INVESTIGATING THE LIFESTYLE OF *PSEUDOMONAS AERUGINOSA* USING CHEMICAL GENOMICS

By

HANNAH DOHERTY

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

DOCTOR OF PHILOSOPHY

College of Life and Environmental Sciences

School of Biosciences

University of Birmingham

September 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

ABSTRACT

Chemical genomics is a powerful tool for functionally linking genotype to phenotype on a genome wide level, elucidating unknown gene function, potential drug targets and antimicrobial resistance determinants. Antimicrobial resistance is threatening the treatment of *Pseudomonas aeruginosa* infections. Therefore, research is required to evaluate the lifestyle of this bacterium and develop new therapies for treatment. This study utilised chemical genomics to evaluate and compare two *P. aeruginosa* strains, PAO1 and PA14. Due to the limited analysis software within the field, this study produced ChemGAPP, a comprehensive software for the analysis of chemical genomic screening data. Tested against real-world data, ChemGAPP showed success in revealing biologically relevant phenotypes. Following testing, ChemGAPP was employed on the PAO1 and PA14 chemical genomic screening data. Analysis revealed clear differences in the phenotypic responses of PAO1 and PA14, despite their genetic similarity. One major phenotype present in both screens was biotin mediated resistance against the diabetes drug metformin. A TraDIS library within a biotin deficient PA14 mutant was produced to reveal insights into the function of metformin. Genes highlighted within the TraDIS screen suggest that metformin disrupts energy production. This led to the hypothesis that biotin-dependent carboxylases are vital for compensation via alternative forms of energy production. Crucially, further research into the connection between metformin and biotin could lead to the development of a new therapy for *P. aeruginosa* infections. Specifically, a combination therapy of metformin with a biotin synthesis inhibitor.

ACKNOWLEDGMENTS

Manuel Banzhaf, thank you for being an amazing supervisor. The support you have provided me, not just for my project, but also for my career has been invaluable. I appreciate you giving me the freedom to truly explore my project and make it my own. Thank you for setting me up with collaborations that not only improved my knowledge but allowed me to travel and experience science in different environments.

Danesh Moradigaravand, thank you for introducing me to coding and the world of bioinformatics. I appreciate you giving up your time week after week to help me develop a skill that has not only improved my capabilities as a scientist but brings me great joy. Ava Jordan, thank you for being a brilliant student and a wonderful person to work with. I appreciate not just your dedication to the project, your hard work, and your insightful ideas, but the many laughs and long philosophical discussions.

Sotiris Kamarinos, thank you for all your hard work within this project. Thank you for spending months helping me spread out and scrape up TraDIS plates.

Jack Bryant, thank you for all your help, support, and advice. Thank you for being an endless source of knowledge and teaching me so much, including how to fix a bike and engrave labware.

Thank you to the Banzhaf Lab group. It has been fantastic being part of such a fun and kind group. It has been a pleasure to work and spend time with you all.

Thank you to my amazing friends and family. Amelia, Katy, Will, Ben, Charles, and Anna, thank you for cheering me along every step of the way, and for all the fun times which got me through the stressful parts of doing a PhD. Mum, thank you for always believing in me. Thank you for being the best role model and showing me what it means to be strong and hardworking.

Last but not least, Oscar, your endless support means the world to me. Thank you for always being there, whether to celebrate the successes or to make me smile when things were hard. Jag älskar dig.

LIST OF FIGURES

Figure 1: A schematic showing the process of biofilm formation, created with BioRender.com (Based on Thi et al., 2020) 11

Figure 2: Diagram depicting the Las and Rhl QS circuit (Reproduced based on Miller and Basler, 2001). 15

Figure 3: Diagram depicting different antibiotic resistance mechanisms in *Pseudomonas aeruginosa* 18

Figure 4: Diagram depicting the methodology of the Z-score test. 29

Figure 5: Diagram depicting the methodology of the Mann-Whitney plate level test. 30

Figure 6: Diagram depicting the methodology of the condition level tests. 32

Figure 7: The data analysis pipelines of the three ChemGAPP sub-packages. 73

Figure 8: Density plots displaying the distribution of pre- and post- normalised colony sizes for three randomly selected condition plates. 74

Figure 9: Plate matrices depicting the colony sizes for replicate plates of the 20 °C cold shock (20C) condition. 77

Figure 10: Density plot of condition Rif2 (Rifampicin 2 µg/ml), Plate 6, Batch1 colony sizes for replicate plates A, B and C. 79

Figure 11: Joint density and scatter plots of the difference in S-scores between A) the non-curated dataset and the final bootstrapped dataset for each mutant-condition pair

and B) the curated dataset the mean bootstrapped dataset for each mutant-condition pair. 81

Figure 12: Clustered heatmaps displaying the S-scores for various single gene knockout mutants in different conditions 85

Figure 13: ChemGAPP Small provides the option of two different output plot types. 87

Figure 14: Genetic interaction plots produced by ChemGAPP GI 89

Figure 15: Venn diagram showing the number of genes unique to PAO1 and PA14 and the number of conserved genes. 94

Figure 16: Overview of the folate biosynthesis pathway and the targets of sulfamethoxazole and trimethoprim. 98

Figure 17: Stringent response within *P. aeruginosa*.. 100

Figure 18: Pie chart displaying the classification of screening conditions for the PAO1 (left) and PA14 (right) screens. 102

Figure 19: Heatmap-based scatterplots comparing replicate colony sizes. 104

Figure 20: Number of strains with a phenotype at different false discovery rates (blue). 106

Figure 21: ROC curve of benchmark categories for PAO1 and PA14. 108

Figure 22: Distribution of Pearson correlation coefficient r values for the KEGG categories with the highest percentage of gene pairs with r values ≥ 0.4 for PA14. 109

Figure 23: Functions of MSR genes for PAO1 and PA14. 112

Figure 24: t-SNE plot of PAO1 and PA14.. 113

Figure 25: Density plot of Pearson r values for ortholog phenotypic profiles between PAO1 and PA14..... 115

Figure 26: Count of PseudoCAP functions assigned to orthologs with a Pearson r value of 0.4 or higher. 116

Figure 27: Heatmap-based scatterplots comparing the S-scores of ortholog genes across PAO1 and PA14 curated datasets. 118

Figure 28: Pie charts displaying A) the KEGG pathway of anticorrelated ortholog pairs and B) the condition types in which anticorrelated ortholog pairs were elucidated.. 120

Figure 29: Heatmap displaying S-scores and gene clustering for sulfonamide drugs within the PAO1 and PA14 chemical genomic screens. 121

Figure 30: Swarm plots of PAO1 ΔPaSAH (PA0431) and PA14 ΔPaSAH (PA14_05600) fitness ratios compared to their respective parent strains in presence of sulfonamides and trimethoprim. 122

Figure 31: Network analysis of phenotypic profiles with Pearson correlation coefficients (r) greater than 0.4 when compared to PA14_05600..... 124

Figure 32: Spotting plate assay of PA14 and PA14 ΔPaSAH in presence of sulfonamide drugs with and without adenine supplementation. 125

Figure 33: Clustered heatmap of S-scores for biotin synthesis mutants in presence of metformin. 143

Figure 34: Biotin chemical structure. 144

Figure 35: Diagram showing how biotin functions in complex with biotin-dependent carboxylases. 145

Figure 36: The first stage of biotin synthesis producing pimeloyl-ACP. 146

Figure 37: The second stage of biotin synthesis 147

Figure 38: A schematic showing the involvement of GCC and MCC in the acyclic terpene and leucine metabolic pathways. 150

Figure 39: Chemical structure of metformin. 152

Figure 40: BBP operon structures of A) Pseudomonadaceae: *Azomonas*, *Permianibacter*, *Pseudomonas*, *Oblitimonas,* and *Thiopseudomonas* and B) Enterobacteriaceae: *Dickeya*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Vibrio* and *Yersinia* 156

Figure 41: Spotting plates of MOPS minimal media agar to test the effect of metformin on the growth of biotin synthesis deficient PA14 mutants. 157

Figure 42: Spotting plate assay on MOPS minimal media agar to test the effect of metformin on the growth of biotin synthesis deficient *Acinetobacter baumannii* AB5075 mutants. 160

Figure 43: S-scores for biotin-dependent carboxylases in the presence of metformin. 161

Figure 44: MIC assays for PA14 Δ*bioB-H* in MHB with biotin-dependent carboxylase products supplemented..... 163

Figure 45: Transposon insertion maps for each concentration of metformin in relation to the genome sequence of PA14..... 167

Figure 46: Multiple genes were no longer essential in the presence of metformin. . 170

Figure 47: Multiple genes showed increased numbers of transposon insertions. ... 173

Figure 48: Multiple genes showed decreased numbers of transposon insertions across all metformin containing conditions..... 175

Figure 49: Seven genes showed decreased numbers of transposon insertions within the 1x and 2x MIC conditions..... 177

LIST OF TABLES

Table 1: Table displaying the curation thresholds used for each dataset. 39

Table 2: Bacterial strains used within the study. 44

Table 3: Primers used for fragment and vector amplification for production of PA0431 and PA14_05600 knockout vectors. 46

Table 4: PCR reagents and cycle conditions for knockout vector up and down fragments. 48

Table 5: PCR reagents and cycle conditions for linearisation of pEXTk. 48

Table 6: PCR reagents and cycle conditions for Colony check of double recombinants. 51

Table 7: Drug concentration used within *Pa*SAH spot plate assay. 52

Table 8: Primers used for sequencing preparation of TraDIS library. 56

Table 9: PCR cycling conditions for amplification of transposon junction. 63

Table 10: Table showing the MIC values for PA14 and PA14 Δ*bioB-H*. 158

Table 11: Insertion statistics for each sequencing replicate.. 165

LIST OF PAPERS

Chapter 3 is based on the following paper. For this paper, I produced the software, performed all analyses, created all figures, and wrote the paper :

- **Doherty, H.M.**, Kritikos, G., Galardini, M., et al. (2023) ChemGAPP: a tool for chemical genomics analysis and phenotypic profiling. *Bioinformatics*, 39 (4): btad171. doi:10.1093/bioinformatics/btad171.

Further papers published during this PhD include:

- Boelter, G., Bryant, J.A., **Doherty, H.**, et al. (2022) The lipoprotein DolP affects cell separation in Escherichia coli, but not as an upstream regulator of NlpD. *Microbiology (Reading, England)*, 168 (5). doi:10.1099/mic.0.001197.

- Bryant, J.A., Staunton, K.A., **Doherty, H.M**., et al. (2023) *Bam complex associated proteins in Escherichia coli are functionally linked to peptidoglycan biosynthesis, membrane fluidity and DNA replication*. p. 2023.07.05.547807. doi:10.1101/2023.07.05.547807.

ABBREVIATIONS

2-MCC	2-methylcitrate cycle
AACs	Aminoglycoside acetyltransferases
ACC	Acetyl-coa carboxylase
AI	Auto inducer
AMR	Antimicrobial resistance
ANTs	Aminoglycoside nucleotidyltransferases
APHs	Aminoglycoside phosphotransferases
AUC	Area under the curve
BBP	Biotin biosynthesis pathway
BC	Biotin carboxylase
BCCP	Biotin-carboxyl carrier protein
BHI	Brain heart infusion
CAUTIs	Catheter-associated urinary tract infections
CFTR	Cystic fibrosis transmembrane conductance regulator
CrpP	Ciprofloxacin resistance protein, plasmid encoded
CT	Carboxyltransferase
CVCs	Colonised central venous catheters
DHAP	Dihydroxyacetone phosphate
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase
DHP	Dihydropteroate
DHP-PPi	Dihydropteroate pyrophosphate
DHPS	Dihydropteroate synthase
eDNA	Extracellular DNA
ENaC.	Epithelial sodium absorption channels
EPS	Exopolysaccharides
FASII	Type II fatty acid synthesis
FDRs	False discovery rates
G3P	Glycerol-3-phosphate
GCC	Geranyl-coa carboxylase
GCV	Glycine cleavage
GSH	Glutathione
HAIs	Hospital acquired infections
HGT	Horizontal gene transfer
HHQ	2-heptyl-4(1H)-quinolone
HSL	Homoserine lactones
IF3	Initiation factor 3
IMP	Inner membrane protein
IQR	Interquartile range

Iqs	Novel integrated QS system
IQS	2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde
LBA	Luria Bertani agar
LBB	Luria Bertani broth
LogFC	Log2 fold change
MAE	Mean absolute error
MBL	Metallo-b-lactamase
MCC	3-methyl-crotonyl-coa carboxylase
mGDH	Mitochondrial glycerophosphate dehydrogenase
MGEs	Mobile genetic elements
MHB	Mueller Hinton broth
MSR	Multiple stress responsive
NLE	No longer essential
OM	Outer membrane
OMP	Outer membrane proteins
PABA	P-aminobenzoic acid
PBP	Penicillin-binding protein
PC	Pyruvate carboxylase
PCA	Principal component analysis
pGpp	Guanosine triphosphate
PLP	Pyridoxal 5'-phosphate
PMF	Proton motive force
PMM	Plate middle mean
ppGpp	Guanosine tetraphosphate
pppGpp	Guanosine pentaphosphate
PQS	2-heptylhydroxy-1H-quinolin-4-one
pqs	Pseudomonas quinolone signal molecule system
PRA	5-phospho-D-ribosylamine
PRPP	5-phosphoribosylpyrophosphate
RND	Resistance-nodulation-division
ROC	Receiver operating characteristic
ROS	Reactive oxygen species
RSH	RelA/SpoT homologue
rt	Room temperature
SAHs	Small alarmone hydrolases
SASs	Small alarmone synthases
SCV	Small colony variant
SPI	Synthetic physical interaction
SPRI	Solid-phase reversible immobilisation
std	Standard deviation
t-SNE	t-distributed stochastic neighbour embedding

T2SS	Type 2 secretion system
T3SS	Type 3 secretion system
T6SS	Type 6 secretion system
TAE	Tris-acetate-EDTA buffer
THF	Tetrahydrofolate
TraDIS	Transposon directed insertion sequencing
UC	Urea carboxylase
YCC	Acyl-coa carboxylase

TABLE OF CONTENTS

1 GENERAL INTRODUCTION 1

1.1.1 *Pseudomonas aeruginosa* 2

1.1.2 *Pseudomonas aeruginosa* infection pathology 2

1.1.2.1 Chronic *Pseudomonas aeruginosa* infection within cystic fibrosis patients 4

1.1.3 *Pseudomonas aeruginosa* virulence factors 7

1.1.3.1 Biofilm formation 7

1.1.3.2 Quorum Sensing 12

1.1.4 Antimicrobial resistance in *Pseudomonas aeruginosa* 16

1.1.4.1 Intrinsic resistance mechanisms 18

1.1.4.2 Extrinsic resistance mechanism 22

1.1.5 High-throughput chemical genomics and phenotypic profiling 23

1.1.6 A brief history of chemical genomics 24

2 MATERIALS AND METHODS 26

2.1 Production and testing of ChemGAPP pipelines 27

2.1.1 ChemGAPP Big pipeline 27

2.1.1.1 ChemGAPP Big: conversion of Iris files 27

2.1.1.2 ChemGAPP Big: plate normalisation 28

2.1.1.3 ChemGAPP Big: quality control analyses 28

2.1.1.3.1 Quality control analyses: Z-score test 28

2.1.1.3.2 Quality control analyses: Mann-Whitney test 30

2.1.1.3.3 Quality control analyses: condition level tests 31

2.1.1.4 ChemGAPP Big: threshold selection and data removal 32

2.1.1.5 ChemGAPP Big: fitness score calculation 33

2.1.1.6 ChemGAPP Big: benchmarking the curated dataset 35

2.1.1.7 ChemGAPP Big: production of the bootstrapped dataset 35

2.1.2 ChemGAPP Small pipeline 36
2.1.3 ChemGAPP GI pipeline 37
2.2 Comparison of *P. aeruginosa* chemical genomic datasets 38
2.2.1 ChemGAPP Big analysis 38
2.2.2 Calculating replicate reproducibility for PA datasets. 39
2.2.3 False discovery rate analysis of S-score significance 40
2.2.4 Benchmarking of PAO1 and PA14 Datasets 40
2.2.5 PseudoCAP analysis of MSR genes 42
2.2.6 Ortholog analysis 42
2.2.7 t-SNE analysis of PAO1 and PA14 S-scores. 43
2.3 Bacterial strains and storage 44
2.4 Production of *Pa*SAH knockouts in PAO1 and PA14 45
2.4.1 Primer used within Section 2.4 45
2.4.2 Production of knockout vectors 47
2.4.3 Transformation of vectors into DH5α for storage. 49
2.4.3.1 Production of chemically competent DH5α 49
2.4.3.2 Transformation of DH5α with knockout vectors 49
2.4.4 Transformation of PAO1 and PA14 50
2.5 *Pa*SAH spot plate assay 51
2.6 *Pa*SAH small-scale chemical genomic screen 52
2.7 Structure analysis of biotin synthesis operon 53
2.8 Biotin synthesis mutant spotting plates 53
2.9 Broth microdilution MIC analysis 54
2.10 Suppressor mutant screen of PA14 Δ*bioB-H* 55
2.11 Production of transposon library within PA14 Δ*bioB-H* 55

2.11.1 Primers used in section 2.11 55
2.11.2 Preparation of TnTm transposome 56
2.11.3 Production of electrocompetent cells 57
2.11.4 Transformation of electrocompetent PA14 Δ*bioB-H* and library construction 57
2.11.5 Production and validation of SPRI beads 58
2.11.6 DNA gel electrophoresis 59
2.11.7 TraDIS library screening and DNA extraction 60
2.11.8 Quantifying DNA using the Qubit® Fluorometer 60
2.11.9 Shearing genomic DNA 61
2.11.10 Library preparation of metformin exposed TraDIS samples 61
2.11.10.1 End repair and illumina adaptor ligation 61
2.11.10.2 Size selection of adaptor-ligated DNA 62
2.11.10.3 PCR Amplification of transposon junction 62
2.11.10.4 SPRI purification of PCR amplified fragments 63
2.11.10.5 PCR based sequencing index addition 63
2.11.11 TraDIS data analysis 64

3 CHEMGAPP: A TOOL FOR CHEMICAL GENOMIC ANALYSIS AND PHENOTYPIC PROFILING 66

3.1 Introduction 67

3.1.1 The need for a comprehensive data analysis software for chemical genomics. 67
3.1.1.1 Data analysis within high-throughput screens 68
3.1.1.2 Data analysis within small-scale screens 70
3.1.1.3 Data analysis within genetic interaction screens 71

3.2 Results 72

3.2.1 ChemGAPP Pipelines 72
3.2.2 ChemGAPP Big successfully normalises condition plates 73
3.2.3 ChemGAPP Big highlights common errors within chemical genomic screens 75

3.2.3.1 The Z-score normality test can distinguish unequal pinning. 75

3.2.3.2 Mann-Whitney quality control analysis 77

3.2.4 ChemGAPP Big produces robust fitness values. 79

3.2.5 ChemGAPP Big produces reliable and accurate representations of fitness 83

3.2.6 ChemGAPP Small is better suited to small-scale screens 86

3.2.7 ChemGAPP Small produces informative fitness ratios 86

3.2.8 ChemGAPP GI successfully reveals epistasis types 88

3.3 Discussion 90

4 CHEMICAL GENOMIC ANALYSIS AND COMPARISON OF TWO PSEUDOMONAS AERUGINOSA STRAINS. 92

4.1 Introduction 93

4.1.1 Genotypic similarity between PAO1 and PA14 93

4.1.2 Loss of *Pa*SAH induces sulfonamide sensitivity within PA14 but not PAO1. 96

4.1.2.1 Sulfamethoxazole and trimethoprim mechanism of action. 96

4.1.2.2 *Pa*SAH and the stringent response. 99

4.2 Results 101

4.2.1 Chemical genetic screening of two *P. aeruginosa* strains. 101

4.2.2 Chemical genomics produces reliable datasets for the acquisition of phenotype data. 103

4.2.3 Comparison of PAO1 and PA14. 110

4.2.3.1 Comparison of multiple stress responsive genes across PAO1 and PA14. 110

4.2.3.2 PAO1 and PA14 phenotypic profiles can be distinguished by t-SNE. 112

4.2.3.3 PAO1 and PA14 ortholog correlation. 114

4.2.3.4 ∆*Pa*SAH results in Sulfamethoxazole sensitivity within PA14 but not PAO1 120

4.3 Discussion 125

5 BIOTIN MEDIATED METFORMIN RESISTANCE. 141

5.1 Introduction 142

5.1.1 Biotin mediates metformin resistance in *Pseudomonas aeruginosa* 142

5.1.2 Biotin 143

5.1.2.1 Biotin Synthesis 145

5.1.2.2 Biotin-dependent carboxylases in *P. aeruginosa* 147

5.1.3 Metformin 151

5.1.3.1 Metformin as a Quorum Sensing Inhibitor 154

5.2 Results 155

5.2.1 The BBP operon structure differs within *Pseudomonas* 155

5.2.2 Validating biotin mediated metformin resistance in *Pseudomonas aeruginosa* 156

5.2.3 MIC of metformin against *Pseudomonas aeruginosa* 157

5.2.4 Biotin synthesis does not provide metformin resistance in *Acinetobacter baumannii* 158

5.2.5 Loss of individual biotin-dependent carboxylases does not result in metformin sensitivity 160

5.2.6 Supplementation with oxaloacetate reduced sensitivity to metformin. 162

5.2.7 PA14 Δ*bioB-H* did not gain spontaneous suppressor mutations. 163

5.2.8 Constructing a TraDIS library in PA14 Δ*bioB-H* 164

5.2.9 Multiple genes became non-essential in the presence of metformin. 168

5.2.10 Genes with increased numbers of transposon insertions. 171

5.2.11 Genes with decreased numbers of transposon insertions. 174

5.3 Discussion 178

6 FINAL DISCUSSION 189

6.1 General summary 190

6.2 Future work 192

6.3 Final remarks 193

CHAPTER 1

GENERAL INTRODUCTION

1.1.1 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa, a bacterium known for its versatility and adaptability, is able to survive in many different niches. Its vast pan-genome of between 10,000 – 40,000 genes allows it to adapt quickly and effectively (Parkins et al., 2018; Silby et al., 2011). Encoded within its vast genome is the ability to utilise multiple carbon sources, as well as operate as a facultative anaerobe, utilising nitrogen as its terminal electron acceptor. Further to this, *P. aeruginosa* can form biofilms, where communities of bacteria attach themselves to a surface and excrete extracellular matrices to help protect against perturbations, such as antimicrobials or detergents (Lee and Yoon, 2017). Unfortunately, this genetic flexibility means that *P. aeruginosa* is also able to inhabit not just soil and water, but problematic environments, such as on hospital equipment or within the thickened mucus of a cystic fibrosis patients' lung (Pendleton et al., 2013). The colonisation of these environments sadly leads to the development of infections and often death.

1.1.2 *Pseudomonas aeruginosa* infection pathology

P. aeruginosa is an opportunistic pathogen responsible for the development of acute nosocomial infections, and chronic infections within cystic fibrosis patients. Unfortunately, due to its opportunistic nature, infection often occurs in those with weakened immune systems, such as the elderly, people suffering from burns and open wounds, patients on immunosuppressants following organ transplantation or surgery, and cancer sufferers (Gellatly and Hancock, 2013).This, combined with its prevalence

on hospital equipment, is hugely problematic as these are the people likely to be in contact with the colonised hospital equipment. Since *P. aeruginosa* is so prevalent within the hospital reservoir, it is unsurprising that it is one of the most common nosocomial pathogens (Weiner-Lastinger et al., 2020). Within the united states *P. aeruginosa* accounts for 8% of hospital acquired infections (HAI's) (Weiner-Lastinger et al., 2020).

P. aeruginosa can be introduced via multiple routes within the hospital environment, including on catheters, ventilation machinery such as endotracheal tubes, and central venous catheters (Gellatly and Hancock, 2013; Moradali et al., 2017). Insertion of a colonised catheter can lead to the development of catheter-associated urinary tract infections (CAUTIs). There are thought to be over 1 million CAUTI cases and 7500 deaths annually, of which *P. aeruginosa* has been implicated as the causative agent in 12% (Cole et al., 2014; Mittal et al., 2009). With ventilation associated infections, intubation with an endotracheal tube can lead to epithelial damage within the lung. This primes the tube for the formation of biofilms and the development of acute lung infections with a mortality rate of 30% (Gellatly and Hancock, 2013). Colonised central venous catheters (CVCs) or 'central lines' are able to introduce infections into a patient's bloodstream (Gahlot et al., 2014; Guenezan et al., 2018). When being inserted, the endogenous skin flora of the patient or healthcare worker can contaminate the CVC. Therefore, once inserted into the vein these bacteria can enter the blood stream (Gahlot et al., 2014). Alternatively, the CVC can become colonised by bacteria within the body that have travelled from other sites of infection (Gahlot et al., 2014). Although not as prevalent as CAUTIs, CVC associated bloodstream infections still have a significant health and economic impact. These infections result

in 250,000 annual cases amounting to over $45,000 in the USA. Of these 250,000 cases, *P. aeruginosa* is indicated in 16% (Bullard and Dunn, 2001; Zimlichman et al., 2013).

1.1.2.1 Chronic *Pseudomonas aeruginosa* infection within cystic fibrosis patients

Although nosocomial infections account for large numbers of infections and deaths, *P. aeruginosa* infections have a far greater impact on cystic fibrosis sufferers. The leading cause of death in people with cystic fibrosis is cardiorespiratory disease, which accounts for the mortality of 70% of cystic fibrosis patients (Hauser et al., 2011). Cardiorespiratory disease is the result of chronic infections of the lung, and 80% of cystic fibrosis patients are chronically colonised with *P. aeruginosa* by the age of 20 (Bhagirath et al., 2016).

Cystic fibrosis is a condition categorised by various mutation types within the cystic fibrosis transmembrane conductance regulator (CFTR) gene. The disease state and severity is entirely dependent on the kind of mutation which effects the CFTR protein. Nonsense frameshift mutations account for the most severe forms of CF, since these result in the complete loss of the functional CFTR protein (Turcios, 2020). In essence, the varying degree to which the mutation affects the functionality of the CFTR genes correlates to disease severity. The normal function of the CFTR protein involves the transport of chloride and bicarbonate anions across the epithelia of the airway and into the periciliary layer. The transport of these ions is imperative as it facilitates the movement of water into the periciliary layer, where it is able to properly hydrate the

airway surface. Hydration of the surface is vital for effective clearance of the mucus layer by the mucociliary clearance apparatus into the trachea. Here the mucus is either swallowed or exhaled, facilitating the removal of the contained bacteria (Brugha and Davies, 2011).

The loss of anion transport within the CF lung, due to the impairment of CFTR, results in a build-up of ions within the epithelial cells. This is further increased by the retained function of the epithelial sodium absorption channels (ENaC). This increased ion concentration amplifies the osmotic gradient within the epithelial cells, drawing fluid away from the airway surface. This dehydration of the airway results in the build-up of the mucus layer, making it thick and problematic to clear (Turcios, 2020). This results in the inability to clear bacteria from the lung surface. Some evidence suggests this is further amplified by the inhibition of host antimicrobial defensins due to the change in salt concentrations. These defensins normally induce microbial killing by neutrophils. Therefore, inhibition further reduces the clearance of bacteria (Hauser et al., 2011). Additionally, impairment of CFTR channels stimulates the increased expression of asialylated glycolipids on the lung epithelial cell surface, a key receptor for *P. aeruginosa* binding and colonisation (Hauser et al., 2011).

It is the combination of these effects that mean the CF lung is easily colonised by *P. aeruginosa.* Repeat infections within cystic fibrosis patients can eventually result in respiratory failure (Bhagirath et al., 2016; Koch, 2002). During chronic infections, reactive oxygen species (ROS) are produced as a response to increased antibodies against the bacterium. This ROS production is a stimulus for *P. aeruginosa* to begin producing alginate, a major component in the formation of biofilms which protect

against antimicrobial action (Koch, 2002). Further to this, a genotypic and phenotypic overhaul is induced within chronic infection resulting in stimulation of the pathogenic state, making the infection more difficult to treat with antibiotics (Bhagirath et al., 2016; Hauser et al., 2011). This state is characterised by the upregulation of antimicrobial resistance genes and biofilm formation, coupled with the slowing of growth and metabolism (Bhagirath et al., 2016). As such, once chronically infected with *P. aeruginosa* antibiotic treatment is often unable to eradicate the infection. Despite this, antibiotic treatment has been shown to reduce morbidity, regardless of eradication (Hauser et al., 2011).

During this chronic stage of infection lung damage occurs due to a continuous cycle of *P. aeruginosa* induced inflammation, which is further amplified by the pro-inflammatory nature of the CF lung. *P. aeruginosa* induces inflammation via its antigens and the release of proteolytic enzymes, which concurrently stimulate the activation and release of host pro-inflammatory factors. These include TNF-α, IL-1, IL-6, IL-8, complement chemo-attractants, leukotriene B4, and the reduction of anti-inflammatory cytokine IL-10 (Hauser et al., 2011). In response to these pro-inflammatory factors, neutrophils are recruited. However, instead of killing *P. aeruginosa* as intended, they hinder the innate immune response and damage lung tissue. In this instance the neutrophils are stimulated to release proteases, such as elastases and collagenases. These proteases are responsible for the remodelling of lung tissue via the breakdown of the bronchi and bronchiole structural framework (Hauser et al., 2011; Ulrich et al., 2010). Another failing of the neutrophil response is the expelling of large quantities of DNA, typically this would ensnare bacteria for capture and killing. However, within the cystic fibrosis lung, this DNA simply contributes to the thickness and viscosity of the mucus layer,

further reducing effective microbial clearance (Hauser et al., 2011). With continuous failing of the immune response to clear bacterial infection, persistent inflammation leads to a cycle of increased bacterial load and increased lung damage. This can eventually progress to severe cardiorespiratory disease resulting in the need for lung transplantation or even leading to death (Hauser et al., 2011; Meachery et al., 2008).

Clearly, in all their forms *P. aeruginosa* infections are hugely problematic, with significant health, economic and global impacts. *P. aeruginosa's* ability to cause these serious and broad ranging infections is due to its large arsenal of virulence factors that allow it to be a formidable and adaptable pathogen. Therefore, understanding its lifestyle, virulence and resistance factors is imperative.

1.1.3 *Pseudomonas aeruginosa* virulence factors

1.1.3.1 Biofilm formation

Arguably one of the most well-known virulence factors of *P. aeruginosa* is its ability to form robust biofilms. Biofilms comprise of bacteria aggregated in an organised fashion and encased in exopolysaccharides (EPS) which are secreted in order to protect the bacterial community (Rasamiravaka et al., 2015; Thi et al., 2020). Biofilm formation provides a powerful protection against a variety of stresses, including antimicrobial exposure, host immune responses, and temperature changes (Thi et al., 2020). This is due to a variety of reasons, including physical and chemical protection by components of the biofilm, as well as the down regulation of metabolic activity which

antibiotics rely on for their activity (Fleming et al., 2022). This makes *P. aeruginosa* biofilms very difficult to eradicate.

Biofilms are comprised of various extracellular polymetric substances; exopolysaccharides, extracellular DNA (eDNA), proteins and lipids (Thi et al., 2020). These substances formulate the biofilm matrix and contribute up to 90% of the total organic carbon within the structure (Donlan, 2002). Within *P. aeruginosa* three exopolysaccharides have been implicated with being important for formation and structural integrity; alginate, Pel and Psl (Ghafoor et al., 2011). Alginate, although suggested to not be hugely relevant in non-mucoidal strains of *P. aeruginosa*, is thought to be the major component of biofilms within mucoidal strain such as those that colonise CF lungs (Høiby et al., 2010; Thi et al., 2020). The alginate-producing mucoid phenotype is more successful within the CF environment because alginate prevents phagocytosis and clearance since it acts as a scavenger of oxygen-radicals (Høiby et al., 2010; Stapper et al., 2004). It has even been shown that the gene mutation responsible for the mucoid phenotype (mucA) can be induced within the CF lung, due to exposure to ROS species released by polymorphonuclear leukocytes (Høiby et al., 2010). These leukocytes also become important in disease progression as alginate triggers them to release further ROS species. This increases inflammation within the CF lung which, as previously described, leads to the break-down of lung tissue (Jurado-Martín et al., 2021). Alginate also plays a role in antibiotic resistance, since it is believed to bind aminoglycoside antibiotics, preventing them from entering the biofilm (Jurado-Martín et al., 2021).

Pel and Psl are the primary exopolysaccharides within non-mucoidal strains of *P. aeruginosa*. The production of the Pel and Psl polysaccharides is encoded by individual loci made up of 7 and 12 genes, respectively, for their synthesis and export (Colvin et al., 2012). The primary polysaccharide between the two is variable amongst strains of *P. aeruginosa* and their distinctive roles similarly vary with strain (Colvin et al., 2012). For example, PAO1 has been shown to favour Psl whereas Pel is much more important in biofilm maturation within PA14 (Colvin et al., 2012). Although being responsible for different roles within the matrix, they do show a level of redundancy. A study by Colvin et al. showed that when one locus is deleted, mutations tend to appear within the other that lead to compensatory upregulation (Colvin et al., 2012). This is important as it provides robustness in the ability to form biofilms in the event of a deleterious mutation.

The function of Pel was investigated by the development of a tool to visualise Pel within the biofilm. This discovered that Pel is responsible for the cross-linking of eDNA within the biofilm stalk (Jennings et al., 2015). This is due to the positive charge on Pel and the negative charge of eDNA. This means Pel is directly responsible for formation of architectural structures within the biofilm, improving integrity and biofilm robustness. The positive charge of Pel also has further indications for the development of infections. Since many relevant host polymers are also negatively charged they could similarly form ionic bonds with Pel, improving adhesion of biofilms to tissues and host surfaces (Jennings et al., 2015).

The role of Psl has been shown to be social in nature, it is critical for the formation of cell-cell interactions and has intercellular signalling properties. Being located on the

bacterial surface, Psl is able to bind to Psl molecules on other *P. aeruginosa* cells (Ma et al., 2009). This is a vital step for the maturation of the biofilm. Psl is also required within early-stage biofilm formation, where it is responsible for association with the colonisation surface. This is critical as it combines with the cell-cell interaction role to bring cells into the biofilm and keep them bound to the surface such that the biofilm can initiate (Ma et al., 2009).

The formation and dispersal of biofilms occurs in six stages: initial adhesion, early attachment, young biofilm, mature biofilm, dispersal, and return to planktonic phenotype. The process is cyclical with the dispersed cells becoming enhanced for attachment and virulence, leading to the progression of further biofilms, see Figure 1. In the first stage the most important factor is the binding and colonisation of the surface. Here flagella and type IV pili facilitate adhesion to the abiotic surface. This adhesion is also tailored to the surface on which the bacterium aims to bind, with different surfaces eliciting a different proteomic response (Guilbaud et al., 2017). Thus, aiding *P. aeruginosa's* ability to colonise a vast number of surfaces and niches. At this stage in the biofilm forming process the attachment is still reversible. However, upon developing to stage two the binding becomes irreversible and the cells begin to divide (Thi et al., 2020). This division and proliferation of bacteria establishes the microcolonies, which are characteristic of stage three. Leading to development of a young biofilm which is developed in structure (Rasamiravaka et al., 2015; Thi et al., 2020). This young biofilm matures in the next stage, forming larger mushroom-shaped three-dimensional structures. In stages 3 and 4 cell-to-cell interactions increase, leading to the induction of the quorum sensing system and the expression of virulence factors (Thi et al., 2020). This aids the growth of the biofilm but also allows it to become

more robust with increased resistance to antibiotics (Thi et al., 2020). It is important for further infection that the biofilm is able to disperse and spread. Therefore, in stage 5 autolysis occurs within the centre of the biofilm forming a matrix cavity which can disrupt the biofilm enough to allow dispersal of bacterial cells. Since these dispersed cells are responsible for the formation of new biofilms, they must revert to the planktonic stage in stage 1. This reversion incurs a two-hour lag phase, however, once transitioned the dispersed cells are left with enhanced attachment ability and virulence (Thi et al., 2020).

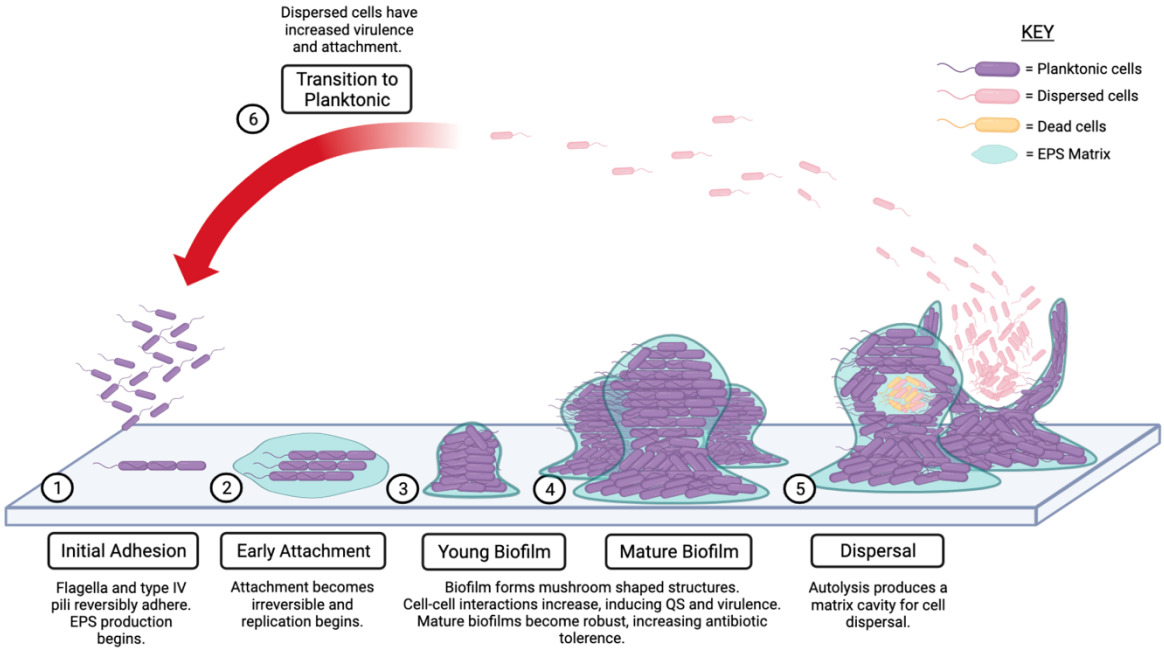


Figure 1: A schematic showing the process of biofilm formation, created with BioRender.com (Based on Thi et al., 2020)

Since *P. aeruginosa* cells form these multi-cell structures and rely on each other for effective virulence and antibiotic tolerance, it is vital that the cells can communicate with each other. The way that *P. aeruginosa* does this is via quorum sensing.

1.1.3.2 Quorum Sensing

P. aeruginosa encodes four systems for quorum sensing, the Las and Rhl homoserine lactones (HSL) based systems, the *Pseudomonas* quinolone signal molecule system (pqs) and the novel integrated QS system (Iqs) (De Kievit, 2009; Jurado-Martín et al., 2021).

Within each of these systems a transcriptional regulator is associated with an auto inducer (AI) and within the Las and Rhl systems these function within a circuit, see Figure 2 (Miller and Bassler, 2001). First, *P. aeruginosa* produces N-(3-oxododecanoyl)-homoserine lactone, the autoinducer synthesised by LasI. This autoinducer is released into the cell population, if cell density is high enough then the quantity of the autoinducer being released will reach a critical threshold. At this threshold it is able to bind to LasR, the transcriptional regulator of the Las system. AI binding to LasR then activates the expression of virulence genes via the binding of the LasR-AI complex to virulence factor gene promoter elements (Miller and Bassler, 2001). Genes which are upregulated include, *lasA* and *lasB*, which encode a protease and elastase, respectively; *toxA* which produces exotoxin A; and *aprA* which encodes alkaline phosphatase (Miller and Bassler, 2001). Exotoxin A is a particularly problematic virulence factor within *P. aeruginosa* infections. Exotoxin A is released by

the Type 2 secretion system (T2SS) and is internalised via host cell clatherin-coated pits following binding to CD91 and α2-macroglobuline receptors. Upon entering host cells, a conformational change occurs allowing exotoxin A to inhibit host protein synthesis and activate caspases involved in apoptosis. The combination of which results in necrosis at the site of colonisation (Jurado-Martín et al., 2021). LasA, LasB and AprA are all proteolytic enzymes and are secreted by the T2SS (LasA/B) and the T1SS (AprA). These enzymes aid in invasion and help initiate acute infections (Jurado-Martín et al., 2021). LasA and LasB, both degrade the host elastin, which is a vital extracellular matrix protein required for the elasticity and recoil of many tissues, including those within the cardiovascular and pulmonary systems (Wang et al., 2021). However, further to this, LasB cleaves other substrates, including epithelial tight-junctions and host proteins involved in the clearance of bacteria (Jurado-Martín et al., 2021). AprA similarly prevents the clearance of bacteria, however, here it degrades complement proteins and cytokines, such that phagocytosis can be avoided. Further to this, AprA function lies within the degradation of host endothelium by targeting fibronectin and laminin (Jurado-Martín et al., 2021). Since the LasR-AI complex also induces LasI expression the system forms a positive feedback loop, heavily increasing the expression of these virulence factors (Miller and Bassler, 2001).

