- JELSBAK, L., JOHANSEN, H. K., FROST, A. L., THOGERSEN, R., THOMSEN, L. E., CIOFU, O., YANG, L., HAAGENSEN, J. A., HOIBY, N. & MOLIN, S. 2007. Molecular epidemiology and dynamics of *Pseudomonas aeruginosa* populations in lungs of cystic fibrosis patients. *Infect Immun*, 75, 2214-24.
- JEON, J. & YONG, D. 2019. Two Novel Bacteriophages Improve Survival in *Galleria mellonella* Infection and Mouse Acute Pneumonia Models Infected with Extensively Drug-Resistant *Pseudomonas aeruginosa*. *Appl Environ Microbiol*, 85.

- JEUKENS, J., FRESCHI, L., VINCENT, A. T., EMOND-RHEAULT, J. G., KUKAVICA-IBRULJ, I., CHARETTE, S. J. & LEVESQUE, R. C. 2017. A pan-genomic approach to understand the basis of host adaptation in *Achromobacter*. *Genome Biol Evol*.
- JIN, F., CONRAD, J. C., GIBIANSKY, M. L. & WONG, G. C. 2011. Bacteria use type-IV pili to slingshot on surfaces. *Proc Natl Acad Sci U S A*, 108, 12617-22.
- JOHNSON, A., MEYER, B. J. & PTASHNE, M. 1978. Mechanism of action of the cro protein of bacteriophage lambda. *Proc Natl Acad Sci U S A*, 75, 1783-7.
- JØRGENSEN, F., BALLY, M., CHAPON-HERVE, V., MICHEL, G., LAZDUNSKI, A., WILLIAMS, P. & STEWART, G. 1999. *RpoS*-dependent stress tolerance in *Pseudomonas aeruginosa*. *Microbiology*, 145 (Pt 4), 835-844.
- JORTH, P., STAUDINGER, B. J., WU, X., HISERT, K. B., HAYDEN, H., GARUDATHRI, J., HARDING, C. L., RADEY, M. C., REZAYAT, A., BAUTISTA, G., BERRINGTON, W. R., GODDARD, A. F., ZHENG, C., ANGERMEYER, A., BRITTNACHER, M. J., KITZMAN, J., SHENDURE, J., FLIGNER, C. L., MITTLER, J., AITKEN, M. L., MANOIL, C., BRUCE, J. E., YAHR, T. L. & SINGH, P. K. 2015. Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host Microbe*, 18, 307-19.
- KATSURA, I. & KUHL, P. W. 1975. Morphogenesis of the tail of bacteriophage lambda. III. Morphogenetic pathway. *J Mol Biol*, 91, 257-73.
- KAZMIERCZAK, B. I., SCHNIEDERBEREND, M. & JAIN, R. 2015. Cross-regulation of *Pseudomonas* motility systems: the intimate relationship between flagella, pili and virulence. *Curr Opin Microbiol*, 28, 78-82.
- KIM, H. J., LEE, H., LEE, Y., CHOI, I., KO, Y., LEE, S. & JANG, S. 2020. The ThiL enzyme is a valid antibacterial target essential for both thiamine biosynthesis and salvage pathways in *Pseudomonas aeruginosa*. *J Biol Chem*, 295, 10081-10091.
- KING, A. M., ADAMS, M.J. LEFKOWITZ, E.J. 2011. *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier.
- KIRCHNER, S., FOTHERGILL, J. L., WRIGHT, E. A., JAMES, C. E., MOWAT, E. & WINSTANLEY, C. 2012. Use of artificial sputum medium to test antibiotic efficacy against *Pseudomonas aeruginosa* in conditions more relevant to the cystic fibrosis lung. *J Vis Exp*, e3857.
- KLAUSEN, M., AÆS-JØRGENSEN, A., MOLIN, S. & TOLKER-NIELSEN, T. 2003. Involvement of bacterial migration in the development of complex multicellular structures in *Pseudomonas aeruginosa* biofilms. *Mol Microbiol*, 50, 61-8.
- KLOCKGETHER, J., CRAMER, N., WIEHLMANN, L., DAVENPORT, C. F. & TUMMLER, B. 2011. *Pseudomonas aeruginosa* Genomic Structure and Diversity. *Front Microbiol*, 2, 150.
- KLUMPP, J., FOUTS, D. E. & SOZHAMANNAN, S. 2013. Bacteriophage functional genomics and its role in bacterial pathogen detection. *Briefings in Functional Genomics*, 12, 354-365.
- KOEBNIK, R. 1999. Membrane assembly of the *Escherichia coli* outer membrane protein OmpA: exploring sequence constraints on transmembrane beta-strands. *J Mol Biol*, 285, 1801-10.
- KONDAKOVA, T., D'HEYGÈRE, F., FEUILLOLEY, M. J., ORANGE, N., HEIPIEPER, H. J. & DUCLAIRIOIR POC, C. 2015. Glycerophospholipid synthesis and functions in *Pseudomonas*. *Chem Phys Lipids*, 190, 27-42.
- KONDO, K., KAWANO, M. & SUGAI, M. 2020. Prophage elements function as reservoir for antibiotic resistance and virulence genes in nosocomial pathogens. *bioRxiv*, 2020.11.24.397166.
- KORNITZER, D., ALTUVIA, S. & OPPENHEIM, A. B. 1991. The activity of the CIII regulator of lambdaoid bacteriophages resides within a 24-amino acid protein domain. *Proc Natl Acad Sci U S A*, 88, 5217-21.
- KROPINSKI, A. M. 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J Bacteriol*, 182, 6066-74.
- KRUPOVIC, M., GHABRIAL, S. A., JIANG, D. & VARSANI, A. 2016. Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch Virol*, 161, 2633-43.

- KRZYWINSKI, M., SCHEIN, J., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res*, 19, 1639-45.
- LAËNNEC, R. 1819. *De l'auscultation médiate ou Traité du Diagnostic des Maladies des Poumon et du Coeur*, Paris: Brosson & Chaudé.
- LAMMENS, E. A. & LAVIGNE, R. 2013. Complete genome sequence of F116-like bacteriophages. *Biosystems*.
- LAMONT, I. L., BEARE, P. A., OCHSNER, U., VASIL, A. I. & VASIL, M. L. 2002. Siderophore-mediated signaling regulates virulence factor production in *Pseudomonasaeruginosa*. *Proc Natl Acad Sci U S A*, 99, 7072-7.
- LEE, C. Y. & IANDOLO, J. J. 1986. Lysogenic conversion of staphylococcal lipase is caused by insertion of the bacteriophage L54a genome into the lipase structural gene. *J Bacteriol*, 166, 385-91.
- LELONG, E., MARCHETTI, A., SIMON, M., BURNS, J. L., VAN DELDEN, C., KÖHLER, T. & COSSON, P. 2011. Evolution of *Pseudomonas aeruginosa* virulence in infected patients revealed in a *Dictyostelium discoideum* host model. *Clin Microbiol Infect*, 17, 1415-20.
- LEONARDI, R. & JACKOWSKI, S. 2007. Biosynthesis of Pantothenic Acid and Coenzyme A. *EcoSal Plus*, 2.
- LIMA-MENDEZ, G., VAN HELDEN, J., TOUSSAINT, A. & LEPLAE, R. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics*, 24, 863-5.
- LIN, S., HANSON, R. E. & CRONAN, J. E. 2010. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nat Chem Biol*, 6, 682-8.
- LOC-CARRILLO, C. & ABEDON, S. T. 2011. Pros and cons of phage therapy. *Bacteriophage*, 1, 111-114.
- ŁOŚ, M. & WĘGRZYŃ, G. 2012. *Advances in Virus Research - Chapter 9 - Pseudolysogeny*.
- LYCZAK, J. B., CANNON, C. L. & PIER, G. B. 2000. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect*, 2, 1051-60.
- LYCZAK, J. B., CANNON, C. L. & PIER, G. B. 2002. Lung infections associated with cystic fibrosis. *Clin Microbiol Rev*, 15, 194-222.
- MANOS, J., HU, H., ROSE, B. R., WAINWRIGHT, C. E., ZABLITSKA, I. B., CHENEY, J., TURNBULL, L., WHITCHURCH, C. B., GRIMWOOD, K., HARMER, C., ANUJ, S. N. & HARBOUR, C. 2013. Virulence factor expression patterns in *Pseudomonas aeruginosa* strains from infants with cystic fibrosis. *Eur J Clin Microbiol Infect Dis*, 32, 1583-92.
- MARCAIS, G., DELCHER, A. L., PHILLIPPY, A. M., COSTON, R., SALZBERG, S. L. & ZIMIN, A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol*, 14, e1005944.
- MARSHALL, B. C., BUTLER, S. M., STODDARD, M., MORAN, A. M., LIOU, T. G. & MORGAN, W. J. 2005. Epidemiology of cystic fibrosis-related diabetes. *J Pediatr*, 146, 681-7.
- MASSIMO CONESE, S. C., SUSANNA D'ORIA, SANTE DI GIOIA AND PASQUALINA MONTEMURRO 2017. Role of Neutrophils in Cystic Fibrosis Lung Disease, Role of Neutrophils in Disease Pathogenesis Maitham Abbas Khajah. IntechOpen.
- MATA, G., FERREIRA, G. M. & SPIRA, B. 2017. *RpoS* role in virulence and fitness in enteropathogenic *Escherichia coli*. *PLoS One*, 12, e0180381.
- MATHEE, K., CIOFU, O., STERNBERG, C., LINDUM, P. W., CAMPBELL, J. I. A., JENSEN, P., JOHNSEN, A. H., GIVSKOV, M., OHMAN, D. E., SOREN, M., HOIBY, N. & KHARAZMI, A. 1999. Mucoïd conversion of *Pseudomonas aeruginosa* by hydrogen peroxide: a mechanism for virulence activation in the cystic fibrosis lung. *Microbiology*, 145 (Pt 6), 1349-1357.
- MATHEE, K., NARASIMHAN, G., VALDES, C., QIU, X., MATEWISH, J. M., KOEHRSEN, M., ROKAS, A., YANDAVA, C. N., ENGELS, R., ZENG, E., OLAVARIETTA, R., DOUD, M., SMITH, R. S., MONTGOMERY, P., WHITE, J. R., GODFREY, P. A., KODIRA, C., BIRREN, B., GALAGAN, J. E. &

- LORY, S. 2008. Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc Natl Acad Sci U S A*, 105, 3100-5.
- MATOS, R. C., LAPAQUE, N., RIGOTTIER-GOIS, L., DEBARBIEUX, L., MEYLHEUC, T., GONZALEZ-ZORN, B., REPOILA, F., LOPES MDE, F. & SERROR, P. 2013. Enterococcus faecalis prophage dynamics and contributions to pathogenic traits. *PLoS Genet*, 9, e1003539.
- MATSUI, H., GRUBB, B. R., TARRAN, R., RANDELL, S. H., GATZY, J. T., DAVIS, C. W. & BOUCHER, R. C. 1998. Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell*, 95, 1005-15.
- MATSUI, H., VERGHESE, M. W., KESIMER, M., SCHWAB, U. E., RANDELL, S. H., SHEEHAN, J. K., GRUBB, B. R. & BOUCHER, R. C. 2005. Reduced three-dimensional motility in dehydrated airway mucus prevents neutrophil capture and killing bacteria on airway epithelial surfaces. *J Immunol*, 175, 1090-9.
- MATSUSHIRO, A., SATO, K., MIYAMOTO, H., YAMAMURA, T. & HONDA, T. 1999. Induction of prophages of enterohemorrhagic *Escherichia coli* O157:H7 with norfloxacin. *J Bacteriol*, 181, 2257-60.
- MATTICK, J. S. 2002. Type IV pili and twitching motility. *Annu Rev Microbiol*, 56, 289-314.
- MCCALLUM, S. J., CORKILL, J., GALLAGHER, M., LEDSON, M. J., HART, C. A. & WALSHAW, M. J. 2001. Superinfection with a transmissible strain of *Pseudomonas aeruginosa* in adults with cystic fibrosis chronically colonised by *P. aeruginosa*. *Lancet*, 358, 558-60.
- MCCALLUM, S. J., GALLAGHER, M. J., CORKILL, J. E., HART, C. A., LEDSON, M. J. & WALSHAW, M. J. 2002. Spread of an epidemic *Pseudomonas aeruginosa* strain from a patient with cystic fibrosis (CF) to non-CF relatives. *Thorax*, 57, 559-60.
- MCGRATH, S. & VAN SINDEREN, D. 2007. *Bacteriophage: Genetics and Molecular Biology*, Caister Academic Press.
- MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res*, 41, W597-600.
- MERCHANT, N., LYONS, E., GOFF, S., VAUGHN, M., WARE, D., MICKLOS, D. & ANTIN, P. 2016. The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLoS Biol*, 14, e1002342.
- MICHEL, B. 2005. After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol*, 3, e255.
- MILLAR, F. A., SIMMONDS, N. J. & HODSON, M. E. 2009. Trends in pathogens colonising the respiratory tract of adult patients with cystic fibrosis, 1985-2005. *J Cyst Fibros*, 8, 386-91.
- MILLER, R. V. & RUBERO, V. J. 1984. Mucoïd conversion by phages of *Pseudomonas aeruginosa* strains from patients with cystic fibrosis. *J Clin Microbiol*, 19, 717-9.
- MITCHELL, A. B., MOURAD, B., BUDDLE, L., PETERS, M. J., OLIVER, B. G. G. & MORGAN, L. C. 2018. Viruses in bronchiectasis: a pilot study to explore the presence of community acquired respiratory viruses in stable patients and during acute exacerbations. *BMC Pulm Med*, 18, 84.
- MITCHELMORE, P. J., RANDALL, J., BULL, M. J., MOORE, K. A., O'NEILL, P. A., PASZKIEWICZ, K., MAHENTHIRALINGAM, E., SCOTTON, C. J., SHELDON, C. D., WITHERS, N. J. & BROWN, A. R. 2017. Molecular epidemiology of *Pseudomonas aeruginosa* in an unsegregated bronchiectasis cohort sharing hospital facilities with a cystic fibrosis cohort. *Thorax*.
- MIYAZAKI, R., MINOIA, M., PRADERVAND, N., SULSER, S., REINHARD, F. & VAN DER MEER, J. R. 2012. Cellular variability of *RpoS* expression underlies subpopulation activation of an integrative and conjugative element. *PLoS Genet*, 8, e1002818.
- MODI, S. R., LEE, H. H., SPINA, C. S. & COLLINS, J. J. 2013. Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature*, 499, 219-22.

- MOJICA, F. J., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & SORIA, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, 60, 174-82.
- MONSON, R., FOULDS, I., FOWERAKER, J., WELCH, M. & SALMOND, G. P. C. 2011. The *Pseudomonas aeruginosa* generalized transducing phage phiPA3 is a new member of the phiKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. *Microbiology (Reading)*, 157, 859-867.
- MONTIE, T. C., DOYLE-HUNTZINGER, D., CRAVEN, R. C. & HOLDER, I. A. 1982. Loss of virulence associated with absence of flagellum in an isogenic mutant of *Pseudomonas aeruginosa* in the burned-mouse model. *Infect Immun*, 38, 1296-8.
- MORGAN, G. J., HATFULL, G. F., CASJENS, S. & HENDRIX, R. W. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J Mol Biol*, 317, 337-59.
- MORILLAS, H. N., ZARIWALA, M. & KNOWLES, M. R. 2007. Genetic causes of bronchiectasis: primary ciliary dyskinesia. *Respiration*, 74, 252-63.
- MORTENSEN, K. L., JENSEN, R. H., JOHANSEN, H. K., SKOV, M., PRESSLER, T., HOWARD, S. J., LEATHERBARROW, H., MELLADO, E. & ARENDRUP, M. C. 2011. *Aspergillus* species and other molds in respiratory samples from patients with cystic fibrosis: a laboratory-based study with focus on *Aspergillus fumigatus* azole resistance. *J Clin Microbiol*, 49, 2243-51.
- MURPHY, J., MAHONY, J., AINSWORTH, S., NAUTA, A. & VAN SINDEREN, D. 2013. Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. *Appl Environ Microbiol*, 79, 7547-55.
- MURRAY, T. S. & KAZMIERCZAK, B. I. 2008. *Pseudomonas aeruginosa* exhibits sliding motility in the absence of type IV pili and flagella. *J Bacteriol*, 190, 2700-8.
- NAGANO, Y., ELBORN, J. S., MILLAR, B. C., WALKER, J. M., GOLDSMITH, C. E., RENDALL, J. & MOORE, J. E. 2010. Comparison of techniques to examine the diversity of fungi in adult patients with cystic fibrosis. *Med Mycol*, 48, 166-76.e1.
- NAKAE, T. 1976. Outer membrane of *Salmonella*. Isolation of protein complex that produces transmembrane channels. *J Biol Chem*, 251, 2176-8.
- NELSON, K. E., WEINEL, C., PAULSEN, I. T., DODSON, R. J., HILBERT, H., MARTINS DOS SANTOS, V. A., FOUTS, D. E., GILL, S. R., POP, M., HOLMES, M., BRINKAC, L., BEANAN, M., DEBOY, R. T., DAUGHERTY, S., KOLONAY, J., MADUPU, R., NELSON, W., WHITE, O., PETERSON, J., KHOURI, H., HANCE, I., CHRIS LEE, P., HOLTZAPPLE, E., SCANLAN, D., TRAN, K., MOAZZEZ, A., UTTERBACK, T., RIZZO, M., LEE, K., KOSACK, D., MOESTL, D., WEDLER, H., LAUBER, J., STJEPANDIC, D., HOHEISEL, J., STRAETZ, M., HEIM, S., KIEWITZ, C., EISEN, J. A., TIMMIS, K. N., DUSTERHOFT, A., TUMMLER, B. & FRASER, C. M. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol*, 4, 799-808.
- NEVES, P. C., GUERRA, M., PONCE, P., MIRANDA, J. & VOUGA, L. 2011. Non-cystic fibrosis bronchiectasis. *Interact Cardiovasc Thorac Surg*, 13, 619-25.
- NHLBI. 2018. *Bronchiectasis* [Online]. Available: www.nhlbi.nih.gov/health-topics/bronchiectasis [Accessed 2020].
- NHS.UK. 2016. *Bronchitis* [Online]. Available: <https://www.nhs.uk/conditions/bronchitis/> [Accessed 12/12/18 2018].
- NHS.UK. 2016 *Treatments for cystic fibrosis* [Online]. Available: <https://www.nhs.uk/conditions/cystic-fibrosis/treatment/> [Accessed 10/12/18 2018].
- NNALUE, N. A., NEWTON, S. & STOCKER, B. A. 1990. Lysogenization of *Salmonella choleraesuis* by phage 14 increases average length of O-antigen chains, serum resistance and intraperitoneal mouse virulence. *Microb Pathog*, 8, 393-402.
- O'BRIEN, S., KUMMERLI, R., PATERSON, S., WINSTANLEY, C. & BROCKHURST, M. A. 2019. Transposable temperate phages promote the evolution of divergent social strategies in *Pseudomonas aeruginosa* populations. *Proc Biol Sci*, 286, 20191794.