The LasR-AI complex is then able to activate the expression of RhlR the transcriptional regulator of the Rhl system. Its autoinducer N-(butryl)-homoserine lactone, then binds to RhlR and this complex binds to further promoter elements to upregulate another set of virulence genes. These genes are *rpoS*, *rhlAB*, lecA, *rh1I* and pyocyanin synthesis genes. *RpoS* encodes the stationary phase RNA polymerase sigma factor σ^s, which is typically involved in the general stress response of bacteria. However, within *P.*

aeruginosa it appears to have a broader role, specifically in the regulation of several virulence factors and quorum-sensing controlled genes (Schuster et al., 2004). A study by Schuster et al. performed a microarray of the transcriptomes of an RpoS mutant compared to the wildtype. They found around 800 genes that were directly or indirectly regulated by RpoS, of which more than 40% had previously been shown to be quorum sensing controlled (Schuster et al., 2004). Clearly there is an interplay between the quorum sensing systems and RpoS, since both were shown to stimulate the other to varying degrees, perhaps enhancing the virulence of *P. aeruginosa*. Similar to the LasR-AI complex, RpoS stimulates the expression of exotoxin A, further increasing the toxicity of infection. Further to this, RpoS also stimulates the production of pyocyanin. Pyocyanin is the redox-active secondary metabolite to which *P. aeruginosa* owes its characteristic blue-green hue. The synthesis of this phenazine is the result of the expression of two operons *phzABCDEFG* 1 and 2 as well as the genes *phzM* and *phzS* (Lau et al., 2004). Pyocyanin holds both pro-inflammatory and free-radical effects, making it capable of inducing an increase in intracellular ROS and H_2O_2, stimulating oxidative stress within host cells (Jurado-Martín et al., 2021). As described previously, the induction of oxidative stress and the pro-inflammatory response is a major cause of decline within cystic fibrosis infections.

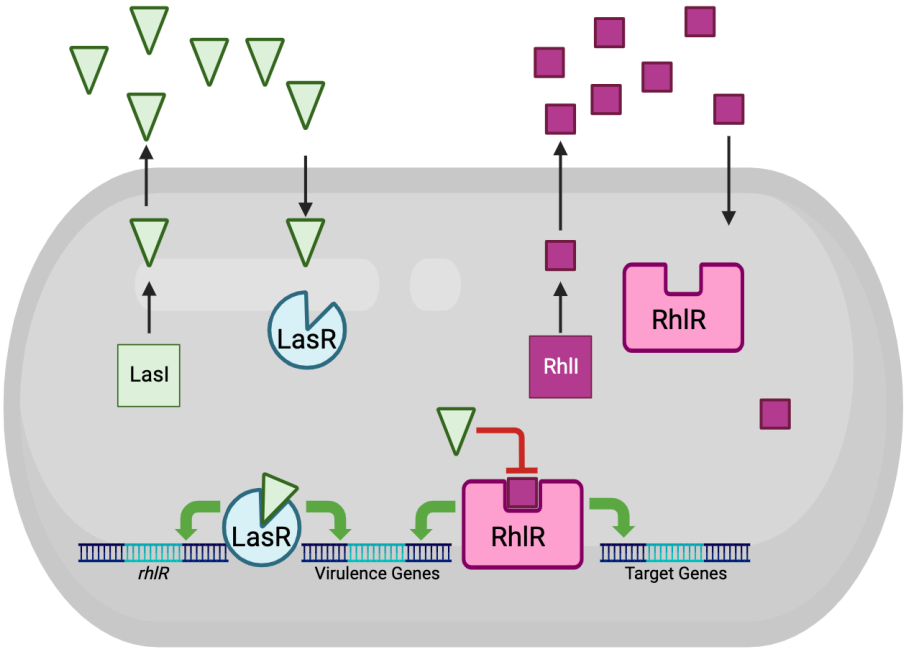


Figure 2: Diagram depicting the Las and Rhl QS circuit (Reproduced based on Miller and Basler, 2001). Created using Biorender.com.

The third quorum sensing system, pqs, differs from the HSL systems as its auto-inducers are alkyl-4(1H)-quinolones (Jurado-Martín et al., 2021). These auto-inducers, including 2-heptylhydroxy-1H-quinolin-4-one (PQS) and 2-heptyl-4(1H)-quinolone (HHQ), are synthesised by enzymes encoded by the *pqsABCDE*, *phnAB*, and *pqsH* genes (Jurado-Martín et al., 2021). The transcriptional regulator protein of the system is PqsR, with which PQS and HHQ are the most commonly associated autoinducers (Jurado-Martín et al., 2021). The pqs system is associated with the Las-Rhl circuit as it is stimulated by the Las system, and in turn induces *rhlI* expression, further enabling the Rhl system (Miller and Bassler, 2001). The pqs system also functions via a positive feedback loop, inducing the expression the *pqsABCDE* operon. A major virulence factor of *P. aeruginosa* is encoded within this operon, PqsE, another contributor to pyocyanin synthesis regulation (Jurado-Martín et al., 2021). Further functions of PqsE

include the upregulation of iron starvation genes and efflux pump genes for antimicrobial efflux. The system also induces the expression of other genes, namely those associated with eDNA release, for the aiding of mature biofilm formation.

The fourth system Iqs has only recently been discovered and only partially characterised. Its auto-inducer has been discovered and is an entirely new kind of signalling molecule, 2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde (IQS) (Jurado-Martín et al., 2021). Due to the Iqs systems recent discovery, a transcriptional regulator for the system has not been elucidated. Iqs regulates virulence factor production via the monitoring of phosphate limitation, and has been shown to be associated with pyocyanin, rhamnolipid and elastase expression (Lee et al., 2013). Additionally, IQS is able to divert host DNA damage repair leading to the halting of cell growth and apoptosis (Jurado-Martín et al., 2021).

This multitude of virulence factors which cause extreme damage to the host and protect *P. aeruginosa* from eradication makes *P. aeruginosa* a formidable pathogen. It is clear then, that with its substantial health and economic impact, successful treatment of these infections is paramount. This is especially true with increasing antimicrobial resistance within *P. aeruginosa*.

1.1.4 Antimicrobial resistance in *Pseudomonas aeruginosa*

Being already intrinsically resistant to multiple antibiotics, horizontally acquired multi-drug resistance is significantly threatening the treatment of *P. aeruginosa* infections. Typically, *P. aeruginosa* infections are treated using various antibiotics depending on

the type of infection and the susceptibility profile. The current medical indications suggest the use of penicillin-β-lactamase-inhibitor combinations, cephalosporins, carbapenems, monobactams, fluoroquinolones, aminoglycosides, and polymyxins (Ibrahim et al., 2020; Pang et al., 2019). Many of these antibiotics are already affected by various intrinsic resistance mechanisms within *P. aeruginosa*. With intrinsic resistance leaving some infections with already limited therapeutic options, acquired resistance only further increases the number of difficult or untreatable infections, see Figure 3. It is clear then that with its multitude of antimicrobial resistance mechanisms that new treatments for *P. aeruginosa* infections are desperately needed.

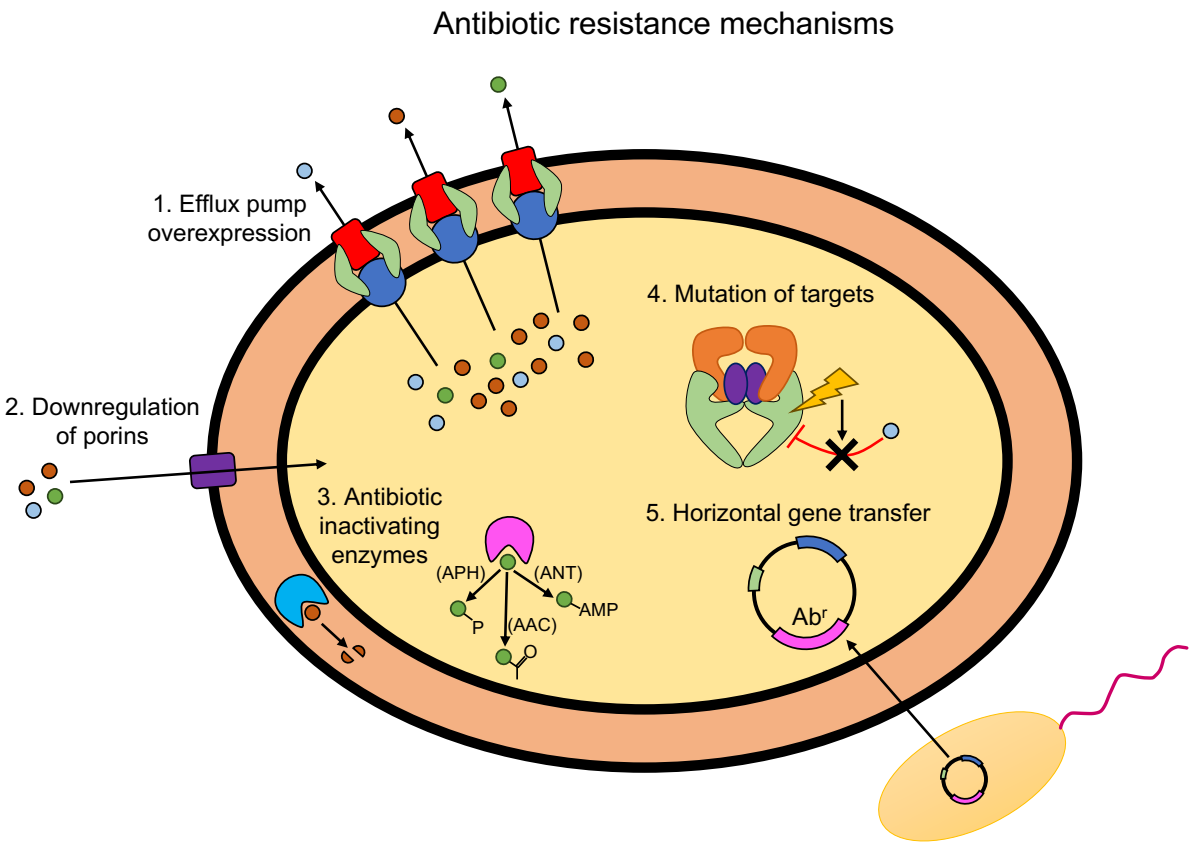


Figure 3: Diagram depicting different antibiotic resistance mechanisms in *Pseudomonas aeruginosa.* APH = aminoglycoside phosphotransferases; AAC = aminoglycoside acetyltransferases; ANT = aminoglycoside nucleotidyltransferases; Ab^r = Antibiotic resistance gene.

1.1.4.1 Intrinsic resistance mechanisms

On a base level, *P. aeruginosa* is intrinsically resistant to multiple classes of antibiotics. There are multiple reasons for this. Firstly, *P. aeruginosa* encodes multiple efflux pumps of different classes, however, the resistance-nodulation-division (RND) pumps are considered the major driver of efflux-based resistance (Pang et al., 2019). Of the

twelve RND pumps within *P. aeruginosa*, four are the primary drug resistance efflux pumps. These are, MexAB-OprM, MexCD-OprJ, MexEF-OprN and MexXY-OprM. The second Mex component of each of these complexes (B/D/F/Y) encodes the inner membrane protein (IMP) portion of the pump, which acts as the transporter. The Opr components are outer membrane proteins (OMP) which act as a channel through the outer membrane and out of the cell. Finally, the first Mex components (A/C/E/X) are periplasmic adaptor proteins which connect the IMP to the OMP (De Kievit et al., 2001; Blanco et al., 2016). Constitutively expressed within the bacterium, MexAB-OprM is considered the major contributor to intrinsic resistance, with substrate specificity to β-lactams, β-lactamase inhibitors, tetracycline, chloramphenicol, quinolones, macrolides, trimethoprim and novobiocin (De Kievit et al., 2001). MexAB-OprM is particularly problematic due to its frequent overexpression within multidrug resistant clinical isolates, often due to mutations in the repressors of the system (Horna et al., 2018). The other three pumps are expressed under specific conditions, and each have their own substrate specificities. MexCD-OprJ and MexEF-OprN expression is typically induced in strains with mutations in the regulators of the pumps (Aeschlimann, 2003). For MexCD-OprJ, these are mutations within the gene for the repressor of the pump NfxB and for MexEF-OprN these are mutations within the gene for its activator MexT, or the repressor NfxC (Fetar et al., 2011; Housseini B Issa et al., 2018). Upon expression, MexCD-OprJ provides resistance to quinolones, macrolides, tetracyclines, chloramphenicol, lincomycin, and some β-lactams. Further to regulator mutations, MexCD-OprJ expression is also induced by benzalkonium chloride and chlorhexidine, and MexEF-OprJ expression by chloramphenicol (Fraud et al., 2008; Fetar et al., 2011). MexXY-OprM provides resistance to the same classes as MexCD-OprJ, with

the exception of novobiocin and the addition of aminoglycosides (Masuda et al., 2000). MexEF-OprN has been shown to efflux fluoroquinolones, chloramphenicol, trimethoprim, and carbapenems (Llanes et al., 2011).

In addition to the efflux of antibiotics, *P. aeruginosa* also intrinsically prevents many antibiotics from entering the cell. It does this by restricting membrane permeability via various outer membrane porins. In fact, the membrane permeability of *P. aeruginosa* is between 12 and 100 times lower than that of *Escherichia coli* DH5α (Pang et al., 2019). This is primarily due to the limited number of non-specific porins (Ude et al., 2021). The major non-specific porin, OprF, has also now been shown to have a limited role in antibiotic translocation (Ude et al., 2021). This may be partially due to >95% of OprF proteins in the outer membrane being in the closed conformation. *P. aeruginosa* also encodes multiple specific porins, however, OprD has been indicated most commonly with antibiotic translocation due to its affinity for carbapenems (Pang et al., 2019). A common means of gaining resistance to carbapenems, is therefore the eradication or reduction of OprD expression (Li et al., 2012). Resistance can also be gained via mutation of the carbapenem binding loops within OprD (Li et al., 2012). This is of particular importance as carbapenem resistance has led to the allocation of *P. aeruginosa* as a critical pathogen on the World Health Organisations priority pathogen list (Asokan et al., 2019). *P. aeruginosa* also encodes gated porins, with OprH being implicated in gentamicin and polymyxin resistance. Reduced expression of OprH leads to the modification of LPS stabilising the membrane against these compounds (Pang et al., 2019).

Finally, *P. aeruginosa* also naturally encodes antibiotic-inactivating enzymes; the major enzymes affecting either β-lactams or aminoglycosides. The major β-lactam degrading enzyme is the β-lactamase encoded by *ampC* (Pang et al., 2019). AmpC is a class C β-lactamase, meaning it functions via a nucleophilic attack on the carbonyl moity of β-lactams to form an acyl ester. A tyrosine within the catalytic site then activates the hydrolysis of the acyl ester by a water molecule, destroying the β-lactam. The nucleophilic attack is modulated by a serine in the catalytic site, which is also activated by this tyrosine residue (Page, 2020). Class C β-lactamases show greatest affinity and activity against first and second generation cephalosporins, and penicillins. Despite their decreased ability to turnover third generation cephalosporins and carbapenems, they can form stable intermediates with these classes sequestering them away from their targets (Page, 2020). Further to this, *P. aeruginosa* encodes aminoglycoside inactivating enzymes from each of the three known classes. The first class is the aminoglycoside phosphotransferases (APHs), which inactivate via phosphorylation of the 3'-hydroxyl of kanamycin, neomycin, and streptomycin. The next class, aminoglycoside acetyltransferases (AACs) acetylate the 3'- and 6'- NH_4 groups of gentamicin, amikacin, kanamycin, tobramycin and netilmicin. Finally, the aminoglycoside nucleotidyltransferases (ANTs) adenylate either of the hydroxyl or amino groups of gentamicin, amikacin, and tobramycin (Pang et al., 2019; Thacharodi and Lamont, 2022).

Clearly, this combination of intrinsic resistance mechanisms makes *P. aeruginosa* a difficult pathogen to treat. Amplifying this is the prevalence of acquired mechanisms of resistance that broaden the resistance profile of many clinical isolates.

1.1.4.2 Extrinsic resistance mechanisms

Antibiotic resistance in bacteria is not limited to intrinsic resistances but can be developed via spontaneous mutation of drug targets and horizontal gene transfer (HGT) of foreign DNA encoding resistance mechanisms (Munita and Arias, 2016). These mechanisms have resulted in resistance to the three major classes of antibiotics prescribed for *P. aeruginosa* infections; fluoroquinolones, aminoglycosides, and β-lactams (Hsu et al., 2005).

One example of antibiotic classes being rendered ineffective in clinical isolates due to target mutation is for the fluoroquinolone class. Fluoroquinolones target the type II topoisomerases DNA gyrase and topoisomerase IV. DNA gyrase, encoded by *gyrA* and *gyrB,* is responsible for the introduction of negative supercoils into DNA as well as releasing torsion in front of replication forks and transcription machinery (Aldred et al., 2014). Topoisomerase IV, encoded by *parC* and *parE*, has roles within DNA replication, separating the daughter chromosomes and removing DNA knots formed following recombination events (Vélez-Cruz and Osheroff, 2004). Multiple studies have shown significant correlations between the development of mutations within *gyrA*, *gyrB*, *parC*, or *parE* and resistance to fluoroquinolones such as ciprofloxacin in clinical isolates (Bruchmann et al., 2013; Wang et al., 2014; Nouri et al., 2016; Arabameri et al., 2021). These mutations occur in the quinolone resistance-determining regions of these genes, reducing the binding affinity of quinolones to the topoisomerases (Bruchmann et al., 2013; Pang et al., 2019).

As aforementioned, horizontal gene transfer is also critical for acquisition of drug resistance. Recently, a novel ciprofloxacin-modifying agent, CrpP (ciprofloxacin

resistance protein, plasmid encoded), was discovered on a plasmid in the nosocomial environment. This further led to the elucidation of multiple *crpP*-like genes on various mobile genetic elements (López et al., 2022). Since these genes are encoded on mobile genetic elements (MGEs), they can be transferred to other *P. aeruginosa* isolates, spreading this resistance further. This has also been the case with carbapenem resistance, where carbapenemases have been encoded on MGEs and spread across multiple strains. Of major concern is that the majority of the acquired carbapenemases belong to the metallo-β-lactamase (MBL) class. These MBLs are problematic due to their innate resistance to β-lactamase inhibitors and ability to hydrolyse nearly all β-lactams (Botelho et al., 2019). Alarmingly, a study found that 20% of isolates within 14 European countries produced carbapenemases (Botelho et al., 2019).

With increasing antimicrobial resistance, it is paramount that we truly understand these bacteria, their resistance mechanisms, and their virulence factors, as well as develop new therapies for their treatment. One method of achieving this is via high-throughput chemical genomics.

1.1.5 High-throughput chemical genomics and phenotypic profiling

For many years now genetic methods have been utilised to understand gene function, with early techniques, such as reverse genetics and chemical genetics, paving the way for high-throughput techniques powerful enough to elucidate gene function on a genome wide level. By screening ordered single-gene knockout libraries in a multitude

of conditional stresses one can functionally link genotype to phenotype across an entire genome. Furthermore, with recent advances in image analysis software, such as Iris, phenotype measurement is not limited to merely growth, but biofilm formation, colony morphology and envelope integrity, amongst others (Kritikos et al., 2017). The true power of high-throughput chemical genomics, however, comes from the hierarchical clustering of mutant phenotypic profiles. A mutants phenotypic profile constitutes its phenotypic response across all tested conditions. By clustering these profiles, one can systematically reconstruct biological pathways and complexes, since mutants with similar functions will harbour similar phenotypic profiles. As such, orphan genes can be functionally clustered to genes with known biology (Nichols et al., 2011). Furthermore, by focusing on genetic manipulation within conditional stresses such as antibiotics, one can discover new drug targets by mining the datasets for genes with conditional essentiality in these conditions. The plethora of information produced by chemical genomics has been displayed in the multiple studies since its introduction.

1.1.6 A brief history of chemical genomics

The first chemical genomic studies were pioneered from gene-gene interaction studies within yeast. These initial studies focused on the fitness effect of double gene knockout mutants in order to elucidate gene function and organisation (Schuldiner et al., 2005; Tong et al., 2001). Parsons et al. progressed this in 2006 by assessing a library of yeast haploid deletion mutants in 82 conditions, discovering many genes involved in drug resistance, as well as potential new gene targets for drug-repurposing (Parsons et al., 2006) . One major find was that the breast cancer drug tamoxifen was effective

at calcium homeostasis disruption (Parsons et al., 2006). In 2008, a study by Hillenmeyer et al., furthered this by testing both homozygous and heterozygous gene deletions within diploid yeast mutants across 400 different conditions, highlighting the truly high-throughput capabilities of the technique (Hillenmeyer et al., 2008). Encouraged by the success of these previous studies, bacterial chemical genomics arose. Various small-scale studies tested the effect of different clinically relevant antibiotics on a non-redundant ordered single-gene deletion transposon library in PA14 (Alvarez-Ortega et al., 2010; Liberati et al., 2006). Other small-scale screens were produced within organisms such as *Acinetobacter baylyi* and *Mycobacterium tuberculosis* (Gomez and Neyfakh, 2006; Lun et al., 2014). The first major chemical genomic screen within bacteria, however, was performed in *Escherichia coli*. Nichols et al., in 2011, assessed the phenotypes of the KEIO collection, an in-frame single-gene knockout mutant library in *E. coli* K-12, within over 300 conditions (Baba et al., 2006; Nichols et al., 2011). The Nichols et al. study successfully produced >10,000 phenotypes and categorised various biologically relevant hits and suggested functions for genes with previously unknown function (Nichols et al., 2011). Given the success of the Nichols et al. screen in *Escherichia coli*, the prospect of screening and functionally annotating less well annotated species became a priority. As such, a parallel screen of the *P. aeruginosa* strains PAO1 and PA14 was performed by Dr Manuel Banzhaf (unpublished), with the aim of characterising orphan genes in both strains and understanding their phenotypic differences. The analysis of these screens is the basis of the doctoral research within this thesis.

CHAPTER 2

MATERIALS AND METHODS

2.1 Production and testing of ChemGAPP pipelines

ChemGAPP was built using Python, with individual scripts packaged and made available on pip for installation. App versions of each pipeline were produced using the Python package Streamlit (Doherty et al., 2023).

2.1.1 ChemGAPP Big pipeline

The ChemGAPP Big pipeline was validated on the chemical genomic screen data from Nichols et al., 2011 that was re-analysed using Iris by Kritikos et al., in 2017 (Kritikos et al., 2017; Nichols et al., 2011). The Kritikos et al., study measured colony integral opacity as the colony size measure, to represent 3-dimensional colony growth (Kritikos et al., 2017).

2.1.1.1 ChemGAPP Big: conversion of Iris files

The Iris files from the Kritikos et al. study were compiled into a dataset by taking the colony size values from each Iris file and processing them as column values within the main dataset. Iris file names were converted into multi-header column names representing the condition, concentration, source plate number, batch number and replicate number. Zero values where not all replicates were also zero were then converted to NaN values, since these likely represent mis-pinned colonies.

2.1.1.2 ChemGAPP Big: plate normalisation

Two stages of plate normalisation were performed on each plate as in Collins et al., 2006 (Collins et al., 2006). Firstly, the outer two edges of the plate were checked via a Wilcoxon rank sum test to determine if normalisation was required for the removal of edge effects. The outer two edges were then compared to the centre of the plate. If a plate had a *p*-value < 0.05 normalisation was performed, since distributions statistically differed. Here, colony sizes were scaled such that the medians for each edge row and column were equal to the plate middle mean (PMM), as such:

$$\boldsymbol{Normalised\ size} = \boldsymbol{Unnormalised\ size} \times (\boldsymbol{PMM}/\boldsymbol{Median\ of\ row\ or\ column})$$

The PMM corresponds to the mean size of colonies, excluding the two outermost rows and columns, that fall within the 40th and 60th percentiles of colony size.

Secondly, all colonies were scaled such that the PMM was equal to the median value of all colonies within the dataset, as such:

$$\boldsymbol{Normalised\ size} = \boldsymbol{Unnormalised\ size} \times (\boldsymbol{Plate\ median}/\boldsymbol{PMM})$$

2.1.1.3 ChemGAPP Big: quality control analyses

2.1.1.3.1 Quality control analyses: Z-score test

The Z-score value describes the distance of a test value from the population mean in measures of standard deviation (std). During the analysis the distribution of the colony sizes was adjusted to a standard normal distribution, with a mean of zero and a std of

1 (Driscoll et al., 2000). Replicate colonies for each replicate plate in each condition were compared using the Z-score test. Outlier colonies (Z-score value >1 or <-1) and missing values were highlighted. The percentage normality for each replicate plate, i.e., the percentage of colonies within the plate which are not outliers or missing values, was then calculated (Fig. 4).

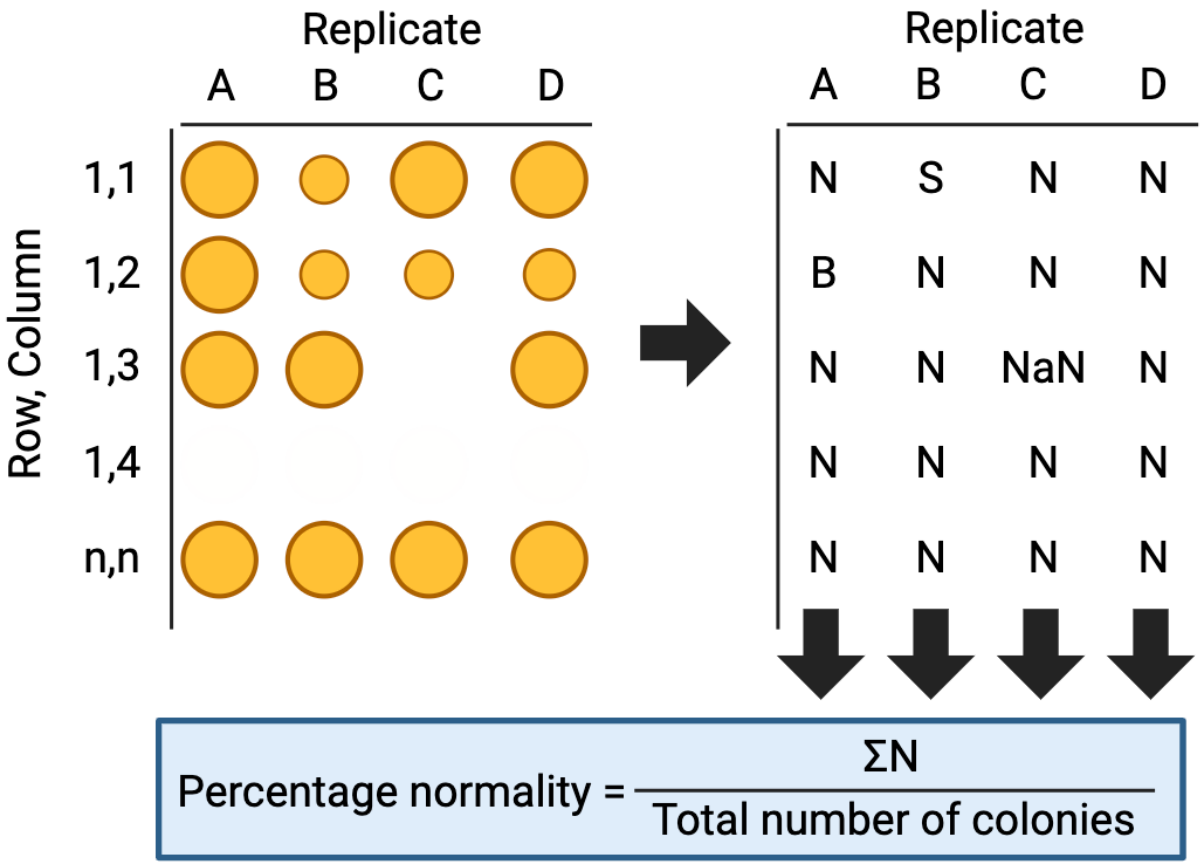


Figure 4: Diagram depicting the methodology of the Z-score test. B = outlier colonies bigger than the mean; S = outlier colonies smaller than the mean; NaN = missing colonies; Row, column = location of colony within condition plate. ΣN = count of colonies labelled N for a given replicate plate. Total number of colonies = total count of colonies for the respective replicate plate (ΣN +ΣB +ΣS +ΣNaN). (Created using Biorender.com).

2.1.1.3.2 Quality control analyses: Mann-Whitney test

The distributions of colony sizes for each replicate plate in the condition of interest were compared to each other replicate plates distribution within the selected condition. The mean p-value was calculated for each set of comparisons which included the replicate of interest. This was performed by averaging the Mann-Whitney p-values of the comparison pairs (Fig. 5).

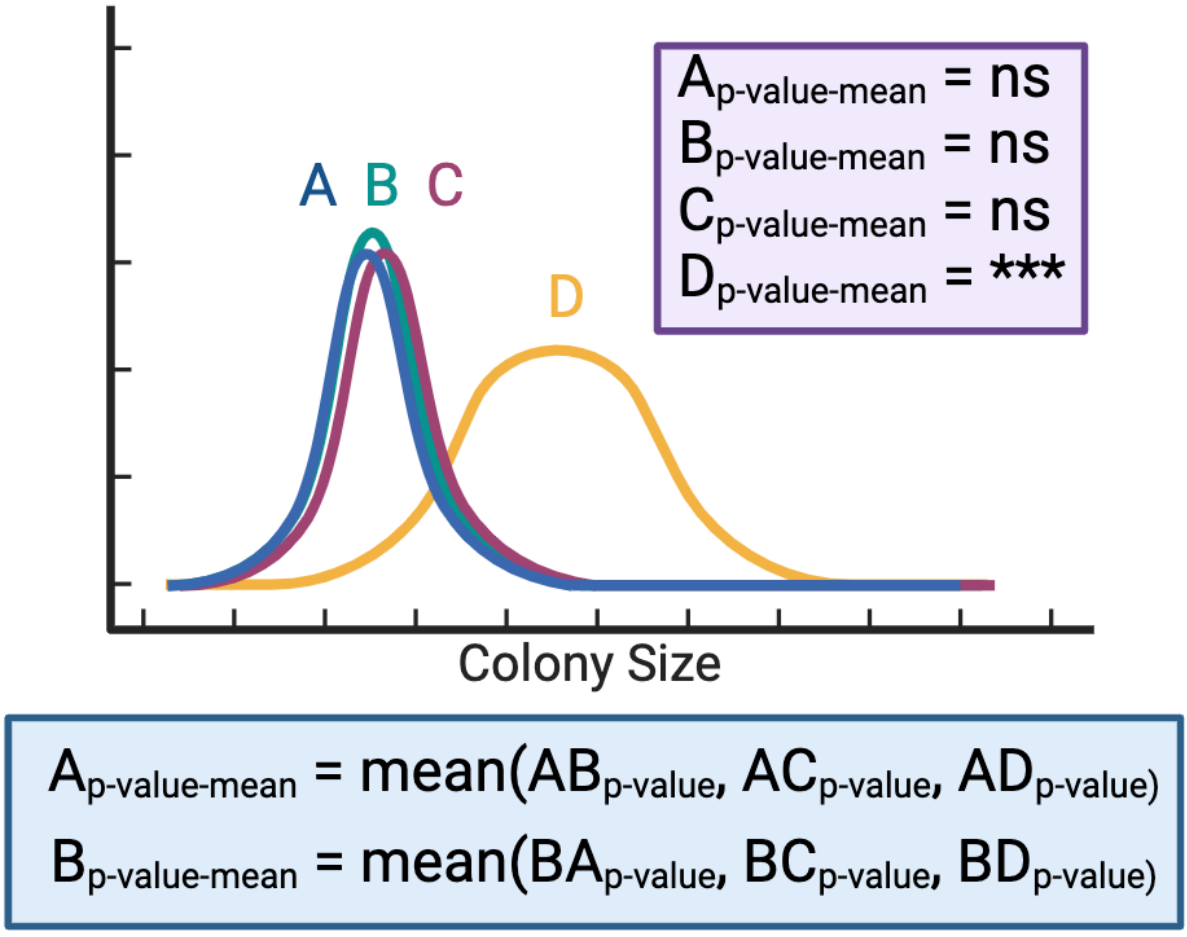


Figure 5: Diagram depicting the methodology of the Mann-Whitney plate level test. $AB_{p\text{-value}}$ = the Mann-Whitney p value for the comparison of the distribution of colony

sizes for replicate plate A versus replicate plate B. *** = significant p-value-mean, demonstrating differing distributions. (Created using Biorender.com).

2.1.1.3.3 Quality control analyses: condition level tests

For the condition level analyses two tests were performed, the condition level Mann-Whitney test and the variance analysis. For the condition level Mann-Whitney test the variance was calculated across all mean p-values for the condition of interest (Fig. 6A). For the variance analysis the variance across replicates of the mutant of interest in the condition of interest was calculated. These values were averaged to find the average variance of the condition (Fig. 6B).

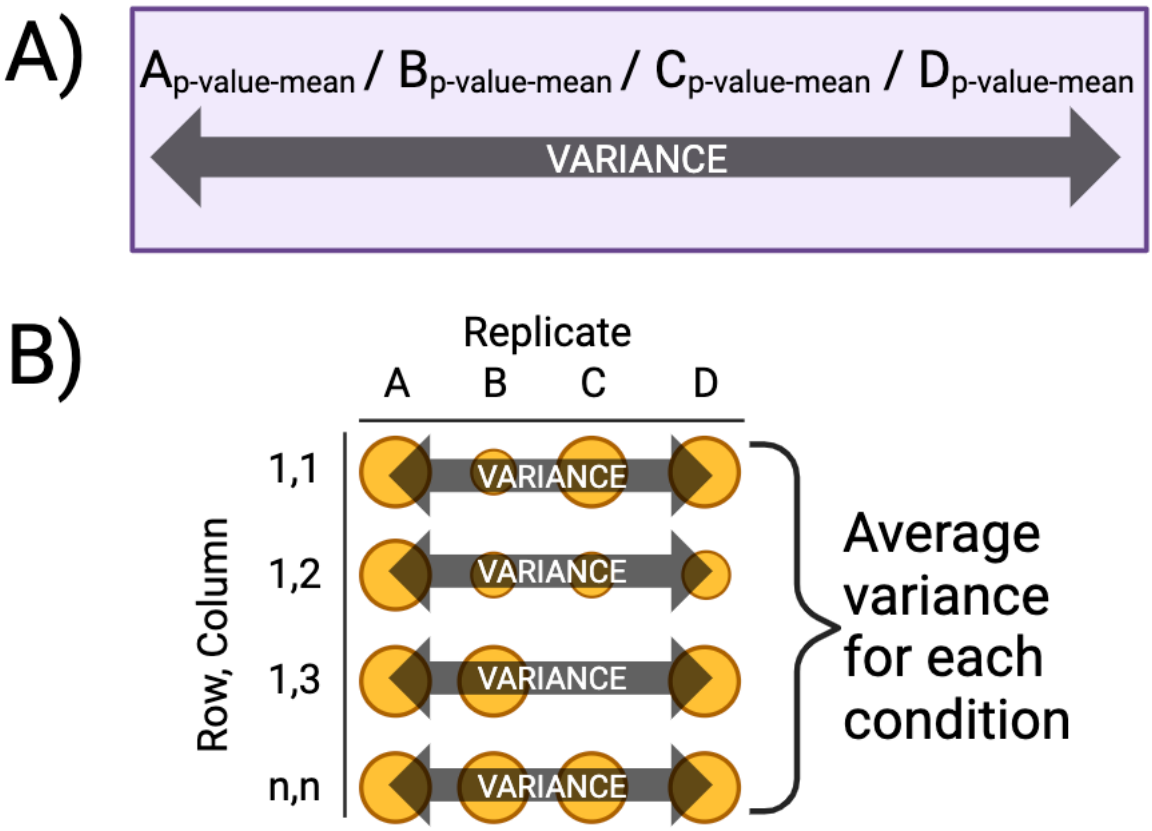


Figure 6: Diagram depicting the methodology of the condition level tests. A) Mann-Whitney condition level test. B) Variance analysis. (Created using Biorender.com)

2.1.1.4 ChemGAPP Big: threshold selection and data removal

The number of plates or conditions which failed the various quality control tests at different thresholds was tested for each quality control analysis. For the Z-score test the thresholds tested were 20%, 30%, 40%, 50%, 60%, 70% and 80% for percentage normality, which is described in 2.1.3.1. Where a replicate plate had a percentage normality lower than the tested threshold it was categorised as failing. Thresholds were generated for the three other tests based on the range of the test values produced from

that particular analysis. Where the mean p-value for the Mann-Whitney test was less than the threshold the replicate was considered as a failure. Where the variance value for the two condition level tests were greater than the threshold these conditions were considered as a failure. For the plate level analyses, the percentage of plates lost (failing plates) within each condition at each of these thresholds was then calculated and plotted for visualisation. For the condition level analyses, the absolute number of conditions lost was calculated and plotted for visualisation. The thresholds chosen for the reanalysed Nichols et al. dataset were 80%, 0.07585, and 0.0128, for normality, Mann-Whitney plate analysis and Mann-Whitney condition analysis. Data was not removed based on the condition level variance analysis. Following threshold selection failing plates and conditions were removed from the dataset to produce a curated dataset which was re-normalised as in 2.1.1.2.

2.1.1.5 ChemGAPP Big: fitness score calculation

Normalised colony sizes were scored using the S-score test, as described in Collins et al., 2006 (Collins et al., 2006). This was:

$$\boldsymbol{S} = (\mu_{\boldsymbol{Exp}} - \mu_{\boldsymbol{Cont}})/\mathbf{sqrt}(\boldsymbol{s}_{\boldsymbol{Var}}/\boldsymbol{n}_{\boldsymbol{Exp}} + \boldsymbol{s}_{\boldsymbol{Var}}/\boldsymbol{n}_{\boldsymbol{Cont}})$$

Where: μ_{Exp} = the mean of the normalised colony sizes for the mutant of interest in the condition of interest; μ_{Cont} = the median of the normalised colony sizes for the mutant of interest in all conditions; n_{Exp} = the number of replicates of the mutant of interest in the condition of interest; n_{Cont} = the median number of replicates for all conditions. S_{Var} is equal to:

$$s_{Var} = \left(var_{Exp} \times (n_{Exp} - 1) + var_{Cont} \times (n_{Cont} - 1)\right) / (n_{Exp} + n_{Cont} - 2)$$

Where: var_{Cont} = the median variance of normalised colony sizes for the mutant of interest in all conditions or a minimum bound; var_{Exp} = the variance of normalised colony sizes for the mutant of interest within the condition of interest or a minimum bound. The minimum bound for var_{Exp} was placed on the expected standard deviation used to calculate var_{Exp}. The average standard deviation for all mutants within the condition of interest, was used to produce a minimum bound variance for var_{Exp}.

The minimum bound for var_{Cont} was equal to μ_{Cont} multiplied by the median relative error of all mutants in all conditions, where:

$$Median\ relative\ error\ =\ standard\ deviation/mean$$

S-scores with infinite values were then converted to NaN values. For each plate, the S-scores were then scaled such that their interquartile range was equal to 1.35, such that outlier values are roughly +/-2. This is due to outliers commonly being described as values which deviate from the first or third quartiles by 1.5 times the interquartile range. Considering the mean of chemical genomic screens is often close to zero due to most mutants displaying no phenotype in most conditions, 1.5 times the interquartile range (IQR) (1.35) is ~2. As such, for the ChemGAPP-reanalysed Nichols et al. dataset significant S-scores were determined as < -3 and > 2. Loss of function mutants (negative S-scores) were given a harsher threshold since these are more common. Both the curated and non-curated datasets were given S-scores before hierarchically clustering for rows and columns using uncentered Pearson Correlation and average linkage in Treeview3 (Keil et al., 2018).

2.1.1.6 ChemGAPP Big: benchmarking the curated dataset

Both the curated and non-curated datasets were compared to each other and the original Nichols et al., dataset. Here, all genes which belonged to the same operon were subjected to cosine similarity analysis for their phenotypic profile of S-scores. An equal number of gene pairs from differing operons were similarly compared. A confusion matrix was produced for thresholds ranging from -1 to 1, thresholds increased incrementally by 0.1. True positives were denoted as genes from the same operon with cosine similarity scores greater than the threshold. False positives were genes from different operons with cosine similarity scores greater than the threshold. True negatives were genes from the different operons with cosine similarity scores less than the threshold. False negatives were genes from the same operon with cosine similarity scores less than the threshold. Confusion matrices at each threshold were used to calculate the sensitivity, specificity and false positive rate which was used to produce a receiver operating characteristic (ROC) curve. The area under the curve (AUC) for each dataset was calculated and compared.

2.1.1.7 ChemGAPP Big: production of the bootstrapped dataset

Both the curated and non-curated datasets were bootstrapped 1000 times with replacement, with 300 plates being randomly selected each time. For each bootstrapped dataset the S-scores were calculated and S-scores for the corresponding mutant-condition pairs across the datasets were averaged. The absolute error for each mutant-condition pair was calculated between the S-score from the original dataset

and the mean bootstrapped S-score. These errors were averaged to compute the mean absolute error (MAE). Due to the nature of replacement some conditions comprised of all the same replicate plate, meaning all colony sizes were identical. This resulted in complications with S-scoring due to var_{Exp} being equal to zero, resulting in the minimum bound also being zero. Furthermore, some sets of genes were lethal in the majority of conditions, resulting in μ_{cont} being equal to zero as well. The combination of these zero values resulted in a small number of abnormally large S-scores. As such, S-scores >125, representing 0.000018%, and 0.000027% of values within the bootstrapped non-curated and bootstrapped curated dataset respectively, were removed as outliers prior to the calculation of absolute errors. For ease of visualisation, S-scores > 125 were not displayed within the scatter plot for the original curated dataset, however these scores were not excluded from absolute error calculations.

2.1.2 ChemGAPP Small pipeline

A small-scale screen of BW25113 and BW25113 ΔenvC was performed by Gabriela Boelter and Michael Alao. BW25113 and BW25113 ΔenvC were pinned onto LBA, and 0.25% SDS + 0.25 mM EDTA supplemented LBA. For each plate three replicates were produced, and plates were grown for 12 hours at 37℃ before imaging. Iris was used to analyse the plate images and the phenotype measured was colony size (Kritikos et al., 2017). Iris files were inputted into ChemGAPP Small. Files were compiled as in 2.1.1.1 and normalised as in 2.1.1.2. Fitness ratios were calculated via two methods for the production of two plot types. For bar plots, the average colony size for each mutant of interest in the condition of interest was divided by the average colony size

for the wildtype in the condition of interest. The significance of the fitness ratio was measured by the 95% confidence interval between replicate plates. For the swarm plot analysis, each individual colony size data point, including wildtype data points, was divided by the mean colony size of the wildtype for the corresponding condition. A one-way ANOVA was then applied to compare each mutants distribution of fitness ratios with the distribution of the wildtype fitness ratios in the condition of interest.

2.1.3 ChemGAPP GI pipeline

Iris files for three genetic interaction case studies were provided by Dr Manuel Banzhaf and Dr Jack Bryant (Banzhaf et al., 2020; Bryant et al., 2023). Iris files were inputted into ChemGAPP GI and compiled as in 2.1.1.1. Strain ID values were added based on coordinate data from the IRIS files and replicate numbers assigned to each set of four colonies. Each set of replicates included one wildtype colony, two single gene knockouts and a double knockout colony, which were arranged in a square on condition plates. For each replicate set, the fitness ratio was calculated for each mutant by dividing the mutant colony size by the wildtype colony size. An expected fitness ratio for double knockout was also calculated using the formula:

$$\boldsymbol{Double\ Expected} = (\Delta \boldsymbol{A})/(\boldsymbol{WT}) * (\Delta \boldsymbol{B})/(\boldsymbol{WT})$$

Data was plotted in a bar chart and significance measured via 95% confidence interval error bars and a one-way ANOVAs between the double observed and double expected fitness ratios.

2.2 Comparison of *P. aeruginosa* chemical genomic datasets

2.2.1 ChemGAPP Big analysis

A chemical genomic screen of *P. aeruginosa* strains PAO1 and PA14 was performed by Dr Manuel Banzhaf. Two single gene transposon insertion libraries were used for the study (Jacobs et al., 2003; Liberati et al., 2006). Both libraries were screened in parallel on ~200 conditions to allow for comparison. A minimum of three replicates were produced for each screening condition and source plate combination. The plates were imaged and then analysed using the image analysis software Iris to convert images to numerical phenotype values. Here, colony size was used as the phenotype. These Iris files were then inputted into ChemGAPP Big for analysis as previously described. PAO1 and PA14 plates were analysed separately and condition plates with a base of MOPS minimal media were analysed separately from condition plates with a base of LB agar. Following analysis, the final curated and non-curated datasets for LBA and minimal media conditions were combined for their respective strains and level of curation. Since some conditions were missing across in the PA14 screen, PAO1 was analysed for all condition plates available for that strain and then for condition plates which matched with PA14. The curation thresholds used for each analysis are displayed in Table 1.

Table 1: Table displaying the curation thresholds used for each dataset.

Dataset	Normality Threshold	Mann Whitney Threshold
PAO1 base dataset LBA	65	0.02559
PAO1 base dataset minimal media	65	0.004741
PAO1 matching dataset LBA	60	0.02562
PAO1 matching dataset minimal media	60	0.006888
PA14 base dataset LBA	60	0.01497
PA14 base dataset minimal media	60	0.01837

2.2.2 Calculating replicate reproducibility for *P. aeruginosa* datasets

The normalised colony sizes for the curated and non-curated base datasets for both PAO1 and PA14 were subject to a replicate reproducibility analysis. Each replicate condition-mutant pair was plotted against each other in a scatterplot-based heatmap, and the Pearson correlation coefficient calculated for each set. The 95% confidence interval for the Pearson correlation coefficient was also calculated.

2.2.3 False discovery rate analysis of S-score significance

After the datasets were given S-scores, the significance of each result at various false discovery rates (FDRs) was determined. FDRs tested were 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1. S-scores with an FDR p-value of ≤ 0.05 were determined as significant and the number of phenotypes at each FDR rate counted. The number of transposon mutants with a certain number of phenotypes at an FDR of 0.05 or 5% was then calculated. Mutants with ≥ 30 phenotypes across the various conditions were categorised as multiple stress responsive (MSR) genes. This analysis was performed on the condition matching datasets.

2.2.4 Benchmarking of PAO1 and PA14 Datasets

The PAO1 and PA14 matching curated dataset were first annotated using gene annotations downloaded from Pseudomonas.com. Each genome was then inputted into eggNOG-mapper 2.1.9 to find further gene annotations and orthologs. If a gene was not annotated by Pseudomonas.com then the eggNOG-mapper gene name was assigned to PAO1. Matching ortholog seeds were then used to determine the PAO1 and PA14 orthologs. Using these seeds, the PAO1 gene name annotations were given to the PA14 genes. Where a gene was unique to PA14, its own annotation was kept. For each dataset, mutants with transposon insertions within the last 20% of the gene were removed. Finally, based on the 5% FDR assignment, mutants with zero phenotypes were also removed from the dataset for this analysis. Four categories were

benchmarked against: replicates, transcriptional units, pathways, and three-letter nomenclature. Replicates were determined as those with the same locus ID but differing mutant IDs, i.e., mutants with different transposon insertion locations for the same gene. Transcriptional units were downloaded from BioCyc and assigned based on locus ID. Pathways were downloaded from Pseudomonas.com and the KEGG pathways from this were used. Pathways with greater than 200 genes within them were excluded. Three-letter nomenclature was determined by selecting the first three letters of gene name annotations. Three-letter nomenclature sets were excluded if they contained less than 3 genes or more than 100 genes. For each category genes with matching category values were compared to each other and a Pearson correlation r value calculated. Then over 50 iterations, randomised pairs were produced for each category and Pearson correlation r values calculated. For each iteration the number of true positives, true negatives, false positives, and false negatives were counted to calculate sensitivity, specificity and false negative rate for r value thresholds ranging from -1 to 1 in increments of 0.1. These values were averaged across each iteration to produce the final ROC curve for each strain for each benchmark. True positives were determined as genes with the same category value with an r value greater than the threshold. True negatives were determined as genes with different category values with an r value lower than the threshold. False positives were determined as genes with different category values with an r value greater than the threshold. False negatives were determined as genes with the same category values with an r value lower than the threshold. The AUC was then calculated for each ROC curve.

2.2.5 PseudoCAP analysis of MSR genes

The genes categorised as MSR genes were extracted for both PAO1 and PA14. The PseudoCAP functions were searched on the Pseudomonas.com database and manually assigned to each gene. If multiple PseudoCAP functions were available only one was chosen, this was chosen based on the KEGG pathway assigned. If the PseudoCAP function matched a KEGG pathway it was chosen. The counts of PseudoCAP functions for each strain was then calculated and plotted in a pie chart.

2.2.6 Ortholog analysis

PA14 was assigned PAO1 gene name annotations as in 2.2.4. Orthologous genes were determined using the eggNOG-mapper 2.1.9 ortholog seeds, mutants with matching seeds were considered orthologs. Mutants with transposon insertion sites in the last 20% of the gene were excluded for each dataset for this analysis. First, ortholog phenotypic profiles were compared using Pearson correlation. A density plot of all r values was then produced and the mean, and 5^{th} and 95^{th} percentiles calculated. Orthologs with r values greater than 0.4 were then assigned functions based on their PseudoCAP functions from Pseudomonas.com. Then, each condition-mutant pair S-score within the curated condition-matching dataset for PAO1 was compared to each equivalent condition-mutant pair ortholog S-score in the PA14 curated dataset and plotted within a scatter plot heatmap. The Pearson correlation coefficient was then calculated between the PAO1 and PA14 ortholog S-scores. The 95% confidence interval was then calculated for the Pearson correlation coefficient. This analysis was

repeated first with the removal of ortholog pairs with non-significant S-scores for the PAO1 mutant, then with the removal of pairs with non-significant S-scores for the PA14 mutant and finally the removal of pairs where neither ortholog had significant S-scores. S-scores were determined as significant if they had a p-value of ≤ 0.05 at an FDR of 5%. PA14 S-scores were then randomly shuffled such that they were matched to a non-orthologous pair from the same condition. Analyses were then rerun with the shuffled data over 10 iterations. Orthologs with anticorrelated S-scores were then extracted. Here orthologs with a difference of at least 5 in their S-scores, where one score was negative and the other positive and both scores were significant were selected. The KEGG categories for each unique gene were assigned and counted, with multiple assignments included. The condition types were counted for each unique condition-gene pair.

2.2.7 t-SNE analysis of PAO1 and PA14 S-scores.

The S-scores for PAO1 and PA14 were subject to t-distributed stochastic neighbour embedding (t-SNE) analysis. The curated S-scores for PAO1 and PA14 condition matching datasets were combined into one dataset and NaN, -inf and inf values imputed to zero. Mutants with the transposon inserted in the last 20% of the gene were then excluded. The dataset was then normalised to a standard deviation of 1 and a mean of 0, before t-SNE analysis with two components, a perplexity of 50 and a learning rate of 1000.

2.3 Bacterial strains and storage

All strains used within the study are listed in Table 2. All strains were stored in 15% glycerol supplemented Luria Bertani broth (LBB) at a temperature of -80 ˚C. Strains were grown for overnight cultures within LBB and appropriate antibiotics, unless stated otherwise. Stock bacteria agar plates were streaked onto Luria Bertani agar (LBA) supplemented with appropriate antibiotics, unless specified otherwise.

Table 2: Bacterial strains used within the study.

Strain Name	Species	Origin	Notes
PAO1	*Pseudomonas aeruginosa*		
PA14	*Pseudomonas aeruginosa*		
PAO1 ∆*Pa*SAH	*Pseudomonas aeruginosa*	This study.	
PA14 ∆*Pa*SAH	*Pseudomonas aeruginosa*	This study.	
DH5α + pEXTk::∆PA14_05600	*Escherichia coli*	This study.	Knockout vector for PA14_05600
DH5α + pEX18Tc::∆PA0431	*Escherichia coli*	This study.	Knockout vector for PA0431
PA14 ∆*bioB-H*	*Pseudomonas aeruginosa*	This study, provided by Dr Erwin Bohn	Biotin synthesis operon deletion strain (*bioB* to *bioH*)
PA14 ∆glpD	*Pseudomonas aeruginosa*	This study, provided by Dr Erwin Bohn	
PA14 *bioA::*MAR2xT7	*Pseudomonas aeruginosa*	(Liberati et al., 2006)	Transposon mutant
PA14 *bioB::*MAR2xT7	*Pseudomonas aeruginosa*	(Liberati et al., 2006)	Transposon mutant
PA14 *bioC::*MAR2xT7	*Pseudomonas aeruginosa*	(Liberati et al., 2006)	Transposon mutant

PA14 *bioD::*MAR2xT7	*Pseudomonas aeruginosa*	(Liberati et al., 2006)	Transposon mutant
PA14 *bioF::*MAR2xT7	*Pseudomonas aeruginosa*	(Liberati et al., 2006)	Transposon mutant
PA14 *bioH::*MAR2xT7	*Pseudomonas aeruginosa*	(Liberati et al., 2006)	Transposon mutant
AB5075 *bioA::*kr	*Acinetobacter baumannii*	(Gallagher et al., 2015)	Transposon mutant
AB5075 *bioB::*kr	*Acinetobacter baumannii*	(Gallagher et al., 2015)	Transposon mutant
AB5075 *bioC::kr*	*Acinetobacter baumannii*	(Gallagher et al., 2015)	Transposon mutant
AB5075 *bioD::kr*	*Acinetobacter baumannii*	(Gallagher et al., 2015)	Transposon mutant
AB5075 *bioF::kr*	*Acinetobacter baumannii*	(Gallagher et al., 2015)	Transposon mutant
BW25113	*Escherichia coli*		
BW25113 ΔenvC	*Escherichia coli*	(Baba et al., 2006)	

2.4 Production of *Pa*SAH knockouts in PAO1 and PA14

Methods described in section 2.4 were performed by Ava Jordan an undergraduate masters student under my supervision.

2.4.1 Primers used within Section 2.4

The primers used within section 2.4 are described in Table 3.

Table 3: Primers used for fragment and vector amplification for production of PA0431 and PA14_05600 knockout vectors.

Strain	Primer Name	Section	Primer Sequence (5'-3')
PAO1	PA0431_up_fragment_fwd	2.4.2	CGGGTACCGAGCTCGGGCCTGGATGCTGCGGCTGTCGTCGCCGTA
PAO1	PA0431_up_fragment_rev	2.4.2	GCTCCCACGATGTCCACGCACCTCAAAATCGAGCACTACCGCCAGTACTACTGA
PAO1	PA0431_down_fragment_fwd	2.4.2	GTACTGGCGGTAGTGCTCGATTTTGAGGTGCGTGGACATCGTGGGAGCGCTCAT
PAO1	PA0431_down_fragment_rev	2.4.2	CTATGACCATGATTACGTGGTGATCGGCTACGGCGACGTGGG
PAO1	PA0431_check_fwd	2.4.4	CGCCAGGGTTTTCCCAGTCACGAC
PAO1	PA0431_check_rev	2.4.4	AGCGGATAACAATTTCACACAGGA
PA14	pEXTk_linearisation_fwd	2.4.2	AGGTCGACTCTAGAGGATCC
PA14	pEXTk_linearisation_rev	2.4.2	TTCCGGCTCGTATAATGTGTGGAATTAGCTTACCTGC
PA14	05600_up_fragment_fwd	2.4.2	AGCTAATTCCACACATTATACGAGCCGGAAATAGCCTCAACGACGCCAT

PA14	05600_up_fragment_rev	2.4.2	TCAGTAGTACTGGCGGTAG TGCTCGATTTTTTCCGTGGA CATCGTGGGAGCGCTCAT
PA14	05600_down_fragment_fwd	2.4.2	ATGAGCGCTCCCACGATGT CCACGGAAAAAATCGAGCA CTACCGCCAGTACTACTGA
PA14	05600_down_fragment_rev	2.4.2	TCGAGCCCGGGGATCCTCT AGAGTCGACCTAGGCTCGG ATCGGCCTG
PA14	05600_check_fwd	2.4.4	GTACAACGACGACTACCTG
PA14	05600_check_rev	2.4.4	CTTCTTGTTGATTCCGACT

2.4.2 Production of knockout vectors

In order to produce clean knockouts of *Pa*SAH (PA0431 and PA14_05600) in PAO1 and PA14, knockout vectors were constructed based on methods described in Huang and Wilks, 2017. For PA14 the vector used was pEXTk and for PAO1 a variation of this vector (pEX18Tc) was used. PCR was used to produce an up fragment with homology to ~800 bp upstream of the gene of interest, and the first ~30 bp of the gene interest, and a down fragment with the last ~30 bp of the gene of interest and ~800 bp of the sequence downstream. The forward primer for the up fragment and the reverse primer for down fragment contained a tail with homology to the relevant vector. The reverse primer for the up fragment and the forward primer for the down fragment, each contained a tail with homology to each other. The primers for this PCR are displayed in Table 3, the PCR reagents and cycle conditions are displayed in Table 4.

Table 4: PCR reagents and cycle conditions for knockout vector up and down fragments.

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	98°C	2 minutes	1
Denaturation	98°C	20 seconds	30
Annealing	62-68°C	10 seconds	
Extension	72°C	25 seconds	
Final Extension	72°C	10 minutes	1
Hold	4°C	∞	
REAGENT		**VOLUME (µL)**	
dH_2O		10.75	
Q5 High-Fidelity DNA Polymerase		0.25	
5x Q5 Reaction Buffer		5	
5x Q5 High GC Enhancer		5	
10 µM Forward Primer		1.25	
10 µM Reverse Primer		1.25	
10 mM dNTPs		0.5	
20 ng/µL Genomic DNA		1	

PCR was then used to linearise pEXTk, see Table 3 and 5. Restriction digestion with New England Biolabs EcoRI-HF in rCutSmart buffer was used to linearize pEX18Tc following the manufacturer's instructions. Linearised plasmids were then fused to the up and down fragments via Gibson assembly, using the NEBuilder HiFi DNA Assembly Master Mix. This was performed following the instructions by the manufacturers for a three-fragment reaction.

Table 5: PCR reagents and cycle conditions for linearisation of pEXTk.

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	98°C	2 minutes	1
Denaturation	98°C	20 seconds	30
Annealing	59-62°C	10 seconds	
Extension	72°C	150 seconds	
Final Extension	72°C	10 minutes	1

Hold	4°C	∞	
REAGENT		**VOLUME (µL)**	
dH_2O		15.75	
Q5 High-Fidelity DNA Polymerase		0.25	
5x Q5 Reaction Buffer		5	
5x Q5 High GC Enhancer		5	
10 µM Forward Primer (pEXTk_linearisation_fwd)		1.25	
10 µM Reverse Primer (pEXTk_linearisation_rev)		1.25	
10 mM dNTPs		0.5	
2 ng/µL pEXTk		1	

2.4.3 Transformation of vectors into DH5α for storage.

2.4.3.1 Production of chemically competent DH5α

Overnight cultures of DH5α were grown at 37 ˚C, under aeration, in a shaking incubator. Luria Bertani broth (50 mL) was inoculated with 50 µL of overnight culture and grown at 37 ˚C until an OD600 of 0.4-0.6. The culture was centrifuged at 4000 rpm for 10 minutes at 4 ˚C and the supernatant discarded. The pellet was resuspended in 25 mL ice-cold 0.1 M $CaCl_2$ before vortexing. The resuspension was incubated on ice for 30 minutes before centrifugation at 4000 rpm for 10 minutes at 4 ˚C. Supernatant was discarded and pellet resuspended in 2 mL of ice-cold 0.1 M $CaCl_2$ before addition of 1 mL 50% glycerol. Cells were then incubated on ice for 30 minutes before aliquoting (100 µL) into microcentrifuge tubes for storage at -80 ˚C.

2.4.3.2 Transformation of DH5α with knockout vectors

To an aliquot (100 µL) of chemically competent DH5α cells, 1 µL of either pEX18Tc_ΔPA0431 or pEXTk_ΔPA14_05600 (10 ng/µL) was added and incubated on ice for 30 minutes. Cells were heat-shocked for 30 seconds in a 42 ˚C water-bath. Cells

were then incubated on ice for a further 2 minutes before 1 mL of LB broth was added. Cells were recovered at 37 ˚C in a shaking incubator for 1 hour. Recovered cells were plated (100 µL) on 15 µg/mL gentamicin LB agar plates for pEXTk_ΔPA14_05600 and 10 µg/mL tetracycline LB agar plates for pEX18Tc_ΔPA0431. Remaining cells were centrifuged for 2 minutes at 13,000 rpm in a tabletop centrifuge and supernatant was discarded. Cells were resuspended in 100 µL of LB broth for further plating. Negative controls containing no plasmid were also plated on LB agar plates supplemented with the appropriate antibiotics. Obtained colonies were selected and re-streaked onto antibiotic supplemented LB agar plates. Overnight cultures were produced, and strains stored in 15% glycerol supplemented LBB at a temperature of -80 ˚C.

2.4.4 Transformation of PAO1 and PA14

Transformation of PAO1 and PA14 was performed as in Huang and Wilks, 2017 (Huang and Wilks, 2017). All processes were performed at room temperature. Overnight cultures (5 mL) were produced for each strain and pelleted in a table-top centrifuge at 13,000 rpm for 1 minute. Cells were then washed twice using sterile 1 mL $MgSO_4$ (1 mM) before finally resuspending in 50 µL $MgSO_4$ (1 mM). Knockout vectors were extracted from DH5α using the QIAprep Spin Miniprep Kit, following the manufacturer's instructions. Knockout vector (5-10 µg) was added to competent cells and electroporated in a 2 mm gap cuvette at 2.2 kV. Cells were immediately recovered in 1 mL Brain Heart Infusion (BHI) broth for three hours at 37 ˚C. Recovered cells were plated on BHI agar plates supplemented with 100 µg/mL tetracycline (pEX18Tc_ΔPA0431) or 20 µg/mL gentamicin (pEXTk_ΔPA14_05600). Plates were

incubated for up to 64 hours at 37 ˚C and single recombinant colonies replated on non-selective LB agar to develop double recombinants. Double recombinants were then counter-selected for on 10% (w/v) sucrose TYS10 plates (pEX18Tc_ΔPA0431) or AZT/IPTG (100 µg/mL & 1 µM) LB agar plates. Double recombinants were then checked for gentamicin or tetracycline sensitivity to confirm double-recombinant status. Sensitive colonies were check by colony PCR for the correct insert, see Table 6. One colony with the correct insertion for each strain was grown in overnight culture and stored in 15% glycerol supplemented LBB at a temperature of -80 ˚C.

Table 6: PCR reagents and cycle conditions for Colony check of double recombinants.

CYCLE STEP	**TEMP**	**TIME**	**CYCLES**
Initial Denaturation	94°C	10 minutes	1
Denaturation	94°C	15 seconds	30
Annealing	64°C	30 seconds	
Extension	68°C	25 seconds	
Final Extension	68°C	5 minutes	1
Hold	4°C	∞	
REAGENT		**VOLUME (µL)**	
dH_2O		11.5	
OneTaq 2X Master Mix with Standard Buffer		12.5	
10 µM Forward Primer (See Table 3)		0.5	
10 µM Reverse Primer (See Table 3)		0.5	
Single colony		n/a	

2.5 *Pa*SAH spot plate assay

Methods described in section 2.5 were performed by Ava Jordan an undergraduate masters student under my supervision.

An overnight culture of PA14 Δ*Pa*SAH was adjusted to an optical density at 600 nm (OD_{600}) of 0.1. PA14 Δ*Pa*SAH culture was then serially diluted 1 in 10 for 8 dilutions. Dilutions (5 µL) were dropped on to LB agar plates supplemented with sulfonamides or trimethoprim with or without purines, Table 7. Plates were incubated at 37 ˚C for 16 hours before imaging.

Table 7: Drug concentration used within *Pa*SAH spot plate assay and small-scale screen.

Drug	**PAO1 concentration**	**PA14 concentration**
Sulfamethoxazole	275 µg/mL	100 µg/mL
Sulfisoxazole	650 µg/mL	350 µg/mL
Trimethoprim	350 µg/mL	25 µg/mL
Adenine	-	0.2 mM

2.6 *Pa*SAH small-scale chemical genomic screen.

Methods described in section 2.6 were performed by Ava Jordan and Michael Alao.

Condition plates were produced as within the PaSAH spot plate assay, using concentrations displayed in Table 7, excluding adenine addition. 60 replicates of PAO1, PA14, PAO1 Δ*Pa*SAH and PA14 Δ*Pa*SAH were pinned from liquid 96-well source plates using an S&P BM6 colony pinning robot, onto 12 LB agar plates. These

12 LB agar plates were incubated overnight at 37 ˚C for use as solid source plates. These colonies were then transferred to three replicate condition plates using the S&P BM6 robot. Plates were incubated at 37 ˚C for 14 hours before imaging and image analysis with Iris (Kritikos et al., 2017). Colony sizes were then analysed using ChemGAPP Small.

2.7 Structure analysis of biotin synthesis operon

Members of the Pseudomonadaceae and Enterobacteriaceae families were searched on BioCyc, and their transcriptional units examined. The biotin synthesis operons were viewed for each species and noted.

2.8 Biotin synthesis mutant spotting plates

MOPS minimal media agar plates (1x MOPS buffer, 1.32mM potassium phosphate dibasic, 40 mM succinate, 1.5% agar), were supplemented with either 40 nM biotin; 400 nM biotin; 40 nM biotin and 100 mM metformin; or 400 nM biotin and 100 mM metformin. Overnight cultures of PA14, PA14 *bioA*::MAR2xT7, PA14 *bioB*::MAR2xT7, PA14 *bioC*::MAR2xT7, PA14 *bioD*::MAR2xT7, PA14 *bioF*::MAR2xT7, PA14 *bioH*::MAR2xT7, PA14 ∆*bioB-H*, PA14 ∆*glpD*, AB5075 *bioA*::kr, AB5075 *bioB*::kr, AB5075 *bioC*::kr, AB5075 *bioD*::kr, and AB5075 *bioF*::kr were grown at 37 ˚C overnight, under aeration, in a shaking incubator. OD600 was obtained for each culture and adjusted to an OD600 of 0.125 within 200 µL of LBB. Cultures were then pipetted

into the first row of a 96-well plate, before serially diluting 20 µL into the subsequent rows of LBB (180 µL). Each column of wells for each strain was spotted (1.5 µL) within a column of each plate. Plates were incubated at 37 ˚C for 16 hours before inspection.

2.9 Broth microdilution MIC analysis

PA14, PA14 Δ*bioB-H*, AB5075, AB5075 *bioA*::kr, AB5075 *bioB*::kr, AB5075 *bioC*::kr, AB5075 *bioD*::kr, and AB5075 *bioF*::kr were plated to gain single colonies on LB agar and incubated overnight at 37 ˚C. Five colonies from each plate were used to inoculate 5 mL of prewarmed Mueller Hinton Broth (MHB) and cultures were grown at 37 ˚C under aeration at 180 rpm for 5 hours. The optical density at 600 nm was measured for each culture and cultures were adjusted to an OD600 of 0.001 in 2 mL of MHB. For biotin carboxylase end-product-supplemented MIC plates, cultures were resuspended in 2 mL MHB + 40 mM acetoacetate, 2 mL MHB + 40 mM succinate, or 2 mL MHB + 40 mM oxaloacetate. A 96-well clear flat bottom plate was set up for each strain. To the first well of each row, excluding the outer edges, 200 µL of appropriate antibiotic was added (2 M Metformin, 1.024 µg/mL Kanamycin, 32 µg/mL Gentamicin), with two rows dedicated to each antibiotic. To the subsequent wells of the rows, 100 µL MHB was added. Antibiotics were then serially diluted (100 µL) into subsequent wells of MHB and the final 100 µL discarded. OD600 adjusted cultures (100µL) were then used to inoculate each well. MHB (200 µL) was then added to the outer edges as a blank. Plates were then incubated for 24 hours at 37 ˚C before optical density was read at 600 nm using a CLARIOstar plate reader. For MOPs minimal media based MIC

assays, the same procedure was repeated but MHB was replaced with MOPs minimal media supplemented with 40 nM biotin.

2.10 Suppressor mutant screen of PA14 Δ*bioB-H*

An overnight culture (5 mL) of PA14 Δ*bioB-H* was grown in MHB at 37 ˚C under aeration. Overnight culture (100 µL) was plated onto MHB plates supplemented with 2x, 4x, 6x, 8x and 10x the MIC of metformin. Initial metformin concentrations were 15.6 mM, 31.2 mM, 46.8 mM, 62.4 mM, and 78 mM, respectively. Plates were incubated for 48 hours at 37 ˚C. Single colonies were purified via streaking from the highest concentration of metformin onto MHB plates of the same concentration and incubated at 37 ˚C overnight. Single purified colonies were resuspended in 1 mL 0.9% NaCl and 100 µL plated onto MHB plates supplemented with 2x, 4x, 6x, 8x and 10x the concentration of metformin from which the colony was purified. This was repeated until colonies were gained on MHB + 250 mM metformin plates. Colonies were then sent for sequencing.

2.11 Production of transposon library within PA14 Δ*bioB-H*

2.11.1 Primers used in section 2.11

The primers used within section 2.11 are described in Table 8.

Table 8: Primers used for sequencing preparation of TraDIS library. -s- = phosphorothioate bond.

Primer	Section	Sequence
TnTm_PCR-1_fwd	2.11.10.3	GAGTCGACCTGCAGGCATGCAAGCTTCAGGGT
TnTm_PCR-1_rev	2.11.10.3	GACTGGAGTTCAGACGTGTGCTCTTCCGATC
TnTm_6.1	2.11.10.5	AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCTCGTACG GCAAGCTTCAGGGTTGAGATGTGTA
TnTm_7.4	2.11.10.5	AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCTTAGCTAG GCAAGCTTCAGGGTTGAGATGTGTA
TnTm_8.2	2.11.10.5	AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCTGCATGC ATGCAAGCTTCAGGGTTGAGATGTGTA
NEBNext Index 2 Primer for Illumina	2.11.10.5	CAAGCAGAAGACGGCATACGAGATACATCGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 4 Primer for Illumina	2.11.10.5	CAAGCAGAAGACGGCATACGAGATTGGTCAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T
NEBNext Index 6 Primer for Illumina	2.11.10.5	CAAGCAGAAGACGGCATACGAGATATTGGCGT GACTGGAGTTCAGACGTGTGCTCTTCCGATC-s-T

2.11.2 Preparation of TnTm transposome

The transposome was produced by combining 100 ng EZ-Tn5™ <DHFR-1> transposon DNA and 2 µL EZ-Tn5™ transposase from the EZ-Tn5™ insertion kit (Lucigen) and 2 µL glycerol. Transposome mix was then briefly vortexed and incubated at 25 ˚C for 30 minutes before storing at -30 ˚C until required.

2.11.3 Production of electrocompetent cells

Overnight cultures of PA14 Δ*bioB-H* were produced and used to inoculate 400 mL prewarmed LB broth. The culture was then incubated at 37 ˚C under aeration at 180 rpm until the culture reached at optical density of 0.5 at 600 nm. The culture was then aliquoted into eight Falcon tubes and centrifuged at 4000 rpm at 4 ˚C for 10 minutes. Supernatant was discarded and pellets combined across four Falcon tubes, each resuspended in 25 mL ice-cold 300mM sucrose buffer. Cultures were then centrifuged again at 4000 rpm at 4 ˚C for 10 minutes, to wash cells, before the supernatant discarded. The cells underwent this wash step three more times. After the final centrifugation supernatant was discarded and the pellets combined in 1 mL ice-cold 300 mM sucrose buffer. This was aliquoted (100 µL) into 10 microcentrifuge tubes and left on ice until required.

2.11.4 Transformation of electrocompetent PA14 Δ*bioB-H* and library construction

The work performed in section 2.11.4 was aided by Sotiris Kamarinos.

Competent cells (100 µL) were incubated on ice for 30 minutes, before addition of 0.2 µL of TnTm transposome. Cells were briefly vortexed and then further incubated on ice for 30-60 minutes. Following incubation, cells were electroporated in ice-cold 1 mm gap electroporation cuvettes at 2.5 kV. Immediately after electroporation, cells were resuspended in 1 mL SOC medium and transferred into 50 mL Falcon tubes. Cells

were recovered for two hours at 37 ˚C with orbital shaking at 180 rpm. Following recovery, cells were plated (100 µL) onto 250 µg/mL trimethoprim LB agar plates using L-shaped spreaders. Plates were incubated at 37 ˚C for 17 hours and colonies pooled by resuspending plates in 20% glycerol LBB and combining in 50 mL Falcon tubes. Cells were stored at -80 ˚C. This procedure was repeated until 1 million colonies were achieved. Once 1 million colonies were achieved, pooled samples were defrosted and pooled together before aliquoting into 1.5 mL microcentrifuge tubes and 50 mL Falcons. Aliquoted samples were then stored at -80 ˚C.

2.11.5 Production and validation of SPRI beads

Solid-phase reversible immobilisation (SPRI) beads were produced following the protocol by Jolivet and Foley, 2020 (Jolivet and Foley, 2020). SPRI beads were then validated against AMPure XP beads (Beckman Coulter) using 100 bp Hyperladder (Bioline). The following steps were carried out for both the homemade SPRI beads and the AMPure SPRI beads. 2 µL 100bp Hyperladder (Bioline, UK) was added to 18 µL UltraPure DNase/RNase-Free Distilled Water (Invitrogen). SPRI beads (30 µL) were added to the DNA mix and mixed by up and down pipetting. SPRI-DNA mix was then incubated for 5 minutes at room temperature (rt). SPRI-DNA mixture was separated on a magnetic stand and supernatant was removed. The supernatant was prepared for DNA gel electrophoresis by the addition of 3.3 µL 6x no SDS Loading Dye. Whilst on the magnetic rack, fresh 70% ethanol (500 µL) was added to the remaining beads and incubated for 1 minute before supernatant was removed. The ethanol wash step was repeated, and beads were left to dry for 3 minutes at 37 ˚C following supernatant

removal. SPRI beads were then rehydrated with 20 µL UltraPure DNase/RNase-Free Distilled Water (Invitrogen) and mixed with the pipette. SPRI beads were left to separate again on the magnetic rack before collecting the supernatant. The supernatant was prepared for DNA gel electrophoresis by the addition of 3.3 µL 6x no SDS Loading Dye.

2.11.6 DNA gel electrophoresis

To produce a 1.5% agarose gel, 1.5 g agarose was added to 100 ml tris-acetate-EDTA buffer (TAE). The mixture was heated until the agarose had fully dissolved. Midori green (7 µl) was then added before pouring into a gel casting plate and leaving to set. The gel was placed in a gel tank and covered with 1% TAE buffer. 5 µl of 100 bp HyperLadder (Bioline, UK) was pipetted into the first well, followed by the AMPure bead supernatant samples and the homemade SPRI bead samples into the subsequent wells. Electrophoresis was carried out at 100V for 60 minutes, to separate DNA fragments by size. The gel was visualised on GelDoc^{TM}XR to determine band sizes. For both the AMPure beads and the homemade SPRI beads bands at 100 and 200 bp were absent within the second supernatant samples. The homemade SPRI beads were therefore determined as acceptable and used within all subsequent DNA immobilisation experiments.

2.11.7 TraDIS library screening and DNA extraction

For the metformin screening conditions, 50 mL of MOPS minimal media was supplemented with 40 nM biotin and either 0, 31.25 mM, 62.5 mM, or 125 mM metformin. The metformin concentrations represented 0, 0.5, 1 and 2 times the MIC of PA14 Δ*bioB-H* within metformin supplemented MOPS minimal media. For each concentration three biological replicates were produced. Cultures were inoculated using samples from the library to a starting OD_{600} of 0.05. Cultures were then incubated at 37 ˚C under aeration with orbital shaking at 180 rpm for 16 hours. The optical density at 600 nm was then calculated for each culture. Samples representing 0.2 OD_{600}/mL were taken from each culture. Samples were spun down and resuspended within 5 mL of MOPs minimal media to a final OD_{600} of 1. Genomic DNA was extracted from each sample using the Neo Biotech Quick Bacteria Genomic DNA Extraction Kit (NB-03-0027) following the instructions for Gram-negative bacteria.

2.11.8 Quantifying DNA using the Qubit® Fluorometer

The concentration of genomic DNA was quantified using the Qubit® dsDNA HS Assay Kit (Invitrogen). Qubit® dsDNA HS Buffer (199 µL per sample) was mixed with 1 µL per sample of Qubit® dsDNA HS Reagent to produce a master mix. A 1:10 dilution of genomic DNA was produced for each sample within Qubit® dsDNA HS Buffer to a final volume of 10 µL. For each genomic DNA sample, 199 µL of master mix was combined with 1 µL of 1:10 genomic DNA. For the Qubit® dsDNA HS Standards, 190 µL of master mix was combined with 10 µL of DNA standard. Qubit samples were vortexed and

incubated for two minutes at room temperature before reading on the Qubit® Fluorometer.

2.11.9 Shearing genomic DNA

Genomic DNA (1 µg) was diluted to a final volume of 500 µL in UltraPure DNase/RNase-Free Distilled Water (Invitrogen). Samples were incubated for 10 minutes on ice before shearing to an average length of 200 bp with a biorupter (Diagenode). The shearing programme consisted of 13 cycles at low intensity of 30 s ON and 90 s OFF. Following fragmentation, samples were condensed to a final volume of 55.5 µL using a vacuum concentrator (Eppendorf, Vacufuge concentrator 5301).

2.11.10 Library preparation of metformin exposed TraDIS samples

2.11.10.1 End repair and illumina adaptor ligation

The NEBNext Ultra I kit was used to repair the ends of the fragmented DNA samples. Fragmented DNA (50 µL) was mixed with 3 µL End Prep Enzyme Mix, 6.5 µL 10 x End Repair Reaction Buffer and 5.5 µL of water. The mixture was then incubated in a thermocycler for 30 minutes at 20 ˚C followed by 30 minutes at 65 ˚C. The NEBNext Adaptor for illumina was ligated to the repaired ends by mixing 15 µL Blunt/TA Ligase Master Mix, 2.5 µL NEBNext Adaptor and 1 µL Ligation Enhancer, with the end prep reaction mixture. This was incubated for 15 minutes at 20 ˚C before the addition of 3 µL USER® enzyme. Mixture was then further incubated at 37 ˚C for 15 minutes.

2.11.10.2 Size selection of adaptor-ligated DNA

In order to select for fragments with a length of 250 bp SPRI beads were utilised. Adaptor-ligated DNA samples were transferred to 1.5 mL low binding microcentrifuge tubes and 13.5 µL UltraPure DNase/RNase-Free Distilled Water (Invitrogen) added to a final volume of 100 µL. SPRI beads were resuspended and then 55 µL added to each DNA sample before mixing via pipette. Samples were incubated for 5 minutes at room temperature before bead separation on a magnetic stand. Supernatant was collected and 25 µL SPRI beads added before incubating again at rt for 5 minutes. SPRI beads were separated again and supernatant discarded. Beads were washed twice for 30 seconds with 200 µL 80% ethanol and left to air dry for 3 minutes. DNA was eluted from the beads by the addition of 17 µL EB Buffer and mixing via pipette. Beads were separated from supernatant using a magnetic stand and the DNA containing supernatant collected.

2.11.10.3 PCR Amplification of transposon junctions

Transposon containing fragments required enrichment for sequencing. As such, a forward primer homologous to the end of the transposon (TnTm_PCR-1_fwd) was paired with a reverse primer (TnTm_PCR-1_rev) that annealed to the illumina adaptor, amplifying the transposon-gDNA junction. SPRI selected DNA (15 µL) was mixed with Q5 Polymerase Master Mix (NEB), 2.5 µL 10µM PCR-1 forward and reverse primers

and 5 µL UltraPure DNase/RNase-Free Distilled Water (Invitrogen). PCR reaction was run following the cycle conditions displayed in Table 9.

Table 9: PCR cycling conditions for amplification of transposon junction.

CYCLE STEP	TEMP	TIME	CYCLES
Initial Denaturation	98°C	3 min	1
Denaturation	98°C	15 seconds	10
Annealing	65°C	30 seconds	
Extension	72°C	30 seconds	
Final Extension	72°C	1 minute	1
Hold	4°C	∞	

2.11.10.4 SPRI purification of PCR amplified fragments

PCR samples were transferred into 1.5 mL microcentrifuge tubes and mixed with 45 µL of resuspended SPRI beads (ratio 0.9:1). Samples were incubated for 5 minutes before separation on a magnetic rack. Supernatant was removed and beads washed twice with 200 µL 80% ethanol for 30 seconds, before air drying for 3 minutes. DNA was eluted as in 2.11.10.2 and collected supernatant transferred to a PCR microcentrifuge tube.

2.11.10.5 PCR-based sequencing index addition

A second PCR step was required to introduce illumina flow-cell adaptors for the binding of fragments to illumina flow cells, such that sequencing primers can anneal during sequencing. Multiplexing barcode sequences were also added to aid computational de-multiplexing of samples sequenced in parallel. PCR reaction was performed as in

2.11.10.3, with a custom inline index forward primer (TnTm_6.1, TnTm_7.4, or TnTm_8.2) and a NEBNext illumina index reverse primer. Furthermore, denaturation, annealing and extension steps were performed for 20 cycles. The final PCR product was purified following the steps outlined in 2.11.10.4, with a final elution volume of 33 µL. The DNA concentrations of the prepped library samples were quantified as in 2.11.8 and diluted to a final concentration of 10 ng/µL in 20 µL before illumina sequencing by Novogene.

2.11.11 TraDIS data analysis

Fastq files containing the sequencing data were first processed using FASTX-Toolkit. The modules 'fastx barcode splitter' and 'fastx trimmer' were used to extract the sequences which began with the relevant multiplexing barcode sequence, and trim the barcode sequence off (Hannon, 2010). Trimmomatic (version 0.36) was then used to filter out sequences shorter than 20 bp (Bolger et al., 2014). The transposon insertion locations were then mapped to the *Pseudomonas aeruginosa* UCBPP-PA14 (NCBI Reference Sequence: NC_008463.1) genome using Bio-Tradis module 'bacteria_tradis' (Barquist et al., 2016). The Bio-Tradis module 'tradis_gene_insert_sites' was then used to count insertions within each *P. aeruginosa* gene and specify which genes were essential, ambiguous, or non-essential. To calculate the $\log_2$ fold change (LogFC) between the metformin conditions and the metformin-free control, the Bio-Tradis module 'tradis_comparison.R' was used. Each set of replicates for each concentration of metformin was compared to the 0 mM metformin set of replicates. Where the LogFC for a gene was ≥ 1.0 or ≤ -1.0 and the

p-value ≤ 0.05, the insertions within the gene were considered as significantly altered compared to the 0 mM control. Where genes were listed as essential by Bio-Tradis::tradis_gene_insert_sites and tradis_essentiality.R in all replicates of the 0 mM metformin condition but listed as ambiguous or non-essential for all replicates in the presence of metformin, then genes were considered as no longer essential in the presence of metformin. KEGG categories were then assigned to each gene.

CHAPTER 3

CHEMGAPP: A TOOL FOR CHEMICAL GENOMIC ANALYSIS AND PHENOTYPIC PROFILING

This chapter is based upon a published article in the journal Bioinformatics, for which I was the first author and corresponding author (see List of Papers). The writing of this paper was supervised by Dr Manuel Banzhaf and Dr Danesh Moradigaravand. Within this paper, I produced the software described, performed all analyses, created all the figures, and wrote the paper (Doherty et al., 2023). My supervisor Dr Manuel Banzhaf has agreed to the inclusion of work based on this paper within this thesis, since I provided the major contributions. Other author contributions included editing the final paper and providing the test data.