- O'CARROLL, M. R., SYRMIS, M. W., WAINWRIGHT, C. E., GREER, R. M., MITCHELL, P., COULTER, C., SLOOTS, T. P., NISSEN, M. D. & BELL, S. C. 2004. Clonal strains of *Pseudomonas aeruginosa* in paediatric and adult cystic fibrosis units. *Eur Respir J*, 24, 101-6.
- O'SULLIVAN, B. P. & FREEDMAN, S. D. 2009. Cystic fibrosis. *Lancet*, 373, 1891-904.
- O'SULLIVAN, D., ROSS, R. P., FITZGERALD, G. F. & COFFEY, A. 2000. Investigation of the relationship between lysogeny and lysis of *Lactococcus lactis* in cheese using prophage-targeted PCR. *Appl Environ Microbiol*, 66, 2192-8.
- OBRITSCH, M. D., FISH, D. N., MACLAREN, R. & JUNG, R. 2004. National surveillance of antimicrobial resistance in *Pseudomonas aeruginosa* isolates obtained from intensive care unit patients from 1993 to 2002. *Antimicrob Agents Chemother*, 48, 4606-10.
- OHGAKI, N. 1994. Bacterial biofilm in chronic airway infection. *Kansenshogaku Zasshi*, 68, 138-51.
- OHNISHI, M., TERAJIMA, J., KUROKAWA, K., NAKAYAMA, K., MURATA, T., TAMURA, K., OGURA, Y., WATANABE, H. & HAYASHI, T. 2002. Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. *Proc Natl Acad Sci U S A*, 99, 17043-8.
- OLIVEIRA, P. H., TOUCHON, M. & ROCHA, E. P. 2014. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res*, 42, 10618-31.
- OLSZAK, T., ZARNOWIEC, P., KACA, W., DANIS-WLODARCZYK, K., AUGUSTYNIAK, D., DREVINEK, P., DE SOYZA, A., MCCLEAN, S. & DRULIS-KAWA, Z. 2015. In vitro and in vivo antibacterial activity of environmental bacteriophages against *Pseudomonas aeruginosa* strains from cystic fibrosis patients. *Appl Microbiol Biotechnol*, 99, 6021-33.
- ORRIOLS, R., HERNANDO, R., FERRER, A., TERRADAS, S. & MONTORO, B. 2015. Eradication Therapy against *Pseudomonas aeruginosa* in Non-Cystic Fibrosis Bronchiectasis. *Respiration*, 90, 299-305.
- OTSUJI, N., SEKIGUCHI, M., IJIMA, T. & TAKAGI, Y. 1959. Induction of phage formation in the lysogenic *Escherichia coli* K-12 by mitomycin C. *Nature*, 184(Suppl 14), 1079-80.
- OVERHAGE, J., SCHEMIONEK, M., WEBB, J. S. & REHM, B. H. 2005. Expression of the *psl* operon in *Pseudomonas aeruginosa* PAO1 biofilms: *PslA* performs an essential function in biofilm formation. *Appl Environ Microbiol*, 71, 4407-13.
- OWEN, S. V., CANALS, R., WENNER, N., HAMMARLÖF, D. L., KRÖGER, C. & HINTON, J. C. D. 2020. A window into lysogeny: revealing temperate phage biology with transcriptomics. *Microb Genom*, 6.
- OZER, E. A., ALLEN, J. P. & HAUSER, A. R. 2014. Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics*, 15, 737.
- PABARY, R., SINGH, C., MORALES, S., BUSH, A., ALSHAFI, K., BILTON, D., ALTON, E. W., SMITHYMAN, A. & DAVIES, J. C. 2016. Antipseudomonal Bacteriophage Reduces Infective Burden and Inflammatory Response in Murine Lung. *Antimicrob Agents Chemother*, 60, 744-51.
- PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T., FOOKES, M., FALUSH, D., KEANE, J. A. & PARKHILL, J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31, 3691-3.
- PALLERONI, N. J. 1992. *The Prokaryotes, A Handbook on the Biology of Bacteria, Ecophysiology, Isolation, Identification and Application*, New York., Springer.
- PASTEUR, M. C., BILTON, D. & HILL, A. T. 2010. British Thoracic Society guideline for non-CF bronchiectasis. *Thorax*, 65 Suppl 1, i1-58.
- PATANKAR, Y. R., LOVEWELL, R. R., POYNTER, M. E., JYOT, J., KAZMIERCZAK, B. I. & BERWIN, B. 2013. Flagellar motility is a key determinant of the magnitude of the inflammasome response to *Pseudomonas aeruginosa*. *Infect Immun*, 81, 2043-52.
- PATTI, G. J., YANES, O. & SIUZDAK, G. 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*, 13, 263-9.

- PATTISON, S. H., ROGERS, G. B., CROCKARD, M., ELBORN, J. S. & TUNNEY, M. M. 2013. Molecular detection of CF lung pathogens: current status and future potential. *J Cyst Fibros*, 12, 194-205.
- PAWLUK, A., BONDY-DENOMY, J., CHEUNG, V. H., MAXWELL, K. L. & DAVIDSON, A. R. 2014. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *mBio*, 5, e00896.
- PEDULLA, M. L., FORD, M. E., HOUTZ, J. M., KARTHIKEYAN, T., WADSWORTH, C., LEWIS, J. A., JACOBS-SERA, D., FALBO, J., GROSS, J., PANNUNZIO, N. R., BRUCKER, W., KUMAR, V., KANDASAMY, J., KEENAN, L., BARDAROV, S., KRIAKOV, J., LAWRENCE, J. G., JACOBS, W. R., JR., HENDRIX, R. W. & HATFULL, G. F. 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell*, 113, 171-82.
- PERRY, J. A. & WRIGHT, G. D. 2013. The antibiotic resistance "mobilome": searching for the link between environment and clinic. *Front Microbiol*, 4, 138.
- PEZZULO, A. A., TANG, X. X., HOEGGER, M. J., ABOU ALAIWA, M. H., RAMACHANDRAN, S., MONINGER, T. O., KARP, P. H., WOHLFORD-LENANE, C. L., HAAGSMAN, H. P., VAN EIJK, M., BANFI, B., HORSWILL, A. R., STOLTZ, D. A., MCCRAY, P. B., JR., WELSH, M. J. & ZABNER, J. 2012. Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung. *Nature*, 487, 109-13.
- PIER, G. B., COLEMAN, F., GROUT, M., FRANKLIN, M. & OHMAN, D. E. 2001. Role of alginate O acetylation in resistance of mucoid *Pseudomonas aeruginosa* to opsonic phagocytosis. *Infect Immun*, 69, 1895-901.
- PIRNAY, J. P., DE VOS, D., COCHEZ, C., BILOCQ, F., VANDERKELEN, A., ZIZI, M., GHYSELS, B. & CORNELIS, P. 2002. *Pseudomonas aeruginosa* displays an epidemic population structure. *Environ Microbiol*, 4, 898-911.
- PLOTKOWSKI, M. C., SALIBA, A. M., PEREIRA, S. H., CERVANTE, M. P. & BAJOLET-LAUDINAT, O. 1994. *Pseudomonas aeruginosa* selective adherence to and entry into human endothelial cells. *Infect Immun*, 62, 5456-63.
- PLUNKETT, G., 3RD, ROSE, D. J., DURFEE, T. J. & BLATTNER, F. R. 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J Bacteriol*, 181, 1767-78.
- POHL, S., KLOCKGETHER, J., ECKWEILER, D., KHALEDI, A., SCHNIEDERJANS, M., CHOUVARINE, P., TUMMLER, B. & HAUSSLER, S. 2014. The extensive set of accessory *Pseudomonas aeruginosa* genomic components. *FEMS Microbiol Lett*, 356, 235-41.
- PRAGMAN, A. A., KIM, H. B., REILLY, C. S., WENDT, C. & ISAACSON, R. E. 2012. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS One*, 7, e47305.
- PRAVIN, S. S. & SHAHUL, H. 2012. *Chapter Twenty-One - Nonfluoroquinolone-Based Inhibitors of Mycobacterial Type II Topoisomerase as Potential Therapeutic Agents for TB*, Academic Press.
- PRZERWA, A., ZIMECKI, M., SWITAŁA-JELEŃ, K., DABROWSKA, K., KRAWCZYK, E., ŁUCZAK, M., WEBER-DABROWSKA, B., SYPER, D., MIEDZYBRODZKI, R. & GÓRSKI, A. 2006. Effects of bacteriophages on free radical production and phagocytic functions. *Med Microbiol Immunol*, 195, 143-50.
- PTASHNE, M. 1992. *A genetic switch - phage λ and higher organisms.*, Cell Press and Blackwell Scientific Publications.
- PURCELL, P., JARY, H., PERRY, A., PERRY, J. D., STEWART, C. J., NELSON, A., LANYON, C., SMITH, D. L., CUMMINGS, S. P. & DE SOYZA, A. 2014. Polymicrobial airway bacterial communities in adult bronchiectasis patients. *BMC Microbiol*, 14, 130.
- QUINN, R. A., PHELAN, V. V., WHITESON, K. L., GARG, N., BAILEY, B. A., LIM, Y. W., CONRAD, D. J., DORRESTEIN, P. C. & ROHWER, F. L. 2016. Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome. *Isme j*, 10, 1483-98.

- RAMIREZ, M., SEVERINA, E. & TOMASZ, A. 1999. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae*. *J Bacteriol*, 181, 3618-25.
- RATJEN, F. A. 2009. Cystic fibrosis: pathogenesis and future treatment strategies. *Respir Care*, 54, 595-605.
- REESE J, URRY. L, CAIN, M., WASSERMAN. S, MINORSKY, P. & JACKSON, R. 2011. *Campbell Biology*, Pearson Benjamin Cummings.
- REHMAT, S. & SHAPIRO, J. A. 1983. Insertion and replication of the *Pseudomonas aeruginosa* mutator phage D3112. *Mol Gen Genet*, 192, 416-23.
- REIS-CUNHA, J. L., BARTHOLOMEU, D. C., MANSON, A. L., EARL, A. M. & CERQUEIRA, G. C. 2019. ProphET, prophage estimation tool: A stand-alone prophage sequence prediction tool with self-updating reference database. *PLoS One*, 14, e0223364.
- REZAEI JAVAN, R., RAMOS-SEVILLANO, E., AKTER, A., BROWN, J. & BRUEGGEMANN, A. B. 2019. Prophages and satellite prophages are widespread in *Streptococcus* and may play a role in pneumococcal pathogenesis. *Nat Commun*, 10, 4852.
- RICE, S. A., TAN, C. H., MIKKELSEN, P. J., KUNG, V., WOO, J., TAY, M., HAUSER, A., MCDUGALD, D., WEBB, J. S. & KJELLEBERG, S. 2009. The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage. *Isme j*, 3, 271-82.
- ROA, M. 1979. Interaction of bacteriophage K10 with its receptor, the lamB protein of *Escherichia coli*. *J Bacteriol*, 140, 680-6.
- RODRIGUEZ-VALERA, F., MARTIN-CUADRADO, A. B. & LÓPEZ-PÉREZ, M. 2016. Flexible genomic islands as drivers of genome evolution. *Curr Opin Microbiol*, 31, 154-160.
- ROGUIN, A. 2006. Rene Theophile Hyacinthe Laennec (1781-1826): the man behind the stethoscope. *Clin Med Res*, 4, 230-5.
- ROLAIN, J. M., FANCELLO, L., DESNUES, C. & RAOULT, D. 2011. Bacteriophages as vehicles of the resistome in cystic fibrosis. *J Antimicrob Chemother*, 66, 2444-7.
- RONCERO, C., DARZINS, A. & CASADABAN, M. J. 1990. *Pseudomonas aeruginosa* transposable bacteriophages D3112 and B3 require pili and surface growth for adsorption. *J Bacteriol*, 172, 1899-904.
- ROSENWASSER, S., ZIV, C., CREVELD, S. G. V. & VARDI, A. 2016. Virocell Metabolism: Metabolic Innovations During Host-Virus Interactions in the Ocean. *Trends Microbiol*, 24, 821-832.
- ROUX, S., ENAULT, F., HURWITZ, B. L. & SULLIVAN, M. B. 2015a. VirSorter: mining viral signal from microbial genomic data. *PeerJ*, 3, e985.
- ROUX, S., HALLAM, S. J., WOYKE, T. & SULLIVAN, M. B. 2015b. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife*, 4.
- ROWLETT, V. W., MALLAMPALLI, V. K. P. S., KARLSTAEDT, A., DOWHAN, W., TAEGTMEYER, H., MARGOLIN, W. & VITRAC, H. 2017. Impact of Membrane Phospholipid Alterations in *Escherichia coli* on Cellular Function and Bacterial Stress Adaptation. *Journal of Bacteriology*, 199, e00849-16.
- RUMBAUGH, K. P., GRISWOLD, J. A., IGLEWSKI, B. H. & HAMOOD, A. N. 1999. Contribution of quorum sensing to the virulence of *Pseudomonas aeruginosa* in burn wound infections. *Infect Immun*, 67, 5854-62.
- RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics*, 16, 944-5.
- SACHA, P., WIECZOREK, P., HAUSCHILD, T., ZÓRAWSKI, M., OLSZAŃSKA, D. & TRYNISZEWSKA, E. 2008. Metallo-beta-lactamases of *Pseudomonas aeruginosa*--a novel mechanism resistance to beta-lactam antibiotics. *Folia Histochem Cytobiol*, 46, 137-42.
- SAGEL, S. D., SONTAG, M. K., WAGENER, J. S., KAPSNER, R. K., OSBERG, I. & ACCURSO, F. J. 2002. Induced sputum inflammatory measures correlate with lung function in children with cystic fibrosis. *J Pediatr*, 141, 811-7.

- SANZ-GARCÍA, F., HERNANDO-AMADO, S. & MARTÍNEZ, J. L. 2018. Mutational Evolution of *Pseudomonas aeruginosa* Resistance to Ribosome-Targeting Antibiotics. *Front Genet*, 9, 451.
- SASSANFAR, M. & ROBERTS, J. W. 1990. Nature of the SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. *J Mol Biol*, 212, 79-96.
- SCHADE, S. Z., ADLER, J. & RIS, H. 1967. How bacteriophage chi attacks motile bacteria. *J Virol*, 1, 599-609.
- SCOTT, A., POTTENGER, S., TIMOFTE, D., MOORE, M., WRIGHT, L., KUKAVICA-IBRULJ, I., JEUKENS, J., LEVESQUE, R. C., FRESCHI, L., PINCHBECK, G. L., SCHMIDT, V. M., MCEWAN, N., RADFORD, A. D. & FOTHERGILL, J. L. 2019. Reservoirs of resistance: polymyxin resistance in veterinary-associated companion animal isolates of *Pseudomonas aeruginosa*. *Vet Rec*, 185, 206.
- SECOR, P. R., MICHAELS, L. A., SMIGIEL, K. S., ROHANI, M. G., JENNINGS, L. K., HISERT, K. B., ARRIGONI, A., BRAUN, K. R., BIRKLAND, T. P., LAI, Y., HALLSTRAND, T. S., BOLLYKY, P. L., SINGH, P. K. & PARKS, W. C. 2017. Filamentous Bacteriophage Produced by *Pseudomonas aeruginosa* Alters the Inflammatory Response and Promotes Noninvasive Infection In Vivo. *Infect Immun*, 85.
- SECOR, P. R., SWEERE, J. M., MICHAELS, L. A., MALKOVSKIY, A. V., LAZZARESCHI, D., KATZNELSON, E., RAJADAS, J., BIRNBAUM, M. E., ARRIGONI, A., BRAUN, K. R., EVANKO, S. P., STEVENS, D. A., KAMINSKY, W., SINGH, P. K., PARKS, W. C. & BOLLYKY, P. L. 2015. Filamentous Bacteriophage Promote Biofilm Assembly and Function. *Cell Host Microbe*, 18, 549-59.
- SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30, 2068-9.
- SEPULVEDA-ROBLES, O., KAMEYAMA, L. & GUARNEROS, G. 2012. High diversity and novel species of *Pseudomonas aeruginosa* bacteriophages. *Appl Environ Microbiol*, 78, 4510-5.
- SHAHID, M., MALIK, A. & SHEEBA 2003. Multidrug-resistant *Pseudomonas aeruginosa* strains harbouring R-plasmids and AmpC beta-lactamases isolated from hospitalised burn patients in a tertiary care hospital of North India. *FEMS Microbiol Lett*, 228, 181-6.
- SHAN, J., JIA, Y., CLOKIE, M. R. & MANN, N. H. 2008. Infection by the 'photosynthetic' phage S-PM2 induces increased synthesis of phycoerythrin in *Synechococcus* sp. WH7803. *FEMS Microbiol Lett*, 283, 154-61.
- SHAO, Y. & WANG, I. N. 2008. Bacteriophage adsorption rate and optimal lysis time. *Genetics*, 180, 471-82.
- SHEN, K., SAYEED, S., ANTALIS, P., GLADITZ, J., AHMED, A., DICE, B., JANTO, B., DOPICO, R., KEEFE, R., HAYES, J., JOHNSON, S., YU, S., EHRLICH, N., JOCZ, J., KROPP, L., WONG, R., WADOWSKY, R. M., SLIFKIN, M., PRESTON, R. A., ERDOS, G., POST, J. C., EHRLICH, G. D. & HU, F. Z. 2006. Extensive genomic plasticity in *Pseudomonas aeruginosa* revealed by identification and distribution studies of novel genes among clinical isolates. *Infect Immun*, 74, 5272-83.
- SILBERT, S., BARTH, A. L. & SADER, H. S. 2001. Heterogeneity of *Pseudomonas aeruginosa* in Brazilian cystic fibrosis patients. *J Clin Microbiol*, 39, 3976-81.
- SILBY, M. W., WINSTANLEY, C., GODFREY, S. A., LEVY, S. B. & JACKSON, R. W. 2011. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev*, 35, 652-80.
- SIRÉN, K., MILLARD, A., PETERSEN, B., GILBERT, M THOMAS P., CLOKIE, M. R. J. & SICHERITZ-PONTÉN, T. 2021. Rapid discovery of novel prophages using biological feature engineering and machine learning. *NAR Genomics and Bioinformatics*, 3.
- SKERRETT, S. J., WILSON, C. B., LIGGITT, H. D. & HAJJAR, A. M. 2007. Redundant Toll-like receptor signaling in the pulmonary host response to *Pseudomonas aeruginosa*. *Am J Physiol Lung Cell Mol Physiol*, 292, L312-22.
- SMITH, D. L., ROOKS, D. J., FOGG, P. C., DARBY, A. C., THOMSON, N. R., MCCARTHY, A. J. & ALLISON, H. E. 2012. Comparative genomics of Shiga toxin encoding bacteriophages. *BMC Genomics*, 13, 311.