3.1 Introduction

3.1.1 The need for a comprehensive data analysis software for chemical genomics.

Chemical genomics holds its power in the plethora of data it produces. By screening genome-wide knockout libraries in many conditions, phenotypic profiles can be produced with enough condition variety to expose functional similarities between genes. However, by producing large numbers of screening plates, the data analysis workload is extensive. Furthermore, despite the rise in studies utilising these techniques, no dedicated software exists for the comprehensive analysis of the screening data. Previous studies have implemented a variety of methods for data analysis, and these methods are often performed using in-house scripts or adaptations of packages used for similar techniques (Collins et al., 2006; Dixon et al., 2009; French et al., 2016; Nichols et al., 2011; Shiver et al., 2016, 2021). Image analysis software, including gitter and Iris, have been produced to tackle the conversion of plate images to numerical phenotype data. However, these software are not able to convert the

values to fitness scores (Wagih and Parts, 2014; Kritikos et al., 2017). Alternatively, many studies have implemented a fitness scoring software, EMAP Toolbox (https://sourceforge.net/projects/emap-toolbox/); however the tool has now been deprecated for a number of years (Collins, 2006; Collins et al., 2006; Nichols et al., 2011). Additionally, having been produced to handle genetic interaction data in yeast, it does not offer a standard strategy for condition-based assays. Further to this, none of these software options offer quality control steps to filter detrimental data. Finally, since most of these options require knowledge of coding languages such as R or MATLAB to implement, they are inaccessible for users with limited computational expertise (French et al., 2016; Shiver et al., 2021, 2016). Therefore, this doctoral research study aimed to produce an easy-to-use software for chemical genomic data analysis. As such, ChemGAPP (Chemical Genomics Analysis and Phenotypic Profiling) was developed.

3.1.1.1 Data analysis within high-throughput screens

Current pipelines for data analysis within high-throughput screens comprise of multiple steps encompassing the conversion of phenotype images to numerical values, the normalisation of data and the scoring of fitness.

A commonly used software for the first stage of data analysis is Iris, an image analysis software with capabilities to convert multiple phenotypes into numerical data points (Kritikos et al., 2017). This software was developed for the evaluation of phenotypes additional to colony growth, with the aim of analysing the 'unresponsive to growth' segment of the genome. This is important as for many gene deletions the phenotypic response is unrelated to growth defects. To this end, Iris is able to analyse biofilm

formation, colony morphology, envelope biosynthesis, sporulation, and reporter activity (Kritikos et al., 2017). On top of this, Iris developed a new method of measuring colony growth. Iris considers both colony size in pixels and colony opacity to produce a 3D measure of colony growth. This value is called the colony integral opacity (Kritikos et al., 2017). Due to Iris's many advanced and valuable features, the output of this software was chosen as the input for ChemGAPP. Despite this, as long as the data is formatted in the Iris format, any image analysis software can be used.

Since each experimental plate constitutes a single Iris file, these files need to be compiled into a single dataset before normalisation. Normalisation is performed to remove noise and standardise all conditions to make them comparable. Finally, fitness scores need to be assigned to the normalised colony phenotype data. Many studies use S-scores as proxies for fitness (Nichols et al., 2011; Shiver et al., 2021, 2016). Described by Collins et al., the S-score test is an alteration of the Student's t-test (Collins et al., 2006). The S-score test compares the mean phenotype value of a set of replicates within a specific condition, with the median phenotype value of the given mutant across all conditions. The result of the test is a statistical measure of fitness which describes the distance from this median. When considering colony size as the phenotype measure, positive scores indicate increased fitness, and negative scores show decreased fitness. Comparing mutants to themselves across all conditions as a simulated wildtype is more powerful than comparing to a true wildtype since mutants often have intrinsic differences in fitness to the wildtype. Thus, by using S-scores comparison to a true baseline is achieved. Finally, since S-scores correlate to statistical significance, they provide a robust and informative measurement of how strong a phenotype is across various conditions. Following S-score assignment, data is typically

hierarchically clustered and visualised to allow for data mining. This allows for the elucidation of orphan gene function via their association with genes of known biology; antibiotic resistance genes due to conditional essentiality within antibiotic conditions; and potential drug targets.

3.1.1.2 Data analysis within small-scale screens

Often, genome-wide screens are not required for confined research hypotheses. In many cases understanding specific gene mutants across specific conditions is more beneficial. In such cases, where a limited number of conditions are used, and conditions are expected to produce phenotypes for most mutants, S-scoring cannot be used as a measure of fitness. This is due to S-scores relying on a normal distribution of data, where most conditions do not result in a phenotypic change. In large-scale screens data is usually normally distributed as most genes are unimportant in most conditions. Furthermore, there are often sufficient mutants and conditions to allow for a normal distribution. Therefore, in the case of small-scale screens, analysis comprises of the same phenotype conversion and normalisation but differs in the assignment of fitness scores. In these instances, it is best to compare a mutant to either a wildtype within the same condition, or itself in a base media without conditional supplementation. Here a fitness ratio is produced instead. These data can then be used to test and describe specific phenotypic responses, answering questions from confined hypotheses.

3.1.1.3 Data analysis within genetic interaction screens

Often, studies aim to understand epistasis between two genes, however, epistasis may only occur within specific conditions. Therefore, chemical genomics can also be applied to genetic interaction screens to understand genetic interactions across a series of conditions. This can equally aid in the understanding of conditional essentiality, gene functionality and gene interaction networks. For genetic interaction screens, two single gene mutants are plated in an array with their parental strain and a double-knockout mutant. Again, phenotypes are measured and analysed using an image analysis software like Iris. Since comparisons occur within plates, normalisation across plates is not required in these instances. Similar to small-scale screens, fitness ratios are utilised in genetic interaction screens. Fitness ratios for the single and double knockout mutants are calculated against the wildtype colony. An expected ratio is also then calculated from the product of the two single knockout fitness ratios. This expected fitness ratio can then be compared to the observed fitness ratio of the double knockout mutant. This allows for the interpretation of the epistasis type between the two genes. Epistasis is defined as instances where the observed fitness of a double knockout mutant is significantly different to the expected fitness ratio (Mani et al., 2008). There are two types of epistasis, positive epistasis, and negative epistasis. Positive epistasis is also known as alleviating epistasis since the double knockout alleviates the fitness defects of the single knockouts. In this instance, the observed fitness ratio is higher than the expected ratio. Negative epistasis describes the instance of two genes synergistically reducing fitness and is displayed by the double knockout having decreased fitness compared the expected fitness ratio (Mani et al., 2008).

3.2 Results

3.2.1 ChemGAPP Pipelines

Since there are multiple types of chemical genomic screens, it was decided that ChemGAPP would comprise of three applications and sub-packages, see Figure 7. By doing this, each technique could have its own dedicated app, tailored to the analysis needs of each screening type. These three apps are ChemGAPP Big, ChemGAPP Small and ChemGAPP GI. ChemGAPP Big is used for the analysis of large-scale chemical genomic screens. Therefore, it was validated using the data from the Nichols et al. chemical genomic screen (Nichols et al., 2011). This screen analysed the effect of over 300 conditions on the *Escherichia coli* KEIO collection, an in-frame single-gene knockout mutant library in *E. coli* K-12 (Baba et al., 2006; Nichols et al., 2011). This dataset was chosen since it represents the first major chemical genomic screen within bacteria and successfully produced >10,000 phenotypes. Furthermore, the study characterised various biologically relevant phenotypes and revealed the function of multiple orphan genes (Nichols et al., 2011). ChemGAPP Small was built for the analysis of small-scale screens. Therefore, ChemGAPP Small was tested against data from a small-scale screen of *E. coli* BW25113 Δ*envC* in membrane stressor conditions. Finally, ChemGAPP GI, produced for genetic interaction screen analysis, was tested against data from three genetic interaction screens (Bryant et al., 2023; Banzhaf et al., 2020).

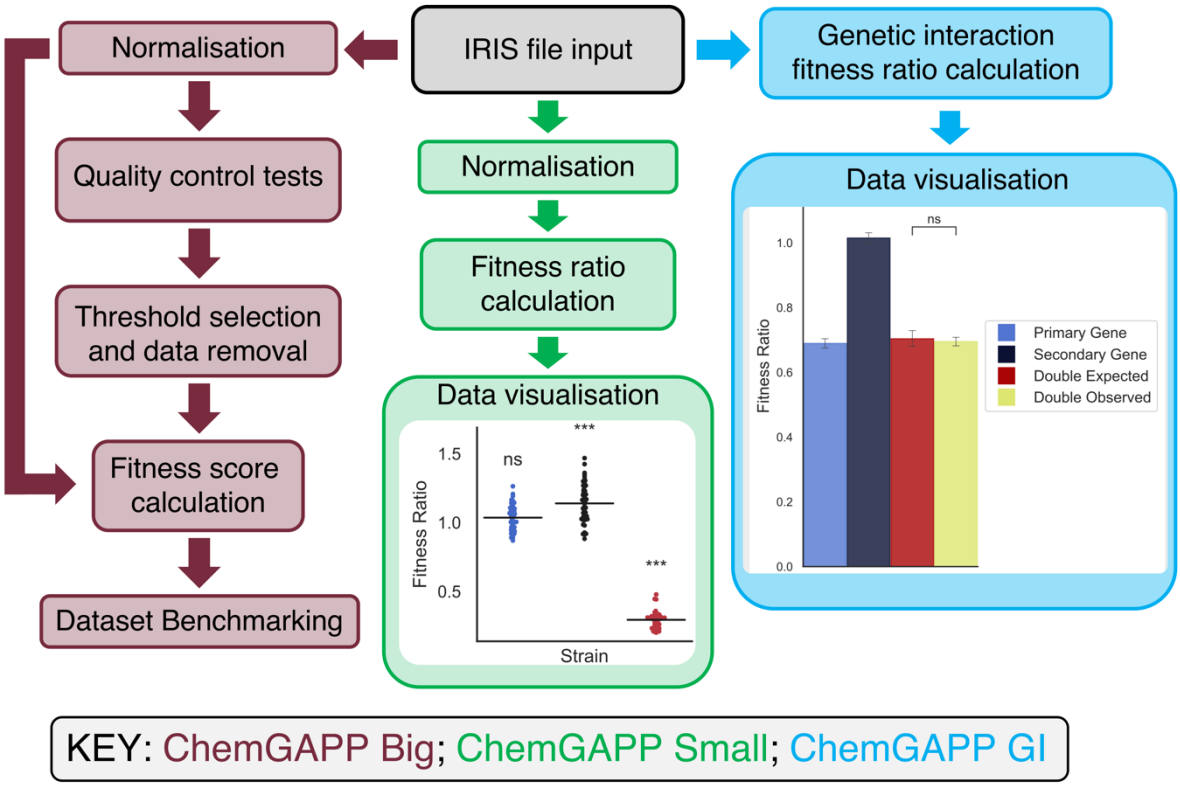


Figure 7: The data analysis pipelines of the three ChemGAPP sub-packages. Lefthand pathway represents ChemGAPP Big pipeline, middle pathway represents ChemGAPP Small pipeline and righthand pathway depicts ChemGAPP GI pipeline.

3.2.2 ChemGAPP Big successfully normalises condition plates.

One of the critical features of ChemGAPP Big is the scaling of the data such that all condition plates are comparable. This is performed via a two-step normalisation. This is a critical step since normalisation removes much of the noise created during the screening process. One of the issues that produces noise within chemical genomic screens is the edge effect (Baryshnikova et al., 2010). The edge effect is characterised as the increased growth of the outermost colonies due to reduced nutrient competition

(Baryshnikova et al., 2010). Noise is further introduced by the differing baseline effects of screening conditions. Some conditions are more likely to result in generalised growth deficits than others, such as a higher concentration of an antibiotic versus a lower concentration. ChemGAPP successfully removed this noise from the Nichols et al. data via its two-step normalisation process, see Figure 8. The distribution of colony sizes for three randomly selected condition plates were compared pre- and post-normalisation. Before normalising, the conditions had significantly different distributions and varying mean peaks. Following normalisation, distributions no longer differed, and the mean peaks were adjusted to become comparable.

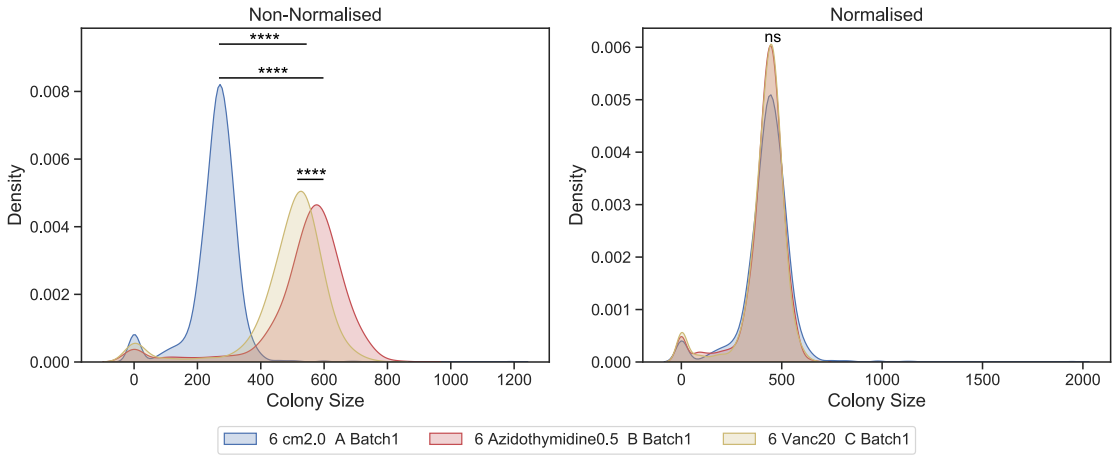


Figure 8: Density plots displaying the distribution of pre- and post- normalised colony sizes for three randomly selected condition plates. Cm2.0 = Chloramphenicol 2 µg/mL; Azidothymidine0.5 = Azidothymidine 0.5 ng/mL; Vanc20 = Vancomycin 20 µg/mL. **** = p-value <= 1.00e-04, ns: p-value > 0.05.

3.2.3 ChemGAPP Big highlights common errors within chemical genomic screens

Since high-throughput chemical genomic screens produce vast quantities of data, it was instrumental that ChemGAPP Big was able to perform effective quality control analyses. ChemGAPP Big performs four different statistical tests to pinpoint common errors within large-scale chemical genomic screens. Frequently occurring errors include inverted images, mislabelled plates, and mutants being unequally pinned between replicates or missed entirely. Furthermore, artefacts can be produced by Iris analysis since its detection algorithm occasionally fails to detect colonies. This can affect whole conditions, especially where a condition changes the colour of the agar. In these instances, colonies can become indistinguishable from the agar background and cannot be visualised by Iris. As such ChemGAPP big includes not just plate level quality controls, which compare replicate plates to each other, but condition level tests which look at condition suitability as a whole. The four tests are the Z-score normality test, the Mann-Whitney plate level test, the Mann-Whitney condition level test, and the condition variance test.

3.2.3.1 The Z-score normality test can distinguish unequal pinning.

To test the four quality control analyses and detect detrimental plates within the Nichols et al. dataset, the Z-score normality test was first applied. This technique is particularly useful for highlighting plate effects due to unequal pinning and plate inversions. The Z-score test compares each mutant of interest across all replicate plates and highlights

outliers as well as missing data points. The percentage normality is then defined as the percentage of colonies within a plate which are not outliers or missing values.

One condition with two replicate plates that had a reduced normality percentage were replicates C and D of the 20 ˚C cold shock condition. Replicate C had a percentage normality of 64.84% and replicate D had a score of 70.53%. Conversely, replicate plates A and B had high percentage normality scores of 90.43% and 95.05%, respectively. Representations of colony sizes for each replicate plate are displayed in Figure 9. It is clear from Figure 9 that unequal pinning has occurred within plates C and D. The colony sizes within the upper segment of plate D are consistently larger than those in the lower segment. This could be due to uneven agar surface, with the upper section being thicker than the lower, resulting in less contact to the lower portions of the plate by the robot's pinning attachment. On plate C, however, there is an inner circle of colonies with increased size compared to the outer segments of the plate, excluding the corners.

The conclusions of this visual inspection are further evidenced by the percentages of larger and smaller colonies within plates C and D. For C, 17.9% of colonies were classified as larger than the mean replicate colony size, and 17.9% were smaller. For D, these values were 15.69% and 13.74%, respectively.

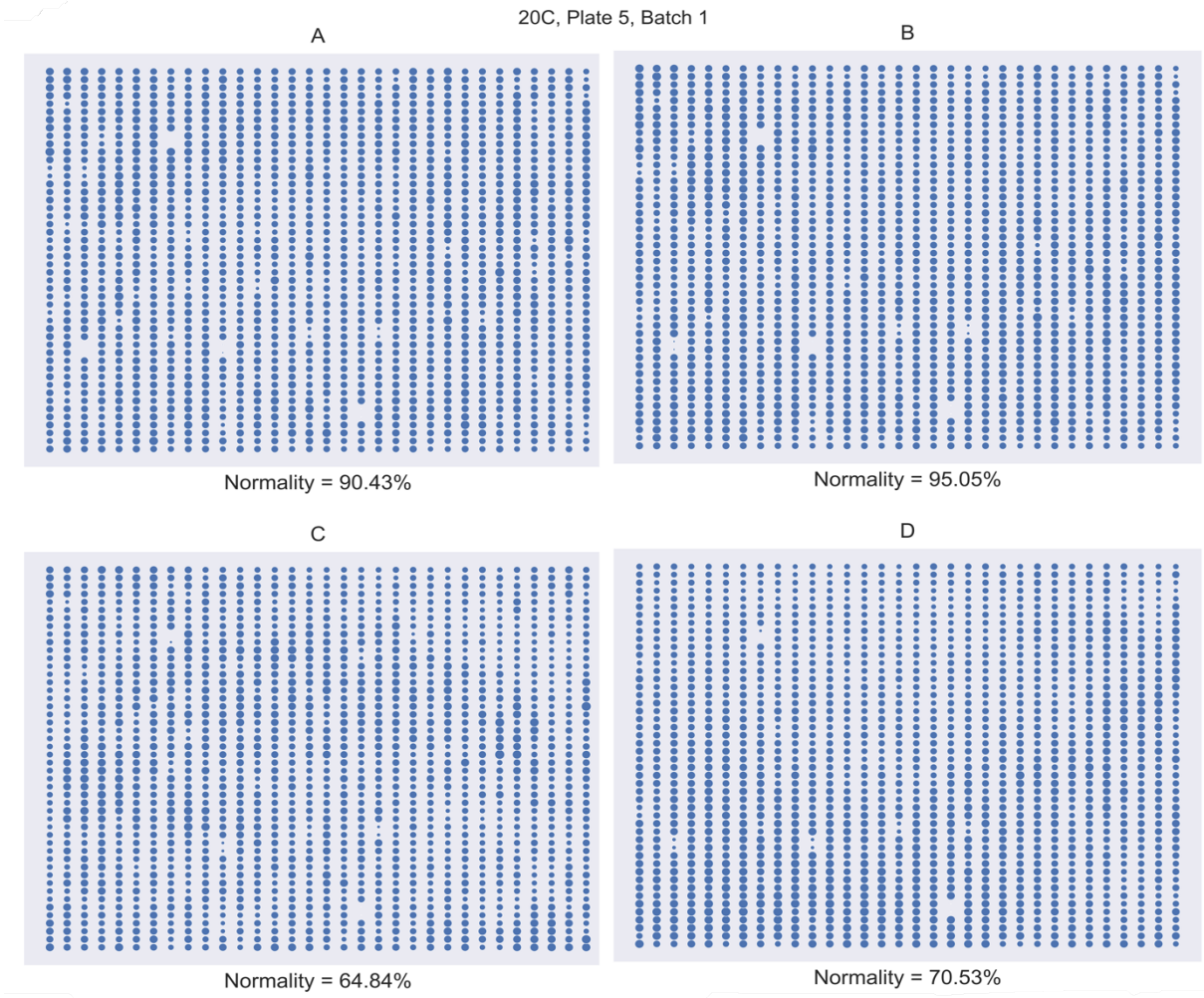


Figure 9: Plate matrices depicting the colony sizes for replicate plates of the 20 °C cold shock (20C) condition. The percentage normality scores determined by the Z-score test are displayed below each matrix. Circles represent the normalised colony size values and positions reflect the positions within the original plates.

3.2.3.2 Mann-Whitney quality control analysis

Similar to the Z-score test, the Mann-Whitney test can capture noise within condition plates. The Mann-Whitney test compares the distribution of colony sizes for each

replicate plate to each other and calculates the mean Mann-Whitney p-value for each plate. A significant mean p-value suggest that the distribution of the plate differs from that of the other replicates. The Mann-Whitney test is therefore adept at identifying mislabelled plates, as these will have extremely different colony size distributions. However, it is also capable of pinpointing other errors, such as unequal or missed pinning, colony spreading due to wet plates, and undetected colonies.

The Mann-Whitney test was performed on the Nichols et al. dataset and replicate plates with statistically significant differences in distributions highlighted. One such condition, 2 μg/ml rifampicin, had one replicate plate with a differing distribution, see Figure 10. Replicate C had a mean Mann–Whitney P-value of 0.0018. The density plot for replicate C showed an increased number of larger colonies than replicates A and B and there is a large reduction in the main peak. A further investigation into these conditions revealed that replicate C had a 5.25-fold higher number of missing colonies (21) than both A and B (4). This could be due to either mis-pinning of colonies or missed detection by Iris. As such, the Mann-Whitney test can accurately detect non-reproducible replicate plates based on multiple defect types.

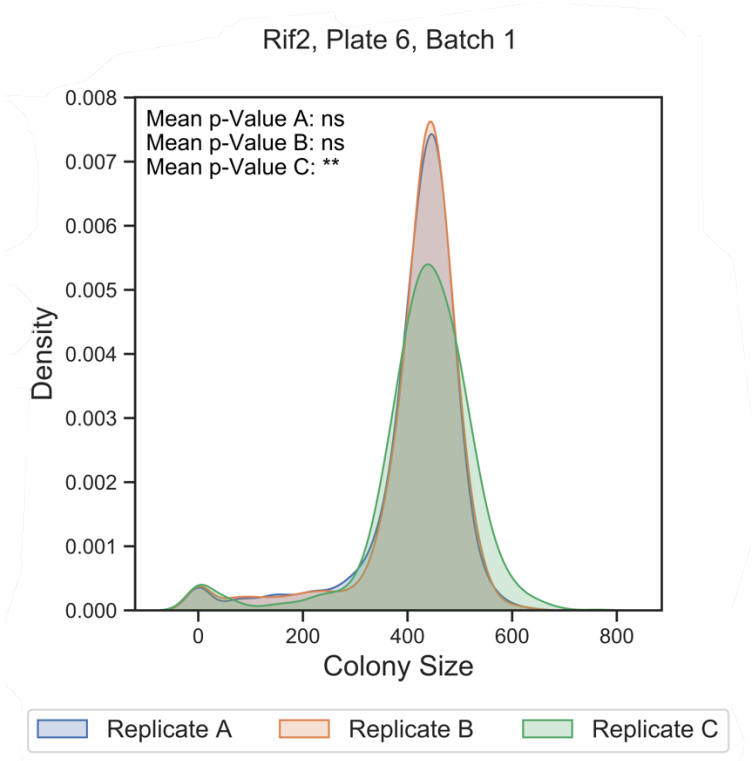


Figure 10: Density plot of condition Rif2 (Rifampicin 2 µg/ml), Plate 6, Batch1 colony sizes for replicate plates A, B and C. Distribution of C is significantly different from A and B by Mann–Whitney test. ** = 0.001 < p-value ≤ 0.01; ns = non-significant.

3.2.4 ChemGAPP Big produces robust fitness values.

Understanding and standardising how the measured phenotypes relate to fitness, underpins the power of chemical genomic screens. The critical statistical value within chemical genomics is the S-score, a measure used as a proxy for fitness. S-scores represent the difference in the fitness of a mutant within a specific condition compared to itself across all conditions, factoring in the variance across conditions. Scores correlate with statistical significance, meaning the more positive or negative a score is

the more significant the change in fitness. Negative scores represent reduced growth, and positive scores indicate improved growth. It is therefore vital that ChemGAPP Big produces robust S-scores which are not affected by the intrinsic noise within chemical genomic screens.

In order to test the robustness of the S-scores a bootstrapped dataset was produced. The curated and non-curated datasets were bootstrapped 1000 times, each time selecting 300 conditions with replacement. S-scores were calculated for each bootstrapped dataset and averaged to produce the final bootstrapped dataset. The mean absolute error (MAE) was then calculated between the original dataset S-scores and the final bootstrapped dataset S-scores, see Figure 11A/B. The MAE for the non-curated dataset was 0.0622, meaning the bootstrapped S-scores and the original S-scores only differed, on average, by 0.0622. This value constitutes just 4.94% of the standard deviation (1.257) of the original non-curated dataset. In comparison, the curated dataset MAE was 0.0657. With a standard deviation of 1.256, this MAE constituted just 5.23% of the base error within the dataset. The MAE was therefore negligible for both datasets, demonstrating the robust nature of the S-scores produced by ChemGAPP Big.

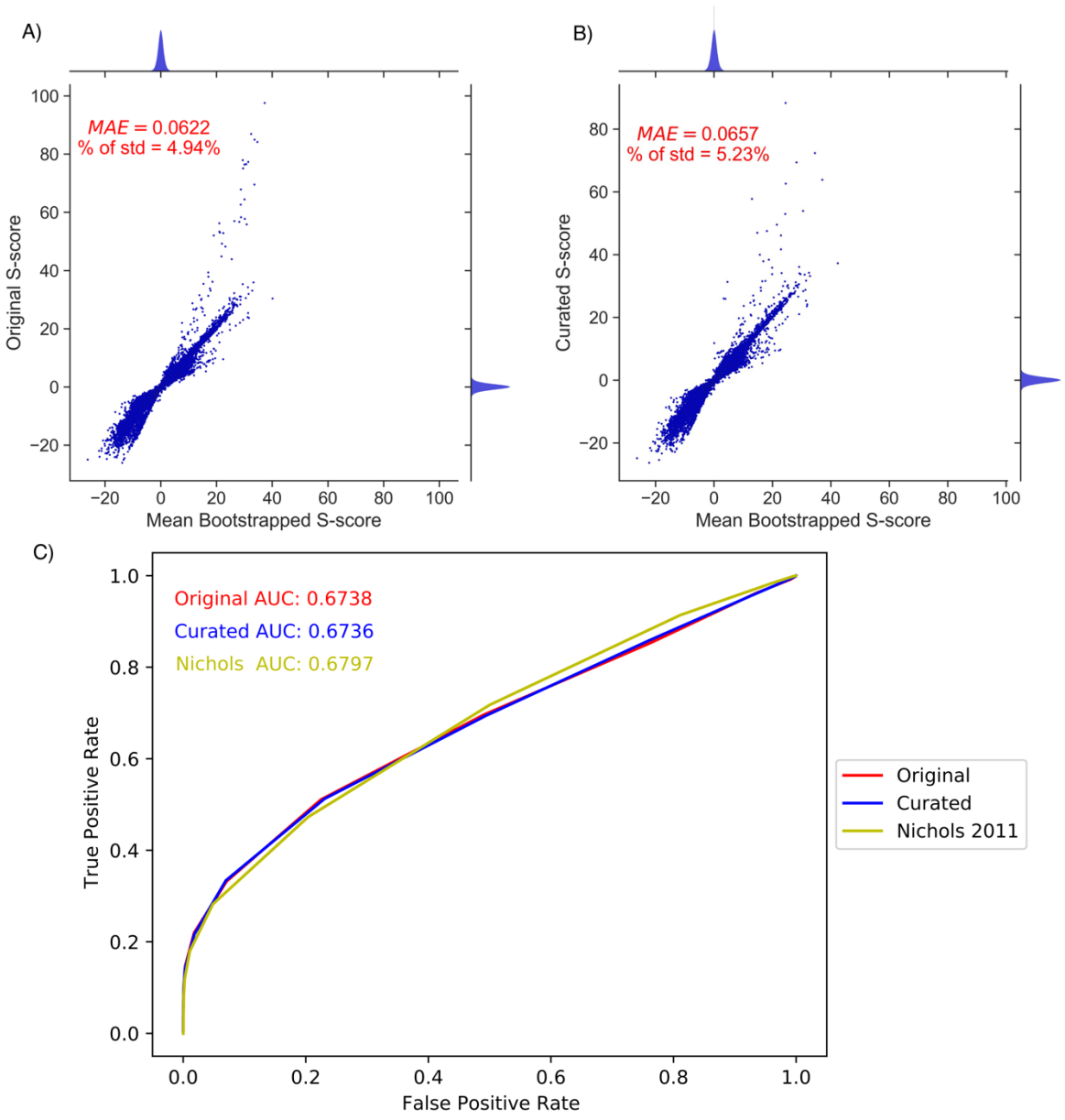


Figure 11: Joint density and scatter plots of the difference in S-scores between A) the non-curated dataset and the final bootstrapped dataset for each mutant-condition pair and B) the curated dataset the mean bootstrapped dataset for each mutant-condition pair. Density plots represent the distribution of S-scores within each dataset. For ease of visualization, outlier bootstrapped S-scores >125 were excluded, representing a negligible 0.000018% (A) and 0.000026% (B) of all values. Outlier curated S-scores

>125 were also excluded, representing a negligible percentage of all values (0.00056%). MAE, mean absolute error; % of std, percentage of the standard deviation for the original dataset that the MAE constitutes. (C) Receiver operating characteristic curve with AUC values for the non-curated dataset (red line), the curated dataset (blue line), and the KEIO dataset from Nichols et al. (yellow line).

Although both datasets produced robust and reliable S-scores, it was important to benchmark the datasets against each other following curation. In general, the curated dataset should perform better than the non-curated dataset having removed detrimental conditions. Further to this, it is imperative that ChemGAPP assigned fitness values as effectively as previous software. As such, both the non-curated and curated dataset, as well as the original Nichols et al. dataset were benchmarked. This was performed via a cosine similarity analysis of phenotypic profiles between genes from the same operon and those from different operons. Since genes within the same operon often have similar functions, the phenotypic profiles of genes from the same operon should have higher cosine similarity scores than genes from different operons. Using this concept, we can determine the effectiveness of ChemGAPP Big S-score assignment. By evaluating the cosine similarity scores at different thresholds, we can calculate the true positive rate and the false positive rate for the dataset. Here, the true positive rate was determined as the probability that mutants with cosine similarity scores above the threshold were from the same operon and the false positive rate being the probability that mutants with lower cosine similarity scores were from different operons. These rates were then plotted in an ROC curve and the AUC was calculated.

Having tested each dataset, no difference was observed between the AUC scores, suggesting each dataset performed equally well. The AUC for the curated and non-curated datasets were 0.6736 and 0.6738, respectively, versus an AUC of 0.6797 for Nichols et al. (Fig. 11C). Therefore, ChemGAPP is equally powerful in its assignment of fitness scores as EMAP, the software used by Nichols et al.. Genes within the same operon do not always have the same function. Often, as in the case of repressors and their targets, genes may have directly opposite functions. Therefore, using cosine similarity between operonic genes is not a perfect model and AUC scores ~0.67 are acceptable. However, to further prove the effectiveness of ChemGAPP Big, it was vital to explore if biologically relevant phenotypes were still displayed within the dataset.

3.2.5 ChemGAPP Big produces reliable and accurate representations of fitness

The aim of chemical genomic screens is to reveal biologically relevant phenotypes; therefore, it is critical that the fitness scores produced by ChemGAPP Big are accurate and reliable. If ChemGAPP Big S-scores are reliable they should recapitulate the results found within the original Nichols et al. dataset. Within the Nichols et al. screen, they discovered that deletion of the glycine cleavage (GCV) system within *E. coli* resulted in sulfonamide sensitivity. The GCV system contributes to one of the two branches in the tetrahydrofolate biosynthesis pathway and is responsible for converting tetrahydrofolate (THF) to 5,10-methylene THF. Sulfonamide drugs inhibit the THF biosynthesis pathway, thus GCV mutation is synergistic to sulfonamide

mechanism of action. The study performed further validation experiments to confirm this phenotype as true (Nichols et al., 2011). This phenotype was recapitulated within the ChemGAPP Big analysed dataset, with S-scores between -3 and -9, see Figure 12A.

Another method of validating ChemGAPP Big fitness scores is by searching for known synthetic lethal pairs. A major contributor to *E. coli* antibiotic resistance is the efflux pump AcrAB-TolC. AcrAB-TolC is known to transport multiple classes of antibiotics, including tetracyclines and chloramphenicol (Chowdhury et al., 2019; Elkins and Nikaido, 2002). Therefore, with deletion of the efflux pump components, sensitivity to substrates should be observed. This is recapitulated following S-score assignment by ChemGAPP Big. The Δ*acrA*, Δ*acrB* and Δ*tolC* mutants all showed sensitivity to chloramphenicol, tetracycline, minocycline and puromycin, with S-scores between -3 and -18 (Fig. 12B).

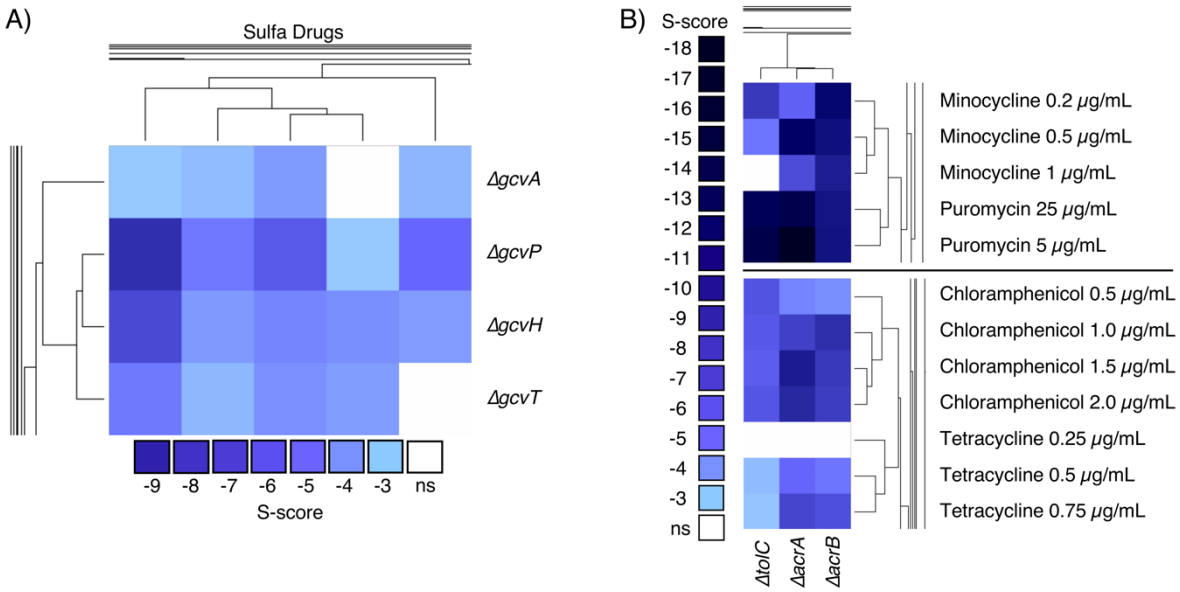


Figure 12: Clustered heatmaps displaying the S-scores for various single gene knockout mutants in different conditions. (A) The effect of sulfonamide (Sulfa) drugs on the GCV system. Conditions from left to right: Sulfamonomethoxine 100 µg/ml; Sulfamethoxazole 100 µg/ml; Sulfamethoxazole 200 µg/ml; Sulfamethoxazole 300 µg/ml; Sulfamonomethoxine 50 µg/ml. (B) The effect of AcrAB-TolC substrates on the AcrAB-TolC system mutants. ns = non-significant.

The acquisition and retainment of significant S-scores for the AcrAB-TolC and GCV systems further evidences the effectiveness of ChemGAPP Big to produce S-scores capable of revealing biologically relevant phenotypes.

3.2.6 ChemGAPP Small is better suited to small-scale screens

High-throughput and small-scale chemical genomic screens differ significantly in their design. Firstly, in large-scale screens, the aim is to have an unbiased approach, with many mutants tested against a manifold of stresses. By screening this way, a normal distribution of colony sizes is achieved as, in most conditions, mutants are unlikely to display a phenotype. Contrastingly, in small-scale screens single or limited conditions are tested, often chosen with the assumption that a phenotype will be displayed. Therefore, the conjecture surrounding normal distributions of colony size does not hold true in small-scale screens and different analysis is required. In such cases, ChemGAPP Big is not fit for purpose.

As such, ChemGAPP small was developed for the analysis of small-scale chemical genomic screens. In place of S-scores, ChemGAPP Small produces fitness ratios for the mutants against themselves on base media or against a wildtype on the corresponding condition. This allows users to robustly analyse smaller datasets without any constraints on the number of mutants or conditions screened.

3.2.7 ChemGAPP Small produces informative fitness ratios

ChemGAPP Small was validated against a small-scale screen produced on the observation within the Nichols et al. screen that Δ*envC* had decreased fitness in membrane stress conditions (Nichols et al., 2011). Within a combination of 0.5% SDS and 0.5 mM EDTA, Δ*envC* had an S-score of -8.5, indicating heavily reduced fitness.

For the current study, a small-scale screen was produced testing Δ*envC* on LB and LB + 0.25% SDS and 0.25 mM EDTA. For both conditions, Δ*envC* showed significantly reduced fitness ratios compared to the wildtype. However, within the membrane stress condition, this fitness ratio was more significantly decreased, with a ratio of 0.28 (Fig. 13). This result was reminiscent of the results achieved within the Nichols et al. screen, evidencing the capability of ChemGAPP Small to accurately interpret small-scale screening results.

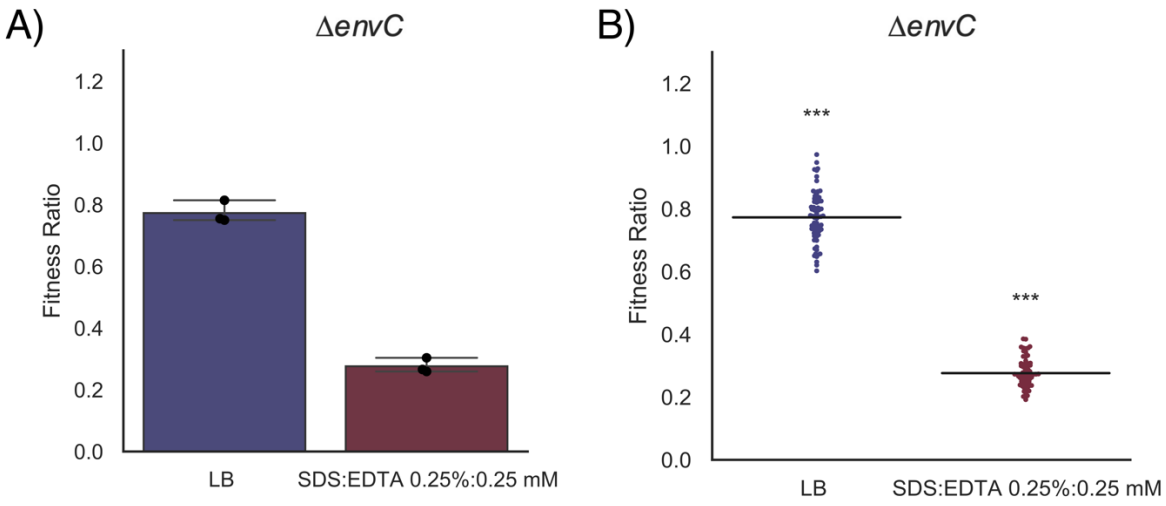


Figure 13: ChemGAPP Small provides the option of two different output plot types. (A) Bar plot output of ChemGAPP Small for Δ*envC* in LB, and 0.25% SDS + 0.25 mM EDTA, 95% confidence intervals is represented by the error bars. (B) Swarm plot output of ChemGAPP small for Δ*envC* in LB, and 0.25% SDS + 0.25 mM EDTA. ns = $p > 0.05$; *** = $0.0001 < p \leq 0.001$.

3.2.8 ChemGAPP GI successfully reveals epistasis types

The final package, capable of analysing genetic interaction screening data, is ChemGAPP GI (ChemGAPP Genetic Interactions). Genetic interaction studies aim to determine the type of epistasis between two genes; therefore, it is vital that ChemGAPP GI is able to accurately distinguish these epistasis types. In order to test this, ChemGAPP GI was employed against three previously validated screens with known epistasis types. Firstly, data from a study evaluating epistasis between the outer membrane (OM) lipoprotein *nlpI* and the peptidoglycan machineries revealed that *nlpI* did not genetically interact with *dacB,* the gene encoding penicillin-binding protein (PBP) 4 (Banzhaf et al., 2020). The Banzhaf et al. study showed a difference of 0.002 in the double observed and double expected fitness ratios. The data from this screen was reanalysed using ChemGAPP GI and the difference between the double observed and double expected fitness ratios was again insignificant at 0.011 (Fig. 14A). Within the Banzhaf et al. screen they also revealed a negative genetic interaction between *nlpI* and PBP1B encoding gene *mrcB*, with a ratio difference of 0.62. This result was similarly achieved following ChemGAPP GI analysis, with a significant difference in the double observed and expected ratios of 0.52, see Figure 14B. Finally, a study by Bryant et al. described a positive genetic interaction between the OM lipoprotein *bamB* and the DNA replication activator *diaA* (Bryant et al., 2023). This data was similarly reanalysed using ChemGAPP GI and revealed a significant increase in the double observed fitness ratio from the double expected fitness ratio, displaying a positive interaction (Fig, 14C). Therefore, it is clear that ChemGAPP GI is capable of accurately predicating all epistasis types.

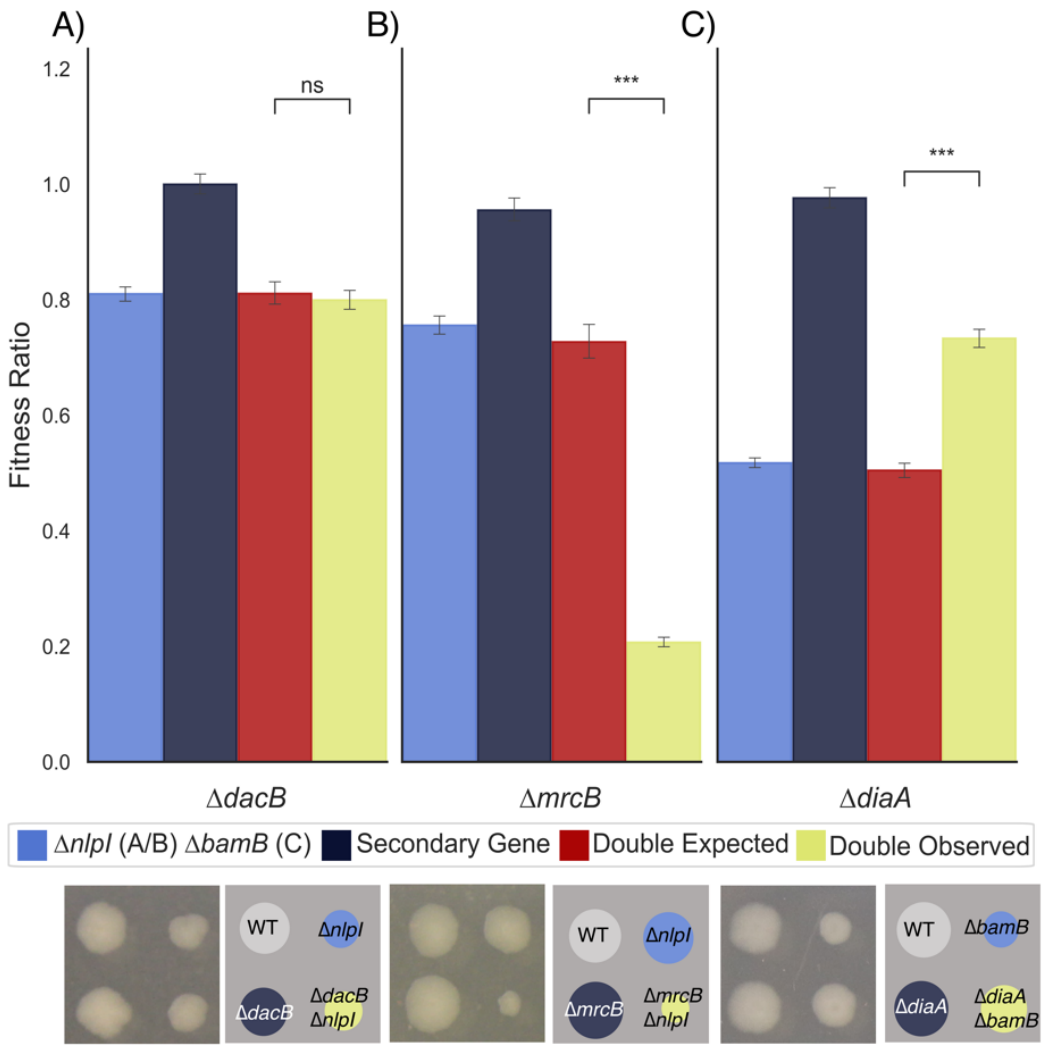


Figure 14: Genetic interaction plots produced by ChemGAPP GI (A) Difference between the double expected and double observed fitness ratios is non-significant between *dacB* and *nlpI*, indicating no epistasis (B) negative epistasis between *mrcB* and *nlpI*. Shown by significant reduction in double observed ratio from double expected (C) Increased fitness within double observed compared to double expected, indicating positive epistasis between *diaA* and *bamB*. ns = $p > 0.05$; *** = $0.0001 < p \leq 0.001$.

3.3 Discussion

Designed to bridge the gap for an easy-to-use and comprehensive chemical genomics analysis package, ChemGAPP is fully suited to purpose. Having designed separate modules for multiple chemical genomic screening types, ChemGAPP allows a wider scientific audience to engage in chemical genomic techniques. The current lack of software had limited the field to researchers with computational expertise who were able to write and implement in-house scripts for analysis. ChemGAPP, however, is easy-to-use, accessible, and flexible since it is available in two formats. Firstly, the software can be run through a Streamlit GUI, meaning no knowledge of coding or programming is required for end users, widening ChemGAPP's scope. Secondly, the package is available as command line modules, giving the user flexibility to choose the features they wish to implement.

In a recent study, another group similarly designed a software for the analysis of plate-based high-throughput screens with the aim of replacing deprecated analysis software (Klemm et al., 2022). However, this software, ScreenGarden, was not designed specifically for chemical genomic screens and was instead tested on Synthetic Physical Interaction (SPI) screen data (Klemm et al., 2022). Additionally, instead of S-scores, ScreenGarden calculates log growth ratios and Z-scores which are compared to one or two control plates to produce a fitness value. As previously described, for high-throughput chemical genomic screens, this is not as effective a fitness measure as S-scores which consider phenotype across all conditions. However, should comparison to a control plate or fitness ratios be needed, ChemGAPP Small covers

this requirement. As such, ChemGAPP is a more comprehensive tool and thus more suited to chemical genomic analysis.

In conclusion, this doctoral research has led to the introduction of ChemGAPP a comprehensive, user-friendly wrapper software dedicated to the analysis of a variety of chemical genomic screen types. Each of ChemGAPP's sub-packages were intentionally designed to allow a wider scientific audience to engage in chemical genomic techniques by streamlining each dedicated analysis pathway. Furthermore, development of ChemGAPP allowed for the analysis of two chemical genomic screens within *P. aeruginosa* in a uniform manner, allowing for accurate comparison during this doctoral research.

CHAPTER 4

CHEMICAL GENOMIC ANALYSIS AND COMPARISON OF TWO *PSEUDOMONAS AERUGINOSA* STRAINS.

4.1 Introduction

4.1.1 Genotypic similarity between PAO1 and PA14

Although the *P. aeruginosa* species comprises of thousands of strains, it has two reference strains that are commonly studied within the field, PAO1 and UCBPP-PA14 (henceforth referred to as PA14) (Grace et al., 2022). Both strains were originally isolated from wounds, with PAO1 being first isolated in Melbourne, Australian in 1954 and PA14 being isolated in a hospital in Pennsylvania, USA (Holloway, 1955; Mathee, 2018; Grace et al., 2022). PAO1 and PA14 were the first two published genomes for *P. aeruginosa* and are as such the most widely studied *P. aeruginosa* strains (Stover et al., 2000; Lee et al., 2006; Klockgether et al., 2011).

PAO1 and PA14 have significant similarities in sequence with their highly conserved core genome representing roughly 90% of genes (Dettman and Kassen, 2021; Grace et al., 2022). The major difference described between the two strains is the increased virulence of PA14 compared to PAO1 (Grace et al., 2022). Although PA14 encodes 58 gene clusters which are absent in PAO1, including two pathogenicity islands PAPI-1 and PAPI-2, this difference in virulence is not solely determined by the presence of genes within the accessory genome of PA14 (Lee et al., 2006; Mikkelsen et al., 2011). The 58 clusters encode 478 genes that are absent in PAO1. On the other hand, there are 54 clusters and 234 genes encoded by PAO1 that are not present in PA14, see Figure 15 (Lee et al., 2006). To evaluate the role of the accessory genome in virulence, one study performed a microarray of 285 PA14 and 130 PAO1 specific sequences across a set of 20 *P. aeruginosa* strains with varying virulence. Limited correlation was observed between the presence of specific genes and the virulence of the strains within

C. elegans (Lee et al., 2006). This suggested that presence of specific genes was not predictive of virulence and that transcriptional studies would be required to understand the difference in virulence phenotype (Lee et al., 2006).

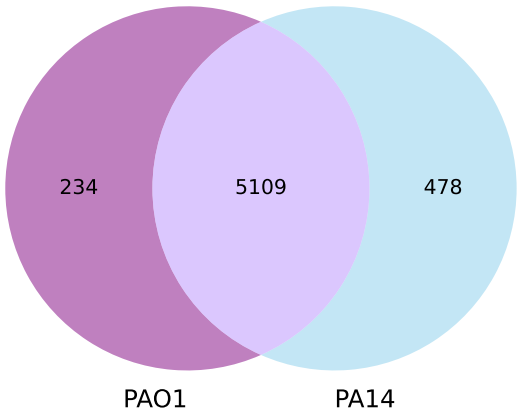


Figure 15: Venn diagram showing the number of genes unique to PAO1 and PA14 and the number of conserved genes. (Based on Grace et al., 2022; Lee et al., 2006)

Later studies focused on the transcription and regulation of genes to highlight differences between PAO1 and PA14. One study performed RNA-Seq and discovered differential expression of 2029 genes, equating to 39% of the *P. aeruginosa* genome. A major upregulation in *qslA* expression was observed between PAO1 and PA14; a 55.8-fold difference was displayed with PA14 barely expressing the gene (Sana et al., 2019). QslA binds to the previously described QS regulator LasR and prevents dimerization and promoter binding. The increased expression of *qslA* within PAO1 led to decreased production of quorum-sensing-stimulated virulence factors in comparison

to PA14. As such, the study concluded that *qslA* expression was a major determinant of virulence differences (Sana et al., 2019).

Another study linked the virulence difference to a frameshift mutation within the PA14 *ladS* gene, which results in overexpression of the Type III Secretion System (T3SS) and a repression of biofilm formation. This T3SS is responsible for excreting the cytotoxic effectors ExoS, ExoU, ExoT and ExoY, partially explaining the increased virulence of PA14 (Mikkelsen et al., 2011).

Other studies have linked filamentous prophages to dissimilarities between PAO1 and PA14, which harbour the prophages Pf4 and Pf5, respectively (Mooij et al., 2007; Rice et al., 2009; Li et al., 2019). Both prophages have integrated into their respective host chromosomes and affect virulence and biofilm formation. It is well documented that prophage genes are highly upregulated during biofilm development and thus heavily effect *P. aeruginosa* in this environment (Webb et al., 2004). For example, Pf4 in PAO1 has been shown to be associated with small colony variant (SCV) formation within biofilms. These cells adapt into a hyper-biofilm-forming phenotype, displaying Pf4 on the cell surface, increasing adhesion and aiding in the dispersal and formation of new biofilms (Webb et al., 2004; Kirisits et al., 2005). Pf4 and SCV has therefore also been associated with increased antibiotic resistance (Mooij et al., 2007). Counter to this, despite its homology to Pf4, no correlation between Pf5 and SCV formation has been observed for PA14 (Mooij et al., 2007; Rice et al., 2009).

It is clear that there are multiple factors which introduce dissimilarities in phenotype within *P. aeruginosa* strains, despite their strong genetic similarities. This study aimed to further characterise the differences in PAO1 and PA14, whilst simultaneously

evaluating the function of unknown genes, antibiotic resistance mechanisms and drug targets within both strains. This study revealed multiple biologically relevant phenotypes and further evaluated two of these hits.

4.1.2 Loss of *Pa*SAH induces sulfonamide sensitivity within PA14 but not PAO1.

The first biologically relevant phenotype that was evaluated was the increased sensitivity to sulfonamide drugs within PA14 with the loss of the small alarmone hydrolase *Pa*SAH. Sensitivity was found to sulfamethoxazole and co-trimoxazole (sulfamethoxazole and trimethoprim combination) but not trimethoprim alone. Furthermore, no sensitivity was seen to any drug within PAO1.

4.1.2.1 Sulfamethoxazole and trimethoprim mechanism of action.

Sulfamethoxazole and trimethoprim both target different steps within the tetrahydrofolate biosynthesis pathway (Masters et al., 2003). Since the two drugs have synergistic activity, they are most commonly used as a co-therapy within the clinical setting under the name co-trimoxazole (Fernández-Villa et al., 2019). THF is a vital one-carbon donor for all living organisms and functions within many biosynthetic pathways, including the synthesis of nucleic acids and certain amino acids. It is known to participate in the conversion of serine to glycine, the catabolism of histidine, and the synthesis of glutamate, methionine, and thymidylate. It is also involved in tRNA

synthesis, making it a major contributor to protein synthesis (Lucock, 2000). On top of its involvement with purine and pyrimidine synthesis, it also regulates associated gene expression (Fernández-Villa et al., 2019).

THF is formed via two methods, de novo synthesis, which is exclusive to bacteria, and via the turnover of dietary folic acid. Sulfamethoxazole targets de novo synthesis, whereas trimethoprim targets both. As such, trimethoprim is used at lower concentrations within co-trimoxazole co-therapy to reduce off target effects within the host (Fernández-Villa et al., 2019). In de novo synthesis, p-aminobenzoic acid (PABA) reacts with dihydropteroate pyrophosphate (DHP-PPi) to form dihydropteroate (DHP), this is mediated by the enzyme dihydropteroate synthase (DHPS). Sulfamethoxazole, as a structural analogue of PABA, competes for DHPS binding at the active site, inhibiting DHP formation, see Figure 16 (Fernández-Villa et al., 2019; Masters et al., 2003). Further to this, sulfamethoxazole is able to form a dead-end complex with DHP-PPi, sequestering it permanently (Minato et al., 2018). Given DHP formation, the pathway then continues via the formation of dihydrofolate (DHF) from DHP via the enzyme dihydrofolate synthase (DHFS). DHP is then converted to THF by dihydrofolate reductase (DHFR). Trimethoprim is a DHFR inhibitor, and it competitively inhibits substrates of DHFR, including DHP and folic acid from the dietary route (Fernández-Villa et al., 2019; Masters et al., 2003).

Despite co-trimoxazole's broad use, it is not currently effective against *P. aeruginosa.* This has previously been attributed to the range of efflux pumps encoded by *P. aeruginosa* (De Kievit, 2009; Llanes et al., 2011). However, the increased sensitivity of PA14 with the loss of *Pa*SAH could suggest a further intrinsic method within this strain.

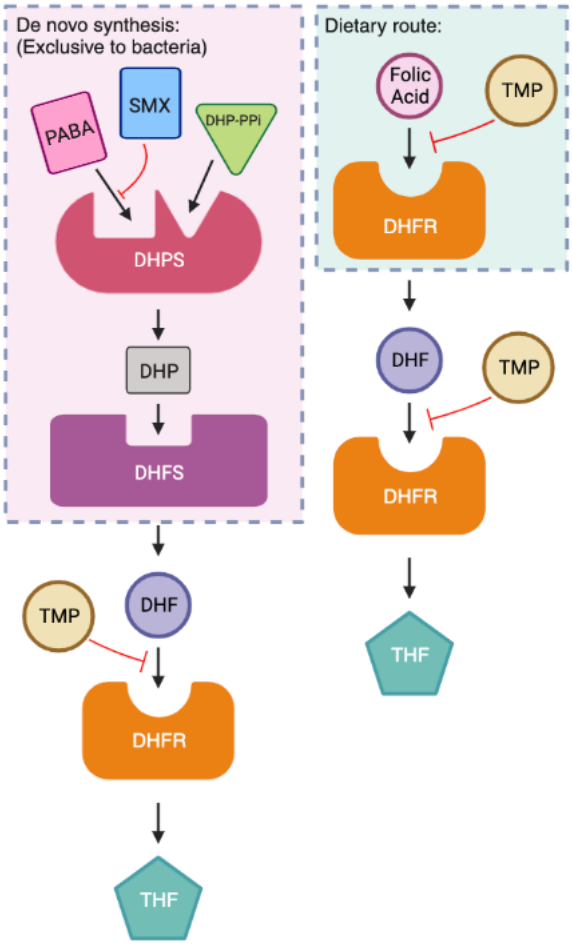


Figure 16: Overview of the folate biosynthesis pathway and the targets of sulfamethoxazole and trimethoprim. Abbreviations: DHP = dihydropteroate, DHP-PPi = dihydropteroate pyrophosphate, DHPS = dihydropteroate synthase, DHF = dihydrofolate, DHFS = dihydrofolate synthase, DHFR = dihydrofolate reductase, PABA = p-aminobenzoic acid, SMX = sulfamethoxazole, THF = tetrahydrofolate, and TMP = trimethoprim (Adapted using Biorender from Fernández-Villa et al., 2019)

4.1.2.2 *Pa*SAH and the stringent response.

When bacteria encounter starvation, it is vital that non-essential biosynthetic pathways are shut down and the transcription of non-essential genes is downregulated. The stringent response comprises of a stress signalling system which responds to this nutrient depletion and mediates cell functions via molecules called alarmones, see Figure 17. These alarmones are guanosine triphosphate (pGpp), guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (pppGpp), known collectively as (pp)pGpp (Irving et al., 2021). However, little is known about pGpp, thus most research has been performed on (p)ppGpp, which includes ppGpp and pppGpp. (p)ppGpp is able to modulate biosynthetic transcription and translation activity in a myriad of ways. Firstly, (p)ppGpp can directly bind to RNAP and regulate promoter activity. By doing so, (p)ppGpp inhibits the transcription of rRNA and ribosomal proteins and enhances amino acid biosynthesis gene expression (Bange et al., 2021). (p)ppGpp further enhances amino acid availability by binding and inhibiting amino acid decarboxylases involved in the conversion of amino acids to polyamines (Bange et al., 2021). Further to this, (p)ppGpp couples nucleotide synthesis inhibition with amino acid enhancement. During de novo purine synthesis, the enzyme PurF mediates the first catalytic step converting 5-phosphoribosylpyrophosphate (PRPP) to 5-phospho-D-ribosylamine (PRA). (p)ppGpp allosterically binds PurF enabling it to adopt an inactive conformation. In addition to this, (p)ppGpp also binds to and inhibits HPRT, an enzyme in the purine salvage pathway that also converts PRPP. By preventing the flux of PRPP through the purine biosynthesis pathways, it can be repurposed towards histidine and tryptophan biosynthesis (Bange et al., 2021). Within Gram-negative bacteria, (p)ppGpp not only inhibits purine synthesis but enhances purine degradation by binding and

activating PpnN, which is able to hydrolyse IMP and GMP into their base and phosphoribose portions (Bange et al., 2021). Furthermore, much of the GMP, GDP and GTP is sequestered during stringent response to produce (pp)pGpp, further reducing purine pools (Irving et al., 2021).

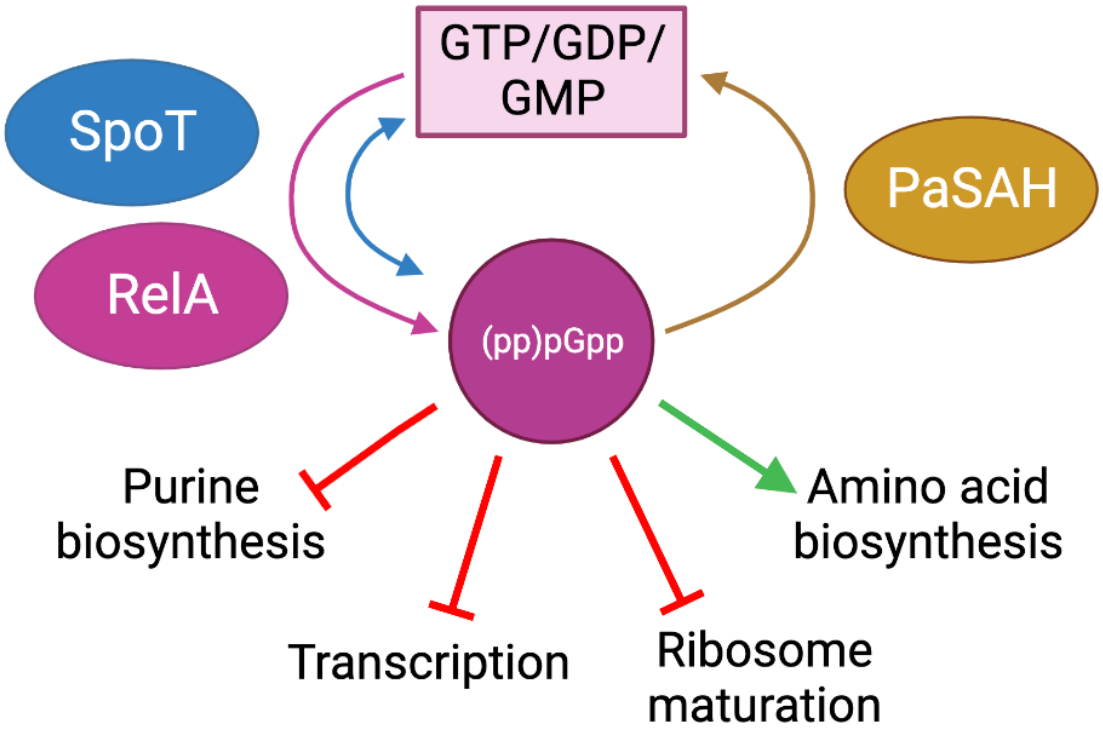


Figure 17: Stringent response within *P. aeruginosa.* SpoT and RelA synthesise (pp)pGpp and SpoT and *Pa*SAH hydrolyse (pp)pGpp back to GTP,GDP and GMP. (pp)pGpp synthesis inhibits purine biosynthesis, transcription and ribosome maturation whilst enhancing amino acid biosynthesis (Created using Biorender).

These alarmones are processed by three types of enzymes, RelA/SpoT homologue (RSH) family enzymes, small alarmone synthases (SASs) and small alarmone hydrolases (SAHs). The latter two families each have a single function, SAS enzymes are involved in the synthesis of (pp)pGpp and SAHs hydrolyse (pp)pGpp. RSH family enzymes encode both a synthetase and hydrolase domain, meaning they can both degrade and synthesise (pp)pGpp to or from GMP,GDP, and GTP (Bange et al., 2021; Irving et al., 2021). *P. aeruginosa* encodes both SpoT and RelA, however, *P. aeruginosa* RelA has an inactive hydrolase domain, meaning it can only perform synthetase activity (Steinchen et al., 2021). It was recently discovered that *P. aeruginosa* encoded a SAH with dual functions, called *Pa*SAH. *Pa*SAH can degrade both (p)ppGpp and (p)ppApp, an alternative phosphorylated purine molecule which is associated with bacterial competition via type 6 secretion systems (T6SS) (Steinchen et al., 2021). As such, *Pa*SAH has been proposed to be a weaker alternative antitoxin within the Tas1-Tis1 toxin-antitoxin pair. (p)ppApp accumulation is toxic to cells, and Tas1, a T6SS effector protein, is the SAS responsible for its production (Steinchen et al., 2021). As such, *Pa*SAH has been implicated in both protection from leaky Tas1 expression as well as the stringent response.

4.2 Results

4.2.1 Chemical genetic screening of two *P. aeruginosa* strains.

Two single-gene transposon mutant libraries of *P. aeruginosa,* representing the strains PAO1 and PA14 (Jacobs et al., 2003; Liberati et al., 2006), were screened against

various stressor conditions in a chemical genomic screen performed by Dr Manuel Banzhaf. Conditions were chosen to provide a wide range of stress types such that the conditional essentiality of genes within the two strains could be investigated and compared. A similar number of conditions were analysed for both PAO1 and PA14, as shown in Figure 18. Although the screens were performed in parallel some plates were excluded at the screening stage for PA14 and PAO1, resulting in a slight disparity in the conditions analysed. For PAO1, 135 base conditions were tested, not counting multiple concentrations, whereas 116 conditions were analysed for PA14. The most common condition types tested for both strains were antibiotics and carbon/nitrogen sources.

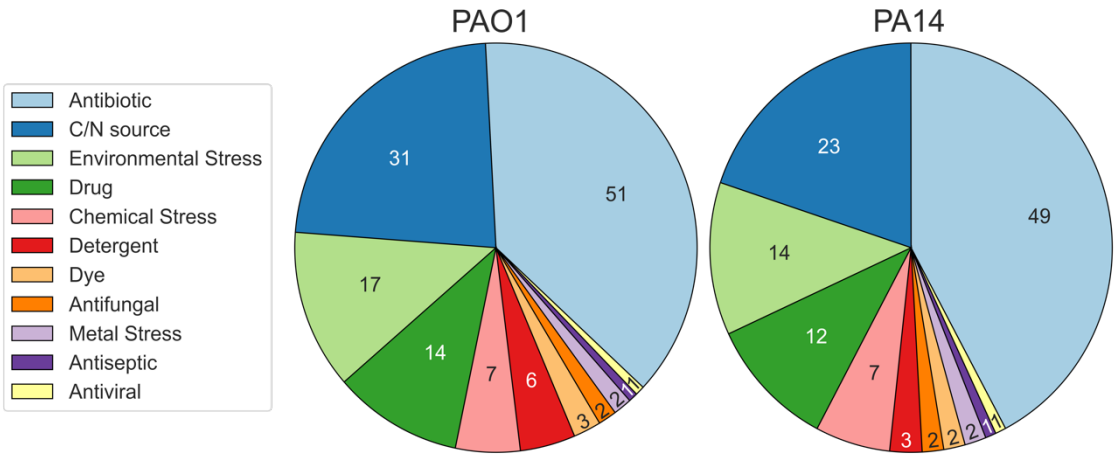


Figure 18: Pie chart displaying the classification of screening conditions for the PAO1 (left) and PA14 (right) screens. C/N source = carbon/nitrogen source.

4.2.2 Chemical genomics produces reliable datasets for the acquisition of phenotype data.

Since high-throughput screening requires the use of robots to pin thousands of plates, errors can occur which create artifacts. Therefore, it is vital to ensure that replicate reproducibility is acceptable before scoring chemical genomic datasets. Outlier replicates, due to artifacts such as pinning errors, can skew scores resulting in the acquisition of false hits. As such, we tested for artifacts by analysing the Pearson correlation coefficient between all sets of replicates. Figure 19 displays the Pearson correlation r values representing how correlated the replicates are, with values closer to 1 being more correlated. Pre-curation PAO1 and PA14 had r values of 0.693 and 0.718, respectively. Following curation by ChemGAPP Big the replicate reproducibility was improved for both PAO1 and PA14, with r values of 0.714 and 0.742, respectively.

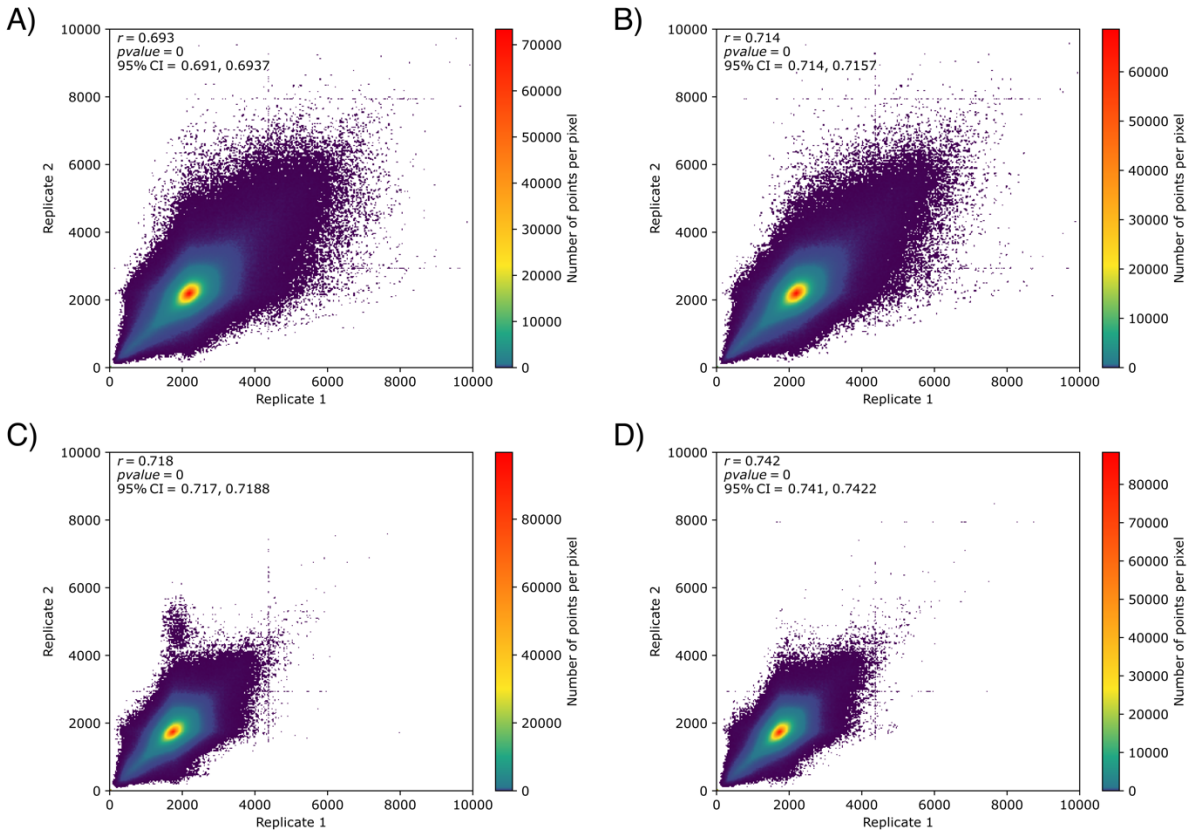


Figure 19: Heatmap-based scatterplots comparing replicate colony sizes. Colour bar represents the number of points per pixel. r = Pearson correlation coefficient score of normalised replicate colony sizes. A) PAO1 non-curated dataset B) PAO1 curated dataset C) PA14 non-curated dataset D) PA14 curated dataset.

Following curation, the datasets were given S-scores as a measure of fitness. In order to determine which scores were significant the 5% FDR cut-off employed by Nichols et al. was utilised. S-scores that were deemed significant were counted as phenotypes. In the PAO1 dataset, 19,898 phenotypes were achieved accounting for 1.23% of all S-scores. For PA14, 9765 phenotypes were achieved, comprising 1.04% of all S-scores. For PAO1 4537 mutants showed at least one phenotype accounting for 48.1% of

mutants. For PA14, 2318 mutants showed at least one phenotype, accounting for 41.8% of mutants, see Figure 20. Of these mutants, some displayed phenotypes in at least 30 conditions and were classified as multiple stress responsive (MSR) genes, see Supplementary Table 1 and 2. The number of MSR genes was 19 for PAO1 and 14 for PA14, representing 0.2% and 0.25% of all mutants respectively. Since both datasets often include more than one transposon mutant for some of the genes, the count of mutants with at least one phenotype was recalculated with mutants for the same gene combined. Following this, the number of mutants with at least one phenotype for PAO1 was 3241. The dataset included 4773 unique genes, meaning 67.9% of mutants showed at least one phenotype. The PA14 dataset included 4435 unique genes and 2066 mutants with at least one phenotype, accounting for 46.5% of all mutants. As such, PAO1 displayed a more responsive genome than PA14 with a difference of 21.4% in mutants with phenotypes.

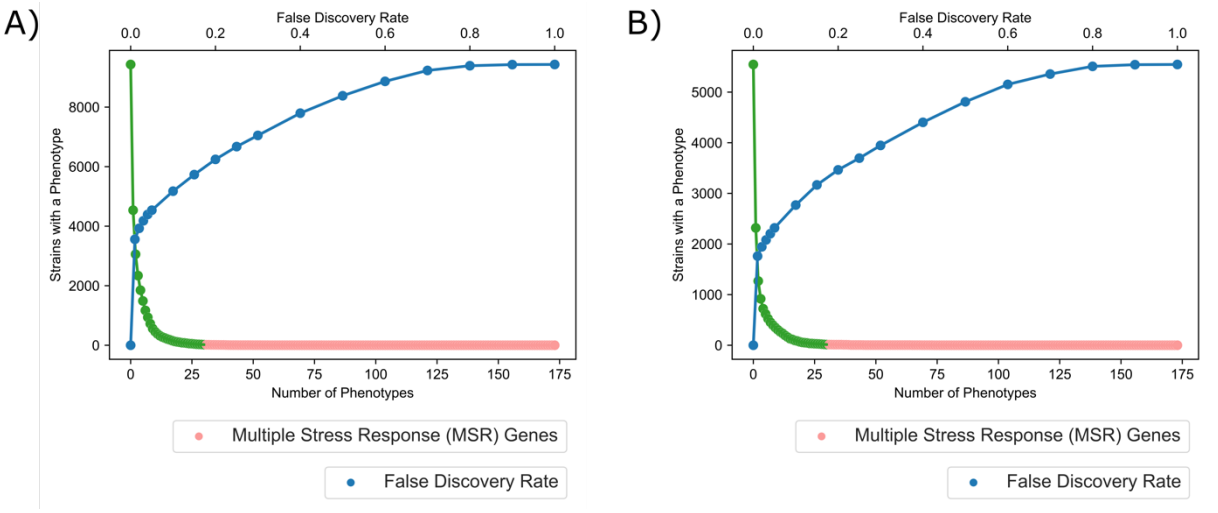


Figure 20: Number of strains with a phenotype at different false discovery rates (blue). At a 5% FDR the number of strains with specific numbers of phenotypes is displayed (green) with multiple stress responsive genes (≥ 30 phenotypes) highlighted in pink. A) PAO1 matching curated dataset B) PA14 matching curated dataset.

It was vital to benchmark both datasets following S-score assignment in order to understand their quality. The datasets were benchmarked against four categories: replicates, transcriptional units, pathways, and 3-letter annotations, see Figure 21. Replicates were defined as mutants with transposon insertions within the same gene but different insertion locations. Transcriptional units were extracted from BioCyc. The KEGG pathways for the PAO1 and PA14 genes were extracted from Pseudomonas.com, however, pathways with more than 200 genes were excluded from the analysis. Finally, the 3-letter annotations were determined as the first three letters of each annotated gene, 3-letter gene categories with less than 3 or more than 100 genes were excluded. Three-letter annotation was chosen instead of protein

complexes due to the limited annotation for complexes within *P. aeruginosa.* However, three-letter nomenclature often represents complexes, pathways or gene functions and therefore was chosen as a proxy. For all benchmarked categories, mutants with transposon insertions in the last 20% of the gene were excluded, as well as mutants which displayed no phenotypes across all conditions. PA14 gene annotation was matched to the PAO1 annotations since it is less well annotated. Pearson correlation coefficients of phenotypic profiles of mutants with corresponding category values were compared to randomised pairs of mutants with different category values to produce ROC curves.

In all categories, PA14 outperformed PAO1, suggesting that in general the PA14 dataset is more reliable than the PAO1 dataset. However, both datasets performed well with replicate mutants showing higher correlation than randomised pairs, with AUC values of 0.661 and 0.724 for PAO1 and PA14, respectively. Both datasets were able to recapitulate transcriptional units, as generally higher correlation values were seen for genes from the same transcriptional units than randomised pairs. This is displayed by their AUC values of 0.61 and 0.672 for PAO1 and PA14, respectively. Similarly, within both datasets, mutants which shared the three-letter nomenclature were predicted with fair accuracy, showing AUC values of 0.602 and 0.638 for PAO1 and PA14, respectively. The only measure which scored poorly and showed poor recapitulation capability was KEGG pathways, with AUC scores of 0.51 for PAO1 and 0.532 for PA14. Since the narrower categories score higher, this is likely due to the broad nature of KEGG pathways and the poor annotation of the *P. aeruginosa* genome. However, in general these benchmarks suggest that the datasets are of sufficient quality and are able to recapitulate known biological interactions. As such we would

expect that if PAO1 and PA14 show similar phenotypic responses this can be captured by comparing various aspects of the datasets and that differences are not due to poor dataset quality.

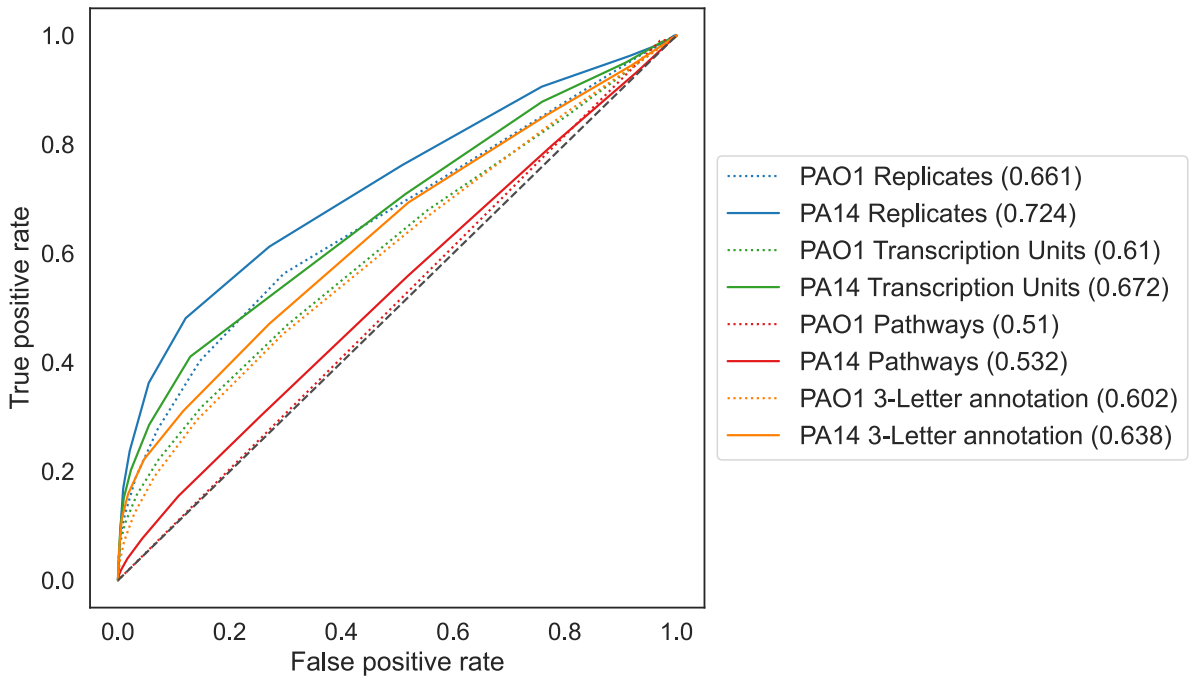


Figure 21: ROC curve of benchmark categories for PAO1 and PA14. Dotted lines represent PAO1, and solid lines represent PA14. Grey dashed line represent random assignment. Legend denotes category and area under the curve in brackets.

Since the KEGG categories were so broad, the distribution of r values for each KEGG category were tested to see if any categories had higher percentages of r values ≥ 0.4. The average percentage of gene pairs with a correlation coefficient ≥ 0.4, excluding pathways with less than 10 comparisons was calculated. For PA14 this was 5.34%.

The top three categories for percentage gene pairs ≥ 0.4 was ‘vitamin B6 metabolism’, ‘pentose phosphate pathway’ and ‘oxidative phosphorylation’, with percentages of 26.7%, 23.1% and 23.0%, respectively (Fig. 22). For PAO1, when excluding pathways with less than 10 comparisons, ‘ubiquinone and other terpenoid-quinone biosynthesis’, ‘starch and sucrose metabolism’, and ‘aminobenzoate degradation’ had the highest percentage of r values ≥ 0.4. However, the percentages were only 11.5%, 10.7%, and 9.5% of gene pairs, respectively. The average here was 3.07%.

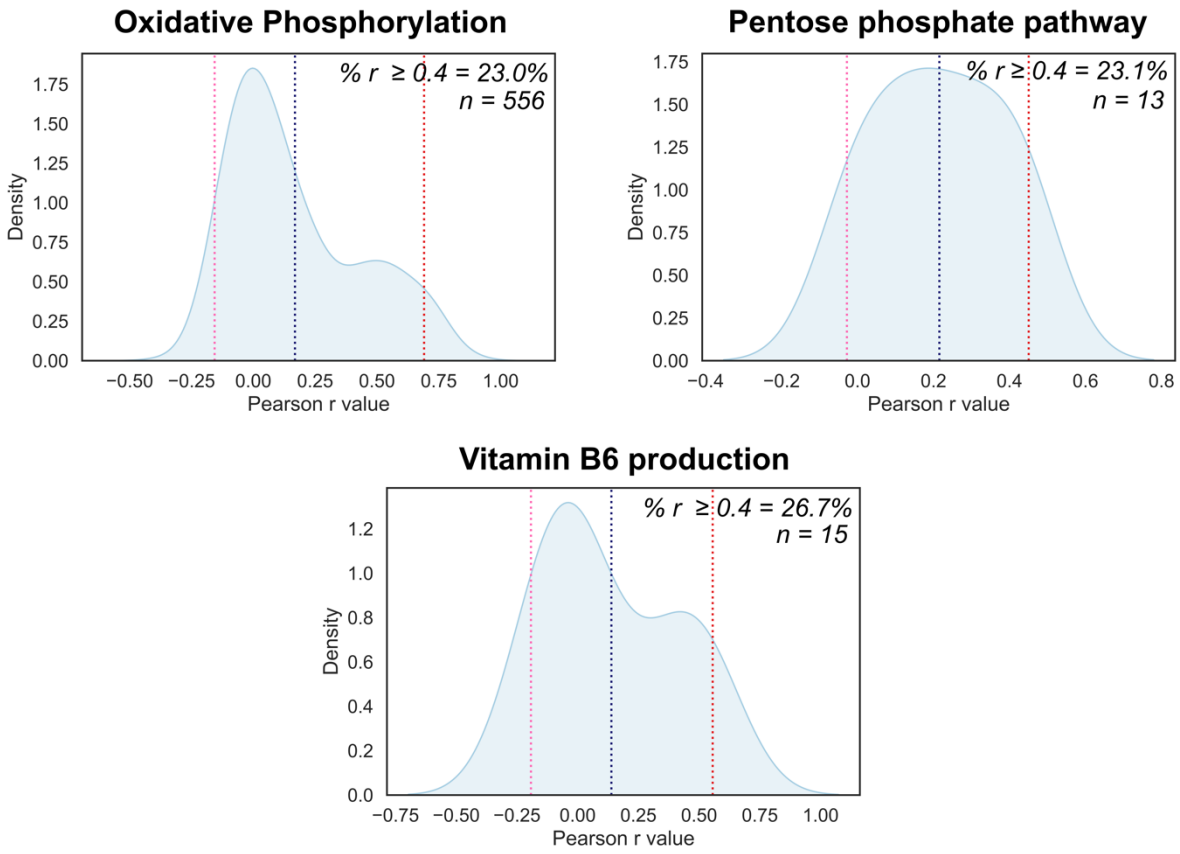


Figure 22: Distribution of Pearson correlation coefficient r values for the KEGG categories with the highest percentage of gene pairs with r values ≥ 0.4 for PA14.