- SPAKOWITZ, A. J. & WANG, Z. G. 2005. DNA packaging in bacteriophage: is twist important? *Biophys J*, 88, 3912-23.
- SPANIER, J. G. & CLEARY, P. P. 1980. Bacteriophage control of antiphagocytic determinants in group A streptococci. *J Exp Med*, 152, 1393-406.
- SRIRAMULU, D. D., LÜNSDORF, H., LAM, J. S. & RÖMLING, U. 2005. Microcolony formation: a novel biofilm model of *Pseudomonas aeruginosa* for the cystic fibrosis lung. *J Med Microbiol*, 54, 667-676.
- SRIVIDHYA, K. V., ALAGURAJ, V., POORNIMA, G., KUMAR, D., SINGH, G. P., RAGHAVENDERAN, L., KATTA, A. V., MEHTA, P. & KRISHNASWAMY, S. 2007. Identification of prophages in bacterial genomes by dinucleotide relative abundance difference. *PLoS One*, 2, e1193.
- STEFFENSKY, M., MÜHLENWEG, A., WANG, Z. X., LI, S. M. & HEIDE, L. 2000. Identification of the novobiocin biosynthetic gene cluster of *Streptomyces spheroides* NCIB 11891. *Antimicrob Agents Chemother*, 44, 1214-22.
- STOKES, H. W., ELBOURNE, L. D. & HALL, R. M. 2007. Tn1403, a multiple-antibiotic resistance transposon made up of three distinct transposons. *Antimicrob Agents Chemother*, 51, 1827-9.
- STOLFA, G. & KOUDELKA, G. B. 2012. Entry and killing of *Tetrahymena thermophila* by bacterially produced Shiga toxin. *mBio*, 4, e00416-12.
- STOVER, C. K., PHAM, X. Q., ERWIN, A. L., MIZOGUCHI, S. D., WARRENER, P., HICKEY, M. J., BRINKMAN, F. S., HUFNAGLE, W. O., KOWALIK, D. J., LAGROU, M., GARBER, R. L., GOLTRY, L., TOLENTINO, E., WESTBROCK-WADMAN, S., YUAN, Y., BRODY, L. L., COULTER, S. N., FOLGER, K. R., KAS, A., LARBIG, K., LIM, R., SMITH, K., SPENCER, D., WONG, G. K., WU, Z., PAULSEN, I. T., REIZER, J., SAIER, M. H., HANCOCK, R. E., LORY, S. & OLSON, M. V. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 406, 959-64.
- STRACHAN, D., GUPTA, R., LIMB, E., HUBBARD, R., GIBSON, J., TATA, L., BURNEY, P., JARVIS, D., CULLINAN, P., HANSELL, A., BAKOLIS, I., GHOSH, R. & SHEIKH, A. 2014. *British Lung foundation - Bronchiectasis statistics* [Online]. British Lung Foundation 2019. Available: <https://statistics.blf.org.uk/bronchiectasis> [Accessed 30/4/19 2019].
- STRAUCH, E., HAMMERL, J. A., KONIETZNY, A., SCHNEIKER-BEKEL, S., ARNOLD, W., GOESMANN, A., PUHLER, A. & BEUTIN, L. 2008. Bacteriophage 2851 is a prototype phage for dissemination of the Shiga toxin variant gene 2c in *Escherichia coli* O157:H7. *Infect Immun*, 76, 5466-77.
- STRESSMANN, F. A., ROGERS, G. B., MARSH, P., LILLEY, A. K., DANIELS, T. W., CARROLL, M. P., HOFFMAN, L. R., JONES, G., ALLEN, C. E., PATEL, N., FORBES, B., TUCK, A. & BRUCE, K. D. 2011. Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *J Cyst Fibros*, 10, 357-65.
- SUBEDI, D., VIJAY, A. K., KOHLI, G. S., RICE, S. A. & WILLCOX, M. 2018. Comparative genomics of clinical strains of *Pseudomonas aeruginosa* strains isolated from different geographic sites. *Sci Rep*, 8, 15668.
- SULLIVAN, M. B., HUANG, K. H., IGNACIO-ESPINOZA, J. C., BERLIN, A. M., KELLY, L., WEIGELE, P. R., DEFRADESCO, A. S., KERN, S. E., THOMPSON, L. R., YOUNG, S., YANDAVA, C., FU, R., KRSTINS, B., CHASE, M., SARRACINO, D., OSBURNE, M. S., HENN, M. R. & CHISHOLM, S. W. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol*, 12, 3035-56.
- SUN, C. L., BARRANGOU, R., THOMAS, B. C., HORVATH, P., FREMAUX, C. & BANFIELD, J. F. 2013. Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol*, 15, 463-70.
- SUN, X., OLIVIER, A. K., LIANG, B., YI, Y., SUI, H., EVANS, T. I., ZHANG, Y., ZHOU, W., TYLER, S. R., FISHER, J. T., KEISER, N. W., LIU, X., YAN, Z., SONG, Y., GOEKEN, J. A., KINYON, J. M., FLIGG, D., WANG, X., XIE, W., LYNCH, T. J., KAMINSKY, P. M., STEWART, Z. A., POPE, R. M., FRANA, T., MEYERHOLZ, D. K., PAREKH, K. & ENGELHARDT, J. F. 2014. Lung phenotype of

- juvenile and adult cystic fibrosis transmembrane conductance regulator-knockout ferrets. *Am J Respir Cell Mol Biol*, 50, 502-12.
- SUTHERLAND, I. W. 2001. The biofilm matrix--an immobilized but dynamic microbial environment. *Trends Microbiol*, 9, 222-7.
- SUTTLE, C. A. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol*, 5, 801-12.
- SWEERE, J. M., VAN BELLEGHEM, J. D., ISHAK, H., BACH, M. S., POPESCU, M., SUNKARI, V., KABER, G., MANASHEROB, R., SUH, G. A., CAO, X., DE VRIES, C. R., LAM, D. N., MARSHALL, P. L., BIRUKOVA, M., KATZNELSON, E., LAZZARESCHI, D. V., BALAJI, S., KESWANI, S. G., HAWN, T. R., SECOR, P. R. & BOLLYKY, P. L. 2019. Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. *Science*, 363.
- TACCETTI, G., BIANCHINI, E., CARIANI, L., BUZZETTI, R., COSTANTINI, D., TREVISAN, F., ZAVATARO, L. & CAMPANA, S. 2012. Early antibiotic treatment for *Pseudomonas aeruginosa* eradication in patients with cystic fibrosis: a randomised multicentre study comparing two different protocols. *Thorax*, 67, 853-9.
- TAKEBE, H., ICHIKAWA, H., IWO, K. & KONDO, S. 1967. Phage induction by ultraviolet radiation in strains of *Escherichia coli* possessing and lacking dark repair capacity. *Virology*, 33, 638-49.
- TARIQ, M. A., EVEREST, F. L., COWLEY, L. A., DE SOYZA, A., HOLT, G. S., BRIDGE, S. H., PERRY, A., PERRY, J. D., BOURKE, S. J., CUMMINGS, S. P., LANYON, C. V., BARR, J. J. & SMITH, D. L. 2015. A metagenomic approach to characterize temperate bacteriophage populations from Cystic Fibrosis and non-Cystic Fibrosis bronchiectasis patients. *Front Microbiol*, 6, 97.
- TARIQ, M. A., EVEREST, F. L. C., COWLEY, L. A., WRIGHT, R., HOLT, G. S., INGRAM, H., DUIGNAN, L. A. M., NELSON, A., LANYON, C. V., PERRY, A., PERRY, J. D., BOURKE, S., BROCKHURST, M. A., BRIDGE, S. H., DE SOYZA, A. & SMITH, D. L. 2019. Temperate Bacteriophages from Chronic *Pseudomonas aeruginosa* Lung Infections Show Disease-Specific Changes in Host Range and Modulate Antimicrobial Susceptibility. *mSystems*, 4.
- TEN HACKEN, N. H., WIJKSTRA, P. J. & KERSTJENS, H. A. 2007. Treatment of bronchiectasis in adults. *Bmj*, 335, 1089-93.
- TERASHIMA, H., KOJIMA, S. & HOMMA, M. 2008. Flagellar motility in bacteria structure and function of flagellar motor. *Int Rev Cell Mol Biol*, 270, 39-85.
- TIMISCHL, B., DETTMER, K., KASPAR, H., THIEME, M. & OEFNER, P. J. 2008. Development of a quantitative, validated capillary electrophoresis-time of flight-mass spectrometry method with integrated high-confidence analyte identification for metabolomics. *Electrophoresis*, 29, 2203-14.
- TINGPEJ, P., SMITH, L., ROSE, B., ZHU, H., CONIBEAR, T., AL NASSAFI, K., MANOS, J., ELKINS, M., BYE, P., WILLCOX, M., BELL, S., WAINWRIGHT, C. & HARBOUR, C. 2007. Phenotypic characterization of clonal and nonclonal *Pseudomonas aeruginosa* strains isolated from lungs of adults with cystic fibrosis. *J Clin Microbiol*, 45, 1697-704.
- TINSLEY, C. R., BILLE, E. & NASSIF, X. 2006. Bacteriophages and pathogenicity: more than just providing a toxin? *Microbes Infect*, 8, 1365-71.
- TOTTEN, P. A., LARA, J. C. & LORY, S. 1990. The rpoN gene product of *Pseudomonas aeruginosa* is required for expression of diverse genes, including the flagellin gene. *J Bacteriol*, 172, 389-96.
- TOUSSAINT, A. & RICE, P. A. 2017. Transposable phages, DNA reorganization and transfer. *Curr Opin Microbiol*, 38, 88-94.
- TSAO, Y. F., TAYLOR, V. L., KALA, S., BONDY-DENOMY, J., KHAN, A. N., BONA, D., CATTOIR, V., LORY, S., DAVIDSON, A. R. & MAXWELL, K. L. 2018. Phage Morons Play an Important Role in *Pseudomonas aeruginosa* Phenotypes. *J Bacteriol*, 200.
- TWORT, F. 1915. An investigation on the nature of ultra-microscopic viruses. *The Lancet*, 186, 1241-3.

- VALENZA, G., TAPPE, D., TURNWALD, D., FROSCHE, M., KÖNIG, C., HEBESTREIT, H. & ABELE-HORN, M. 2008. Prevalence and antimicrobial susceptibility of microorganisms isolated from sputa of patients with cystic fibrosis. *J Cyst Fibros*, 7, 123-7.
- VALLET, I., DIGGLE, S. P., STACEY, R. E., CAMARA, M., VENTRE, I., LORY, S., LAZDUNSKI, A., WILLIAMS, P. & FILLOUX, A. 2004. Biofilm formation in *Pseudomonas aeruginosa*: fimbrial cup gene clusters are controlled by the transcriptional regulator MvaT. *J Bacteriol*, 186, 2880-90.
- VALPUESTA, J. M. & CARRASCOSA, J. L. 1994. Structure of viral connectors and their function in bacteriophage assembly and DNA packaging. *Q Rev Biophys*, 27, 107-155.
- VAN EWIJK, B. E., VAN DER ZALM, M. M., WOLFS, T. F., FLEER, A., KIMPEN, J. L., WILBRINK, B. & VAN DER ENT, C. K. 2008. Prevalence and impact of respiratory viral infections in young children with cystic fibrosis: prospective cohort study. *Pediatrics*, 122, 1171-6.
- VESES-GARCIA, M., LIU, X., RIGDEN, D. J., KENNY, J. G., MCCARTHY, A. J. & ALLISON, H. E. 2015. Transcriptomic analysis of Shiga-toxigenic bacteriophage carriage reveals a profound regulatory effect on acid resistance in *Escherichia coli*. *Appl Environ Microbiol*, 81, 8118-25.
- VICA PACHECO, S., GARCÍA GONZÁLEZ, O. & PANIAGUA CONTRERAS, G. L. 1997. The *lom* gene of bacteriophage lambda is involved in *Escherichia coli* K12 adhesion to human buccal epithelial cells. *FEMS Microbiol Lett*, 156, 129-32.
- VIDAKOVIC, L., SINGH, P. K., HARTMANN, R., NADELL, C. D. & DRESCHER, K. 2018. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nat Microbiol*, 3, 26-31.
- WAGNER, P. L., NEELY, M. N., ZHANG, X., ACHESON, D. W., WALDOR, M. K. & FRIEDMAN, D. I. 2001. Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. *J Bacteriol*, 183, 2081-5.
- WAGNER, P. L. & WALDOR, M. K. 2002. Bacteriophage control of bacterial virulence. *Infect Immun*, 70, 3985-93.
- WALDOR, M. K. 1998. Bacteriophage biology and bacterial virulence. *Trends Microbiol*, 6, 295-7.
- WARK, P. A., TOOZE, M., CHEESE, L., WHITEHEAD, B., GIBSON, P. G., WARK, K. F. & MCDONALD, V. M. 2012. Viral infections trigger exacerbations of cystic fibrosis in adults and children. *Eur Respir J*, 40, 510-2.
- WATERS, E. M., NEILL, D. R., KAMAN, B., SAHOTA, J. S., CLOKIE, M. R. J., WINSTANLEY, C. & KADIOGLU, A. 2017. Phage therapy is highly effective against chronic lung infections with *Pseudomonas aeruginosa*. *Thorax*, 72, 666-667.
- WEBB, J. S., LAU, M. & KJELLEBERG, S. 2004. Bacteriophage and phenotypic variation in *Pseudomonas aeruginosa* biofilm development. *J Bacteriol*, 186, 8066-73.
- WECKWERTH, W. 2003. Metabolomics in systems biology. *Annu Rev Plant Biol*, 54, 669-89.
- WHITCHURCH, C. B., TOLKER-NIELSEN, T., RAGAS, P. C. & MATTICK, J. S. 2002. Extracellular DNA required for bacterial biofilm formation. *Science*, 295, 1487.
- WIEDENBECK, J. & COHAN, F. M. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev*, 35, 957-76.
- WIEHLMANN, L., WAGNER, G., CRAMER, N., SIEBERT, B., GUDOWIUS, P., MORALES, G., KOHLER, T., VAN DELDEN, C., WEINEL, C., SLICKERS, P. & TUMMLER, B. 2007. Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*, 104, 8101-6.
- WILKE, M., BUIJS-OFFERMAN, R. M., AARBIOU, J., COLLEDGE, W. H., SHEPPARD, D. N., TOUQUI, L., BOT, A., JORNA, H., DE JONGE, H. R. & SCHOLTE, B. J. 2011. Mouse models of cystic fibrosis: phenotypic analysis and research applications. *J Cyst Fibros*, 10 Suppl 2, S152-71.
- WILLIAMS, D., EVANS, B., HALDENBY, S., WALSHAW, M. J., BROCKHURST, M. A., WINSTANLEY, C. & PATERSON, S. 2015. Divergent, coexisting *Pseudomonas aeruginosa* lineages in chronic cystic fibrosis lung infections. *Am J Respir Crit Care Med*, 191, 775-85.
- WILLIAMS, H. T. 2013. Phage-induced diversification improves host evolvability. *BMC Evol Biol*, 13, 17.

- WILLNER, D., FURLAN, M., HAYNES, M., SCHMIEDER, R., ANGLY, F. E., SILVA, J., TAMMADONI, S., NOSRAT, B., CONRAD, D. & ROHWER, F. 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One*, 4, e7370.
- WILSON, R., ROBERTS, D. & COLE, P. 1985. Effect of bacterial products on human ciliary function in vitro. *Thorax*, 40, 125-31.
- WINSTANLEY, C., LANGILLE, M. G., FOTHERGILL, J. L., KUKAVICA-IBRULJ, I., PARADIS-BLEAU, C., SANSCHAGRIN, F., THOMSON, N. R., WINSOR, G. L., QUAIL, M. A., LENNARD, N., BIGNELL, A., CLARKE, L., SEEGER, K., SAUNDERS, D., HARRIS, D., PARKHILL, J., HANCOCK, R. E., BRINKMAN, F. S. & LEVESQUE, R. C. 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res*, 19, 12-23.
- WINSTANLEY, C., O'BRIEN, S. & BROCKHURST, M. A. 2016. *Pseudomonas aeruginosa* Evolutionary Adaptation and Diversification in Cystic Fibrosis Chronic Lung Infections. *Trends Microbiol*, 24, 327-37.
- WOLFGANG, M. C., KULASEKARA, B. R., LIANG, X., BOYD, D., WU, K., YANG, Q., MIYADA, C. G. & LORY, S. 2003. Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*, 100, 8484-9.
- WOO, T. E., LIM, R., HEIRALI, A. A., ACOSTA, N., RABIN, H. R., MODY, C. H., SOMAYAJI, R., SURETTE, M. G., SIBLEY, C. D., STOREY, D. G. & PARKINS, M. D. 2019. A longitudinal characterization of the Non-Cystic Fibrosis Bronchiectasis airway microbiome. *Sci Rep*, 9, 6871.
- WOO, T. E., LIM, R., SURETTE, M. G., WADDELL, B., BOWRON, J. C., SOMAYAJI, R., DUONG, J., MODY, C. H., RABIN, H. R., STOREY, D. G. & PARKINS, M. D. 2018. Epidemiology and natural history of *Pseudomonas aeruginosa* airway infections in non-cystic fibrosis bronchiectasis. *ERJ Open Res*, 4.
- WORKENTINE, M. L., SIBLEY, C. D., GLEZERSON, B., PURIGHALLA, S., NORGAARD-GRON, J. C., PARKINS, M. D., RABIN, H. R. & SURETTE, M. G. 2013. Phenotypic heterogeneity of *Pseudomonas aeruginosa* populations in a cystic fibrosis patient. *PLoS One*, 8, e60225.
- XIA, J., PSYCHOGIOS, N., YOUNG, N. & WISHART, D. S. 2009. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res*, 37, W652-60.
- XIANG, Y., MORAIS, M. C., COHEN, D. N., BOWMAN, V. D., ANDERSON, D. L. & ROSSMANN, M. G. 2008. Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage phi29 tail. *Proc Natl Acad Sci U S A*, 105, 9552-7.
- XU, J., HENDRIX, R. W. & DUDA, R. L. 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Mol Cell*, 16, 11-21.
- XU, J., HENDRIX, R. W. & DUDA, R. L. 2013. A balanced ratio of proteins from gene G and frameshift-extended gene GT is required for phage lambda tail assembly. *J Mol Biol*, 425, 3476-87.
- XU, Y. J., WANG, C., W.E., H. & ONG, C. N. 2014. Recent developments and applications of metabolomics in microbiological investigations. *TrAC Trends in Analytical Chemistry*, 56, 37-42.
- YOURGENOME. 2020. Available: <https://www.yourgenome.org/copyright> [Accessed 2020].
- ZABURLIN, D., MERCANTI, D. J. & QUIBERONI, A. 2017. A fast PCR-based method for the characterization of prophage profiles in strains of the *Lactobacillus casei* group. *J Virol Methods*, 248, 226-233.
- ZHANG, Y. & LEJEUNE, J. T. 2008. Transduction of bla(CMY-2), tet(A), and tet(B) from *Salmonella enterica* subspecies *enterica* serovar Heidelberg to *S. Typhimurium*. *Vet Microbiol*, 129, 418-25.