Centre dark blue dotted line represents mean r value, lefthand pink dotted line represents 5th percentile and righthand red dotted line represents 95th percentile.

4.2.3 Comparison of PAO1 and PA14.

Despite their large genetic cross-over, PAO1 and PA14 have been shown to have differing lifestyles and pathogenicity. Therefore, this study aimed to further understand and characterise the differences between PAO1 and PA14 by evaluating their phenotypic responses to multiple conditions. In addition, by using two closely related strains, the ability of chemical genomics as a predictive tool between strains will be evaluated. As such the PAO1 and PA14 datasets were compared via multiple methods to evaluate the phenotypic similarity of PAO1 and PA14.

4.2.3.1 Comparison of multiple stress responsive genes across PAO1 and PA14.

If PAO1 and PA14 are both genetically and phenotypically similar we would expect orthologous MSR genes to be elucidated within the screen since their functions would be equally important within both strains. As such, the functions of the MSR genes for PAO1 and PA14 were collected for comparison. PseudoCAP functions were extracted from Pseudomonas.com and assigned to the MSR genes. For PAO1, the most common PseudoCAP categories were ‘Hypothetical’ genes and genes involved in ‘Motility and Attachment’, with 5 genes for both categories (Fig. 23). Contrary to this, the most common category for PA14 was ‘Transport of small molecules’ with 3 genes.

The only common categories were ‘Hypothetical’, ‘Cell wall/LPS/Capsule’, ‘Protein secretion/export apparatus’ and ‘Putative enzymes’. Despite this crossover, none of the genes were orthologous across the two strains. Furthermore, when calculating the Pearson correlation coefficient between the pathway counts for PAO1 and PA14 a non-significant correlation of -0.17 was achieved, meaning the two strains do not correlate. The mutant with the most phenotypes in PAO1 was the peptidoglycan biosynthesis gene *murA* with phenotypes in 78 conditions. In PA14 a mutant of the putative enzyme gene PA14_34540 displayed the most phenotypes, with a phenotype in 73 conditions. One MSR gene which was present in the PA14 dataset was a PA14 unique gene, PA14_15600. PA14_15600 is a hypothetical gene but has a bacteriophage T5 Orf172 domain, suggesting it could be from phage origin. The lack of crossover in the MSR genes suggests that although PAO1 and PA14 are genetically very similar, that their conditional phenotypic responses differ. Thus, providing evidence that simply examining genetic sequence is insufficient to understand the nuance of *P. aeruginosa* strain lifestyles.

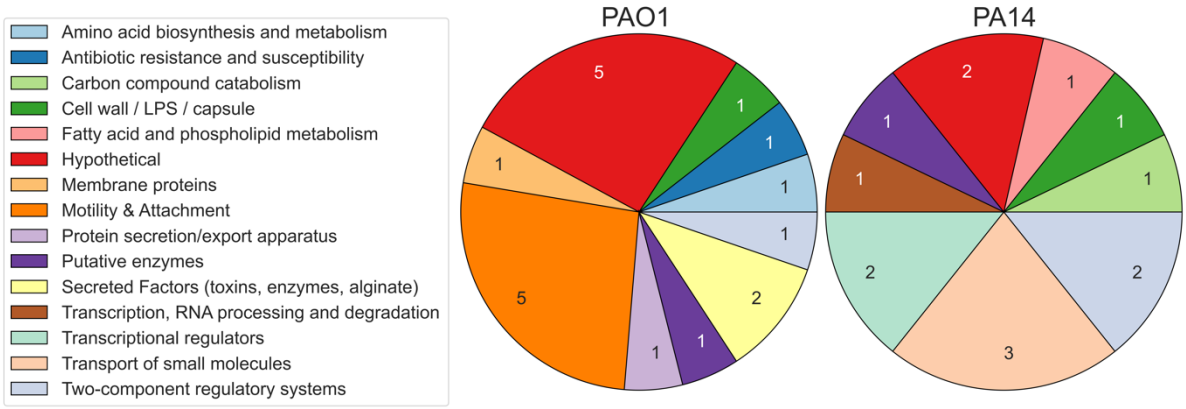


Figure 23: Functions of MSR genes for PAO1 and PA14. Functions based on PseudoCAP functions. Each gene has only one function assigned.

4.2.3.2 PAO1 and PA14 phenotypic profiles can be distinguished by t-SNE

Since chemical genomic datasets are produced using multiple conditions, the datasets have large dimensionality, with each condition representing one dimension of the data. Therefore, it is often helpful to use dimension reduction tools to analyse and understand these datasets. At first, a principal component analysis (PCA) was conducted to attempt to reduce the dimensionality of the data. However, following analysis only 12.7% of the variance was explained by the first two principal components. As such, it was decided that a non-linear dimensionality reduction tool should be attempted instead. Therefore, the t-SNE algorithm was applied to both curated matching datasets for PAO1 and PA14. As such, the dataset was reduced from 174 dimensions to two. This allowed each phenotypic profile to be represented by a single point on a two-dimensional plot, see Figure 24. If PAO1 and PA14 had strong phenotypic similarity then the data points for the two strains would have a clear

overlap. However, from Figure 24, there is a clear separation between data points for PAO1 and PA14. This provides further evidence of the disparity between the phenotypic response of PAO1 and PA14.

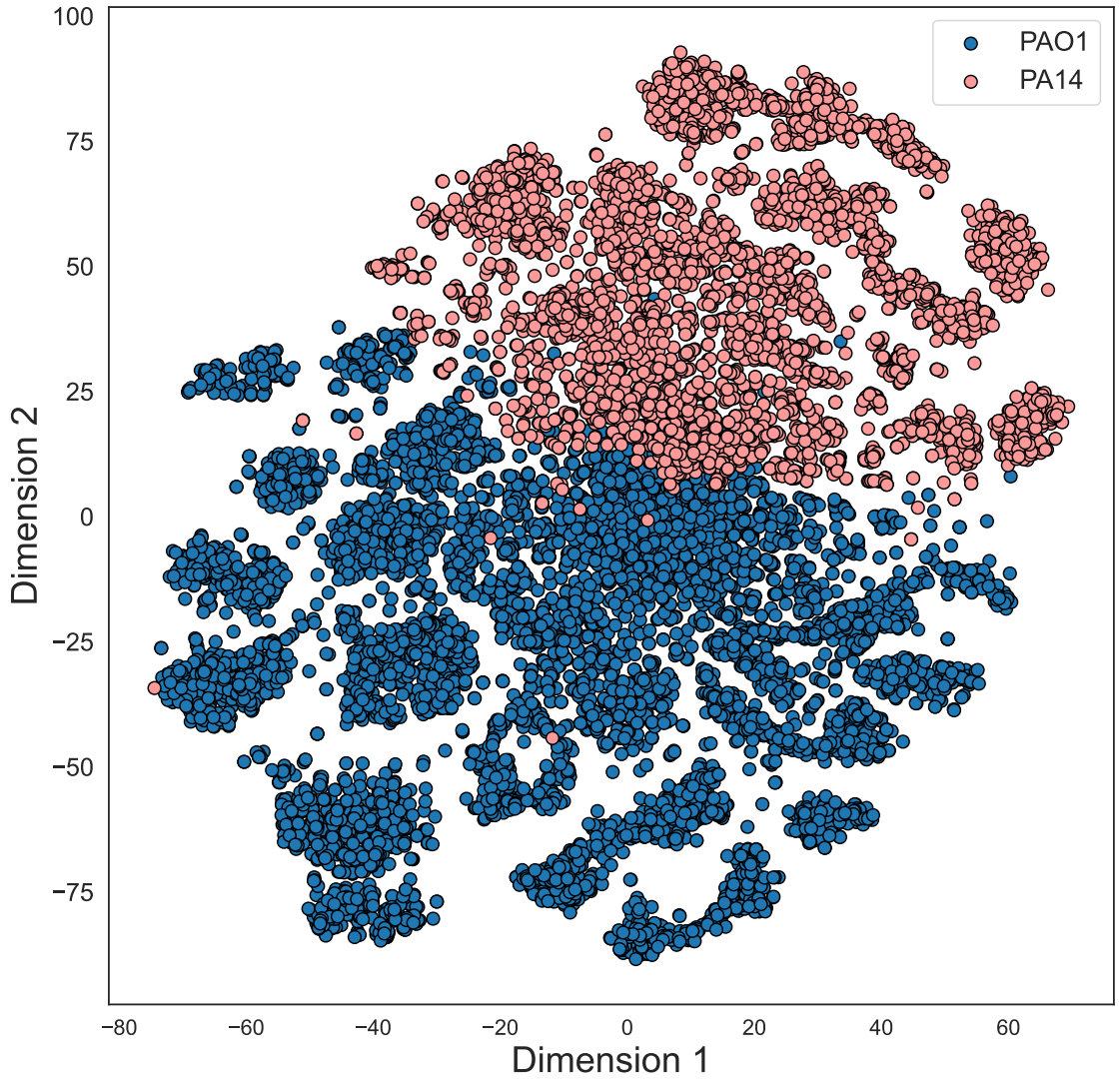


Figure 24: t-SNE plot of PAO1 and PA14. Blue dots represent PAO1 phenotypic profiles and pink dots represent PA14 phenotypic profiles.

4.2.3.3 PAO1 and PA14 ortholog correlation.

Following the separation of PAO1 phenotypic profiles from PA14 via t-SNE, the phenotypic profiles of orthologs were compared to see if any particular genes had similarity across the datasets. Since PA14 is not as well annotated as PAO1, all PA14 orthologs were given the PAO1 annotation. Only genes with a known gene name which were present in both datasets were included and any mutant with a transposon insertion within the last 20% of the gene or without any phenotypes were excluded. The Pearson correlation coefficient was then calculated between the phenotypic profiles of each ortholog. Where multiple transposon mutants of the same gene were in the dataset, all ortholog comparisons were included. A density plot of the Pearson correlation coefficient r values was produced to evaluate the distribution of r values, see Figure 25. The mean r value was 0.075, and the 5^{th} and 95^{th} percentiles were -0.126 and 0.324, respectively. This highlights that the majority of orthologs show little to no correlation across the two strains.

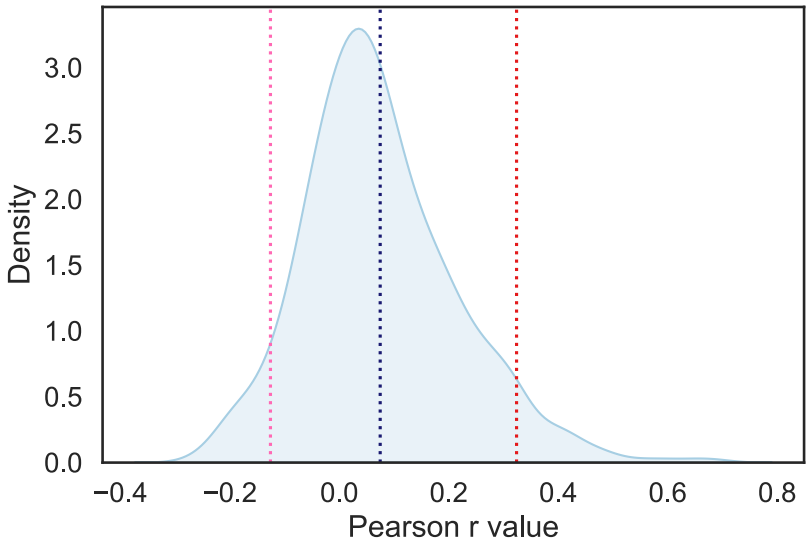


Figure 25: Density plot of Pearson r values for ortholog phenotypic profiles between PAO1 and PA14. Centre dark blue dotted line represents mean r value, lefthand pink dotted line represents 5th percentile and righthand red dotted line represents 95th percentile.

A small number of genes (22) within this analysis had an ortholog r value greater than 0.4, see Supplementary Table 3. The PseudoCAP functions of these genes were assigned to each gene and the count of each function type calculated, see Figure 26. The most common function was transport of small molecules, with 5 genes associated. Two of these genes included MexB and OprM, two members of the major efflux pump MexAB-OprM. Although MexA was not included, as it had an r value of 0.362, this value was within the top 5% of r values. Furthermore, MexB showed the highest correlation of all orthologs within this analysis, with an r value of 0.68. The next two highest classes were “Amino acid biosynthesis and metabolism” and “Carbon compound catabolism”. Again, with the second highest condition type within the screen being carbon or nitrogen limitation, this could be explained by the conditions chosen.

It may be that gene disruption is leading to auxotrophy and thus strong phenotypes within these limitation conditions carry more weight for the correlation.

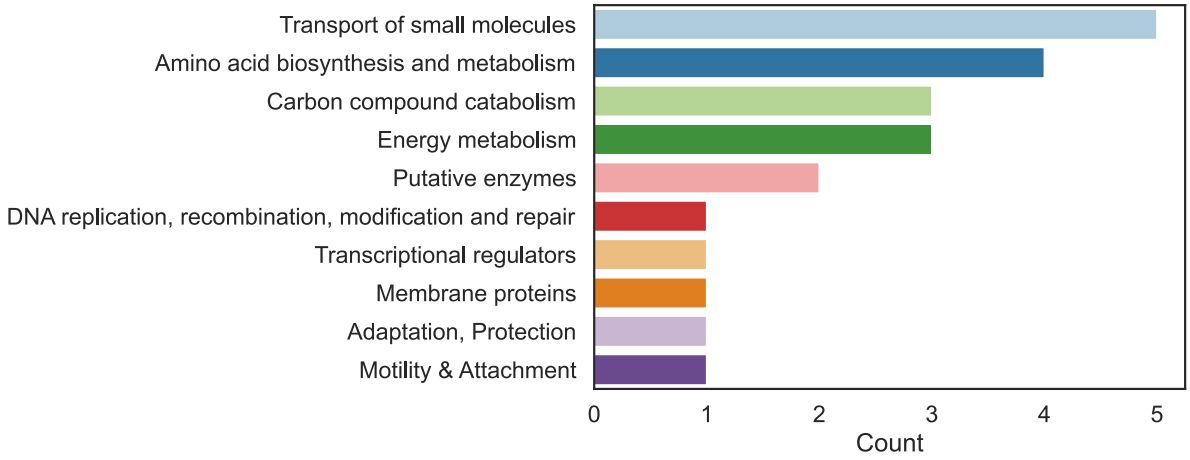


Figure 26: Count of PseudoCAP functions assigned to orthologs with a Pearson r value of 0.4 or higher.

In order to test if the carbon and nitrogen limitation conditions were solely responsible for correlation between “Amino acid biosynthesis and metabolism” and “Carbon compound catabolism” orthologs, “C/N source” conditions were removed, and the analysis rerun. Following this, only one of four “Amino acid biosynthesis and metabolism” genes remained and two of the three “Carbon compound catabolism” genes. These were gshA, and prpC and prpB, respectively.

It is clear from the comparison of full phenotypic profiles that individual conditions can hold significant weight in the correlation calculation. Therefore, to look at similarity on a more fine-tuned scale the S-scores for each condition-mutant ortholog pair were compared. Orthologs were compared to scores from the same condition and then the

Pearson correlation coefficient was calculated. When considering all S-scores, there was no correlation between PAO1 and PA14, with a Pearson r value of 0.033. The vast majority of S-scores within chemical genomic screens are non-significant since most genes will be irrelevant in most conditions. As such, the correlation when non-significant values were removed was tested. When pairs with a non-significant PAO1 score were removed the correlation increased to 0.203 (Fig. 27B), and when non-significant PA14 S-score pairs were removed the r value increased to 0.261 (Fig. 27C). When only considering scores which are significant in both PAO1 and PA14, the S-scores become significantly more correlated with a score of 0.601. However, only 894 ortholog pairs were considered for this analysis, representing only 0.07% of ortholog pairs. To evaluate the significance of the correlations, randomised pairs of genes from the same conditions were subject to the same analysis, see Figure 27E. No correlation was found for any of the gene pair comparison tests except when both sets of non-significant scores were removed. Here, the correlation was 0.19, which although limited, suggests that much of the correlation seen within the original data is due to random chance. Therefore, it is unlikely there is a correlation between the S-scores for PAO1 and PA14. Further suggesting their phenotypic dissimilarity.

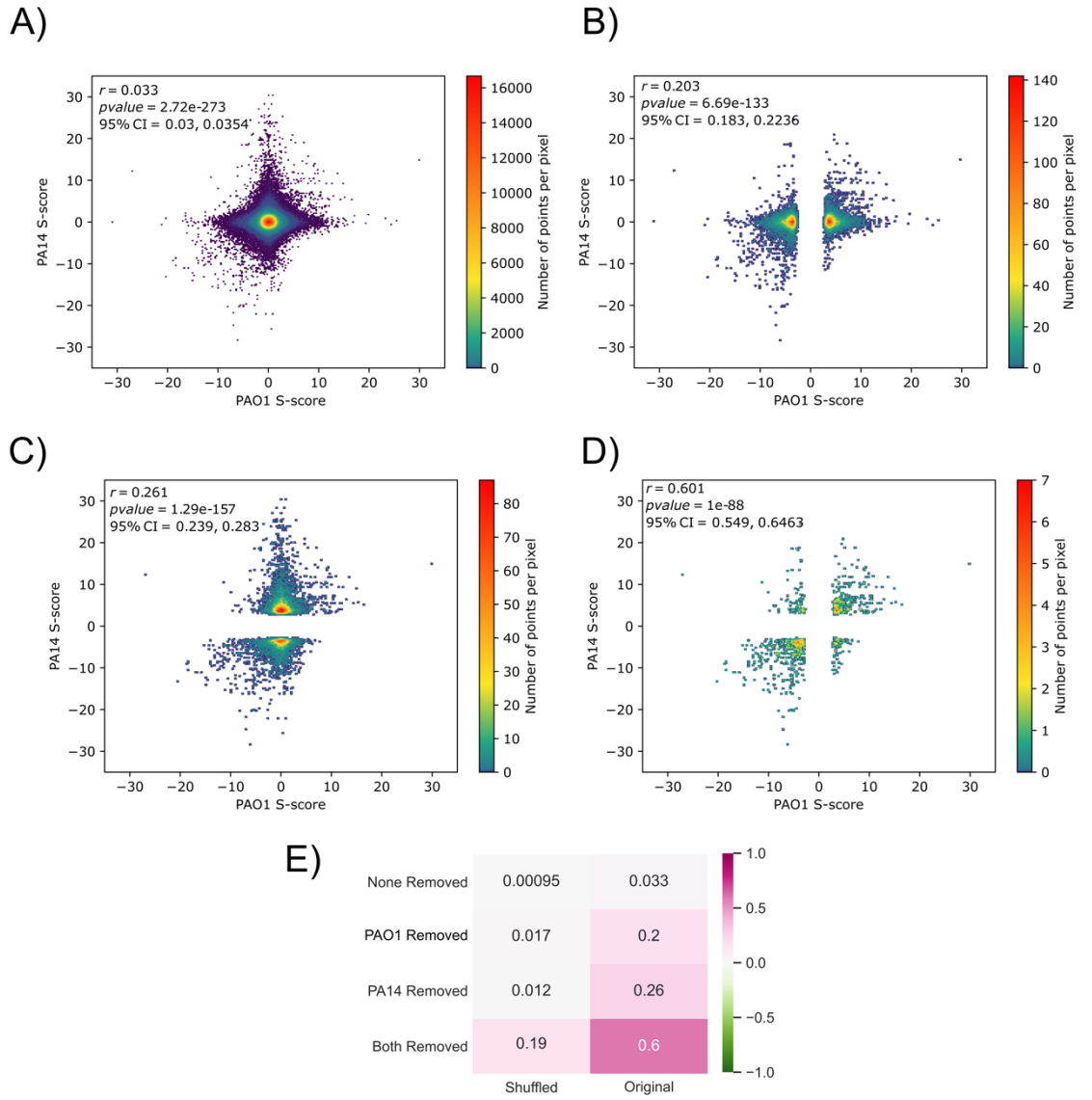


Figure 27: Heatmap-based scatterplots comparing the S-scores of ortholog genes across PAO1 and PA14 curated datasets. Colour bars for A-D) represent the number of points per pixel and for E) the Pearson r value. r = correlation score of homologue S-scores. A) all values B) excluding non-significant scores within the PAO1 dataset C) excluding non-significant scores within the PA14 dataset D) only S-scores significant in both datasets. Significant S-scores are scores with an FDR of <5%. E) PA14 scores were shuffled to create random non-orthologous pairs and the Pearson correlation coefficient re-calculated for comparison to the original dataset.

Of particular interest within this analysis, were the gene-condition pairs which were anticorrelated between PAO1 and PA14. As such, gene-condition ortholog pairs with a difference of at least 5 in their S-scores, where one score was significantly negative and the other significantly positive were extracted. Overall, 150 unique gene-condition pairs were found, see Supplementary Table 4. Where multiple transposon mutants appeared for the same gene-condition pair, these were collated for counts. The KEGG pathways for each of the gene-condition pairs was then analysed (Figure 28A), as well as the condition types (Fig. 28B). If no KEGG pathway was assigned to a gene it was left out of the KEGG analysis, this included 24 genes. The most common KEGG pathways were ‘Protein families: signalling and cellular processes’ and ‘Amino acid metabolism’, with 28 and 22 genes respectively. Within the ‘Protein families: signalling and cellular processes’ category the subcategories were ‘Bacterial motility proteins’, ‘Transporters’, ‘Two-component system’, ‘Exosome’, ‘Secretion system’, and ‘Antimicrobial resistance genes’. Here, ‘Transporters’ and ‘Bacterial motility proteins’ were the most common. Within the ‘Amino acid metabolism’ category, the most common sub-category was ‘Histidine metabolism’. However, this is likely due to the most common condition type being ‘carbon or nitrogen source’, with 52 counts. The next most common condition type, however, was ‘Antibiotic’, with 48 gene pairs. Since strains display differences in phenotypic responses to antibiotics, it is vital this is considered within the clinical setting and when considering new drug targets.

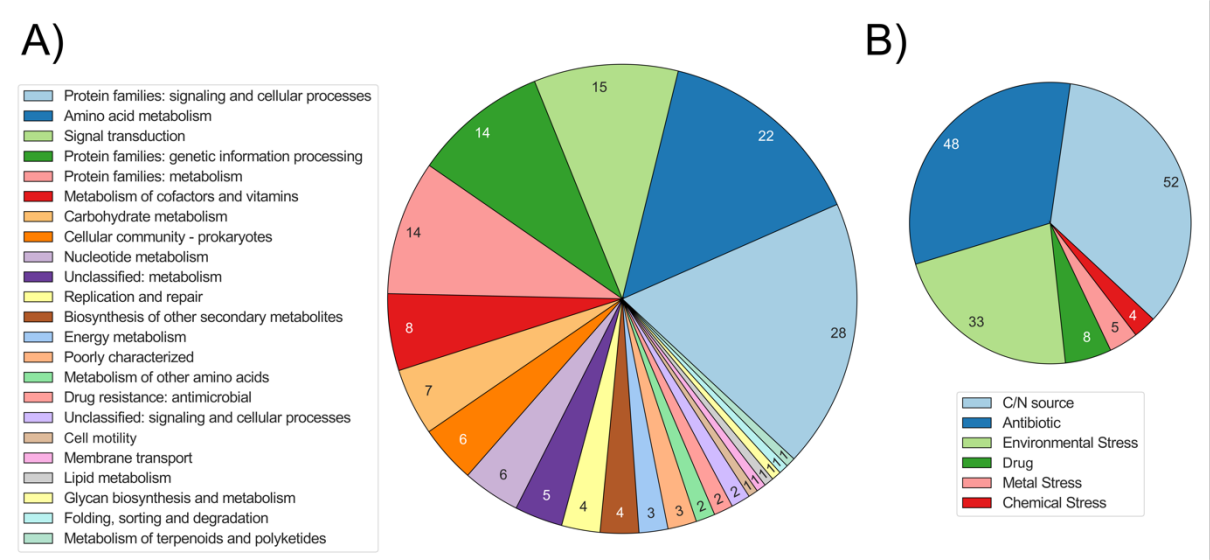


Figure 28: Pie charts displaying A) the KEGG pathway of anticorrelated ortholog pairs and B) the condition types in which anticorrelated ortholog pairs were elucidated. Where multiple KEGG pathways were assigned to a gene, all categories were included.

4.2.3.4 Δ*Pa*SAH results in sulfamethoxazole sensitivity within PA14 but not PAO1

One mutant which showed differing responses between PAO1 and PA14 was Δ*Pa*SAH, encoded by PA0431 and PA14_05600, respectively, see Figure 29. Within the presence of sulfamethoxazole, the PA14_05600 transposon mutant displayed a negative S-score of -9.25. Within the presence of a combination of sulfamethoxazole and trimethoprim (co-trimoxazole), ΔPA14_05600 had an S-score of -11.24. However, in trimethoprim alone, ΔPA14_05600 had a non-significant score of -0.49. For ΔPA0431, no S-score was significant with scores of -0.58, -0.63 and 0.45 for sulfamethoxazole, co-trimoxazole and, trimethoprim, respectively.

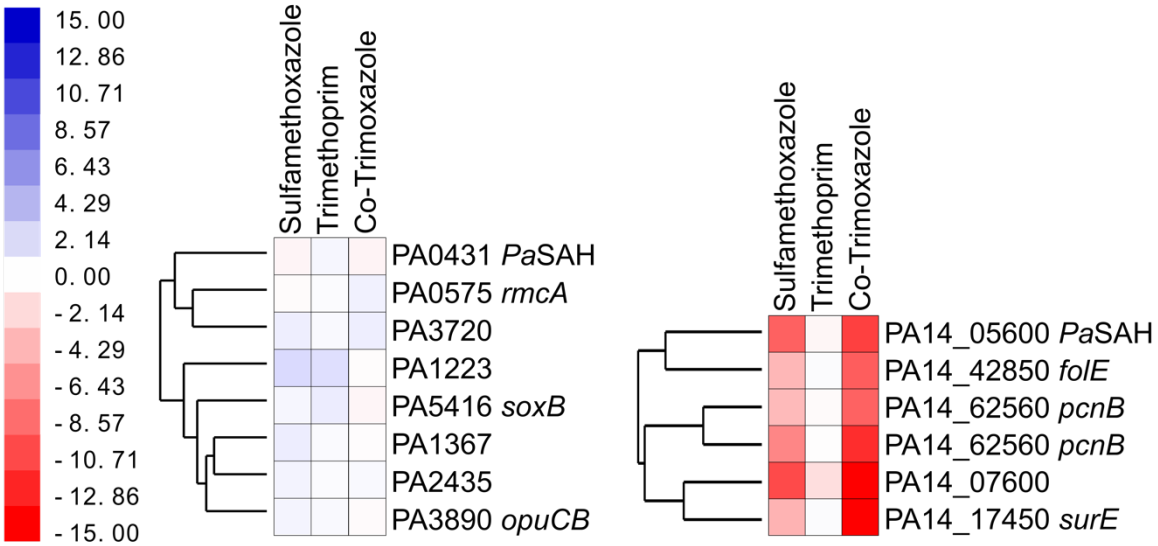


Figure 29: Heatmap displaying S-scores and gene clustering for sulfonamide drugs within the PAO1 and PA14 chemical genomic screens. Clustering represents the uncentered Pearson correlation clustering from the whole PAO1 and PA14 datasets. Clusters only relate to the genes; conditions are not clustered. Colours represent S-scores, with blue representing positive scores and red representing negative scores. Drug concentrations used were 100 µg/mL for sulfamethoxazole and 1 µg/mL trimethoprim, these concentrations were also used for the combination of the two drugs.

In order to validate this phenotypic result, clean knockouts of PA0431 and PA14_05600 were produced and screened within a small-scale chemical genomic screen. The wildtype strains, PAO1 and PA14 were plated in single colony format with PAO1 Δ*Pa*SAH and PA14 Δ*Pa*SAH on sulfonamide and trimethoprim condition plates.

Colony sizes were measured using Iris and fitness ratios calculated using ChemGAPP Small. For PAO1 loss of *Pa*SAH did not affect fitness compared to the wildtype in any of the conditions, see Figure 30. However, for PA14, in both sulfamethoxazole and sulfisoxazole a significant fitness defect was seen for the *Pa*SAH mutant compared to the wildtype. Although there is a slightly significant difference between PA14 Δ*Pa*SAH and PA14 in trimethoprim, the average fitness ratio was 1.05. These results not only confirm the high-throughput screening results but show resistance to not just sulfamethoxazole but other sulfonamide drugs as well.

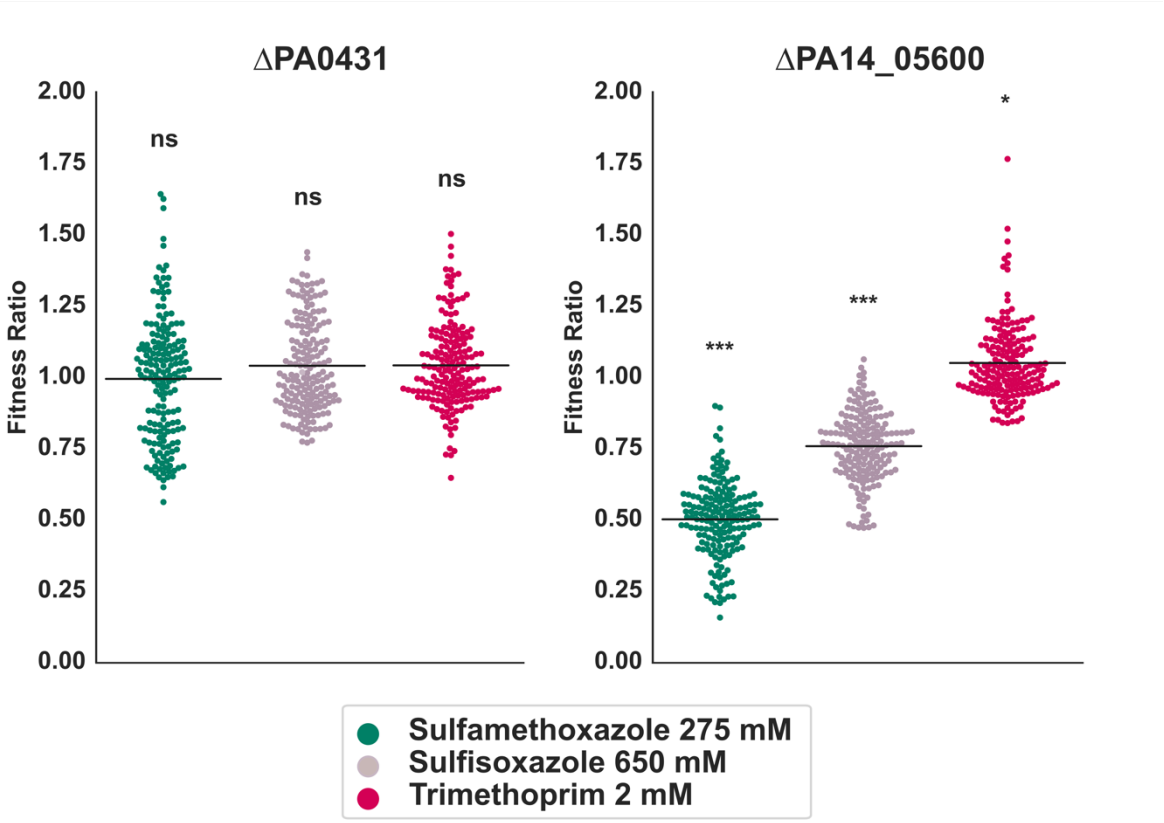


Figure 30: Swarm plots of PAO1 ΔPaSAH (PA0431) and PA14 ΔPaSAH (PA14_05600) fitness ratios compared to their respective parent strains in presence of sulfonamides and trimethoprim. ns = $p > 0.05$; * = $0.05 < p \leq 0.01$; *** = $0.0001 < p \leq 0.001$.

Following hierarchical clustering of the phenotypic profiles (Fig. 29), ΔPA14_05600 was clustered with Δ*folE*, Δ*pcnB*, ΔPA14_07600 (*folK2*) and Δ*surE*. To find more mutants profiles with similarity to ΔPA14_05600, a network analysis of mutants with a Pearson r value greater than 0.4, when compared to ΔPA14_05600, was produced (Fig. 31). From this analysis, five additional genes were found to have similarity to the phenotypic profile of ΔPA14_05600. These were, Δ*phoQ*, Δ*purH*, Δ*mexA*, and Δ*mexB*. Within both analyses, Δ*folE2* was the most highly associated with ΔPA14_05600, with a r value of 0.66. Within the network analysis the next highest r values were for Δ*pcnB* (0.54), Δ*purH* (0.54), and Δ*folK2* (0.54). Based on the KEGG pathway for folate biosynthesis, PA14_07600 is likely a *folK* paralog. PurH, the final enzyme in the de novo purine synthesis pathway catalyses the formation of IMP (Costa Brandão Cruz et al., 2019). Finally, *pcnB* is 8 bases upstream of *folK*, and they are likely co-transcribed, meaning transposon disruption of *pcnB* may have also disrupted *folK* (Green and Matthews, 2007). The loss of *PaSAH* likely results in the inappropriate triggering of the stringent response due to the reduced ability to degrade (p)ppGpp. One major role of (p)ppGpp is the inhibition of purine biosynthesis. Since Δ*purH* showed a similar phenotypic response to sulfonamides to PA14 ΔPA14_05600, this may suggest that the synthesis of purines is particularly important in the presence of sulfonamides. Therefore, this study hypothesises that since THF biosynthesis has a role in nucleic acid production, and loss of *Pa*SAH may lead to increased purine degradation and reduced synthesis, that there is a combinatorial effect leading to a lethal reduction in purines.

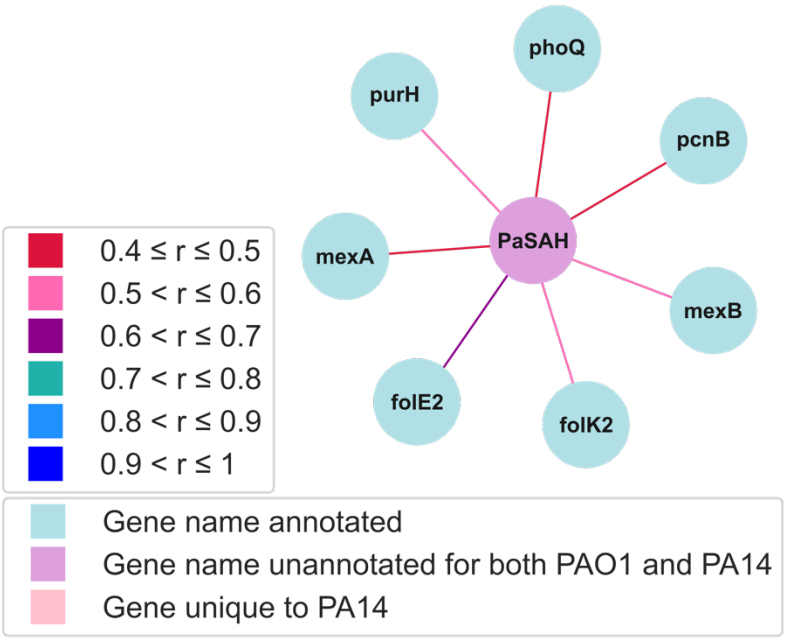


Figure 31: Network analysis of phenotypic profiles with Pearson correlation coefficients (r) greater than 0.4 when compared to PA14_05600. Only connections to PA14_05600 are shown.

In order to test the hypothesis that PA14 Δ*Pa*SAH is due to a deficit in purines, spotting plates were produced supplementing the sulfonamide and trimethoprim conditions with 0.2 mM adenine (Fig. 32). Although the presence of adenine did improve growth of PA14 Δ*Pa*SAH within sulfonamide conditions, it equally recovered the wildtype. Therefore, it is possible that recovery is due to the association of purine loss with sulfonamide treatment, rather than stringent response associated purine reduction.

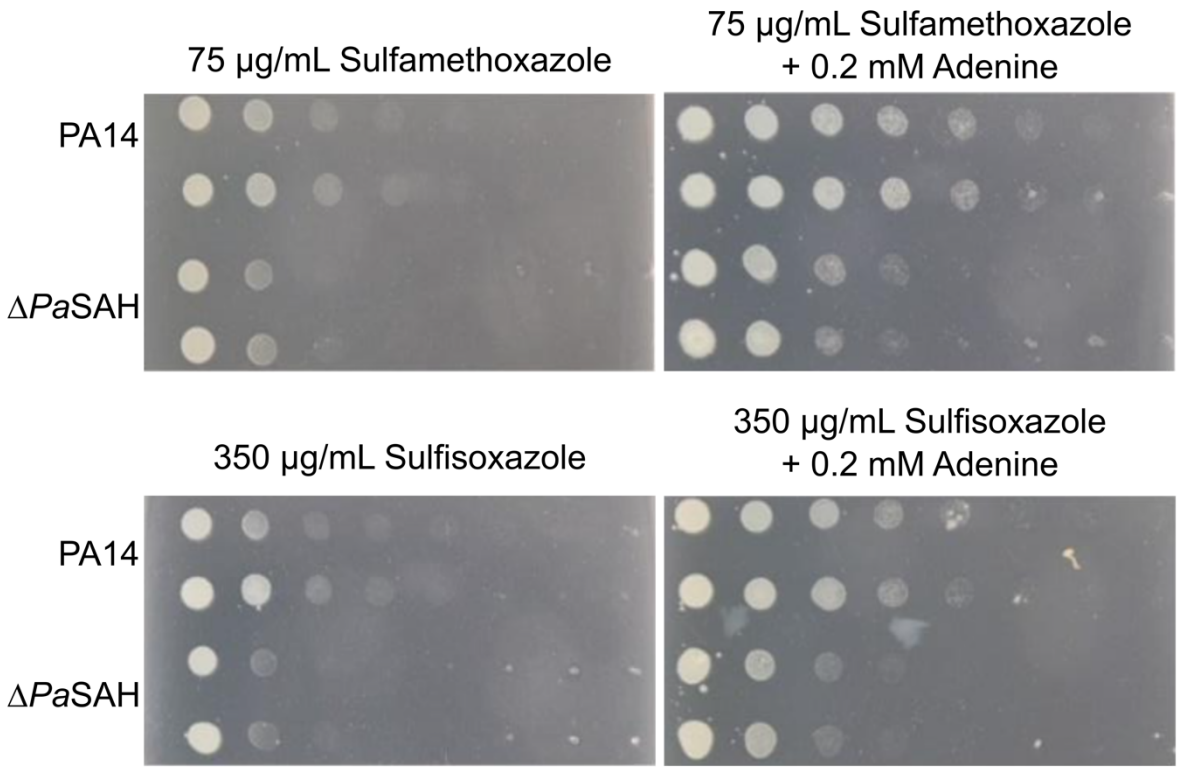


Figure 32: Spotting plate assay of PA14 and PA14 ΔPaSAH in presence of sulfonamide drugs with and without adenine supplementation.

4.3 Discussion

Although there are multiple published high-throughput chemical genomic screens, there are currently no published studies using chemical genomics to compare multiple bacterial strains or species to one another. The only related published studies comparing genome-wide deletions were performed in yeast strains. First, one study compared genome-wide genetic interaction profiles for *Saccharomyces cerevisiae* and *Saccharomyces pombe* (Ryan et al., 2012). However, the compared screens were not produced within the same study, the *S. cerevisiae* set was computationally produced

from various screens. As such, the two screens were scored using different methods and both screens had to be scaled to each other (Ryan et al., 2012). Another study screened and compared genome-wide single gene knockout libraries for four strains of yeast (Galardini et al., 2019). Although screened in parallel, the screen only used 38 conditions. The unique nature of the current study is in the parallel screening of two closely related bacterial strains in a wide range of conditions. Parallel screening removes much of the noise that comes from screens being performed by different people. Furthermore, it reduces independent variables and makes the data more comparable without relying on scaling of datasets. Within the current study, source plates of bacteria were stored and grown in the same environment, condition plates were produced from the same chemical stocks on the same day, and plates were screened at the same time. Additionally, not just the same conditions were used but the same concentrations, meaning all conditions can be directly compared. This is often not the case for the comparison of screens performed in different groups, since research questions and choice of conditions may differ.