- ZHANG, Y., ZHANG, C., DU, X., ZHOU, Y., KONG, W., LAU, G. W., CHEN, G., KOHLI, G. S., YANG, L., WANG, T. & LIANG, H. 2019. Glutathione Activates Type III Secretion System Through Vfr in *Pseudomonas aeruginosa*. *Front Cell Infect Microbiol*, 9, 164.
- ZHAO, X., SHEN, M., JIANG, X., SHEN, W., ZHONG, Q., YANG, Y., TAN, Y., AGNELLO, M., HE, X., HU, F. & LE, S. 2017. Transcriptomic and Metabolomics Profiling of Phage-Host Interactions between Phage PaP1 and *Pseudomonas aeruginosa*. *Front Microbiol*, 8, 548.
- ZHOU, Y., LIANG, Y., LYNCH, K. H., DENNIS, J. J. & WISHART, D. S. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res*, 39, W347-52.
- ZISSLER, J. 1967. Integration-negative (int) mutants of phage lambda. *Virology*, 31, 189.
- ZULIANELLO, L., CANARD, C., KÖHLER, T., CAILLE, D., LACROIX, J. S. & MEDA, P. 2006. Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*. *Infect Immun*, 74, 3134-47.
- LEPLAE, R., LIMA-MENDEZ, G. & TOUSSAINT, A. 2010. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res*, 38, D57-61.
- NHS.UK. 2016. *Bronchitis* [Online]. Available: <https://www.nhs.uk/conditions/bronchitis/> [Accessed 12/12/18 2018].
- WINSTANLEY, C., LANGILLE, M. G., FOTHERGILL, J. L., KUKAVICA-IBRULJ, I., PARADIS-BLEAU, C., SANSCHAGRIN, F., THOMSON, N. R., WINSOR, G. L., QUAIL, M. A., LENNARD, N., BIGNELL, A., CLARKE, L., SEEGER, K., SAUNDERS, D., HARRIS, D., PARKHILL, J., HANCOCK, R. E., BRINKMAN, F. S. & LEVESQUE, R. C. 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. *Genome Res*, 19, 12-23.

11 APPENDICES

Appendix A

Table 1 Longitudinal Pa isolates from BR patients collected by the Freeman hospital Newcastle upon Tyne, UK

BRONCHIECTASIS STORE REF	PATIENT CODE	DATE OF ISOLATION FROM PATIENT
141	1	12/01/2011
550	1	07/06/2012
915	1	25/03/2013
1082	1	20/08/2013
1192	1	16/09/2013
1311	1	18/11/2013
1493	1	26/03/2014
1758	1	15/10/2014
1795	1	27/10/2014
1890	1	05/12/2014
1948	1	23/01/2015
1980	1	25/02/2015
1998	1	12/03/2015
2033	1	24/04/2015
2074	1	22/05/2015
2101	1	09/06/2015
2107	1	26/06/2015
2165	1	12/08/2017
2223	1	18/09/2015
2264	1	16/11/2015
2302	1	03/03/2016
2375	1	06/07/2016
2437	1	07/09/2016
2459	1	07/10/2016
83A	2	22/10/2008
89B	2	22/10/2008
150	2	03/03/2011
253	2	31/08/2011
604	2	01/08/2012
796	2	31/08/2011
844	2	30/01/2013
1544	2	09/05/2014
1952	2	13/02/2015
1956A	2	06/02/2015
1957B	2	06/02/2015
1970	2	16/02/2015
2012	2	15/04/2015

2173	2	12/08/2015
2338	2	10/05/2016
2351	2	15/06/2016
2514	2	16/11/2016
2547	2	01/12/2016
125	3	19/05/2010
209	3	27/07/2011
213	3	21/07/2011
240A	3	12/08/2011
241B	3	12/08/2011
242	3	30/08/2011
317	3	03/11/2011
454A	3	19/03/2012
455B	3	19/03/2012
664A	3	16/08/2012
665B	3	16/08/2012
903	3	14/03/2013
1200	3	24/09/2013
1297	3	14/11/2013
1451	3	12/02/2014
1487A	3	22/03/2014
1501B	3	22/03/2014
1592	3	10/07/2014
1623A	3	04/08/2014
1624	3	04/08/2014
1737	3	12/10/2014
1762	3	02/10/2014
1765	3	03/10/2014
1806	3	05/11/2014
1983	3	04/03/2015
2009A	3	15/04/2015
2010B	3	15/04/2015
2168	3	31/07/2015
2193	3	03/09/2015
2203	3	13/09/2015
2212	3	02/10/2015
2226	3	12/10/2015
2284	3	29/12/2015
2299	3	25/02/2016
2334	3	29/04/2016
945	4	24/04/2013
1427	4	05/02/2014
1434	4	11/02/2014
1481	4	26/03/2014
1620	4	05/08/2014
1954	4	11/02/2015

2207	4	16/09/2015
79	5	22/10/2008
208	5	26/07/2011
464	5	14/03/2012
509	5	16/04/2012
631	5	09/07/2012
681	5	12/09/2012
722	5	29/10/2012
900	5	04/03/2013
1243	5	16/10/2013
1301	5	12/11/2013
1348	5	03/12/2013
1430	5	12/02/2014
1555	5	21/05/2014
1627	5	06/08/2014
1846	5	19/11/2014
1986	5	05/03/2015
2015	5	08/04/2015
2434	5	24/08/2016
2492	5	26/10/2016
102	6	09/01/2009
727	6	01/11/2012
804	6	13/12/2012
852	6	11/02/2013
997	6	13/06/2013
1017	6	14/05/2013
1279	6	07/11/2013
1515	6	04/04/2014
1622	6	04/08/2014
1669	6	28/08/2014
1738	6	15/10/2014
1825	6	12/11/2014
1911	6	09/12/2014
2024A	6	30/04/2015
2034B	6	30/04/2015
2035	6	30/04/2015
2088	6	28/05/2015
2146	6	22/07/2015
2373	6	06/07/2016
2480	6	07/10/2016
153	7	15/03/2011
255	7	01/09/2011
264	7	08/09/2011
491	7	11/04/2012
618	7	25/07/2012
694	7	08/10/2012

942	7	25/04/2013
1208	7	02/10/2013
1690	7	03/09/2014
1993	7	04/03/2015
2183	7	27/08/2015
2422	7	17/08/2016
2456	7	20/09/2016
73	8	21/10/2008
205	8	27/07/2011
657	8	08/08/2012
845	8	30/01/2013
1122	8	28/08/2013
1821	8	06/11/2014
2051	8	07/05/2015
2098	8	12/06/2015
2247	8	03/11/2015
2592	8	19/12/2016
144	9	02/02/2011
289	9	27/09/2011
2041	9	29/04/2015
2249	9	06/11/2015
2298	9	26/02/2016
2439	9	07/09/2016
2523	9	11/11/2016
910	10	15/03/2013
1465	10	07/03/2014
1984	10	06/03/2015
200	11	21/07/2011
494	11	12/04/2012
701	11	08/10/2012
916	11	24/03/2013
988	11	07/06/2013
1048	11	29/07/2013
178	12	12/07/2011
234	12	01/08/2011
346	12	30/11/2011
887	12	04/03/2013
927	12	03/04/2013
1424	13	05/02/2014
1609	13	29/07/2014
1812	13	05/11/2014
2332	13	06/05/2016
204	14	20/07/2011
607	14	01/08/2012
1542	14	07/05/2014
2612	14	11/01/2017

285	15	23/09/2011
560	15	20/06/2012
944	15	24/04/2013
1091	15	21/08/2013
311	16	18/10/2011
319	16	04/11/2011
2170	16	20/08/2015
2522	16	09/11/2016
459	17	21/03/2012
686	17	19/09/2012
2153	17	23/07/2015
2506	17	02/11/2016
14A	18	12/03/2008
15B	18	12/03/2008
146	18	02/02/2011
312A	19	17/10/2011
313B	19	17/10/2011
1061	19	08/07/2013
728	20	31/10/2012
2346	20	08/06/2016
2521	20	09/11/2016
123	21	20/05/2010
254	21	31/08/2011
360	21	06/12//2011
199	22	20/07/2011
368	22	18/01/2012
948	22	17/04/2013
197	23	20/07/2011
231A	23	04/08/2011
232B	23	04/08/2011
136	24	13/10/2010
511	24	18/04/2012
288	25	03/10/2011
949	25	17/04/2013
244	26	31/08/2011
1071	26	05/07/2013
8	27	30/10/2007
53	27	06/09/2008
257	28	01/09/2011

Table 2 Pa samples from University of Liverpool from co-colonised BR patients with

Isolate ID	Description	
C12	Pa sample A from patient 42	
C13	Pa sample B from patient 42	
C86	Pa sample A from patient 73	
C87	Pa sample B from patient 73	
C107	Pa sample A from patient 84	
C108	Pa sample B from patient 84	
C109	Pa sample A from patient 85	
C110	Pa sample B from patient 85	
C125	Pa sample A from patient 92	
C127	Pa sample B from patient 92	
C129	Pa sample C from patient 92	
A100	Pa sample A from patient 148	
A106	Pa sample B from patient 148	

Table 3 LES phage lysogens in PAO1 obtained from the University of Liverpool

Isolate ID	<i>Pseudomonas aeruginosa</i> strain used as host	Prophages present in genome	Description	Origin
PAO1	N/A	N/A	Wild-type PAO1	Winstanley strain collection: well studied laboratory reference strain (Stover <i>et al.</i> , 2000)
LES ϕ 2	PAO1	LES ϕ 2	LES ϕ 2 lysogen	Winstanley strain collection. Constructed by Chloe James (James <i>et al.</i> , 2012).
LES ϕ 3	PAO1	LES ϕ 3	LES ϕ 3 lysogen	Winstanley strain collection. Constructed by Chloe James (James <i>et al.</i> , 2012).
LES ϕ 4	PAO1	LES ϕ 4	LES ϕ 4 lysogen	Winstanley strain collection. Constructed by Chloe James (James <i>et al.</i> , 2012).
LES ϕ 2+ ϕ 3+ ϕ 4	PAO1	LES ϕ 2, 3 and 4	Triple LES phage lysogen	Winstanley strain collection. Constructed by Chloe James (James <i>et al.</i> , 2012).
LES ϕ 3+ ϕ 4	PAO1	LES ϕ 3 and 4	Double LES phage lysogen	Winstanley strain collection. Constructed by Chloe James (James <i>et al.</i> , 2012).

Table 4 Isolates from the IPCD chosen for induction experiments to obtain temperate phages to form lysogens with PA14

Isolate_ID	IPCD_ID	Phage_clade	Isolate_city	Human_Isolation_site	HumanPathology
1	108	clade1	unknown	Sputum	Bronchiectasis
2	148	clade1	Sherbrooke	Sputum	Cystic fibrosis
3	1330	clade1	Quebec	Sputum	Cystic fibrosis
4	32	clade1	Montreal	Sputum	Cystic fibrosis
5	1323	clade1	Quebec	Sputum	Cystic fibrosis
6	1584	clade1	Chicoutimi	Sputum	Cystic fibrosis
7	177	clade1	Sherbrooke	Nasopharyngeal swab	Cystic fibrosis
8	1450	clade1	Saint-Roch-de-Achigan	infected lung	Cystic fibrosis
9	170	clade2	Sherbrooke	Sputum	Cystic fibrosis
10	1319	clade2	Quebec	Sputum	Cystic fibrosis
11	1627	clade2	Calgary (CACFC)	Sputum	Cystic fibrosis
12	1325	clade2	Quebec	Sputum	Cystic fibrosis

13	1632	clade2	Calgary (CACFC)	Sputum	Cystic fibrosis
14	164	clade2	Sherbrooke	Nasopharyngeal swab	Cystic fibrosis
15	355	clade2	Brisbane	Sputum	Cystic fibrosis
16	1592	clade2	Chicoutimi	Sputum	Cystic fibrosis
17	1572	clade2	Montreal	Sputum	Cystic fibrosis
18	922	clade3	Hobart	Sputum	Cystic fibrosis
19	948	clade3	Hobart	Sputum	Cystic fibrosis
20	971	clade3	Hobart	Sputum	Cystic fibrosis
21	976	clade3	Hobart	Sputum	Cystic fibrosis
22	1257	clade3	Melbourne	Sputum	Cystic fibrosis
23	1258	clade3	Hobart	Sputum	Cystic fibrosis
24	1360	clade3	unknown	Throat swab	Cystic fibrosis
25	1408	clade3	Hannover	Throat swab	Cystic fibrosis
26	1579	clade3	unknown	Sputum	Cystic fibrosis
27	103	clade3	Unknown	Sputum	Pneumonia
28	346	clade4	Brisbane	Sputum	Cystic fibrosis
29	940	clade4	Hobart	Sputum	Cystic fibrosis
30	73	clade4	Montreal	Sputum	Cystic fibrosis
31	1069	clade4	Nottingham	Sputum	Cystic fibrosis
32	74	clade4	Montreal	Sputum	Cystic fibrosis
33	1085	clade4	Nottingham	Sputum	Cystic fibrosis
34	1589	clade4	Chicoutimi	Sputum	Cystic fibrosis
35	1517	clade4	Rouyn-Noranda	Throat	Cystic fibrosis
36	1072	clade4	Nottingham	Sputum	Cystic fibrosis
37	128	clade5	Sherbrooke	Sputum	Cystic fibrosis
38	1333	clade5	Quebec	Sputum	Cystic fibrosis
39	1067	clade5	Nottingham	Sputum	Cystic fibrosis
40	1080	clade5	Nottingham	Sputum	Cystic fibrosis
41	1335	clade5	Quebec	Sputum	Cystic fibrosis
42	905	clade5	Dunedin	Sputum	Cystic fibrosis
43	1436	clade5	Lund	Throat swab	Cystic fibrosis
44	1629	clade5	Calgary (CACFC)	Sputum	Cystic fibrosis

45	1515	clade5	Rouyn-Noranda	Throat	Cystis fibrosis
46	644	clade6	Leiden	Sputum	Non-CF
47	1363	clade6	Munich	Throat swab	Cystic fibrosis
48	124	clade7	Sherbrooke	Sputum	Cystic fibrosis
49	810	clade7	Birmingham	Sputum	Cystic fibrosis
50	1328	clade7	Quebec	Sputum	Cystic fibrosis
51	1412	clade7	Hannover	Throat swab	Cystic fibrosis
52	1574	clade7	Montreal	Sputum	Cystic fibrosis
53	545	clade7	Seattle	Throat	Cystic fibrosis
54	1604	clade7	Chicoutimi	Sputum	Cystic fibrosis
55	159	clade7	Sherbrooke	Nasopharyngeal swab	Cystic fibrosis
56	715	clade7	Vancouver	Sputum	Cystic fibrosis
57	1451	clade7	Montreal-Nord	infected lung	Cystic fibrosis
58	43	clade8	Montreal	Sputum	Cystic fibrosis
59	112	clade8	Sherbrooke	Sputum	Cystic fibrosis
60	1308	clade8	Quebec	Sputum	Cystic fibrosis
61	1309	clade8	Quebec	Sputum	Cystic fibrosis
62	1307	clade8	Quebec	Sputum	Cystic fibrosis
63	1341	clade8	Quebec	Sputum	Cystic fibrosis
64	1370	clade8	Munich	Throat swab	Cystic fibrosis
65	1433	clade8	Hannover	Throat swab	Cystic fibrosis
66	1593	clade8	Chicoutimi	Sputum	Cystic fibrosis
67	29	clade8	Montreal	Sputum	Cystic fibrosis
68	1488	clade8	Rouyn-Noranda	Sputum	Cystis fibrosis
69	334	clade8	Brisbane	Sputum	Cystic fibrosis

Table 5 PAO1 lysogens containing phages from Pa clinical CF and BR isolates made by Dr Adnan Tariq

Lysogen	Bacterial strain	Disease etiology	Putative phage	Origin
CF47	PAO1	Paediatric Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF53	PAO1	Paediatric Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF124	PAO1	Paediatric Cystic Fibrosis	JBD24-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF165	PAO1	Paediatric Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF187	PAO1	Paediatric Cystic Fibrosis	H70-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF24	PAO1	Adult Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF52	PAO1	Adult Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF121	PAO1	Adult Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF177	PAO1	Adult Cystic Fibrosis	D3112-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
CF74	PAO1	Adult Cystic Fibrosis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR152	PAO1	< 10 year Bronchiectasis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR233	PAO1	< 10 year Bronchiectasis	D3112-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR299	PAO1	< 10 year Bronchiectasis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR327	PAO1	< 10 year Bronchiectasis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR332	PAO1	< 10 year Bronchiectasis	D3112-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR141	PAO1	> 10 years Bronchiectasis	D3112 and LPB1 like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)

BR228	PAO1	> 10 years Bronchiectasis	LPB1 and D3112-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR243	PAO1	> 10 years Bronchiectasis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR313	PAO1	> 10 years Bronchiectasis	F10-like	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)
BR200	PAO1	> 10 years Bronchiectasis	D3112	Lysogen made by colleague Adnan Tariq (Tariq et al., 2019)

Appendix B

Preparation of Artificial Sputum Medium (ASM)

ASM was prepared by addition of 4g DNA from fish sperm to 250 ml sterile water very slowly over a period of several hours. The DNA takes several hours to completely dissolve and can be stirred overnight at room temperature.

5g of mucin from porcine stomach (type II) was slowly added to 250 ml sterile water until the mucin was dissolved completely. The solution can be stirred overnight at 4 °C.

Each essential and non-essential L-amino acid, with the exception of L- tyrosine and L- cysteine was added in to 100 ml sterile water. 0.25 g of L-cysteine was dissolved in 25 ml of 0.5 M potassium hydroxide (Mr 56.11 g/mol) and 0.25 g of L-tyrosine in 25 ml sterile water.

Then 5.9 mg of diethylenetriaminepentaacetic acid (DTPA), 5 g NaCl and 2.2 g of KCl was dissolved in 100 ml of sterile water and combined with the DNA, Mucin, L-amino acids, DTPA, NaCl and KCl in a 1 litre bottle and 5ml of egg yolk emulsion. Then the bottle was filled to approximately 850 ml with sterile water.

pH was adjusted to 6.9 with 1 M Tris (pH 8.5; Mr 121.14) and the volume was brought to 1 litre with sterile water.

Sterilization of the ASM was done by filtration using a Vacuubrand ME 2 diaphragm vacuum pump and Millipore Steritop filter units with a pore and neck size of 0.22 µm and 45 mm

Unfiltered and filtered ASM should be stored at 4 °C in the dark. Using fresh ASM is recommended however, it can be kept under these conditions for a maximum of one month.

Appendix D

UHPLC-MS protocol

Metabolite profiling of the urine samples were acquired on a Dionex 3000 Ultra High Pressure Liquid chromatography (UHPLC) system hyphenated to the Q-Exactive classic high resolution mass spectrometer system (ThermoScientific, Bremen, Germany). All solvents and ionization agent used were of analytical grade or higher unless stated.

The chromatographic separation was performed on a Water Acquity Ethylene Bridge Hybrid (BEH) Amide analytical column (1 x 150mm) with particle size of 1.7micron at a flow rate of 100 μ L/min, the column temperature was set to 45oC. The Binary buffer system was as follows: Buffer A was MilliQ water and Buffer B was ACN, both with 10mM ammonium formate adjusted to pH 3.5 using formic acid.

The LC gradient profile was as follows: T:0 min: 95%(B), T: 2min 60% (B) T: 5min 40%(B), T:7.5 min 40%(B), T:7.6min 95%(B), T:10 min 90% (B).

The Heated spray ionization source (HESI) was set to the following parameters: Sheath gas flow rate of 50, the Aux gas flow rate was set to 13 and the sweep gas flow rate was 3. The Spray voltage of set to 3.5kV with a Capillary temperature of 275oC. The Aux gas heater temperature was adjusted to 425oC.

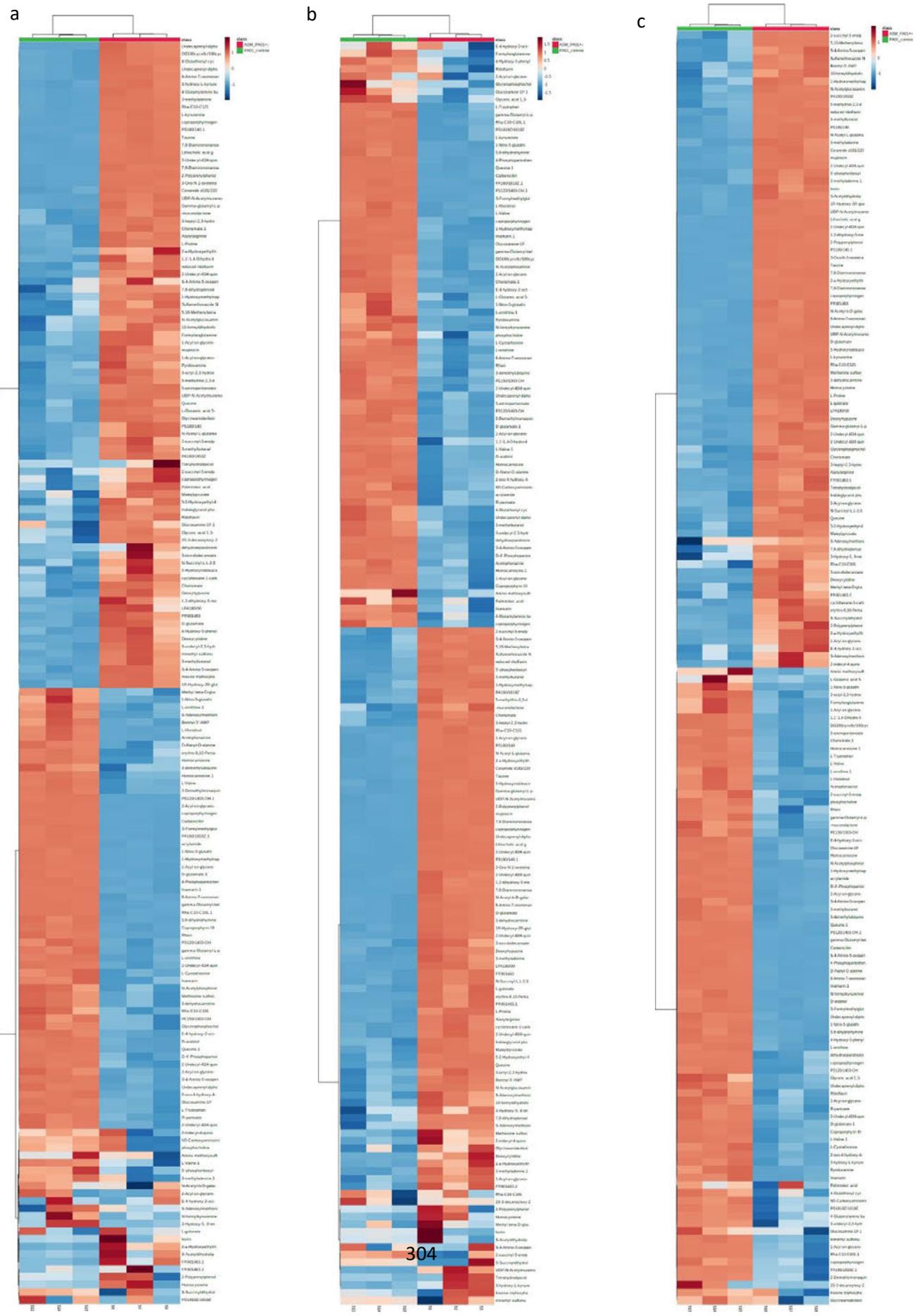
MS1 mass acquisition: The mass acquisition range was as follows: 75-1000 m/z units at a mass resolution of 35,000 at approximately 7.6 scans per second, microscan: 1, lock mass: off. The AGC was set to 1e6 and the ion injection time was 100mS-1. The data was acquired on both Positive and Negative mode polarity (independently), the setting for the negative mode was the same as positive ion mode except the voltage was set to 2.5kV.

Peak table generation and alignment were performed using compound discoverer 2.1 (ThermoScientific, Bremen, Germany) with an alignment window of 0.25min, mass tolerance: 5ppm and an signal intensity threshold of 200,000 counts with a signal to noise ratio of 5:1. Protonated adduct only.

3 μ L injection was applied.

The system was prime with a minimum of 10 sequential injections of pooled QC to stabilised the HESI source and to check for chromatographic stability before initialling the batch analysis.

Appendix E



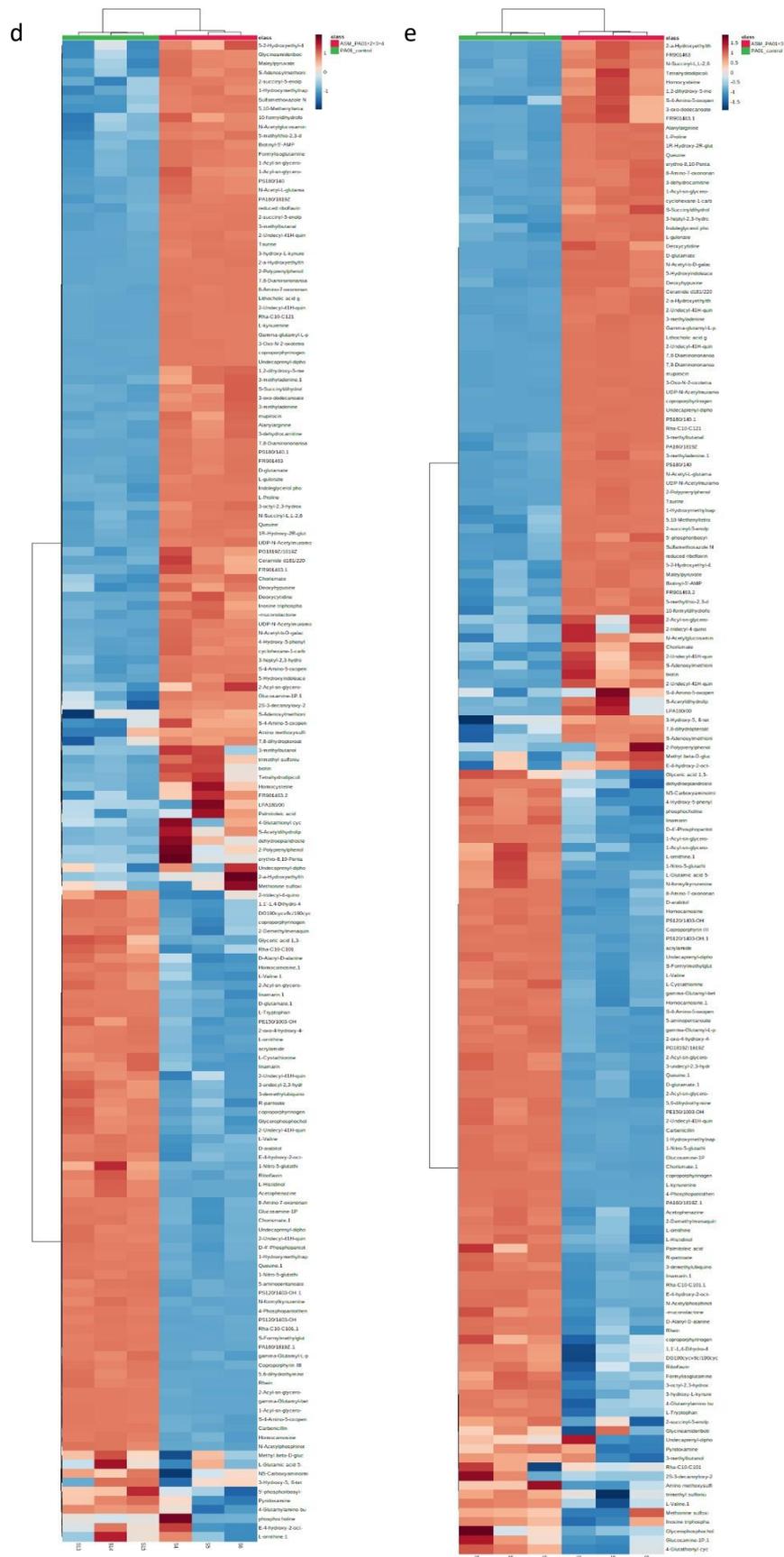


Figure 1 - Clustering result shown as heatmap. Representing the up and down regulation of each significant metabolite between a) the LES-2 lysogen and the control. b) the LES-3 lysogen and the control. c) the LES-4 lysogen and the control. d) the LES-2+3+4 lysogen and the control. e) the LES-3+4 lysogen and the control.

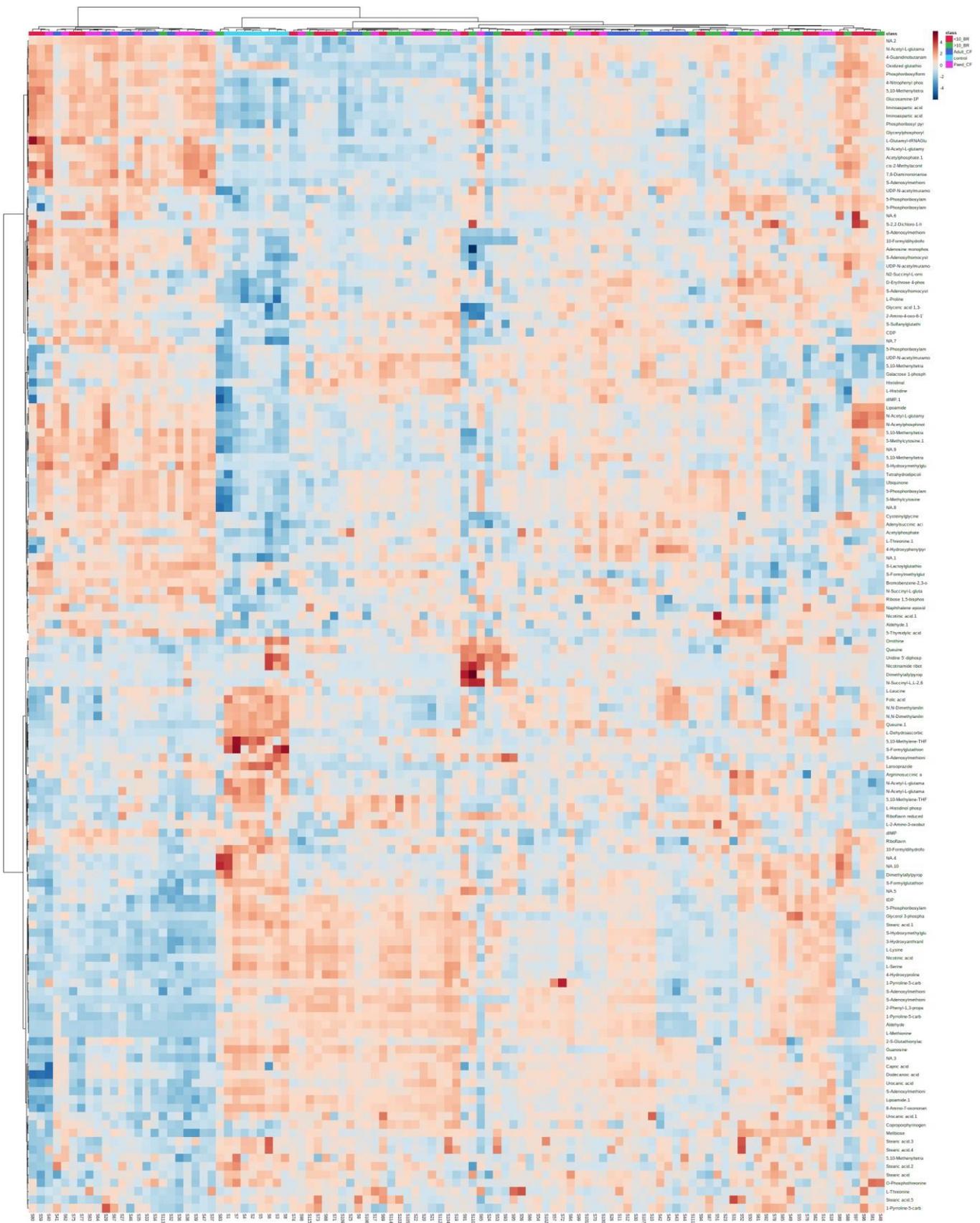
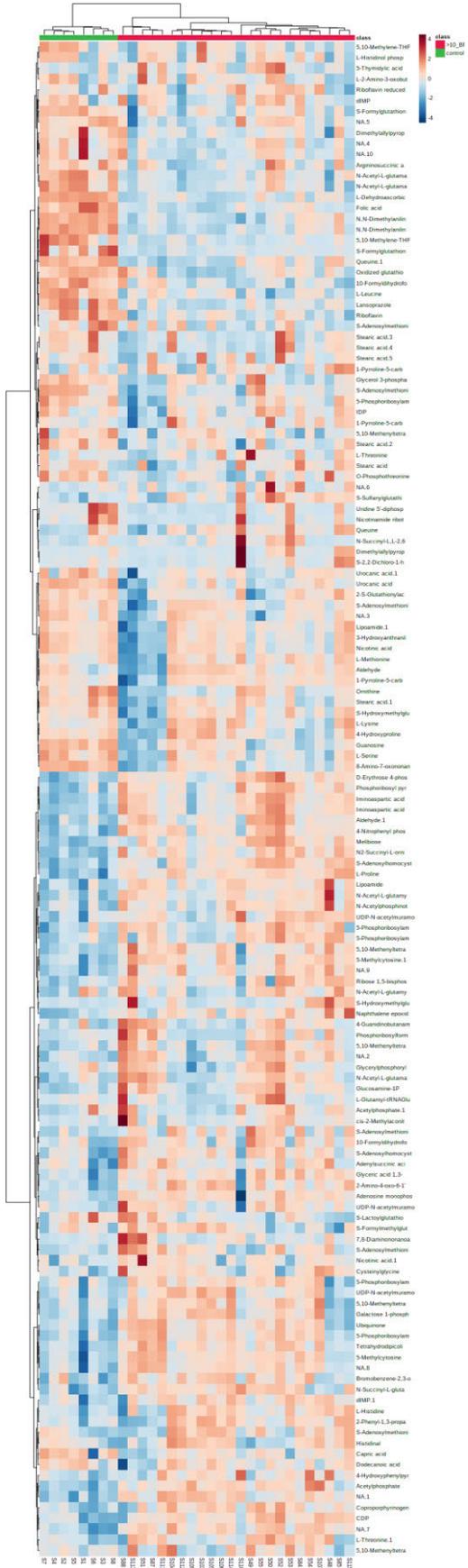


Figure 2 - Clustering result shown as heatmap. Representing the up and down regulation of each significant metabolite between all lysogen groups consisting of PAO1 containing full complements of phages induced from the four groups Adult CF, Paed CF, >10 years BR and >10 years BR.

a



b

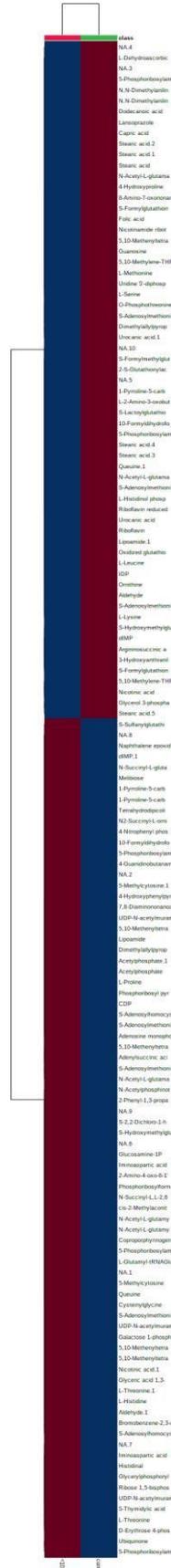


Figure 4 - Clustering result shown as heatmap. Representing the up (red) and down (blue) regulation of each significant metabolite between a) lysogens in the Old BR (>10years) group compared to the PAO1 control. b) lysogens in the young BR (< 10years) group compared to the PAO1 control

Appendix F

Table 1 – Lysogens associated with samples numbers from the metabolomics analysis of lysogens containing LES phages from chapter 5.

SAMPLE	LYSOGEN	GROUP
S1	LES3	ASM_PA01+3
S2	LES3	ASM_PA01+3
S3	LES5	ASM_PA01+3
S4	LES2+3+4	ASM_PA01+2+3+4
S5	LES2+3+4	ASM_PA01+2+3+4
S6	LES2+3+4	ASM_PA01+2+3+4
S7	LES2	ASM_PA01+2
S8	LES2	ASM_PA01+2
S9	LES2	ASM_PA01+2
S10	LES3+4	ASM_PA01+3+4
S11	LES3+4	ASM_PA01+3+4
S12	LES3+4	ASM_PA01+3+4
S13	PAO1 control	PA01_control
S14	PAO1 control	PA01_control
S15	PAO1 control	PA01_control
S16	LES4	ASM_PA01+4
S17	LES4	ASM_PA01+4
S18	LES4	ASM_PA01+4

Table 2 – Lysogens associated with samples numbers from the metabolomics analysis of lysogens containing phages from different clinical groups from chapter 5.