Despite the unique nature of the screens and the reduction of noise, it was still important that both screens were of high quality, such that they could be accurately compared. Both datasets had high replicate reproducibility scores post curation of 0.714 and 0.742 for PAO1 and PA14, respectively. These results are slightly lower than those achieved within the Nichols et al screen, where the E.coli KEIO collection had a replicate reproducibility score of 0.77. Despite this, it is not unexpected that the *P. aeruginosa* scores are lower than that achieved for *E. coli* due to the motile tendencies of *P. aeruginosa*. It is known that media with limited nitrogen sources, like many used within the current study, can induce swarming motility in *P. aeruginosa*

(Overhage et al., 2007). It was frequently observed across multiple conditions that some mutants of PA14 and PAO1 would swarm, resulting in variable colony sizes between replicates. This occurred in multiple antibiotic supplemented conditions, likely because swarming can induce increased resistance to antibiotics (Overhage et al., 2008). Further to this, Iris has difficulty detecting the true edge of these colonies, sometimes taking the shallow outer edge made of swarming cells and other times taking the more concentrated core of the colony. This further resulted in large differences in replicate colony sizes. The PAO1 dataset had worse replicate reproducibility than that of PA14. This is likely due to the swarming being more prevalent amongst the PAO1 mutants. Despite this, both datasets had high replicate reproducibility. This is especially true considering replicate colonies were biological replicates pinned from multiple source plates, meaning there will be natural variance. Having some noise from natural variance also improves the robustness of the data and the reliability of the scores. For example, if there are highly conserved results between replicates, by including this natural variance we know these are true phenotypic responses and not simply a clonal response.

The datasets were both further benchmarked to test their quality and predictive ability. Again, both datasets scored well for most of the benchmarking categories, however, PAO1 performed worse than PA14. This is likely due to the quality of the base libraries. Although unpublished, groups who have worked with the library have speculated that there are misannotated mutants within the PAO1 screen. This is likely due to the library having been produced in 2003 and replicated multiple times over its lifetime, leaving it vulnerable to human error. However, with the PA14 library being produced only three years later in 2006, this could also be true for the PA14 library. Within the benchmark

analysis the highest AUC scores for PAO1 and PA14 were for the replicate analysis. This is understandable as phenotype scores should be the most similar between mutants of the same gene. However, with replicate analysis AUC values of 0.661 (PAO1) and 0.724 (PA14), values could be considered as lower than expected. Since replicates were classified as mutants with transposons inserted in the same gene but different locations, this could account for much of the variability between replicates. Transposon libraries are not as robust as clean knockout libraries such as the KEIO collection since transposon position and orientation can entirely change the effect on the function of the gene. For example, if the transposon is inserted at the end of the gene, it is less likely to result in a full disruption of the gene and the product may retain some functionality (Liberati et al., 2006). Additionally, even if the gene is inserted at the very start of the gene, if alternative in-frame start codons are available then translation can still occur. Again, this could result in functioning proteins. Furthermore, transposons can orient in the same direction of the target gene or in the opposite direction, which can result in polar downstream effects (Jacobs et al., 2003; Liberati et al., 2006). Considering this, the replicate benchmark scores were acceptable for both PAO1 and PA14.

For the other three benchmark analyses, similar AUC scores were achieved as the previous yeast study (Galardini et al., 2019). In the 2019 study, when analysing based on protein complex, AUC values were 0.56, 0.62, 0.7 and 0.81 for the four strains. For PAO1 and PA14, protein complex data was not readily available and three-letter nomenclature was used as a proxy. Here the AUC scores were 0.602 and 0.638 for PAO1 and PA14, respectively. The *P. aeruginosa* values are closer to the midrange values for the yeast strains. However, these lower values could again be attributed to

the use of a transposon library over a knockout library as was used in the 2019 study. Furthermore, having to rely on three-letter nomenclature over curated protein complex data may not be as accurate at defining truly related genes.

Galardini et al. also benchmarked the four yeast strains against KEGG pathways. Similar to the current study, reduced AUC values were obtained for the strains based on KEGG pathways, with a range of 0.53 to 0.65 (Galardini et al., 2019). This provides more support for KEGG pathways not being the most effective measure of benchmarking due to categories being too broad. Since the datasets showed poor ability in recapitulating KEGG pathways in general, each individual category was examined to see if any particular categories had higher percentages of correlation. As expected, based on the AUC values, PAO1 performed worse than PA14. However, for both strains, the categories with the highest percentage of r values ≥ 0.4 were associated with metabolism or energy generation. This is not entirely unexpected based on the phenotype measured. By measuring colony size as a measure of fitness, genes directly associated with growth are likely to have stronger phenotypes. Many pathways are activated or more important in different stages of growth and have an effect on physiology that does not affect growth rate (Kritikos et al., 2017). For example, Kritikos et al., reanalysed the Nichols et al. dataset and found multiple phenotypes for genes which were unresponsive in over 300 conditions when measuring only colony size (Kritikos et al., 2017). Further to this, they analysed biofilm phenotypes within *P. aeruginosa* and discovered new connections within known pathways, as well as new roles for phenazines in early biofilm development (Kritikos et al., 2017). Therefore, it would be potentially beneficial in future screens to measure multiple different phenotypes in order to gain a more complete understanding of the organism and

perhaps improve the ability of screens to recapitulate pathways. This is especially important if a non-growth-related characteristic is particularly important within a specific organism. For example, another screen within *Bacillus subtilis* also focused on multiple phenotypes, here they looked at growth and sporulation (Koo et al., 2017). For *B. subtilis,* sporulation is a fundamental aspect of its developmental pathway, but cannot be accurately targeted using growth assays. Using a colorimetric based phenotypic screen, Koo et al., highlighted 73 genes with connections to sporulation which were previously unknown. Of these, 33 genes were poorly characterised and 12 have gone on to be functionally annotated (Koo et al., 2017). Therefore, there could be a host of phenotypic information being missed simply due to the chosen phenotype.

Despite the screens only characterising responsive-to-growth phenotypes, both screens produced large numbers of phenotypes. The number of phenotypes achieved for chemical genomics screens is variable within the literature. The number of phenotypes gained within the current study was comparable to that of the Nichols et al. dataset, which found a phenotype in 49% of strains (Nichols et al., 2011). Another study, however, which performed a chemical genomic screen on a panel of *E. coli* strains instead of a knockout library reported only 77/266 strains (28.9%) showing at least one phenotype (Galardini et al., 2017). In contrast, a study in yeast, reported an extraordinary response rate with 97% of genes showing at least one phenotype (Hillenmeyer et al., 2008). However, the Hillenmeyer et al. study screened over 400 conditions, significantly more than tested in the current study. Furthermore, many of these phenotypes occurred in severe stresses, which are more likely to induce a response than not (Hillenmeyer et al., 2008; Nichols et al., 2011). PAO1 showed a higher percentage of responsive genes at 67.9% compared to 46.5% for PA14. This

may be because of the larger number of replicates within the PAO1 screen. For most genes within this library there were multiple mutants with transposons in the same gene but in different positions. As previously mentioned, transposon location can affect the phenotypic response of a mutant. Therefore, since phenotypes were collated for mutants of the same gene, it may be that there were more opportunities for ideal transposon placement within the library. Nevertheless, both datasets produced an acceptable number of phenotypes.

Some genes showed phenotypes not just in one condition but multiple. A previous study defined genes with greater than 30 phenotypes as multiple stress responsive genes (Nichols et al., 2011). The gene functions of the various MSR genes within the current study were fairly broad, evidencing that there was a diverse range of conditions within the screens. Since one of the aims of this study was to compare the two strains and see if they are phenotypically similar, the MSR genes were compared. Given phenotypic similarity, one would expect that similar genes or pathways would be found across the two strains. This hypothesis assumes that if genes are equally important then responses would be achieved in the same conditions for both strains. However, there was no cross over in genes and limited cross over in pathways, which showed no correlation. In PAO1 it appears that motility is much more important across multiple conditions than in PA14. This may be an artifact of the increased prevalence of swarming within the PAO1 screening plates. Conversely, this increased swarming on screening plates as a response to stress, in combination with the MSR status of the pilA/M/X/Y1 genes, may indicate that swarming is more important for survival in PAO1 than PA14. Some evidence for this comes from a study which looked at transcriptional differences between PAO1 and PA14 (Sana et al., 2019). This study published a

dataset of the transcriptional data for PAO1 and PA14. Of the MSR genes, only information for *pilM* was available within this dataset. Based on this study, *pilM* was transcribed 1.8-fold more in PAO1 than PA14 and on average the pil genes were transcribed 1.5-fold higher in PAO1 (Sana et al., 2019). Further to this, various studies have looked at the importance of the pil genes for motility in PAO1 and PA14. The results of these studies has concluded that although the pil genes are essential to motility in PAO1, they are not in PA14 (Köhler et al., 2000; Rashid and Kornberg, 2000; Toutain et al., 2005; Overhage et al., 2007). Furthermore, PA14 encodes a second type 4 pilus within the PAPI-1 pathogenicity island, which PAO1 does not encode (Carter et al., 2010). This type 4b pilus has been implicated in the conjugative transfer of the PAPI-1 pathogenicity island between *P. aeruginosa* strains (Carter et al., 2010). Although studies have shown deletion of PAPI-1 does not affect swimming or twitching motility (He et al., 2004; Harrison et al., 2010), this second pilus could potentially play a role with the loss of the conserved type 4 pilus. Since phenotypic differences in motility are well evidenced in the literature, this provides clear support for the ability of chemical genomic screens to distinguish true phenotypic differences between strains.

In PAO1, the MSR gene with the highest number of phenotypes was *murA*. MurA is the first catalytic step for the production of UDP-MurNAc which is incorporated into peptidoglycan. It is therefore very important in the formation of the bacterial cell-wall (Elamin et al., 2017). Considering the number of cell wall affecting conditions within the screen from cell-wall targeting antibiotics to surfactants and detergents, it is unsurprising that *murA* deletion resulted in multiple phenotypes. Unfortunately, Δ*murA* was not present within the PA14 library, therefore it is impossible to conclude if the same phenotypic responses would have occurred in PA14. For PA14, the gene with

the highest number of phenotypes was ΔPA14_34540, an uncharacterised gene. This gene was annotated based on structural similarity as *soxA*, however another gene within PA14 (PA14_71500) has been annotated as *soxA* and these genes only share 23.73% amino acid sequence identity. Unfortunately, little information surrounding this gene is available within the literature. One paper, focusing on the genome wide identification of antimicrobial resistance (AMR) determinants, found that loss of PA14_34540 led to increased resistance to piperacillin, cefepime, cefotaxime and ceftazidime (Dötsch et al., 2009). This could explain the multiple phenotypes displayed, considering many of the conditions within the screen were antibiotics. Further to this, all phenotypes displayed by ΔPA14_34540 were positive S-scores, implying improved fitness after deletion. This conforms with the gain of resistance seen within the Dötsch et al. paper. Within this AMR study, another MSR gene, PA14_15600, was also described. Similar to ΔPA14_34540, ΔPA14_15600 showed increased resistance upon deletion (Dötsch et al., 2009). A further paper similarly displayed increased resistance to ceftazidime in a ΔPA14_15600 strain (Alvarez-Ortega et al., 2010). Interestingly, this gene is suggested to come from T5 bacteriophage origin, based on its T5orf172 domain, and is unique to PA14. As previously described, the literature has shown that prophages and phage genes can have a profound impact on the phenotypic responses of *P. aeruginosa* strains. Therefore, PA14_15600 could perhaps be a further determinant of the phenotypic differences between PAO1 and PA14.

Clearly there are differences within the MSR genes, but it was important to look at the similarity of the datasets as a whole. Given that the data had 174 dimensions t-SNE was employed to reduce the dimensions to two. By doing this each phenotypic profile was able to be represented as a single datapoint within a scatterplot, where points

which are closer represent more similar data. Hypothetically, if the two strains had similar phenotypic profiles for the various orthologs, it would be expected that a large amount of crossover would occur. This is because t-SNE preserves relationships within the data, focusing on keeping similar data close together. This means one can better visualise the structure of the data (van der Maaten and Hinton, 2008). However, following the t-SNE analysis, the two datasets formed two clusters which could be linearly separated, with only a few datapoints crossing over. As such, it was concluded that the PAO1 and PA14 orthologs did not correlate in their phenotypic profiles.

To confirm the results found within the t-SNE analysis, the phenotypic profiles of orthologs were compared via a Pearson correlation analysis. Again, very little correlation was seen between orthologs, with only 22 genes showing a correlation greater than 0.4. Of these genes, 5 were classified as transporters of small molecules. The most prominent of these were MexB and OprM, with r values of 0.68 and 0.58, respectively. Since MexAB-OprM is the major efflux pump in both PAO1 and PA14, and the majority of conditions used within the screens were antibiotics, it is logical that increased correlation is seen. Additionally, the MexAB-OprM substrates are conserved across PAO1 and PA14, meaning conserved fitness defects are observed in the presence of its substrates with component loss. Therefore, these conditions likely hold high weight in the correlation analysis. Further to this, mexB is highly transcribed in both PAO1 and PA14 (Sana et al., 2019). This may explain the similar importance and thus similar phenotypic responses for loss of *mexB* in both PAO1 and PA14, especially since Δ*mexB* showed the highest r value between the screens.

The next most common categories were ‘amino acid biosynthesis and metabolism’, and ‘carbon compound catabolism’. Genes within these categories being more highly correlated is understandable. As previously described, these genes are likely to show responsive-to-growth phenotypes as they are directly related to the cells ability to grow. This is especially true since many genes in these categories lead to auxotrophy, meaning often correlation occurs strongly within conditions that do not cater to this auxotrophy. To somewhat reduce this response, minimal media conditions were analysed separately from the LBA based conditions and the two datasets were concatenated together. This was based on an observation made by Wu et al. when reanalysing the Nichols et al. dataset. Upon reanalysis, when minimal media conditions were removed, the precision was greatly increased for benchmark analyses recapitulating protein complexes and EcoCyc pathways (Wu et al., 2020). Although in the current study the minimal media conditions were included within the screen, S-scores were not as extreme when analysed separately, meaning they had less weight to skew the data. Despite this, removal of these conditions was also tested to see if loss reduced the r values of the ‘amino acid biosynthesis and metabolism’, and ‘carbon compound catabolism’ genes. When these conditions were removed one ‘amino acid biosynthesis and metabolism’ gene, and two ‘carbon compound catabolism’ genes remained, *gshA*, and *prpB* and *prpC*. GshA is an enzyme within the glutathione (GSH) biosynthesis pathway (Van Laar et al., 2018). In a previous study, which evaluated the same PAO1 *gshA* transposon mutant as this study, it was shown that *gshA* disruption results in increased sensitivity to oxidative stress and thiol-reactive antibiotics. Further to this, impairment in biofilm formation as well as swarming and swimming motility was demonstrated, and the mutant produced less pyocyanin than the PAO1 wildtype (Van

Laar et al., 2018). The retainment of *gshA* following removal of carbon and nitrogen limitation conditions is therefore likely due to the multiple phenotypes shown within the Van Laar et al. study. Therefore, GshA likely played a role in more condition types than simply the induction of auxotrophy. Although these studies are only focused on PAO1, based on the Sana et al. dataset, *gshA* is transcribed almost 2-fold higher in PA14, so is likely equally important. This would perhaps explain the similarity between the two ortholog phenotypic profiles. PrpB and PrpC are involved in the 2-methylcitrate cycle (2-MCC) stabilising the levels of propionate and propionyl-CoA within the cell. The 2-MCC pathway feeds into the TCA cycle via the conversion of propionyl-CoA to pyruvate in a series of enzymatic steps involving PrpB and PrpC, amongst others (Cui et al., 2022). Therefore, PrpC and PrpB have a central role in carbon and energy metabolism. However, it has recently been discovered that they may also play a role in virulence via the transcriptional regulator PmiR. PmiR responds to levels of 2-methylisocitrate within the cell. 2-methylisocitrate is formed by PrpC and PrpD and converted to pyruvate and succinate by PrpB (Cui et al., 2022). Since PmiR was only recently described, its annotation was not in the dataset and therefore did not appear in this analysis. However, by comparing the phenotypic profiles of the PmiR loci within PAO1 and PA14 a Pearson r value of 0.61 was calculated, showing phenotypic similarity between the mutants.

Despite the analysis of minimal media conditions separately, it is clear that singular conditions can still hold weight. This is especially true since Pearson correlation was used as the similarity measure, and it is known that Pearson correlation analyses are sensitive to outliers (Rousselet and Pernet, 2012). As such, in future analyses of phenotypic profiles perhaps correlation metrics less sensitive to outliers should be

used to compare strains. As an alternative to investigating phenotypic profiles as a whole, and to avoid this skew, individual gene-condition ortholog pairs were compared instead. Following this analysis, no correlation was found between the two datasets yet again. Even after removing non-significant values for either PAO1 or PA14, the correlation did not get higher than 0.26. Finally removing all non-significant values, although resulting in a higher correlation of 0.6, was unreliable since random pairs gained a correlation of 0.19. This was likely the result of the loss of the majority of the data. Further to this, this removed much of the varying phenotypes where one strain showed a phenotype, and the other did not. Often these genes were of the most interest since they may better explain the phenotypic differences between PAO1 and PA14. Of particular interest were the anticorrelated genes, where one strain showed a loss of fitness and the other a gain.

The most common KEGG category within the anticorrelated gene set was 'Protein families: signalling and cellular processes', and one of the most common subcategories was 'Bacterial motility proteins'. This was due to the fact that many of the *pil* operon genes were anticorrelated, including *pilG*, *pilI*, *pilK*, *pilM*, *pilO*, *pilQ*, *pilS*, and *pilX*. Δ*pilI* and Δ*pilM* showed decreased fitness only for PA14. Δ*pilG*, Δ*pilK*, Δ*pilS*, and Δ*pilX* showed a decreased fitness for only PAO1. Finally, Δ*pilO* and Δ*pilQ* showed decreased fitness for both PAO1 and PA14 within different conditions. For the conditions in which PA14 showed decreased fitness and PAO1 showed improved fitness, four out of five condition-gene pairs were within minimal media supplemented with glutamate as the sole nitrogen source. Glutamate, when supplied as the sole nitrogen source, is known to induce swarming motility within PAO1 (Köhler et al., 2000). This is because glutamate metabolism is correlated with the production of

rhamnolipids, which are used as biosurfactants to aid swarming motility (Soberón-Chávez et al., 2005; Overhage et al., 2007). Given that it was Δ*pilM,* Δ*pilO,* and Δ*pilQ* in the glutamate condition and PilM/N/O/P/Q are required for pilin assembly, one would expect full loss of motility in this condition (Ayers et al., 2009; Leighton et al., 2018). However, one study showed that in a swarming motility deficient PAO1 mutant (Δ*pilA,* Δ*fliC*), that rhamnolipid presence stimulated sliding motility (Murray and Kazmierczak, 2008). This may explain the increased fitness in PAO1, as these colonies may have spread due to sliding motility and been perceived as larger by Iris. Furthermore, mutants of the pil genes still showed swarming motility in multiple conditions despite disruption, based on visual inspection of plate images. Again, this could be due to the transposon insertion locations potentially resulting in functional type 4 pilus. This could explain the differences between PAO1 and PA14. However, without validation using clean knockouts this cannot be confirmed.

Another discovery from the screen, was the difference in phenotypic response for the *P. aeruginosa* small alarmone hydrolase *Pa*SAH within sulfamethoxazole containing conditions. Here, sensitivity was observed in PA14 but not PAO1. This was confirmed by the production of clean knockouts and testing on both spot plates and in a small-scale screen, suggesting that phenotypes within the screen are accurate. The initial hypothesis for the sensitivity of Δ*Pa*SAH in the presence of sulfonamides was the two-front depletion of purine biosynthesis. This was supported by the clustering of PA14 Δ*Pa*SAH with the Δ*purH* mutant as well as the folate biosynthesis mutants Δ*folE2* and Δ*folK2*. However, following supplementation with purines, despite improvement being seen in both the wildtype and PA14 Δ*Pa*SAH, the gap in fitness was not closed. This suggests there is an alternative mechanism for the sensitivity of PA14 Δ*Pa*SAH to

sulfonamides. Alternatively, it may be that providing adenine is not the correct supplement. IMP may be a better option since it is the product of PurH activity, which PA14 Δ*Pa*SAH correlated with. Furthermore, during the stringent response purine degradation occurs via the degradation of IMP and GMP. Therefore, IMP would potentially be more central to the response and complement PA14 Δ*Pa*SAH to wildtype fitness.

An alternative, albeit similar hypothesis for the mechanism of sensitivity, could be the depletion of GTP levels. During stringent response, much of the GTP is converted to pppGpp, sequestering it for signalling responses. GTP also directly feeds into the folate biosynthesis pathway, the target of sulfonamide drugs. Further evidence of this could be the sensitivity seen in the presence of sulfonamides but not trimethoprim, despite both drugs targeting the same pathway. Sulfonamides irreversibly form dead-end complexes, sequestering folate biosynthesis intermediates, whereas trimethoprim reversibly competes for its target (Minato et al., 2018). Typically, the products of the folate biosynthesis pathway undergo a series of reactions to produce IMP and then a further set of reactions to reformulate GTP (Minato et al., 2018). It may be that conversion of GTP to pppGpp, with the reduced turnover back to GTP from loss of *Pa*SAH, as well as the wasted flow through of GTP in a sulfonamide targeted folate biosynthesis pathway is resulting in cell death. Since trimethoprim does not sequester intermediates, it may not deplete the GTP pool sufficiently enough for cell death. Therefore, in future experiments supplementation with GTP would be beneficial, however, due to its expense an alternative method would be preferred. Since IMP can be converted into GTP this would still be a potential candidate for supplementation for either hypothesis.

An alternative reason that sensitivity in the presence of trimethoprim and sulfonamide differs may be that trimethoprim is more actively effluxed than sulfonamides. Based on Dreier et al. in 2015, MexAB-OprM can efflux both trimethoprim and sulfamethoxazole, but trimethoprim can also be effluxed by other efflux pumps as well (Dreier and Ruggerone, 2015). Therefore, it may be that internal levels of trimethoprim are not sufficient to produce a phenotypic response in a PA14 Δ*Pa*SAH strain. Despite the lack of a conclusive mechanism of Δ*Pa*SAH sensitivity, the validation of these phenotype supports the ability of chemical genomics screening to accurately reveal phenotypic differences between similar strains.

Another major discovery within the chemical genomic screen was that biotin synthesis conferred resistance to metformin in both PAO1 and PA14. Since this has therapeutic implications, with potential for a biotin-synthesis inhibitor and metformin co-therapy an entire chapter was dedicated to the evaluation of this phenotype.

CHAPTER 5

BIOTIN MEDIATED METFORMIN RESISTANCE.

5.1 Introduction

5.1.1 Biotin mediates metformin resistance in *Pseudomonas aeruginosa*

Another major discovery within the *P. aeruginosa* screen was the susceptibility of biotin synthesis mutants to the diabetes drug metformin, see Figure 33. Although both PAO1 and PA14 showed this phenotype, the effect was more profound within PA14. As previously described, antimicrobial resistance within *P. aeruginosa* has led to the increasing need for new therapeutics. In recent years, attempts to produce new antibiotics have been fraught with failure. This is due to reduced private investment, minimal compounds reaching clinical trials, and many discovered or engineered compounds being derivatives of current classes of antibiotics (Beyer and Paulin, 2020; Miethke et al., 2021). Therefore, the repurposing of currently approved drugs has become an appealing prospect for the medical community (Schein, 2020). As such, repurposing metformin as an antibiotic in combination with a biotin synthesis inhibitor could pose a new therapeutic option for the treatment of *P. aeruginosa*.

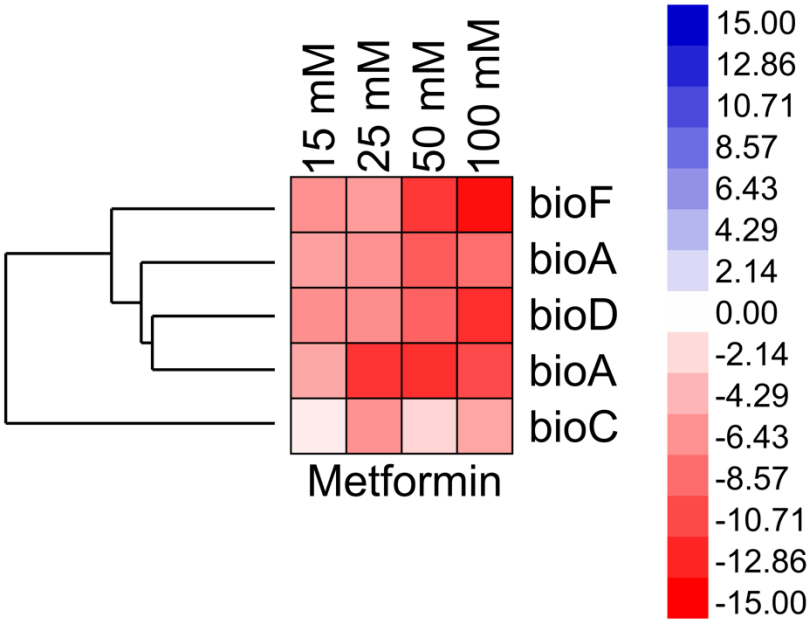


Figure 33: Clustered heatmap of S-scores for biotin synthesis mutants in presence of metformin. Clustering represents the uncentered Pearson correlation clustering from the whole PA14 dataset. Clusters only relate to the genes; conditions are not clustered. Blue squares represent positive S-scores, and red squares represent negative S-scores. Genes show decreased fitness for all conditions.

5.1.2 Biotin

Biotin (vitamin B7) is an essential vitamin, which functions as an enzyme cofactor for carboxylation, decarboxylation and transcarboxylation within multiple biological processes and pathways such as, fatty acid biosynthesis, the citric acid cycle, gluconeogenesis, and amino acid metabolism. Biotin consists of a valeryl side chain

attached to a bicyclic ring comprised of a ureido ring and a tetrathiophene ring, see Figure 34 (Lin and Cronan, 2011).

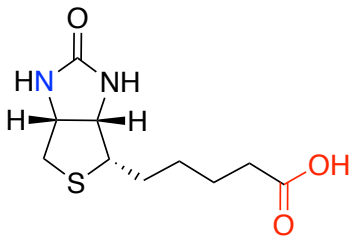


Figure 34: Biotin chemical structure. Red = carboxylic acid group. Blue = position 8 Nitrogen (N8) (Taken from Doherty, 2020)

Biotin functions in complex with biotin-dependent carboxylases. These carboxylases form this complex via an amide bond between biotin's valeryl side chain carboxylic acid group and the biotin-dependent enzyme's biotin-carboxyl carrier protein (BCCP) domain (Lin and Cronan, 2011; Tong, 2013). The binding of biotin to the BCCP domain allows the valeryl chain to function as a swinging arm that can translocate between the biotin carboxylase (BC) domain and the carboxyltransferase (CT) domain of the biotin-dependent carboxylase. The N8 nitrogen of the ureido ring becomes carboxylated by the BC domain such that the CO_2 moiety can be transferred from N8 at the CT domain to the biotin-dependent carboxylases substrate, see Figure 35 (Lin and Cronan, 2011; Tong, 2013).

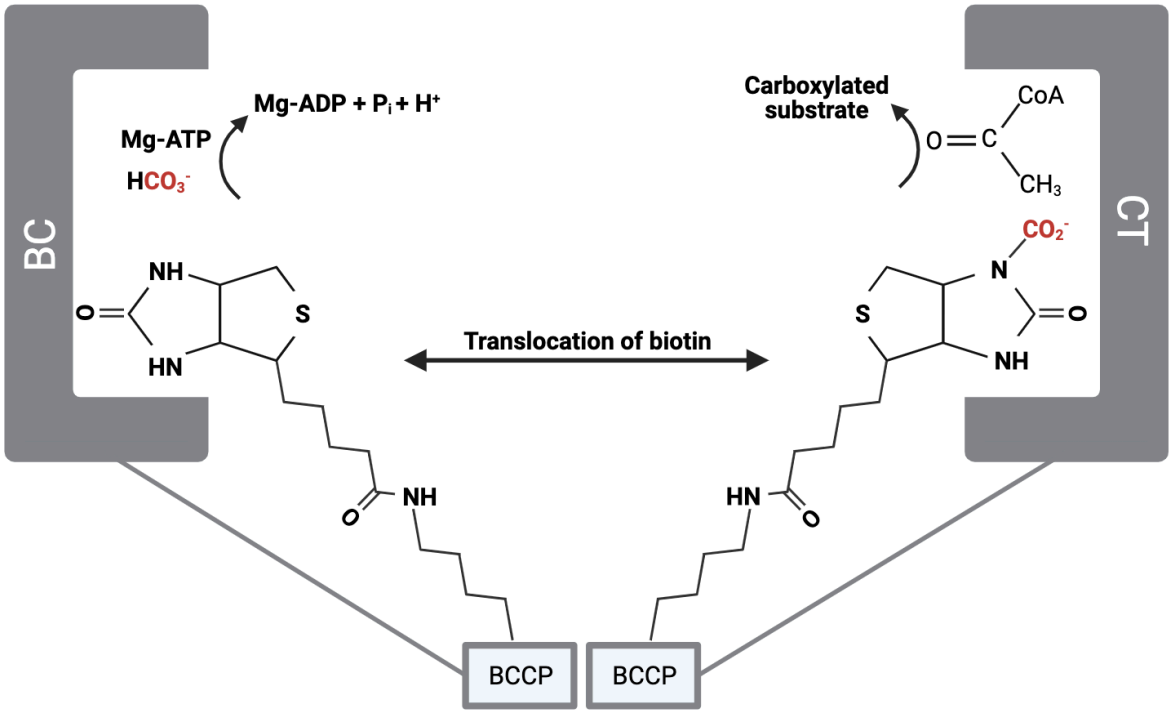


Figure 35: Diagram showing how biotin functions in complex with biotin-dependent carboxylases. Biotin is attached to the biotin-dependent carboxylase by its valeryl side chain at the BCCP domain. Biotin is translocated between the BC domain, where it picks up a CO_2 moiety, and the CT domain where it carboxylates the substrate. (Reproduced based on Tong, 2013) Created using Biorender.com.

5.1.2.1 Biotin biosynthesis

Biotin, despite being an essential co factor within all three domains of life, is only synthesised by bacteria, fungi, and plants (Hu and Cronan, 2020). The biotin biosynthesis pathway (BBP) has not been described within *P. aeruginosa*; however, it is well characterised in *E. coli* and *M. tuberculosis*. Therefore, the *P. aeruginosa* pathway can be inferred from the elucidation of these pathways (Lin and Cronan, 2011;

Salaemae et al., 2011). Biotin synthesis is divided into two stages, where the first stage involves the formation of pimeloyl-ACP, see Figure 36 (Lin and Cronan, 2011; Salaemae et al., 2011).

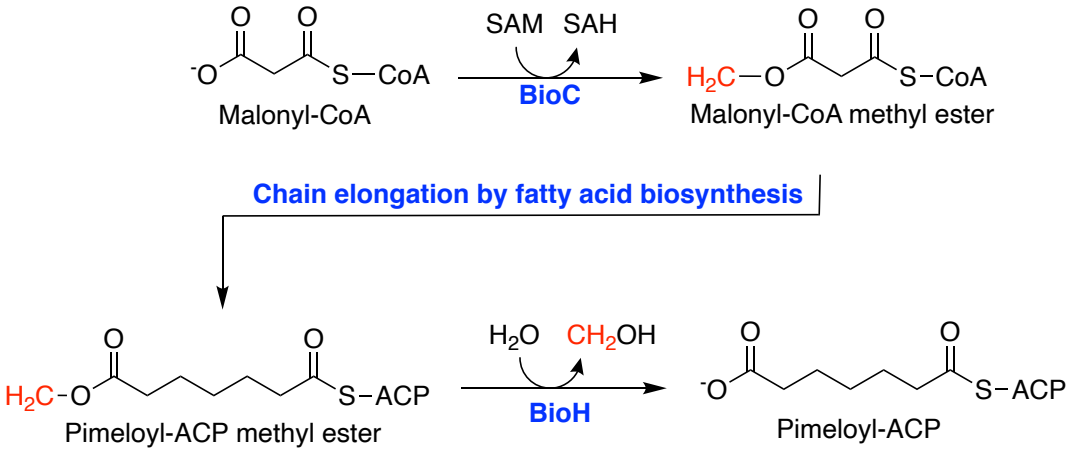


Figure 36: The first stage of biotin synthesis producing pimeloyl-ACP. (Based on Lin and Cronan, 2011)

Pimeloyl-ACP is then passed onto the second stage of biotin synthesis where the bicyclic rings are formed (Fig. 37). BioF, an 8-amino-7-oxononanoate synthase, is the first enzyme in this stage. BioF catalyses a decarboxylative condensation reaction between pimeloyl-ACP and L-alanine, forming 8-amino-7-oxonanoate (Lin and Cronan, 2011). Following this, BioA transaminates 8-amino-7-oxonanoate, using SAM as an amino donor to form 7,8-diamino-nonanoate (Lin and Cronan, 2011). Cyclisation of the ureido ring is then performed in an ATP dependent reaction, by BioD, to form dethiobiotin. The tetrathiophene ring is the final element to be formed in order to

produce fully synthesised biotin. This final reaction is performed by BioB and requires two SAM molecules (Lin and Cronan, 2011).

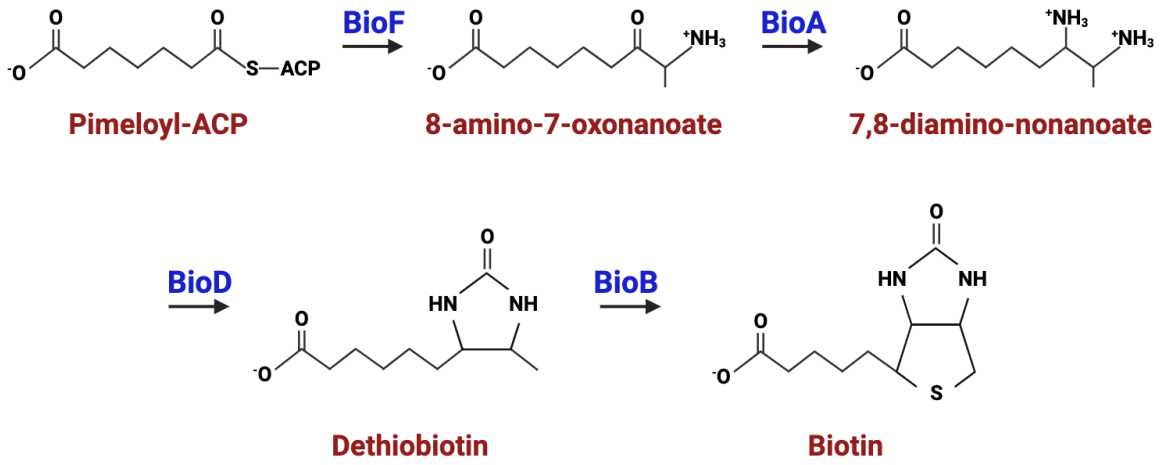


Figure 37: The second stage of biotin synthesis. Labels in blue represent biotin biosynthesis enzymes and labels in red represent intermediate names (Based on Lin and Cronan, 2011) Created using Biorender.com.

5.1.2.2 Biotin-dependent carboxylases in *P. aeruginosa*

Biotin function within specific organisms relates to the biotin-dependent carboxylases an organism produces. Some bacteria, such as *E. coli* only produce one biotin-dependent carboxylase. However, *P. aeruginosa* encodes six, namely acetyl-CoA carboxylase (ACC), acyl-CoA carboxylase (YCC), 3-methyl-crotonyl-CoA carboxylase (MCC), geranyl-CoA carboxylase (GCC), pyruvate carboxylase (PC), and urea carboxylase (UC). This vast profile of carboxylases provides a multitude of different functions for the cell (Tong, 2013).

Acetyl-CoA carboxylase, functions in the first committed step of the type II fatty acid synthesis (FASII) pathway, converting acetyl-CoA to malonyl-CoA (Tong, 2013). This pathway is involved in producing fatty acid chains which are required for phospholipid biosynthesis. Therefore, the function of ACC is vital for cell membrane integrity. This is because the cell membrane is largely comprised of these phospholipids and the majority of these are glycerolipids, which contain two fatty acid chains (Parsons and Rock, 2011). Further to this, the products of the fatty acid biosynthesis pathway have been shown to be important for quorum sensing in bacteria, an integral part of *P. aeruginosa's* virulence (Polyak et al., 2012).

Acyl-CoA carboxylase, the second biotin-dependent carboxylase, is only categorised by its presence within the *P. aeruginosa* genome. It has yet to be characterised biochemically and is therefore designated with the generic name for carboxylases which carboxylate CoA esters of organic acids (Tong, 2013). Little more is known about it other than its multi-domain structure formed of subunits which combine to a size of approximately 120 kDa (Tong, 2013).

The third and fourth biotin-dependent carboxylases are 3-methyl-crotonyl-CoA carboxylase (MCC) and geranyl-CoA carboxylase (GCC) (Tong, 2013). Both MCC and GCC share a similar subunit formation with an α subunit and a CT domain containing β subunit. Bridging this CT domain is a BT domain which allows it's interaction with the BC-BCCP domain for carboxylation of substrates (Tong, 2013). MCC and GCC are mentioned together since they form part of the same pathway, the metabolism of acyclic terpenes to acetoacetate whilst releasing acetyl-CoA, Figure 38 (Aguilar et al., 2008; Tong, 2013). GCC functions within the second step of the acyclic terpene

pathway, following the conversion of the acyclic terpene to cis-geranyl-CoA by the dehydrogenase AtuD. GCC carboxylates cis-geranyl-CoA to form isohexenyl-glutaconyl-CoA. This then undergoes hydratase and lyase activity before two rounds of β-oxidation to produce 3-methyl-crotonyl-CoA. It is this product that is carboxylated by MCC, forming 3-methyl-glutaconyl-CoA which continues along the leucine catabolism pathway to the formation of acetoacetate (Aguilar et al., 2008; Tong, 2013). The function of GCC and MCC allows *P. aeruginosa* to use acyclic terpenes as a carbon source (Tong, 2013). As shown in Figure 38, the substrate of MCC is also the second intermediate in the leucine catabolism pathway, produced from the dehydrogenation of leucine by LiuA.

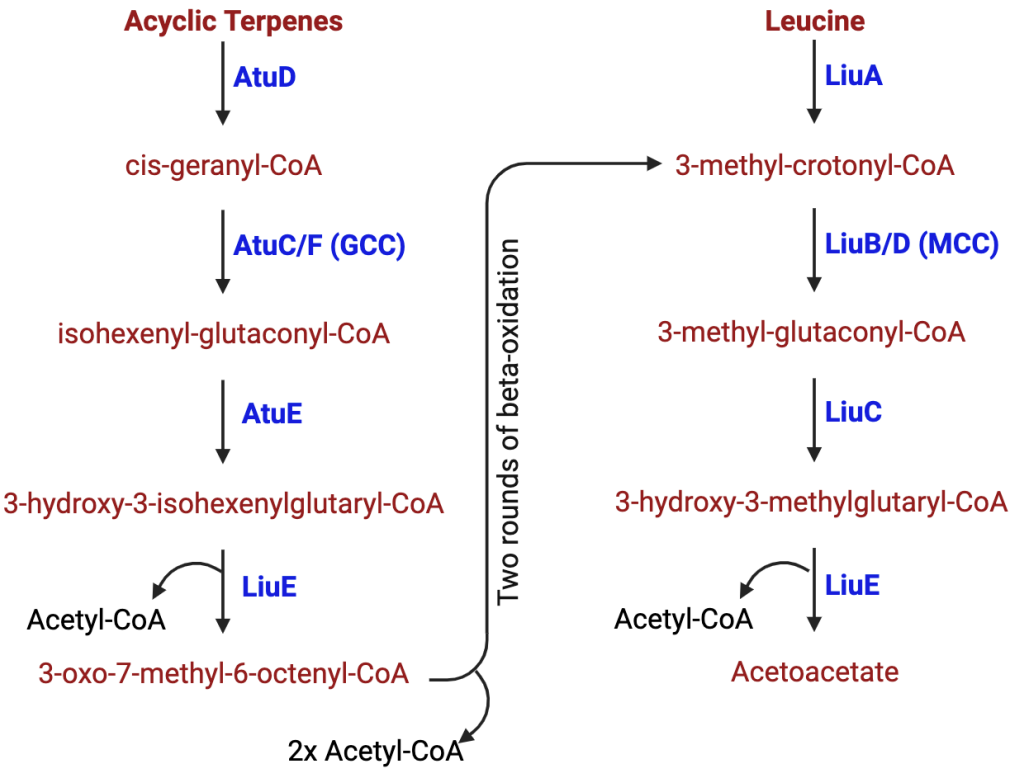


Figure 38: A schematic showing the involvement of GCC and MCC in the acyclic terpene and leucine metabolic pathways. (Based on Aguilar et al., 2008). Created using Biorender.com.

Across the whole acyclic terpene catabolism process, four acetyl-CoA molecules are released (Aguilar et al., 2008). As previously mentioned, acetyl-CoA is integral for fatty acid biosynthesis, but not only this, it is vital in many other biological processes. The major consumption of acetyl-CoA being within the TCA cycle, where condensation with oxaloacetate produces citrate, which feeds into the cycle and generates energy for the cell (Krivoruchko et al., 2015). This suggests that the acyclic terpene catabolism pathway could be an alternative route for energy generation, further enhancing *P. aeruginosa's* adaptability to multiple environmental niches and conditions.

The fifth biotin-dependent carboxylase is pyruvate carboxylase a two-subunit protein with a BC domain within the α subunit and a BCCP and CT domain within the β subunit (Tong, 2013). PC plays a significant anapleurotic role within the cell since it converts pyruvate to oxaloacetate. Therefore, replenishing the TCA cycle when intermediates have been drained by alternative pathways, including gluconeogenesis and fatty acid biosynthesis (Lai et al., 2006; Tong, 2013). As a major route of energy production within the cell, the TCA cycle is vital for bacterial survival. The reduction of electron carriers such as NAD+ and FAD during the various dehydrogenase reactions of the cycle, interconnects with the oxidative phosphorylation pathway to produce large quantities of ATP (Berg et al., 2002). Furthermore, oxaloacetate produces the precursors for many vital biomolecules such as nucleotides and proteins (Berg et al., 2002).