SAMPLE ID	LYSOGEN ID	CLINICAL/AETIOLOGICAL GROUP
S1	PA01	control
S2	PA01	control
S3	PA01	control
S4	PA01	control
S5	PA01	control
S6	PA01	control
S7	PA01	control
S8	PA01	control
S9	CF24	Adult_CF
S10	CF24	Adult_CF
S11	CF24	Adult_CF
S12	CF24	Adult_CF
S13	CF47	Paed_CF
S14	CF47	Paed_CF
S15	CF47	Paed_CF
S16	CF47	Paed_CF

S17	CF47	Paed_CF
S18	CF53	Paed_CF
S19	CF53	Paed_CF
S20	CF53	Paed_CF
S21	CF53	Paed_CF
S22	CF53	Paed_CF
S23	CF53	Paed_CF
S24	CF52	Adult_CF
S25	CF52	Adult_CF
S26	CF52	Adult_CF
S27	CF74	Adult_CF
S28	CF74	Adult_CF
S29	CF74	Adult_CF
S30	CF74	Adult_CF
S31	CF121	Adult_CF
S32	CF121	Adult_CF
S33	CF121	Adult_CF
S34	CF121	Adult_CF
S35	CF124	Paed_CF
S36	CF124	Paed_CF
S37	CF124	Paed_CF
S38	CF124	Paed_CF
S39	CF124	Paed_CF
S40	CF124	Paed_CF
S41	CF177	Adult_CF
S42	CF177	Adult_CF
S43	CF177	Adult_CF
S44	CF177	Adult_CF
S45	CF177	Adult_CF
S46	CF177	Adult_CF
S47	CF177	Adult_CF
S48	BR141	>10_BR
S49	BR141	>10_BR
S50	BR141	>10_BR
S51	BR141	>10_BR
S52	BR141	>10_BR
S53	BR141	>10_BR
S54	BR141	>10_BR
S55	BR141	>10_BR
S56	BR152	<10_BR
S57	BR152	<10_BR
S58	BR152	<10_BR
S59	BR152	<10_BR
S60	BR152	<10_BR
S61	CF187	Paed_CF
S62	CF187	Paed_CF

S63	CF187	Paed_CF
S64	CF187	Paed_CF
S65	CF187	Paed_CF
S66	CF187	Paed_CF
S67	CF187	Paed_CF
S68	BR223	<10_BR
S69	BR223	<10_BR
S70	BR223	<10_BR
S71	BR223	<10_BR
S72	BR223	<10_BR
S73	BR223	<10_BR
S74	BR223	<10_BR
S75	BR322	<10_BR
S76	BR322	<10_BR
S77	BR322	<10_BR
S78	BR322	<10_BR
S79	BR322	<10_BR
S80	BR322	<10_BR
S81	BR322	<10_BR
S82	BR322	<10_BR
S83	BR313	>10_BR
S84	BR313	>10_BR
S85	BR313	>10_BR
S86	BR313	>10_BR
S87	BR313	>10_BR
S88	BR313	>10_BR
S89	BR299	<10_BR
S90	BR299	<10_BR
S91	BR299	<10_BR
S92	BR299	<10_BR
S93	BR327	<10_BR
S94	BR328	<10_BR
S95	BR329	<10_BR
S96	BR330	<10_BR
S97	BR331	<10_BR
S98	CF165	Paed_CF
S99	CF165	Paed_CF
S100	CF165	Paed_CF
S101	CF165	Paed_CF
S102	CF165	Paed_CF
S103	BR200	>10_BR
S104	BR200	>10_BR
S105	BR200	>10_BR
S106	BR200	>10_BR
S107	BR200	>10_BR
S108	BR228	>10_BR

S109	BR228	>10_BR
S110	BR228	>10_BR
S111	BR228	>10_BR
S112	BR228	>10_BR
S113	BR243	>10_BR
S114	BR243	>10_BR
S115	BR243	>10_BR

Appendix G

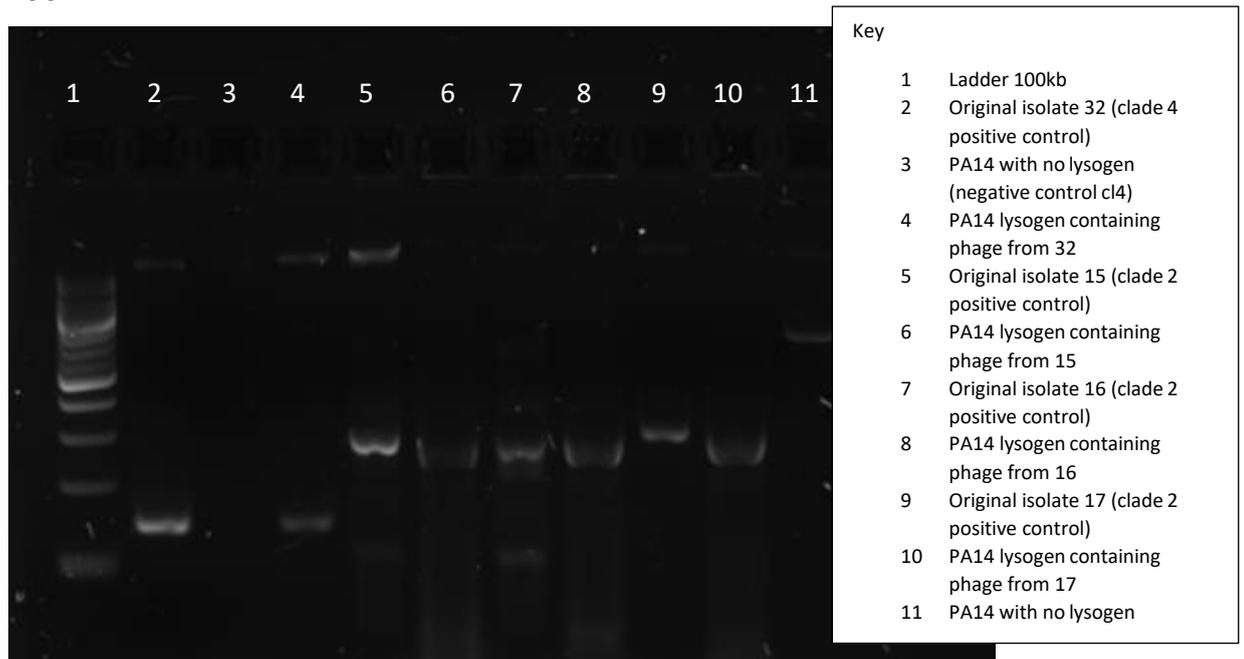


Figure 1- Electrophoresis gel showing the results of the PCR indicating phage infection clade 2. This gel shows if the selected phage gene marker is present. This indicates that the phage integrated into the PA14 isolate, compared to the PA14 negative control without the phage integrated.

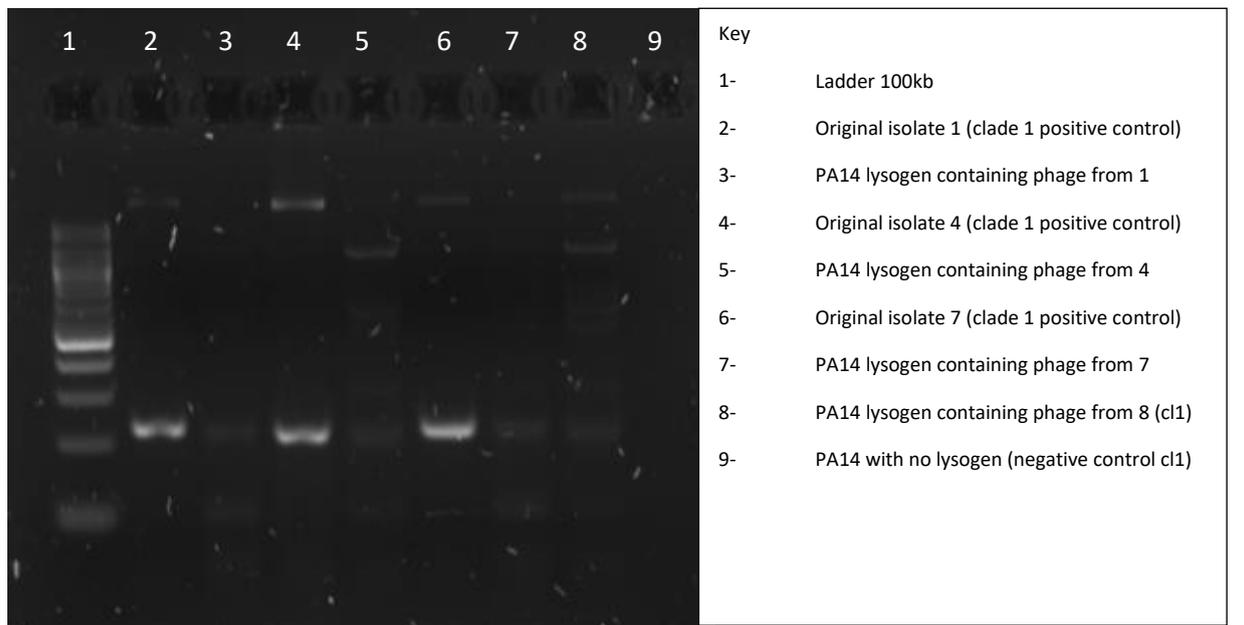


Figure 2- Electrophoresis gel showing the results of the PCR indicating phage infection clade 1. This gel shows if the selected phage gene marker is present. This indicates that the phage integrated into the PA14 isolate, compared to the PA14 negative control without the phage integrated

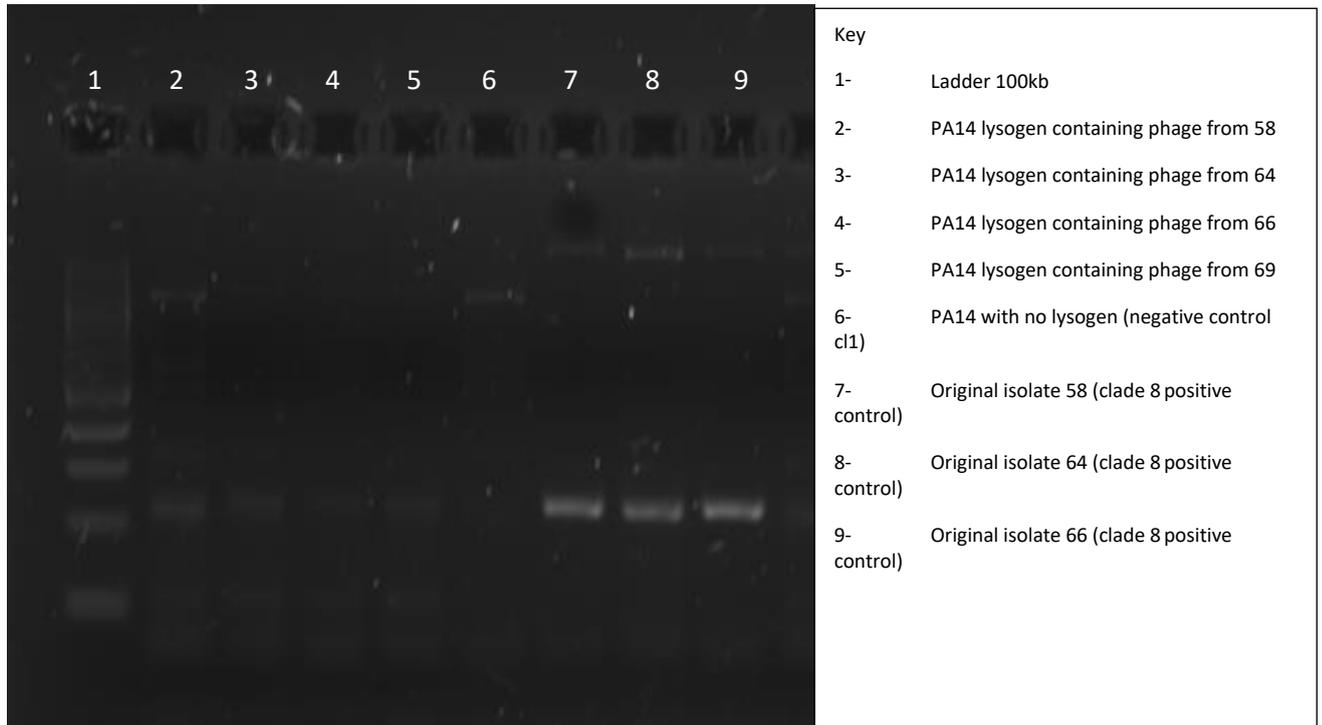
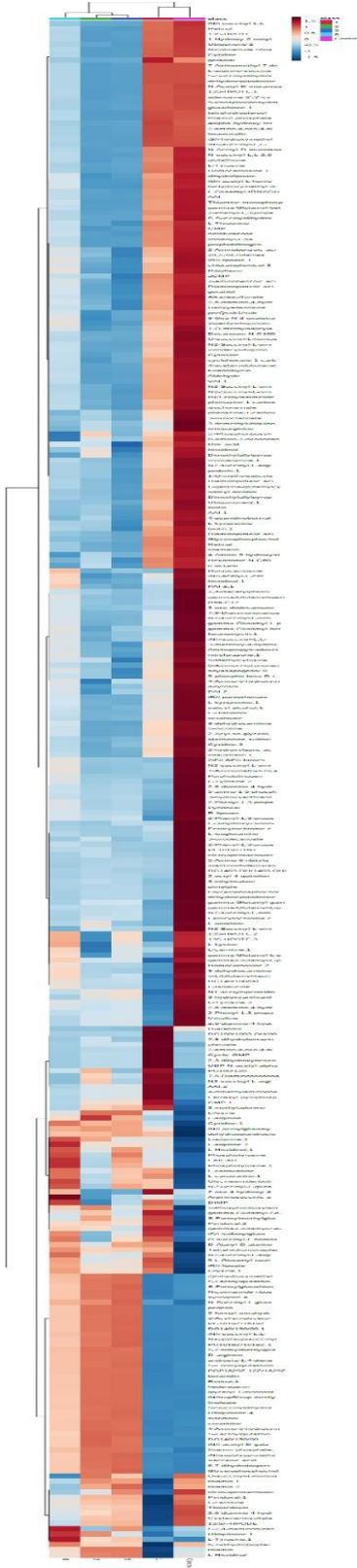


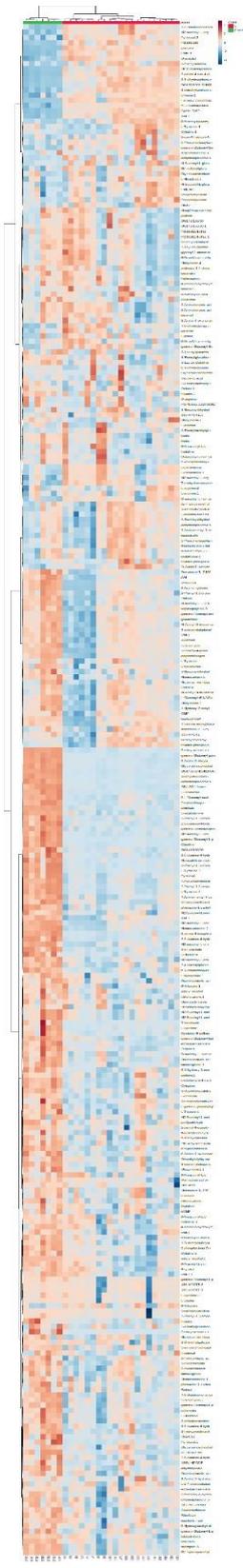
Figure 3- Electrophoresis gel showing the results of the PCR indicating phage infection clade 8. This gel shows if the selected phage gene marker is present. This indicates that the phage integrated into the PA14 isolate, compared to the PA14 negative control without the phage integrated.

Appendix H

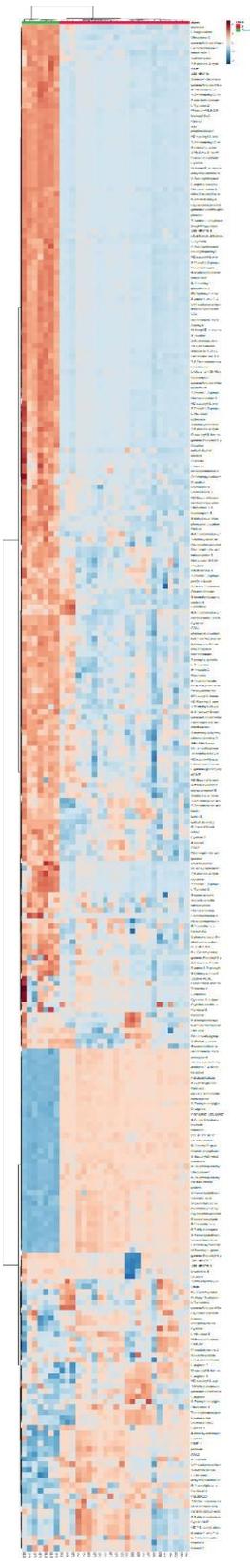
a



b



c



Appendix I

Table 1 Lysogens associated with sample numbers from the metabolomics analysis of lysogens containing phages from different clades (phage types) from chapter 7.

SAMPLE	LYSOGEN	CLADE
S1	1	1
S2	1	1
S3	1	1
S4	1	1
S5	1	1
S6	1	1
S7	1	1
S8	1	1
S10	1	1
S11	4	1
S12	4	1
S14	4	1
S16	4	1
S17	4	1
S18	4	1
S19	4	1
S20	5	1
S21	5	1
S22	5	1
S24	5	1
S25	5	1
S26	5	1
S27	5	1
S29	5	1
S32	5	1
S33	7	1
S34	7	1
S35	7	1
S36	7	1
S39	7	1
S40	7	1
S41	8	1
S42	8	1
S43	8	1
S44	8	1
S45	8	1
S46	8	1
S47	8	1
S49	8	1

S50	17	2
S51	17	2
S52	17	2
S53	17	2
S55	17	2
S56	17	2
S57	17	2
S59	12	2
S62	12	2
S63	12	2
S64	12	2
S66	12	2
S67	12	2
S69	15	2
S70	15	2
S71	15	2
S72	15	2
S73	15	2
S74	15	2
S75	16	2
S76	16	2
S77	16	2
S78	16	2
S79	16	2
S80	37	5
S81	37	5
S82	37	5
S83	37	5
S84	37	5
S85	37	5
S86	37	5
S87	37	5
S88	37	5
S90	42	5
S91	42	5
S93	42	5
S94	42	5
S95	42	5
S96	43	5
S97	43	5
S98	43	5
S99	43	5
S100	43	5
S101	43	5
S102	44	5
S103	44	5
S105	44	5
S106	44	5

S107	44	5
S108	44	5
S109	44	5
S111	44	5
S112	58	8
S113	58	8
S114	58	8
S115	58	8
S116	58	8
S117	58	8
S118	58	8
S119	58	8
S120	58	8
S121	64	8
S122	64	8
S123	64	8
S124	64	8
S125	64	8
S126	64	8
S127	64	8
S129	66	8
S130	66	8
S131	66	8
S132	66	8
S133	66	8
S134	66	8
S135	PA14	Control
S136	PA14	Control
S137	PA14	Control
S138	PA14	Control
S139	PA14	Control
S140	PA14	Control
S141	PA14	Control
S142	PA14	Control
S143	PA14	Control
S144	PA14	Control
S145	PA14	Control
S146	PA14	Control
S147	PA14	Control
S148	PA14	Control
S149	PA14	Control
S150	PA14	Control
S151	PA14	Control
S152	PA14	Control

Appendix J

Table 1 - Full list of Pa samples taken longitudinally and clinical details for patients in chapter 8. The samples highlighted in yellow are the isolates selected for sequencing and genome analysis (NR = not recorded?).

Bronchiectasis store ref	Patient code	Date of isolation from patient	IV antibiotics	FEV %	admissions
141	A	12/01/2011	NR	43.00	NR
550	A	07/06/2012	NR	38.00	08/06/2012
915	A	25/03/2013	NR	NR	23/03/2013
1082	A	20/08/2013	Meropenem	NR	20/08/2013
1192	A	16/09/2013	NR	NR	NR
1311	A	18/11/2013	Tazocin	NR	16/11/2013
1493	A	26/03/2014	Meropenem	36.00	27/03/2014
1758	A	15/10/2014	Ceftazidime	NR	30/10/2014
1795	A	27/10/2014	Ceftazidime	NR	30/10/2014
1890	A	05/12/2014	NR	NR	07/12/2014
1948	A	23/01/2015	NR	NR	NR
1980	A	25/02/2015	NR	NR	NR
1998	A	12/03/2015	Ceftazidime	NR	12/03/2015
2033	A	24/04/2015	NR	NR	NR
2074	A	22/05/2015	NR	NR	NR
2101	A	09/06/2015	NR	NR	NR
2107	A	26/06/2015	Tazocin	NR	26/06/2015
2165	A	12/08/2015	NR	NR	NR
2223	A	18/09/2015	Meropenem & Tazocin	NR	18/09/2015
2264	A	16/11/2015	Colistin	NR	16/11/2015
2302	A	03/03/2016	Meropenem	NR	25/02/2016

2375	A	06/07/2016	NR	NR	NR
2437	A	07/09/2016	NR	NR	NR
2459	A	07/10/2016	Meropenem	NR	06/10/2016
83a	B	22/10/2008	NR	NR	NR
89b	B	22/10/2008	NR	NR	NR
150	B	03/03/2011	NR	NR	NR
253	B	31/08/2011	NR	NR	NR
604	B	01/08/2012	NR	NR	NR
796	B	31/08/2012	NR	NR	NR
844	B	30/01/2013	NR	NR	NR
1544	B	09/05/2014	NR	NR	NR
1952	B	13/02/2015	NR	NR	NR
1956a	B	06/02/2015	NR	NR	NR
1957b	B	06/02/2015	NR	NR	NR
1970	B	16/02/2015	NR	NR	NR
2012	B	15/04/2015	NR	NR	NR
2173	B	12/08/2015	NR	NR	NR
2338	B	10/05/2016	NR	NR	NR
2351	B	15/06/2016	NR	NR	NR
2514	B	16/11/2016	NR	NR	NR
2547	B	01/12/2016	NR	NR	NR
125	C	19/05/2010	NR	34.40	NR
209	C	27/07/2011	Meropenem	NR	21/07/2011
213	C	21/07/2011	Meropenem	NR	21/07/2011
240a	C	12/08/2011	NR	NR	NR
241b	C	12/08/2011	Ceftazidime	NR	12/08/2011
242	C	30/08/2011	Timentin & Colomycin	NR	05/09/2011

317	C	03/11/2011	Timentin & Colomycin	26.85	19/10/2011
454	C	19/03/2012	Ceftazidime	31.00	19/03/2012
455	C	19/03/2012	Ceftazidime	31.00	19/03/2012
664a	C	16/08/2012	initially clavulanic acid/ticarcillin. switched to fosfomycin, colomycin IV for 14 days commenced taurolin nebs after trial	Measured 0.71	02/08/2012
665b	C	16/08/2012	initially clavulanic acid/ticarcillin. switched to fosfomycin, colomycin IV for 14 days commenced taurolin nebs after trial	Measured 0.71	02/08/2012
903	C	14/03/2013	Fosfomycin and Colomycin	NR	22/04/2013
1200	C	24/09/2013	Colomycin & Tazocin	NR	25/09/2013
1297	C	14/11/2013	Ceftazidime and Gentamicin	Measured 0.65	09/12/2013
1451	C	12/02/2014	NR	NR	NR
1487a	C	22/03/2014	Tazocin & Gentamicin	NR	17/03/2014
1501b	C	22/03/2014	Tazocin & Gentamicin	NR	17/03/2014
1592	C	10/07/2014	Ceftazidime, Tazocin & Gentamicin	Measured 0.6	10/07/2014
1623a	C	04/08/2014	Ceftazidime, Tazocin & Gentamicin	Measured 0.6	10/07/2014
1624b	C	04/08/2014	Ceftazidime, Tazocin & Gentamicin	Measured 0.6	10/07/2014
1737	C	12/10/2014	NR	NR	NR
1762	C	02/10/2014	NR	NR	NR
1765	C	03/10/2014	NR	NR	NR
1806	C	05/11/2014	Ceftazidime, Tazocin & Gentamicin - Plasma Exchange	NR	03/11/2014
1983	C	04/03/2015	NR	NR	NR
2009a	C	15/04/2015	Tazocin & Gentamicin	NR	16/04/2015
2010b	C	15/04/2015	Tazocin & Gentamicin	NR	16/04/2015
2168	C	31/07/2015	NR	NR	NR
2193	C	03/09/2015	Gentamicin & Aztreonam	NR	04/09/2015

2203	C	13/09/2015	Gentamicin & Aztreonam	NR	04/09/2015
2212	C	02/10/2015	PLASMAPHARESIS and IV IMMUNOGLOBULIN treatment	NR	05/10/2015
2226	C	12/10/2015	PLASMAPHARESIS and IV IMMUNOGLOBULIN treatment	NR	05/10/2015
2284	C	29/12/2015	Colomycin & Ceftazidime	NR	03/12/2015
2299	C	25/02/2016	Tazocin	NR	17/03/2016
2334	C	29/04/2016	Meropenem, Colomycin & Aztreonam	NR	26/03/2016
102	D	09/01/2009	NR	NR	NR
727	D	01/11/2012	NR	NR	NR
804	D	13/12/2012	NR	NR	NR
852	D	11/02/2013	NR	NR	NR
997	D	13/06/2013	NR	NR	NR
1017	D	14/05/2013	NR	NR	NR
1279	D	07/11/2013	NR	NR	NR
1515	D	04/04/2014	NR	NR	NR
1622	D	04/08/2014	NR	NR	NR
1669	D	28/08/2014	NR	NR	NR
1738	D	15/10/2014	NR	NR	NR
1825	D	12/11/2014	NR	NR	NR
1911	D	09/12/2014	NR	NR	NR
2034a	D	30/04/2015	NR	NR	NR
2035b	D	30/04/2015	NR	NR	NR
2088	D	28/05/2015	NR	NR	NR
2146	D	22/07/2015	NR	NR	NR
2373	D	06/07/2016	NR	NR	NR
2480	D	07/10/2016	NR	NR	NR
79	E	22/10/2008	NR	NR	NR

208	E	26/07/2011	NR	NR	NR
464	E	14/03/2012	NR	NR	NR
509	E	16/04/2012	NR	NR	NR
631	E	09/07/2012	NR	NR	NR
681	E	12/09/2012	NR	NR	NR
722	E	29/10/2012	NR	NR	NR
900	E	04/03/2013	NR	NR	NR
1243	E	16/10/2013	NR	NR	NR
1301	E	12/11/2013	NR	NR	NR
1348	E	03/12/2013	NR	NR	NR
1430	E	12/02/2014	NR	NR	NR
1555	E	21/05/2014	NR	NR	NR
1627	E	06/08/2014	NR	NR	NR
1846	E	19/11/2014	NR	NR	NR
1986	E	05/03/2015	NR	NR	NR
2015	E	08/04/2015	NR	NR	NR
2434	E	24/08/2016	NR	NR	NR
2492	E	26/10/2016	NR	NR	NR
945	F	24/04/2013	NR	67.10	NR
1427	F	05/02/2014	Tazocin	NR	11/02/2014
1434	F	11/02/2014	Tazocin	NR	11/02/2014
1481	F	26/03/2014	NR	NR	NR
1620	F	05/08/2014	Meropenem	63.00	21/10/2014
1954	F	11/02/2015	Meropenem	NR	26/03/2015
2207	F	16/09/2015	Meropenem	NR	12/10/2015
144	G	02/02/2011	NR	67.00	NR
289	G	27/09/2011	NR	Measured 1.35	NR

2041	G	29/04/2015	NR	63.00	NR
2249	G	06/11/2015	Tazocin	NR	04/12/2015
2298	G	26/02/2016	Tazocin	NR	14/04/2016
2439	G	07/09/2016	NR	NR	NR
2523	G	11/11/2016	Ceftazidime	NR	11/11/2016
200	H	21/07/2011	NR	99.00	NR
494	H	12/04/2012	NR	99.00	NR
701	H	08/10/2012	NR	104.00	NR
916	H	24/03/2013	NR	85.00	NR
988	H	07/06/2013	NR	NR	NR
1048	H	29/07/2013	NR	NR	NR
285	I	23/09/2011	Meropenem	53.00	22/09/2011
560	I	20/06/2012	NR	56.00	NR
944	I	24/04/2013	NR	57.00	NR
1091	I	21/08/2013	NR	53.00	NR

Appendix K

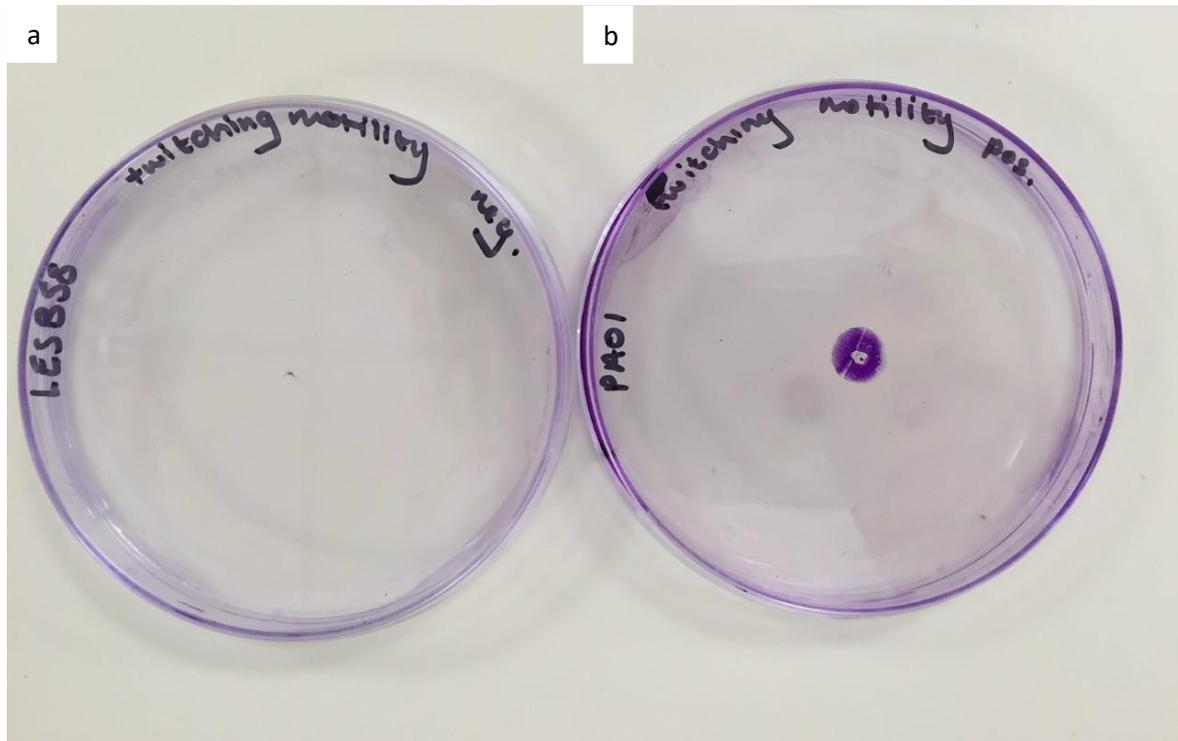


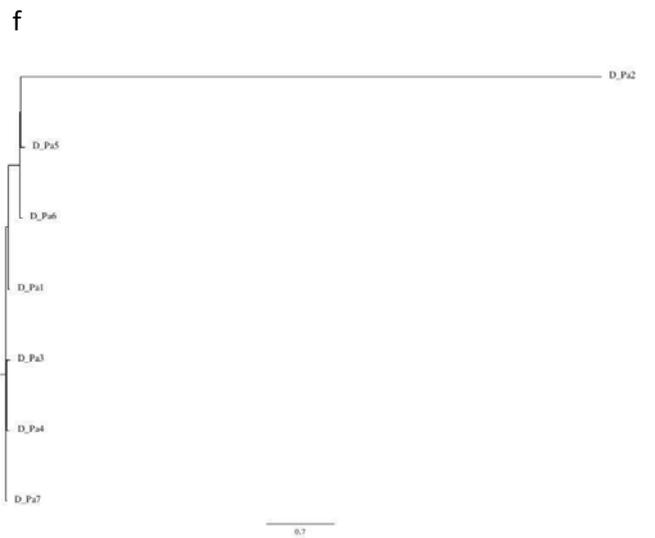
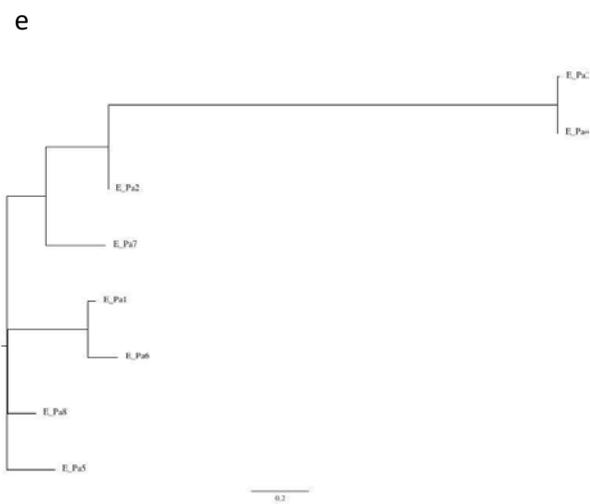
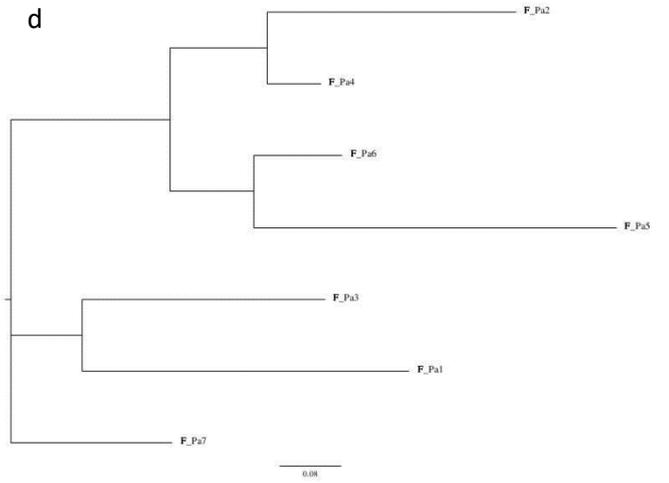
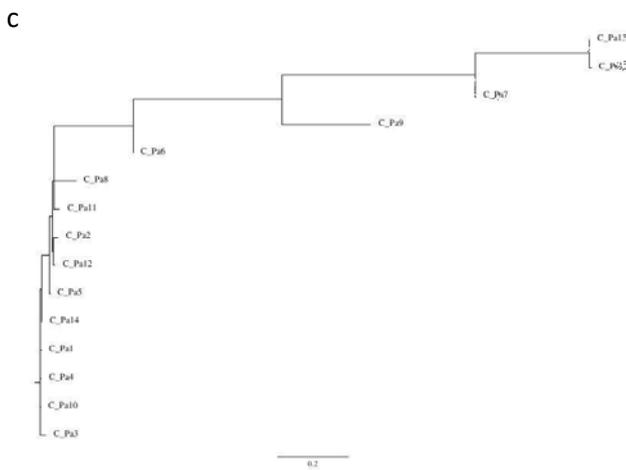
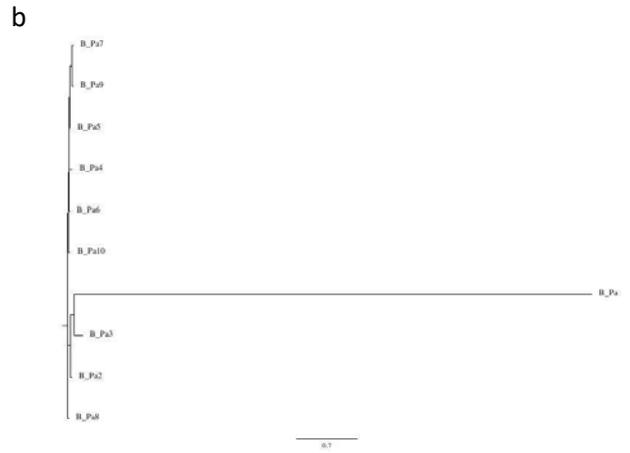
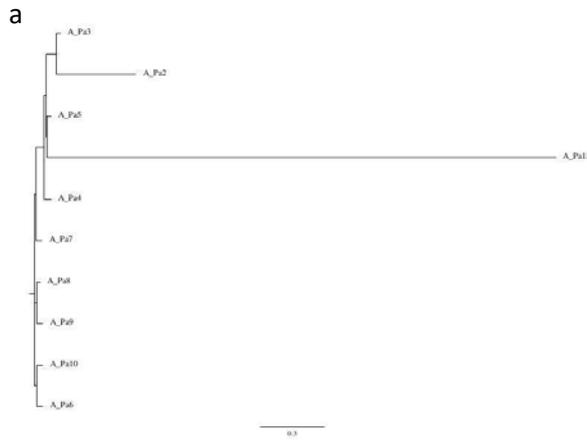
Figure 1 – Twitching motility assay positive and negative controls. a) LESB58 negative control. b) PAO1 positive control.

Appendix L

Table 1 – Quality control of sequencing that was performed by the University of Laval. A subset of 60 genomes sequenced

Isolate ID	File Name	Contigs	Scaffolds	Genome Size (bp)	Longest Scaffold (bp)	N50	No. of Raw reads	EC Reads	% reads passing EC	Raw nt	EC nt	% nt passing EC	Raw cov	Median cov	10th percentile cov	bases >= Q40
997	Run68-E10_S77	104	103	6820161	829654	234418	514972	499486	96.99	139166629	110491016	79.39	16.20	18	9	6682052
1669	Run68-E11_S85	60	59	6896205	682745	346740	505550	497561	98.42	137210157	112811563	82.22	16.36	19	11	6837521
1911	Run68-E12_S93	45	43	6896984	665469	346610	621640	612771	98.57	173073745	142063334	82.08	20.60	24	15	6874896
2010b	Run68-E6_S45	30	29	6445306	1154037	355471	650676	622468	95.66	171137996	138254059	80.79	21.45	24	15	6421758
2334	Run68-E7_S53	2113	2110	8813740	50229	8808	427052	418110	97.91	115413819	95740099	82.95	10.86	12	5	7764721
727	Run68-E9_S69	45	36	6897006	1007384	428920	663604	638618	96.23	188071152	148761001	79.10	21.57	25	16	6877711
2034a	Run68-F1_S6	55	47	6860188	654549	410497	720822	665380	92.31	202350555	155653658	76.92	22.69	26	18	6842980
2035b	Run68-F2_S14	60	51	6859758	664388	402629	776082	673961	86.84	212995229	151046627	70.92	22.02	25	16	6829999
2088	Run68-F3_S22	49	41	6901904	1020128	402629	964486	860840	89.25	267611588	197952544	73.97	28.68	33	23	6891238

Appendix M



Appendix N

Table 1 -Gene annotation results from the prophage expansion region in Pa isolate 5 compared to isolate 4 from patient D. Translated gene products from the regions of expansion were queried against the Pseudomonas database using BLASTP?

Gene ID	Prokka identification	Pseudomonas database identification
HHALDBOO_00002	hypothetical protein	recombinase family protein [Stenotrophomonas maltophilia]
HHALDBOO_00008	hypothetical protein	ATP-binding protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00004	hypothetical protein	DNA-binding protein [Stenotrophomonas sp. TD3]
BFMGCJGK_00003	Lysozyme RrrD	lysozyme [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00017	hypothetical protein	DNA methylase N-4 [Rubrivivax sp. SCN 71-131]
BFMGCJGK_00021	hypothetical protein	major capsid protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00013	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00007	hypothetical protein	antirepressor [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00006	hypothetical protein	DUF2793 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00001	hypothetical protein	DUF2924 domain-containing protein [Burkholderia multivorans]
BFMGCJGK_00011	hypothetical protein	DUF2163 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00010	hypothetical protein	hypothetical protein [Stenotrophomonas sp. TD3]
HHALDBOO_00011	hypothetical protein	isoleucyl-tRNA synthetase [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00012	hypothetical protein	helix-turn-helix domain-containing protein [Burkholderia multivorans]
HHALDBOO_00014	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00015	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00016	hypothetical protein	RNA polymerase subunit sigma [Rubrivivax sp. SCN 71-131]

HHALDBOO_00018	hypothetical protein	site-specific DNA-methyltransferase [<i>Pseudomonas aeruginosa</i>]
----------------	----------------------	--

HHALDBOO_00019	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00020	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00021	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00022	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00023	hypothetical protein	DUF3489 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00024	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00025	hypothetical protein	elements of external origin [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00001	hypothetical protein	site-specific DNA-methyltransferase [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00002	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00004	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00003	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00005	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00007	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00008	hypothetical protein	tail assembly protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00009	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00010	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00012	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00013	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00014	hypothetical protein	tape measure protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00015	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00016	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00017	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00018	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00019	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BFMGCJGK_00020	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00005	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

HHALDBOO_00006	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
HHALDBOO_00009	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

Table 2 Gene annotation results from the prophage expansion region in Pa isolate 2 compared to isolate 1 from patient B.

Gene ID	Prokka identification	Pseudomonas database identification
AJGGILGI_00001	Sulfurtransferase TusE	TusE/DsrC/DsvC family sulfur relay protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00002	Protein TusB	sulfurtransferase complex subunit TusB [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00003	Protein TusC	sulfurtransferase complex subunit TusC [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00004	Putative sulfurtransferase DsrE	sulfurtransferase complex subunit TusD [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00005	hypothetical protein	24K membrane protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00006	Modulator of FtsH protease YccA	Bax inhibitor-1/YccA family protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00010	hypothetical protein	DNA cytosine methyltransferase [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00011	hypothetical protein	hypothetical protein IPC1270_24600 [<i>Pseudomonas aeruginosa</i>] 71%
AJGGILGI_00015	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00016	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00018	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00019	hypothetical protein	BcepMu gp16 family phage-associated protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00020	hypothetical protein	helix-turn-helix domain-containing protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00021	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00022	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00023	Nucleoid-associated protein YejK	nucleoid-associated protein ndpA [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00024	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00032	hypothetical protein	phage tail protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00033	hypothetical protein	phage tail fiber protein [<i>Pseudomonas aeruginosa</i>]
AJGGILGI_00038	hypothetical protein	phage virion morphogenesis protein [<i>Pseudomonas aeruginosa</i>]

AJGGILGI_00051	hypothetical protein	ImmA/IrrE family metallo-endopeptidase [Pseudomonas chlororaphis]
AJGGILGI_00052	hypothetical protein	hypothetical protein Q091_05264 [Pseudomonas aeruginosa C52]
AJGGILGI_00053	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]
FFHAKPON_00002	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]
FFHAKPON_00003	hypothetical protein	NO HIT
FFHAKPON_00007	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]
FFHAKPON_00008	hypothetical protein	hypothetical protein PACL_0046 [Pseudomonas aeruginosa]
FFHAKPON_00010	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]
FFHAKPON_00011	hypothetical protein	hypothetical protein [Pseudomonas sp. HMSC05H02]
FFHAKPON_00012	hypothetical protein	DNA-binding protein [Pseudomonas aeruginosa]
FFHAKPON_00014	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]
FFHAKPON_00015	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]
FFHAKPON_00023	hypothetical protein	phage tail protein [Pseudomonas aeruginosa]
FFHAKPON_00024	hypothetical protein	phage tail protein [Pseudomonas aeruginosa]
FFHAKPON_00029	hypothetical protein	DUF2971 domain-containing protein [Pseudomonas aeruginosa]
FFHAKPON_00030	hypothetical protein	phage virion morphogenesis protein [Pseudomonas aeruginosa]
FFHAKPON_00043	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]

Table 3 - Gene annotation results from the prophage expansion region in Pa isolate 3 compared to isolate 1 from patient F.

Gene ID	Prokka identification	Pseudomonas database identification
LPHANECM_00014	hypothetical protein	integrase [Pseudomonas aeruginosa]
IMJGBJEL_00051	hypothetical protein	ImmA/IrrE family metallo-endopeptidase [Pseudomonas chlororaphis]
IMJGBJEL_00023	Nucleoid-associated protein YejK	nucleoid-associated protein ndpA [Pseudomonas aeruginosa]
IMJGBJEL_00053	hypothetical protein	hypothetical protein [Pseudomonas aeruginosa]

Table 4 - Gene annotation results from the prophage expansion region in Pa isolate 2 compared to isolate 1 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
OCLENNDD_00072	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00073	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00075	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00076	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00077	hypothetical protein	putative phage protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00078	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00079	hypothetical protein	siphovirus Gp157 family protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00080	hypothetical protein	LytTR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00081	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00082	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00083	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00084	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00085	hypothetical protein	class I SAM-dependent methyltransferase [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00086	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00087	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00088	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00089	hypothetical protein	phage protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00090	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00091	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00092	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00093	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00094	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00095	hypothetical protein	DUF2591 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00096	Putative defective protein IntQ	site-specific integrase [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00097	putative ABC transporter ATP-binding protein YbiT	ABC-F family ATPase [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00098	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00099	FMN-dependent NADH-azoreductase 1	FMN-dependent NADH-azoreductase [<i>Pseudomonas aeruginosa</i>]

OCLENNDD_00100	HTH-type transcriptional regulator DmlR	LysR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
NJNDLIIL_00007	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
NJNDLIIL_00009	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> phage vB_Pae_CF74b]
OCLENNDD_00039	hypothetical protein	hypothetical protein PputGB1_3424 [<i>Pseudomonas putida</i> GB-1]
NJNDLIIL_00046	hypothetical protein	hypothetical protein BPPAER656_00910 [<i>Pseudomonas</i> phage YMC11/02/R656]
NJNDLIIL_00055	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00045	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00068	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00069	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00070	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
OCLENNDD_00071	hypothetical protein	hypothetical protein PAK_P500124 [<i>Pseudomonas</i> phage PAK_P5]

Table 4 - Gene annotation results from the prophage expansion region in Pa isolate 3 compared to isolate 2 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
JALDOAGK_00068	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00069	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00070	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00071	hypothetical protein	hypothetical protein PAK_P500124 [<i>Pseudomonas</i> phage PAK_P5]
JALDOAGK_00072	hypothetical protein	hypothetical protein BPPAER656_00080 [<i>Pseudomonas</i> phage YMC11/02/R656]
JALDOAGK_00073	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00075	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> phage vB_Pae_CF23a]
JALDOAGK_00076	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00077	hypothetical protein	putative phage protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00078	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00079	hypothetical protein	siphovirus Gp157 family protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00080	hypothetical protein	LytTR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00081	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> virus D3]

JALDOAGK_00082	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00083	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00084	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00085	hypothetical protein	class I SAM-dependent methyltransferase [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00086	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00087	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00088	hypothetical protein	phage protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00089	hypothetical protein	phage protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00090	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00091	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00092	hypothetical protein	ead/Ea22-like family protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00093	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00094	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00095	hypothetical protein	DUF2591 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00096	Putative defective protein IntQ	site-specific integrase [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00097	putative ABC transporter ATP-binding protein YbiT	ABC-F family ATPase [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00098	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
NJNDLIL_00007	hypothetical protein	hypothetical protein T266_27315 [<i>Pseudomonas aeruginosa</i> VRFPA05]
JALDOAGK_00099	FMN-dependent NADH-azoreductase 1	FMN-dependent NADH-azoreductase [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00100	HTH-type transcriptional regulator DmiR	LysR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
NJNDLIL_00009	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
NJNDLIL_00046	hypothetical protein	hypothetical protein BPPAER656_00910 [<i>Pseudomonas</i> phage YMC11/02/R656]
NJNDLIL_00055	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
JALDOAGK_00039	hypothetical protein	hypothetical protein PputGB1_3424 [<i>Pseudomonas putida</i> GB-1]
JALDOAGK_00045	hypothetical protein	hypothetical protein BPPAER656_00910 [<i>Pseudomonas</i> phage YMC11/02/R656]

Table 5 - Gene annotation results from the prophage expansion region in Pa isolate 4 compared to isolate 3 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
JALDOAGK_00069	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00108	Branched-chain amino acid transport system 2 carrier protein	branched-chain amino acid transport system II carrier protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00109	Phosphoethanolamine transferase EptA	phosphoethanolamine--lipid A transferase [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00110	hypothetical protein	pyrroloquinoline quinone biosynthesis protein PqqF [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00111	hypothetical protein	porin [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00112	hypothetical protein	GfdT protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00113	hypothetical protein	hybrid sensor histidine kinase/response regulator [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00032	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00101	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00102	hypothetical protein	GrpB family protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00103	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	5-carboxymethyl-2-hydroxymuconate isomerase [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00104	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00105	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00106	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00107	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

Table 6 - Gene annotation results from the prophage expansion region in Pa isolate 7 compared to isolate 5 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
FLCCLKDO_00101	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00110	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00111	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00112	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

FLCCLKDO_00113	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00102	hypothetical protein	GrpB family protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00103	5-carboxymethyl-2-hydroxymuconate Delta-isomerase	5-carboxymethyl-2-hydroxymuconate isomerase [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00104	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00105	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00106	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00107	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00108	Branched-chain amino acid transport system 2 carrier protein	branched-chain amino acid transport system II carrier protein [<i>Pseudomonas aeruginosa</i>]
FLCCLKDO_00109	Phosphoethanolamine transferase EptA	phosphoethanolamine--lipid A transferase [<i>Pseudomonas aeruginosa</i>]

Table 7 - Gene annotation results from the prophage expansion region in Pa isolate 13 compared to isolate 11 from patient C.

Gene ID	Prokka identification	<i>Pseudomonas</i> database identification
KMGOMMIA_00070	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00071	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00072	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00073	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00075	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> phage vB_Pae_CF23a]
KMGOMMIA_00076	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00077	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00078	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00079	hypothetical protein	siphovirus Gp157 family protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00080	hypothetical protein	LytTR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00081	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> virus D3]
KMGOMMIA_00082	hypothetical protein	hypothetical protein PAJU2_gp44 [<i>Pseudomonas</i> phage PAJU2]
KMGOMMIA_00083	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00084	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

KMGOMMIA_00085	hypothetical protein	class I SAM-dependent methyltransferase [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00086	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00087	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00088	hypothetical protein	hypothetical protein Q049_06387 [<i>Pseudomonas aeruginosa</i> BWHPSA044]
KMGOMMIA_00089	hypothetical protein	phage protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00090	hypothetical protein	hypothetical protein BN405_2-10_Ab1_orf_38 [<i>Pseudomonas</i> phage vB_PaeM_C2-10_Ab1]
KMGOMMIA_00091	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00092	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00093	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00094	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00095	hypothetical protein	DUF2591 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00096	Putative defective protein IntQ	site-specific integrase [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00097	putative ABC transporter ATP-binding protein YbiT	ABC-F family ATPase [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00098	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00099	FMN-dependent NADH-azoreductase 1	FMN-dependent NADH-azoreductase [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00100	HTH-type transcriptional regulator DmlR	LysR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00034	hypothetical protein	HNH endonuclease [<i>Pseudomonas aeruginosa</i>]
AFPKENJN_00014	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AFPKENJN_00023	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AFPKENJN_00059	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> phage vB_Pae_CF74b]
AFPKENJN_00061	hypothetical protein	hypothetical protein BPPAER656_00580 [<i>Pseudomonas</i> phage YMC11/02/R656]
KMGOMMIA_00039	hypothetical protein	hypothetical protein PputGB1_3424 [<i>Pseudomonas putida</i> GB-1]
KMGOMMIA_00045	hypothetical protein	hypothetical protein BPPAER656_00910 [<i>Pseudomonas</i> phage YMC11/02/R656]
KMGOMMIA_00067	hypothetical protein	hypothetical protein CTT40_03668 [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00068	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
KMGOMMIA_00069	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

Table 8 - Gene annotation results from the prophage expansion region in Pa isolate 14 compared to isolate 13 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
BGLBLFFF_00045	hypothetical protein	hypothetical protein BPPAER656_00910 [<i>Pseudomonas</i> phage YMC11/02/R656]
BGLBLFFF_00067	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]

BGLBLFFF_00068	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00069	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00070	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00071	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00072	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00073	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00075	hypothetical protein	hypothetical protein [Pseudomonas phage vB_Pae_CF23a]
BGLBLFFF_00076	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00077	hypothetical protein	putative phage protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00078	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00079	hypothetical protein	siphovirus Gp157 family protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00080	hypothetical protein	LytTR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00081	hypothetical protein	hypothetical protein [Pseudomonas virus D3]
BGLBLFFF_00082	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00083	hypothetical protein	hypothetical protein PAJU2_gp44 [Pseudomonas phage PAJU2]
BGLBLFFF_00084	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00085	hypothetical protein	class I SAM-dependent methyltransferase [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00086	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00087	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00088	hypothetical protein	hypothetical protein X778_24940 [Pseudomonas aeruginosa VRFPA07]
BGLBLFFF_00089	hypothetical protein	phage protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00090	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00091	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00092	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00093	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00094	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00095	hypothetical protein	DUF2591 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00096	Putative defective protein IntQ	site-specific integrase [<i>Pseudomonas aeruginosa</i>]
AFPKENJN_00014	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00097	putative ABC transporter ATP-binding protein YbiT	ABC-F family ATPase [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00098	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00099	FMN-dependent NADH-azoreductase 1	FMN-dependent NADH-azoreductase [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00100	HTH-type transcriptional regulator DmlR	LysR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
AFPKENJN_00023	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
AFPKENJN_00059	hypothetical protein	hypothetical protein [Pseudomonas phage vB_Pae_CF74b]
AFPKENJN_00061	hypothetical protein	hypothetical protein [<i>Pseudomonas aeruginosa</i>]
BGLBLFFF_00034	hypothetical protein	HNH endonuclease [<i>Pseudomonas aeruginosa</i>]

BGLBLFFF_00039	hypothetical protein	hypothetical protein PputGB1_3424 [<i>Pseudomonas putida</i> GB-1]
----------------	----------------------	---

Table 9 - Gene annotation results from the prophage expansion region in Pa isolate 4 compared to isolate 3 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
LEJODOHM_00001	C4-dicarboxylate TRAP transporter large permease protein DctM	TRAP transporter large permease [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00063	Multiple antibiotic resistance protein MarR	MarR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00002	C4-dicarboxylate TRAP transporter small permease protein DctQ	TRAP transporter small permease [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00003	C4-dicarboxylate-binding periplasmic protein DctP	DctP family TRAP transporter solute-binding subunit [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00004	C4-dicarboxylate transport transcriptional regulatory protein DctD	sigma-54-dependent Fis family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00005	C4-dicarboxylate transport sensor protein DctB	sensor histidine kinase [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00057	hypothetical protein	hypothetical protein [<i>Pseudomonas</i> phage Dobby]
LEJODOHM_00060	Multidrug export protein EmrB	DHA2 family efflux MFS transporter permease subunit [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00061	Multidrug export protein EmrA	HlyD family efflux transporter periplasmic adaptor subunit [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00062	Solvent efflux pump outer membrane protein SrpC	efflux transporter outer membrane subunit [<i>Pseudomonas aeruginosa</i>]

Table 10 - Gene annotation results from the prophage expansion region in Pa isolate 6 compared to isolate 6 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
LEJODOHM_00001	C4-dicarboxylate TRAP transporter large permease protein DctM	TRAP transporter large permease [<i>Pseudomonas aeruginosa</i>]
KEJLMHHE_00005	Octopine transport system permease protein OccM	ABC transporter permease [<i>Pseudomonas aeruginosa</i>]
KEJLMHHE_00006	hypothetical protein	methyltransferase [<i>Pseudomonas aeruginosa</i>]

LEJODOHM_00002	C4-dicarboxylate TRAP transporter small permease protein DctQ	TRAP transporter small permease [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00003	C4-dicarboxylate-binding periplasmic protein DctP	DctP family TRAP transporter solute-binding subunit [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00004	C4-dicarboxylate transport transcriptional regulatory protein DctD	sigma-54-dependent Fis family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
LEJODOHM_00005	C4-dicarboxylate transport sensor protein DctB	sensor histidine kinase [<i>Pseudomonas aeruginosa</i>]
KEJLMHHE_00001	hypothetical protein	DUF502 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
KEJLMHHE_00002	Octopine permease ATP-binding protein P	ABC transporter ATP-binding protein [<i>Pseudomonas aeruginosa</i>]
KEJLMHHE_00003	ABC transporter arginine-binding protein 1	transporter substrate-binding domain-containing protein [<i>Pseudomonas aeruginosa</i>]
KEJLMHHE_00004	Histidine transport system permease protein HisQ	ABC transporter permease [<i>Pseudomonas aeruginosa</i>]

Table 11 - Gene annotation results from the prophage expansion region in Pa isolate 10 compared to isolate 9 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
FEPDKAND_00002	3-oxoacyl-[acyl-carrier-protein] reductase FabG	SDR family oxidoreductase [<i>Pseudomonas aeruginosa</i>]
FIIOPBKG_00003	C4-dicarboxylate-binding periplasmic protein DctP	DctP family TRAP transporter solute-binding subunit [<i>Pseudomonas aeruginosa</i>]
FIIOPBKG_00004	C4-dicarboxylate transport transcriptional regulatory protein DctD	sigma-54-dependent Fis family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]
FIIOPBKG_00005	C4-dicarboxylate transport sensor protein DctB	sensor histidine kinase [<i>Pseudomonas aeruginosa</i>]
FEPDKAND_00003	hypothetical protein	DUF502 domain-containing protein [<i>Pseudomonas aeruginosa</i>]
FEPDKAND_00004	Octopine permease ATP-binding protein P	ABC transporter ATP-binding protein [<i>Pseudomonas aeruginosa</i>]
FEPDKAND_00005	ABC transporter arginine-binding protein 1	transporter substrate-binding domain-containing protein [<i>Pseudomonas aeruginosa</i>]
FEPDKAND_00006	Histidine transport system permease protein HisQ	ABC transporter permease [<i>Pseudomonas aeruginosa</i>]
FEPDKAND_00007	Octopine transport system permease protein OccM	ABC transporter permease [<i>Pseudomonas aeruginosa</i>]
FEPDKAND_00008	hypothetical protein	methyltransferase [<i>Pseudomonas aeruginosa</i>]
FIIOPBKG_00001	C4-dicarboxylate TRAP transporter large permease protein DctM	TRAP transporter large permease [<i>Pseudomonas aeruginosa</i>]
FIIOPBKG_00002	C4-dicarboxylate TRAP transporter small permease protein DctQ	TRAP transporter small permease [<i>Pseudomonas aeruginosa</i>]

Table 12 - Gene annotation results from the prophage expansion region in Pa isolate 7 compared to isolate 5 from patient C.

Gene ID	Prokka identification	Pseudomonas database identification
CNDPFFKM_00022	Formate dehydrogenase nitrate-inducible major subunit	sulfate ABC transporter substrate-binding protein [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00001	hypothetical protein	DNA-binding protein [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00015	hypothetical protein	LuxR family transcriptional regulator [<i>Pseudomonas aeruginosa</i>]

CNDPFFKM_00016	Selenocysteine-specific elongation factor	selenocysteine-specific elongation factor [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00017	L-seryl-tRNA(Sec) selenium transferase	L-seryl-tRNA(Sec) selenium transferase [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00018	Protein FdhE	formate dehydrogenase accessory protein FdhE [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00019	Formate dehydrogenase cytochrome b556(fdo) subunit	formate dehydrogenase subunit gamma [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00020	Formate dehydrogenase-O iron-sulfur subunit	formate dehydrogenase subunit beta [<i>Pseudomonas aeruginosa</i>]
CNDPFFKM_00021	Formate dehydrogenase-O major subunit	formate dehydrogenase-N subunit alpha [<i>Pseudomonas aeruginosa</i>]