Finally, the sixth biotin-dependent carboxylase of *P. aeruginosa* is urea carboxylase. UC is a further example of *P. aeruginosa's* adaptability, permitting the utilisation of urea as a nitrogen source. In *P. aeruginosa* UC catalyses the carboxylation of urea to form allophanate, which is then converted by allophanate hydrolase to ammonia and carbon dioxide (Kanamori et al., 2004; Tong, 2013).

Due to the high number of biotin-dependent carboxylases within *P. aeruginosa*, and the variety of functions they play, they may enact an important role in the resistance to metformin. It may be that the increased number of biotin-dependent carboxylases are the reason why *P. aeruginosa* is affected by metformin with the loss of biotin synthesis when other bacteria, such as *E. coli,* are not. Perhaps one or more of the functions of these pathways compensate for the effect of metformin on the cell and therefore with the loss of biotin synthesis are unable to perform such functions.

5.1.3 Metformin

Metformin, as shown in Figure 39, is an anti-hyperglycaemic type 2 diabetes drug. Its mechanism of action is currently unknown. However, multiple theories of how it reduces blood glucose levels have been put forward. The first theory is that it initially targets the mitochondrial respiratory chain complex I, which is responsible for the first step in the oxidative phosphorylation pathway. During this pathway, complex I oxidises NADH from the TCA cycle to produce two electrons for the reduction of ubiquinone to ubiquinol. Re-oxidisation of ubiquinol by the cytochrome bc1 complex then enables the reduction of oxygen to water at complex IV, via the transfer of electrons. Complexes I,

III and IV, use the redox energy produced to pump protons across the inner mitochondrial membrane, producing a proton gradient (Sharma et al., 2009). The generated proton motive force allows complex V to generate ATP, and thus produce energy for the body (Sharma et al., 2009).

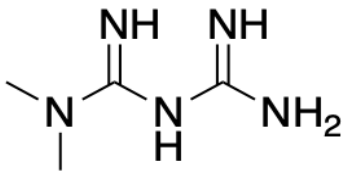


Figure 39: Chemical structure of metformin.

It may be that by inhibiting mitochondrial respiratory chain complex I metformin would lead to a reduction in NADH oxidation, proton transportation across the mitochondrial inner membrane, and rate of oxygen consumption. Ultimately decreasing the proton gradient and in turn proton driven ATP synthesis (Viollet et al., 2012). AMPK activation is triggered by the reduction in the ATP:ADP ratio, inducing the cell to switch from an anabolic to a catabolic state. Therefore, biosynthetic pathways that are heavily ATP-consuming, such as gluconeogenesis, become inhibited. With a reduction in gluconeogenesis, the level of glucose in the blood is decreased (Viollet et al., 2012).

Madiraju et al. proposed a different theory, suggesting that metformin targets mitochondrial glycerophosphate dehydrogenase (mGDH) (Madiraju et al., 2014). mGDH converts glycerol-3-phosphate (G3P) into dihydroxyacetone phosphate (DHAP), via the oxidation of G3P (Mráček et al., 2013). G3P conversion is at the core

of multiple pathways, including glycerol catabolism, gluconeogenesis, and oxidative phosphorylation. Within glycerol catabolism, G3P is the produced from the conversion of glycerol, catalysed by glycerol kinase. Within gluconeogenesis, the DHAP is fed into the cycle to aid glucose production. Finally, G3P conversion to DHAP is coupled with the reduction of FAD^+ to $FADH_2$ within the oxidative phosphorylation pathway, facilitating electron transfer (Mráček et al., 2013). This theory proposes a double-edged sword approach, as metformin would not only reduce blood glucose by directly inhibiting gluconeogenesis, but again also reduce the ATP:ADP ratio, via inhibition of the oxidative phosphorylation pathway (Madiraju et al., 2014; Viollet et al., 2012). Interestingly, homologues of both mitochondrial respiratory chain complex I and mGDH are present within *P. aeruginosa*. The homologue of mitochondrial respiratory chain complex I is encoded by the *nuo* genes which together form the respiratory chain complex I, known as NDH-1 within *P. aeruginosa* (Friedrich and Scheide, 2000; Noda et al., 2006). The homologue of mGDH is encoded by *glpD* (Shuman et al., 2018). Therefore, it may be that one or both of these homologues are the target of metformin within *P. aeruginosa.*

Despite the lack of a known mechanism, metformin's safety and efficacy has been shown through its extensive use around the world. First prescribed within the 1950's, there are now 120 million people prescribed metformin globally (Viollet et al., 2012). This proven safety makes metformin a prime candidate for repurposing as an antibiotic.

5.1.3.1 Metformin as a quorum sensing inhibitor

Recent research has indicated that metformin could already be a potential quorum sensing inhibitor within *P. aeruginosa* (Hegazy et al., 2020). As previously described, quorum sensing is a vital component of virulence and pathogenicity within *P. aeruginosa* since it is responsible for controlling gene expression across communities of *P. aeruginosa* cells (Pang et al., 2019). Therefore, the ability of metformin to not only kill *P. aeruginosa* but reduce pathogenicity would be hugely beneficial in the repurposing of metformin as an antimicrobial agent. The study by Hegazy et al., focused on a variety of QS associated features and the ability of metformin treatment to reduce them. Metformin was given at subinhibitory concentrations of 10 mg/mL. Metformin was able to reduce biofilm formation in PAO1 by 67%, swimming by 44%, swarming by 55% and twitching by 69%, when compared to no treatment. The levels of virulence enzyme production showed a reduction of 43%, 76% and 26% for protease, hemolysin and elastase respectively. Furthermore, a significant decrease (between 2- and 4-fold) in the expression of the QS genes *lasI*, *lasR*, *rhlI*, *rhlR*, *pqsA* and *pqsR* was found. An *in vivo* experiment treating *P. aeruginosa* infected mice with metformin was also performed, unfortunately treatment with metformin did not reduce the ability of *P. aeruginosa* to kill the mice (Hegazy et al., 2020). Therefore, combining metformin with biotin synthesis inhibition could be a solution to enhance the effectiveness of metformin.

Since the antibiotic effect of metformin is unknown within *P. aeruginosa* the major aim of this chapter was to elucidate the mechanism of action of metformin. Further to this,

the second focus was understanding how biotin mediates resistance to metformin killing.

5.2 Results

5.2.1 The BBP operon structure differs within *Pseudomonas*

In a previous unpublished study by Banzhaf et al., it was discovered that in *Escherichia coli*, disruption of the biotin synthesis operon did not result in killing by metformin. To try and understand the reason for this difference in phenotype, the structural differences within the biotin synthesis operon between *Pseudomonadaceae* and *Enterobacteriaceae* were observed and compared. This was performed by visualising representative strain biotin operons for *Enterobacteriaceae* and *Pseudomonadaceae* on BioCyc. The operon structure for the *Enterobacteriaceae* strains showed a conserved gene structure, where bioH is located separately from the rest of the biotin biosynthesis pathway genes. Within *Enterobacteriaceae bioA* is located on the negative strand upstream of *bioB/F/C/D* (Fig. 40). Despite this, *Pseudomonadaceae* shows a different operon structure. Within the *Pseudomonadaceae* family, *bioH* is located within the BBP operon upstream of *bioC*, and *bioA* is located separately (Fig. 40).

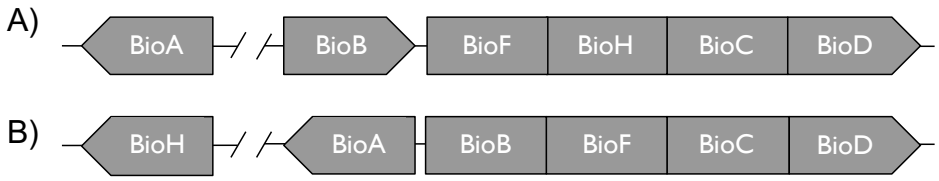


Figure 40: BBP operon structures of A) Pseudomonadaceae: *Azomonas*, *Permianibacter*, *Pseudomonas*, *Oblitimonas,* and *Thiopseudomonas* and B) Enterobacteriaceae: *Dickeya*, *Erwinia*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Vibrio* and *Yersinia*

5.2.2 Validating biotin mediated metformin resistance in *Pseudomonas aeruginosa*

Since chemical genomic screens can occasionally give rise to false hits, it was vital to confirm that disruption in biotin synthesis truly leads to sensitivity to metformin. Therefore, the biotin synthesis transposon mutants as well as the clean knockout PA14 Δ*bioB-H* were tested for their sensitivity to metformin via a spotting plate assay. The biotin synthesis mutants, PA14 (WT), and PA14 Δ*glpD*, were plated on MOPS minimal media agar supplemented with either: 40 nM biotin, 40 nM biotin and 100 mM metformin, 400 nM biotin, or 400 nM biotin and 100 mM metformin (Fig. 41). All strains grew on the 40 nM biotin and 400 nM biotin plates and swarming motility was observed. When the 40 nM biotin plates were supplemented with 100 mM metformin, only PA14 and Δ*glpD* strains grew. This confirms that in the absence of the biotin synthesis pathway that metformin has antimicrobial activity against *P. aeruginosa*. To confirm

that it is the loss of biotin that is resulting in sensitivity to metformin, the strains were plated on excess biotin (400 nM) and metformin (100 mM). Supplementation with excess biotin complemented the growth of the biotin synthesis mutants, confirming biotin's role in resistance.

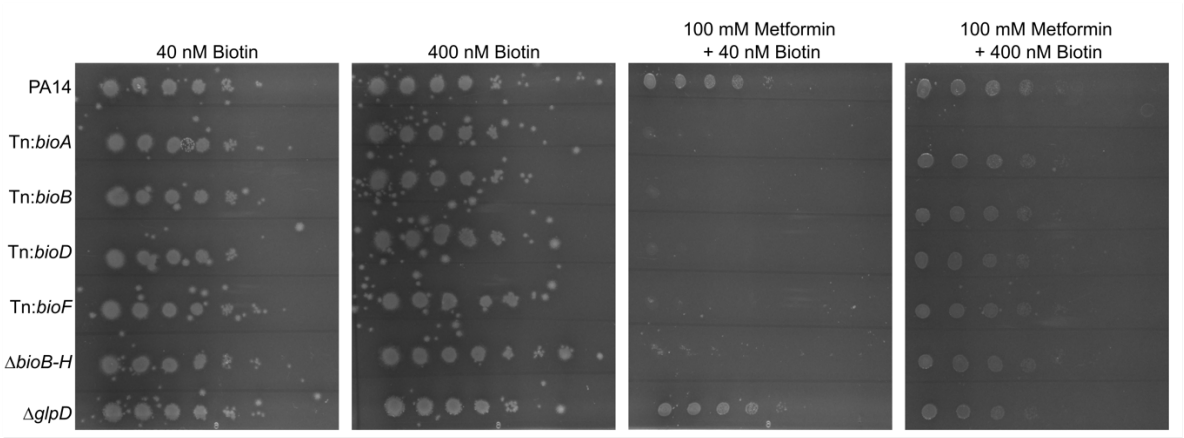


Figure 41: Spotting plates of MOPS minimal media agar to test the effect of metformin on the growth of biotin synthesis deficient PA14 mutants.

5.2.3 MIC of metformin against *Pseudomonas aeruginosa*

Following confirmation that biotin synthesis disruption leads to metformin sensitivity, it was important to determine the MIC at which metformin prevented growth. The MIC for wild type PA14 within MHB was 250 mM metformin, whereas the MIC for PA14 Δ*bioB-H* was 7.8 mM. Therefore, there is a 32x reduction in MIC with the loss of the biotin synthesis pathway for PA14. One hypothesised target of metformin was the NADH dehydrogenase, NDH-1, due to its homology to the proposed human target. Loss of NDH-1 has been shown to cause increased gentamicin and kanamycin

resistance (Hreha et al., 2021). This is due to a disruption in the proton motive force, which these antibiotics require to enter the cell (Hreha et al., 2021). Therefore, the MIC of PA14 and PA14 Δ*bioB-H* for gentamicin and kanamycin in the presence of metformin was tested. Since the MIC of metformin for PA14 Δ*bioB-H* was 7.8 mM the concentration used of metformin was 3.9 mM for both PA14 and PA14 Δ*bioB-H*. The concentration of metformin may have been too high for PA14 Δ*bioB-H* in combination with gentamicin and kanamycin since no growth was observed. However, growth was seen for PA14. A 2-fold decrease in MIC was seen in the presence of metformin for both gentamicin and kanamycin, with MICs of 0.5 µg/ml and 16 µg/ml compared to 1 µg/ml and 32 µg/ml respectively.

Table 10: Table showing the MIC values for PA14 and PA14 Δ*bioB-H*.

Strain	**Metformin (mM)**	**Gentamicin (µg/mL)**	**Gentamicin + 3.9 mM Metformin (µg/mL)**	**Kanamycin (µg/mL)**	**Kanamycin + 3.9mM Metformin (µg/mL)**
PA14	250	1	0.5	32	16
Δ*bioB-H*	7.8	0.5	No growth	64	No growth

5.2.4 Biotin synthesis does not provide metformin resistance in *Acinetobacter baumannii*

Similar to *P. aeruginosa*, *Acinetobacter* has an operon encoded BioH (Cao et al., 2017). Further to this, a study by Carfrae et al., showed that the biotin requirement of

both *P. aeruginosa* and *Acinetobacter baumannii* was much higher than that of *Enterobacteriaceae* species (Carfrae et al., 2020). Due to this, it was theorised that *Acinetobacter baumannii* may respond similarly to *P. aeruginosa* in the absence of a functioning biotin synthesis pathway. Therefore, it was decided that biotin synthesis transposon mutants of *A. baumannii* AB5075 would be similarly tested for sensitivity to metformin in a spotting plate assay. As such, transposon mutants AB5075 *bioA*::kr, AB5075 *bioB*::kr, AB5075 *bioC*::kr, AB5075 *bioD*::kr, and AB5075 *bioF*::kr were tested on MOPS minimal media agar supplemented with either: 40 nM biotin, 40 nM biotin and 100 mM metformin, 400 nM biotin, or 400 nM biotin and 100 mM metformin (Fig. 42) (Gallagher et al., 2015). Similar to *P.* aeruginosa, all strains grew well on 40 nM and 400 nM biotin. However, where no growth was seen with metformin supplementation for *P. aeruginosa*, growth was observed for the *A. baumannii* mutants. Despite this, the colonies were much fainter and therefore growth was reduced, suggesting that metformin is partially affecting the growth of the *A. baumannii* transposon mutants. However, addition of excess biotin (400 nM) did not restore growth to the level of no metformin. There are a few reasons why this may be. Firstly, unlike *P. aeruginosa, Acinetobacter baumannii* does not have a biotin transporter pump (Carfrae et al., 2020). Therefore, its ability to gain exogenous biotin is likely severely reduced compared to *P. aeruginosa* (Carfrae et al., 2020). Thus, although the biotin is present in the media the strains may not be able to utilise it effectively. Secondly, it may be that the loss of biotin synthesis does not infer resistance to metformin within *A. baumannii*, and therefore the loss of biotin synthesis does not induce a sensitivity to metformin.

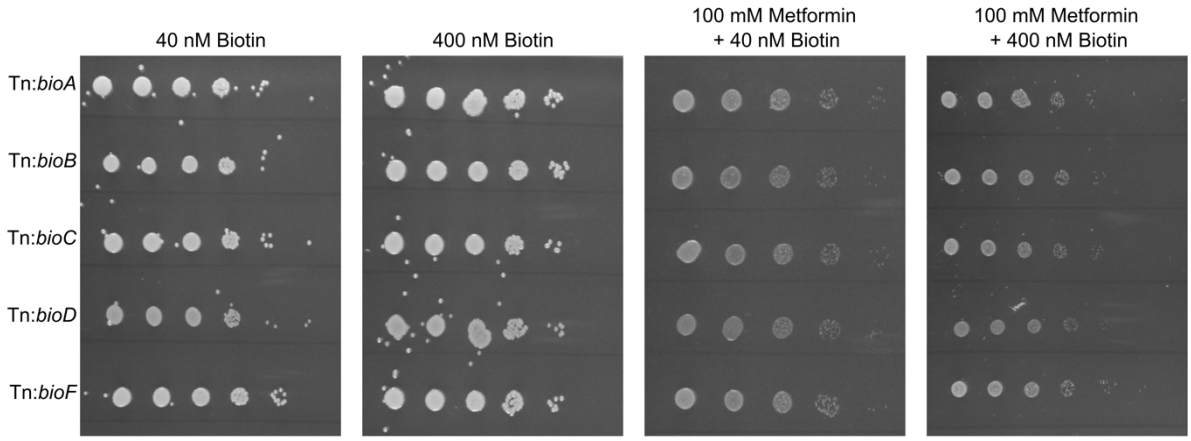


Figure 42: Spotting plate assay on MOPS minimal media agar to test the effect of metformin on the growth of biotin synthesis deficient *Acinetobacter baumannii* AB5075 mutants.

To test this second hypothesis, the MIC of metformin against the *A. baumannii* biotin synthesis mutants as well as the wildtype (AB5075) was tested. Following testing it was concluded that the MIC for all strains was the same at 250 mM metformin. Therefore, it is valid to conclude that unlike in *P. aeruginosa*, the loss of biotin synthesis does not render *A. baumannii* susceptible to killing by metformin.

5.2.5 Loss of individual biotin-dependent carboxylases does not result in metformin sensitivity

In order to understand if any particular biotin-dependent carboxylase was responsible for the resistance to metformin, the S-scores of the biotin-dependent carboxylase

mutants in metformin was analysed (Fig. 43). The only mutant which showed a significant phenotype was one replicate for the GCC alpha subunit in 25 mM metformin. This mutant showed improved fitness upon loss, with an S-score of 4.1. Despite this, the phenotype was not retained in any of the other concentrations or within the alternative transposon mutant of the GCC alpha subunit. This suggests the result may have been a false positive. Therefore, it is more likely that reduced function of multiple or all of the carboxylases due to biotin loss is responsible for susceptibility in metformin.

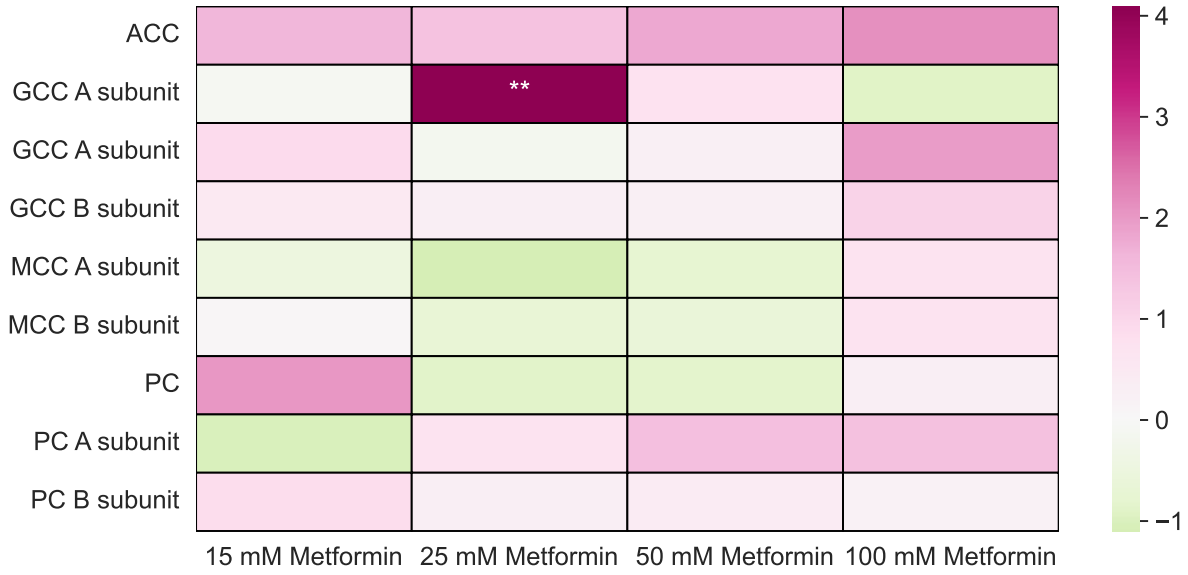


Figure 43: S-scores for biotin-dependent carboxylases in the presence of metformin. ** = 0.001 < 5% FDR p-value ≤ 0.01. A subunit = alpha subunit, B subunit = beta subunit.

5.2.6 Supplementation with oxaloacetate reduced sensitivity to metformin.

Although loss of individual carboxylases did not result in any significant changes in metformin sensitivity, the effect of supplementing PA14 Δ*bioB-H* with biotin-dependent carboxylase products or end points was tested. Contrary to the previous MIC test of PA14 Δ*bioB-H* in MHB, in this analysis the MIC for metformin was lower and not captured in the base MHB plates. However, the MIC for PA14 remained the same, see Figure 44. MHB plates were supplemented with either acetoacetate, oxaloacetate, or succinate. Acetoacetate was chosen since it is the end product of the terpene degradation and leucine degradation pathways in which GCC and MCC are fundamental. Oxaloacetate was chosen since pyruvate carboxylase function involves the conversion of pyruvate to oxaloacetate for entry into the TCA cycle. Finally, succinate was chosen since it is the last product of the TCA cycle and feeds into the oxidative phosphorylation pathway. Supplementation of oxaloacetate resulted in an increased MIC for PA14 Δ*bioB-H*, improving the MIC to 62.5, however growth was severely reduced between 7.8 mM and 31.25 mM metformin. Furthermore, addition of succinate doubled the MIC for the PA14 wildtype and resulted in very slight growth in 0.98 mM metformin for PA14 ΔBioBH. This suggests that addition of succinate is beneficial in general in presence of metformin.

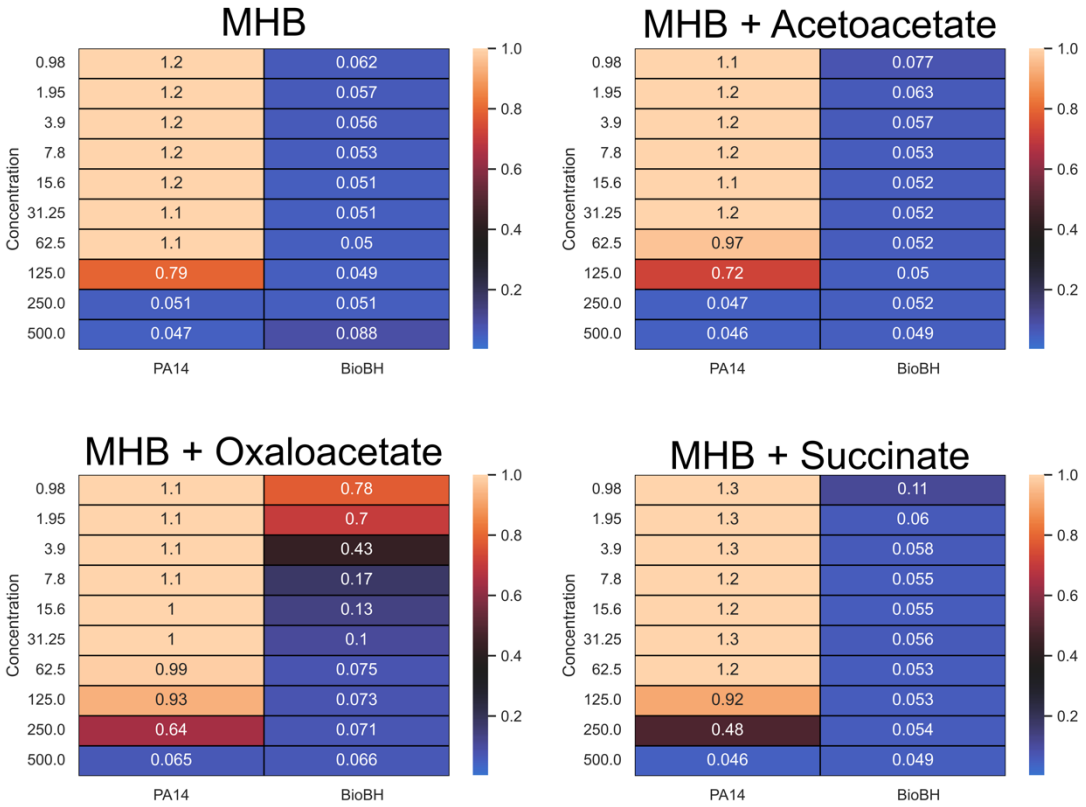


Figure 44: MIC assays for PA14 Δ*bioB-H* in MHB with biotin-dependent carboxylase products supplemented. Values represent optical density at 600 nm (OD600).

5.2.7 PA14 Δ*bioB-H* did not gain spontaneous suppressor mutations

In mechanism of action studies, suppressor mutation screens are often employed to elucidate the target of a drug (Abrahams et al., 2012). When exposing sensitive strains to high concentrations of drugs, spontaneous mutations with increased resistance can be selected for. Often these mutations occur in the target of a drug, reducing the ability of drugs to bind or causing an overexpression of the target to create a dosage effect. Therefore, a suppressor screen was performed using PA14 Δ*bioB-H* in the presence

of increasing concentrations of metformin. Unfortunately, in the first screen, a contaminant was passaged and therefore the experiment was repeated. Across multiple attempts, no suppressor mutants were gained for PA14 Δ*bioB-H*. Since spontaneous mutants with suppression phenotypes could not be produced organically, this study attempted targeted mutagenesis via the production of a Transposon Directed Insertion Sequencing (TraDIS) library.

5.2.8 Constructing a TraDIS library in PA14 Δ*bioB-H*

TraDIS is a powerful technique used for the elucidation of gene essentiality, especially across different conditions (Langridge et al., 2009). For the technique, dense pooled transposon mutant libraries are produced, with each mutant containing a single transposon insertion and the library containing insertions roughly every 6-8 base pairs. When initially producing these libraries, insertions do not typically occur in essential genes, as loss of function should result in lack of growth. As such, we can determine the base essentiality of genes within the laboratory environment. Furthermore, by growing these libraries in different conditions, we can enrich for mutants whose transposon insertion gives them a fitness advantage. As such, mutants with transposon insertions in genes that are important within the tested condition will have reduced fitness and will appear in reduced numbers compared to the base dataset. Conversely, genes where disruption is beneficial in a given condition will show increased numbers of transposon insertions compared to the base library. As such, this study aimed to produce a TraDIS library in PA14 Δ*bioB-H* and expose this library to a range of metformin concentrations. The hypothesis was that gene mutations which

incurred a fitness advantage in metformin may indicate the mechanism of action of metformin.

This study successfully produced a TraDIS library containing roughly 1.2 million mutants. Aliquots of this library were grown in 0, 31.25 mM, 62.5 mM, or 125 mM metformin before the insertions were sequenced. These values represented 0x, 0.5x, 1x and 2x MIC of PA14 Δ*bioB-H* in minimal media supplemented with metformin. For each condition three biological replicates were produced. For each of the replicates the total number of unique insertions was similar, with a range of 407051 – 600397, and an average of 520565 insertions. Furthermore, on average there was an insertion every 13 base pairs, see Table 11.

Table 11: Insertion statistics for each sequencing replicate. % Mapped is a percentage of the matched reads not total reads. UIs = Unique insertions.

MIC	Replicate	Total Reads	% Matched	% Mapped	Unique Insertions	Seq Len/ UIs
0x	1	9051308	86.917	80.540	407051	16.0610
	2	7290687	83.674	82.977	466615	14.0108
	3	8582552	87.475	82.938	510387	12.8092
0.5x	1	19010519	86.992	86.472	429526	15.2206
	2	16164628	84.669	84.268	534884	12.2226
	3	27633533	87.748	84.723	584429	11.1864
1x	1	15447717	87.565	80.492	596669	10.9569
	2	13565468	84.914	87.658	594653	10.9941
	3	21777152	87.758	85.126	600397	10.8889
2x	1	21796612	85.778	76.060	477264	13.6982
	2	16281513	84.600	83.191	481411	13.5802
	3	12866151	87.084	83.791	563494	11.6020

For a library to be useful, insertions need to be dense and evenly spread. To check the density of each library, the number of insertions for each replicate were concatenated before plotting in relation to the PA14 genome using DNAPlotter. Each of the libraries contained a similar density and shared some features, including a large peak of insertions between the 0 and 325,000 bp position, see Figure 45. It was important that there were no large segments of unexplained unmatched insertions. The only small region without insertions appeared in all libraries just following the 650,000 bp position. However, from a previous study that performed a TraDIS screen in the PA14 wildtype, this region contained a set of core essential genes, explaining the lack of insertions (Poulsen et al., 2019). Given that the libraries were of good quality, the genes with alterations in insertion number were evaluated.

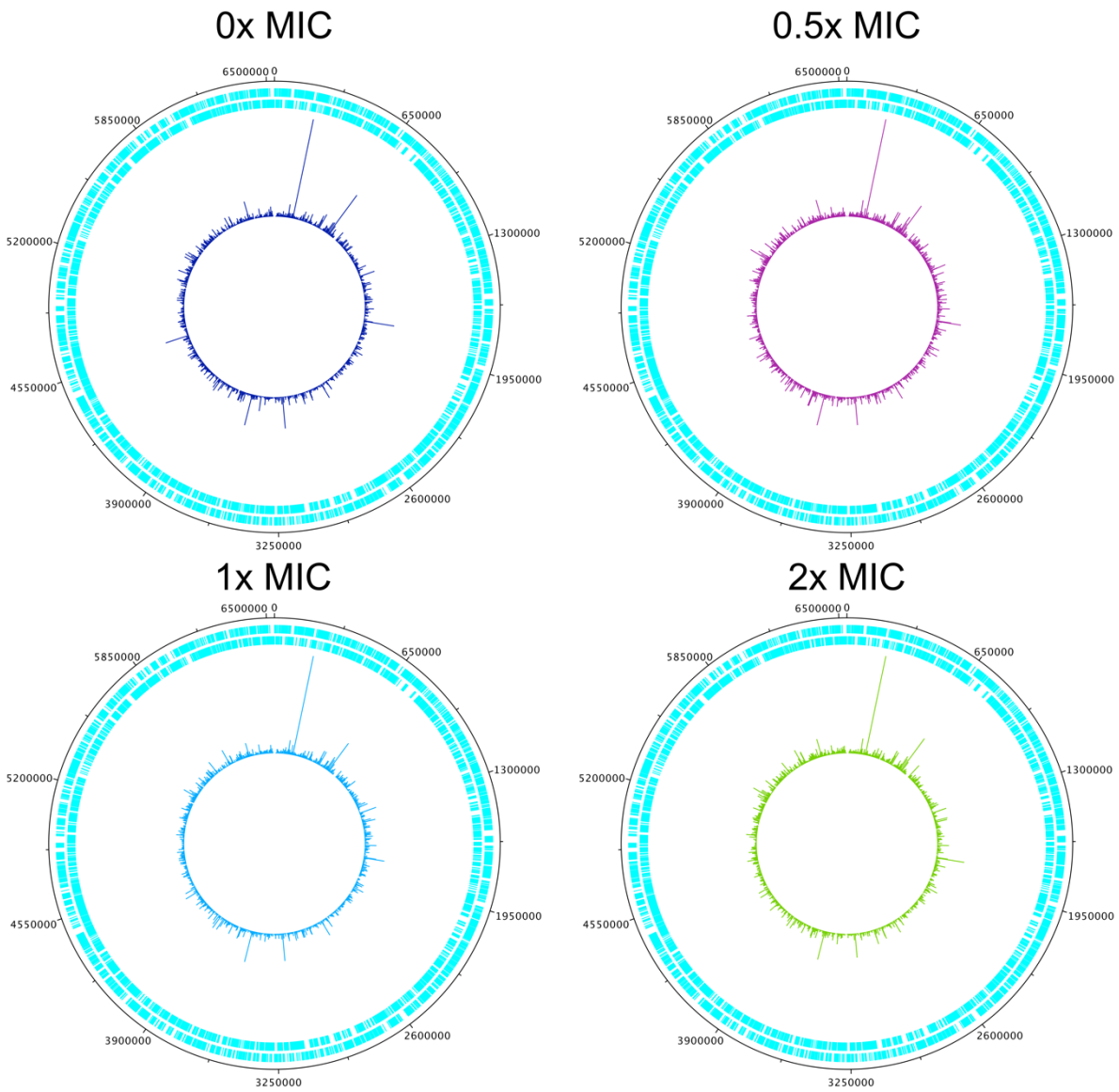


Figure 45: Transposon insertion maps for each concentration of metformin in relation to the genome sequence of PA14. Produced using DNAPlotter.

5.2.9 Multiple genes became non-essential in the presence of metformin.

Looking for genes which showed a loss of essentiality in the presence of metformin could indicate potential targets of metformin. Genes were categorised as essential using the BioTraDIS tradis_essentiality.R module. Genes which were categorised as essential within the 0x MIC condition but not in the metformin containing conditions, were considered as 'no longer essential in the presence of metformin'. In total 43 genes were concluded as no longer essential (NLE), suggesting a positive genetic interaction with biotin synthesis loss in presence of metformin. The KEGG categories for each gene were assigned to elucidate the function of the NLE genes.

The most common categories were genes involved in 'Amino acid metabolism' and 'Nucleotide metabolism' with 9 genes each, see Figure 46B. Within these categories, multiple genes with functions in purine biosynthesis and arginine pathways were found (Fig. 46C). These included *purD*, *purE*, *purF*, *purL* and *purM*, as well as *argA*, *argB*, *argC*, *argF*, *argG* and *argH* (Fig. 46A). The average LogFC for genes within these sets, across all concentrations, was 2.03 for the *pur* genes, and 1.62 for the *arg* genes.

The gene with the highest log-fold change for all concentrations of metformin compared to the no metformin control was *epd* (Fig. 46A). This can also be seen within the transposon insertion plot in Figure 46D, where the metformin containing conditions show increased insertion density. *Epd* encodes D-erythrose 4-phosphate dehydrogenase and is responsible for the formation of 4-phosphoerythronate within the pyridoxal 5'-phosphate (PLP) biosynthesis pathway (Denise et al., 2023). The log fold changes for this gene were 2.76, 3.6 and 4.02 for 0.5x, 1x and 2x MIC, respectively. This was closely followed by *pdxJ* with scores of 2.66, 3.39, and 4.05

(Fig. 46A/D). PdxJ is involved in the biosynthesis of vitamin B6, which follows from the formation of PLP. Two other genes within this pathway, *pdxA* and *pdxH*, were also found within this analysis. Given the proposed function of metformin within humans being associated with inhibition of the NADH oxidizing mitochondrial respiratory complex I, it is interesting that the NAD biosynthesis genes are no longer essential. *nadA*, *nadB* and *nadC* had log fold change values between 1.13 and 3.51, see Figure 46A/D.

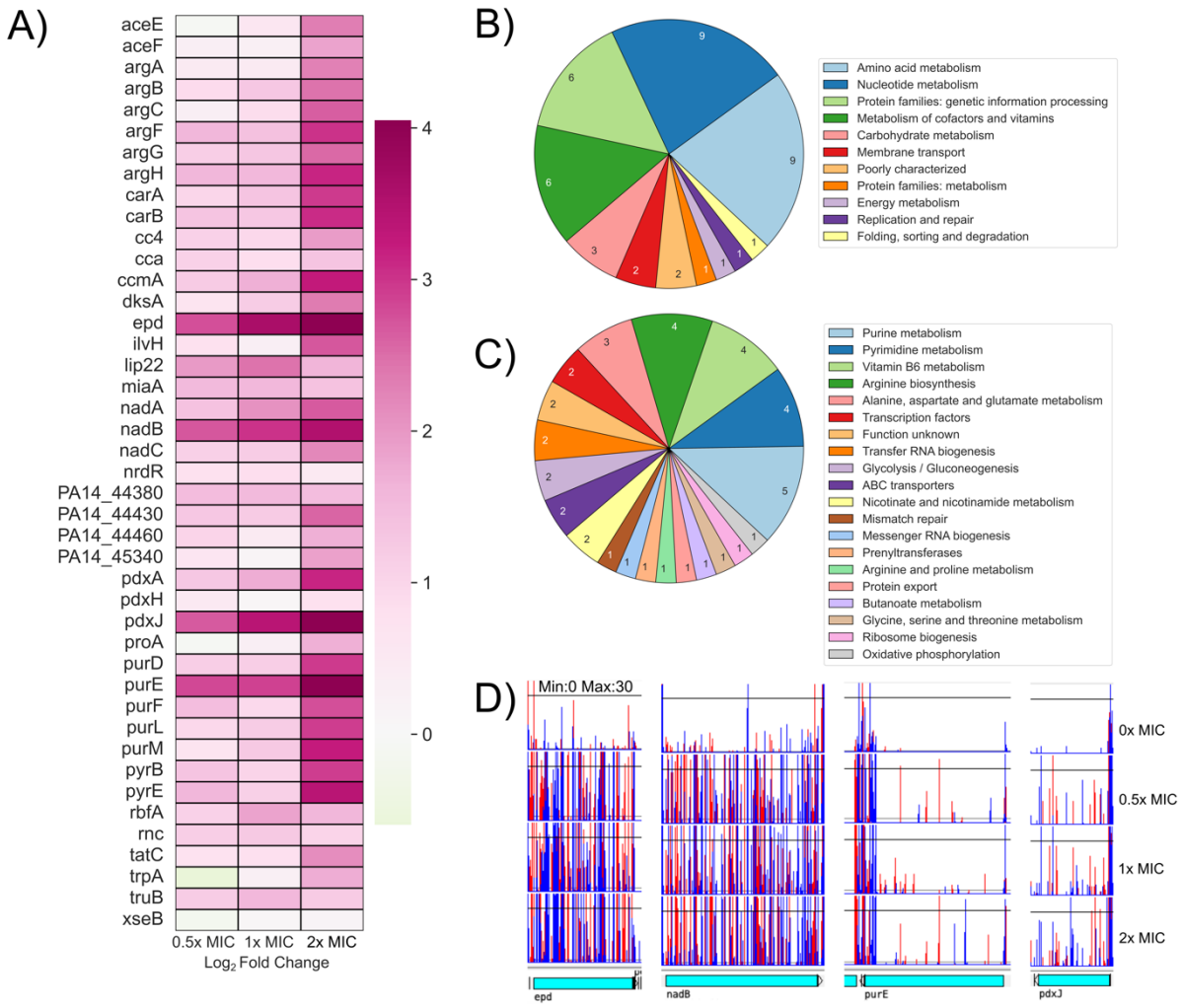


Figure 46: Multiple genes were no longer essential in the presence of metformin. A) heatmap of the genes which were no longer essential. Colour bar represents the log fold change from the 0 mM metformin condition. B-C) pie charts representing the KEGG categories and sub-categories to which the NLE genes are associated. D) TraDIS reads representing transposon insertions within the genes *epd*, *nadB*, *purE*, and *pdxJ*. Colours represent direction of insertion with red representing forward and blue representing reverse. The height of the lines indicate the number of insertions within a given position. Max value indicates the maximum height of insertions contained within the frame. Max insertion height is 30.

5.2.10 Genes with increased numbers of transposon insertions.

Many other genes showed alterations in the number of transposon insertions in the presence of metformin. First, genes with increased numbers of insertions were analysed, see Supplementary Table 5. Genes with a significant $\log_2$ fold change ≥ 1 in all metformin containing conditions were selected and the log fold change values displayed in a heatmap (Fig. 47A). NLE genes described within section 5.2.9 were excluded. 26 genes were elucidated in the analysis. Although not qualifying as NLE genes, some essential genes were found with increased insertions within the analysis. These genes included dxs, gidA, hemD, hscB, lgt, murG, orn, PA14_27200, PA14_56300, PA14_65690, PA14_66090, PA14_66120, PA14_68360, prtR, rpoA, serA, serB, trmE. The genes with the highest LogFC values were *dxs*, *lgt,* PA14_56300, PA14_66090*,* and *rpoA*, with scores ranging between 4.27 and 6.75. Interestingly, *dxs* is involved in the same pathway as *epd*, the NLE gene with the highest LogFC. Dxs is responsible for the formation of deoxyxylulose-5-phosphate from pyruvate and G3P during the PLP biosynthesis pathway (Denise et al., 2023). Lgt is involved in the formation of mature lipoproteins within the cell membrane and thus has roles in growth and virulence (Diao et al., 2021; Heywood and Lamont, 2020; Pailler et al., 2012). PA14_56300 is a hypothetical protein that is most likely a FumC paralog involved in the formation of malate from fumarate in the TCA cycle (Rossi et al., 2018). PA14_66090, is also hypothetical, however may be involved in the synthesis of the LPS inner core or toluene tolerance (King et al., 2009). Finally, RpoA is the alpha subunit of RNA polymerase (Cai et al., 2023). Of note, is the essentiality of all the

highest LogFC genes. Typically, one would not expect insertions in essential genes, therefore the locations within the gene of these insertions were visualised (Fig. 47B). For *dxs*, *rpoA*, *lgt* and PA14_66090, the insertions were mostly increased towards the ends of the genes, where disruption is unlikely to result in loss of function. However, for PA14_56300, *dxs*, *rpoA*, and *lgt* some insertions were also obtained within the centre of the gene. This was most abundant for PA14_56300. One set of non-essential genes within this analysis were the MexAB-OprM efflux pump encoding genes. Despite typically being implicated in providing antimicrobial resistance, in the presence of metformin the absence or loss of function of the efflux pump appears to improve fitness. On average across the three concentrations the MexAB-OprM genes had a LogFC of 1.47. This increase in transposon insertions can be clearly observed within Figure 47C across the entire length of each gene.

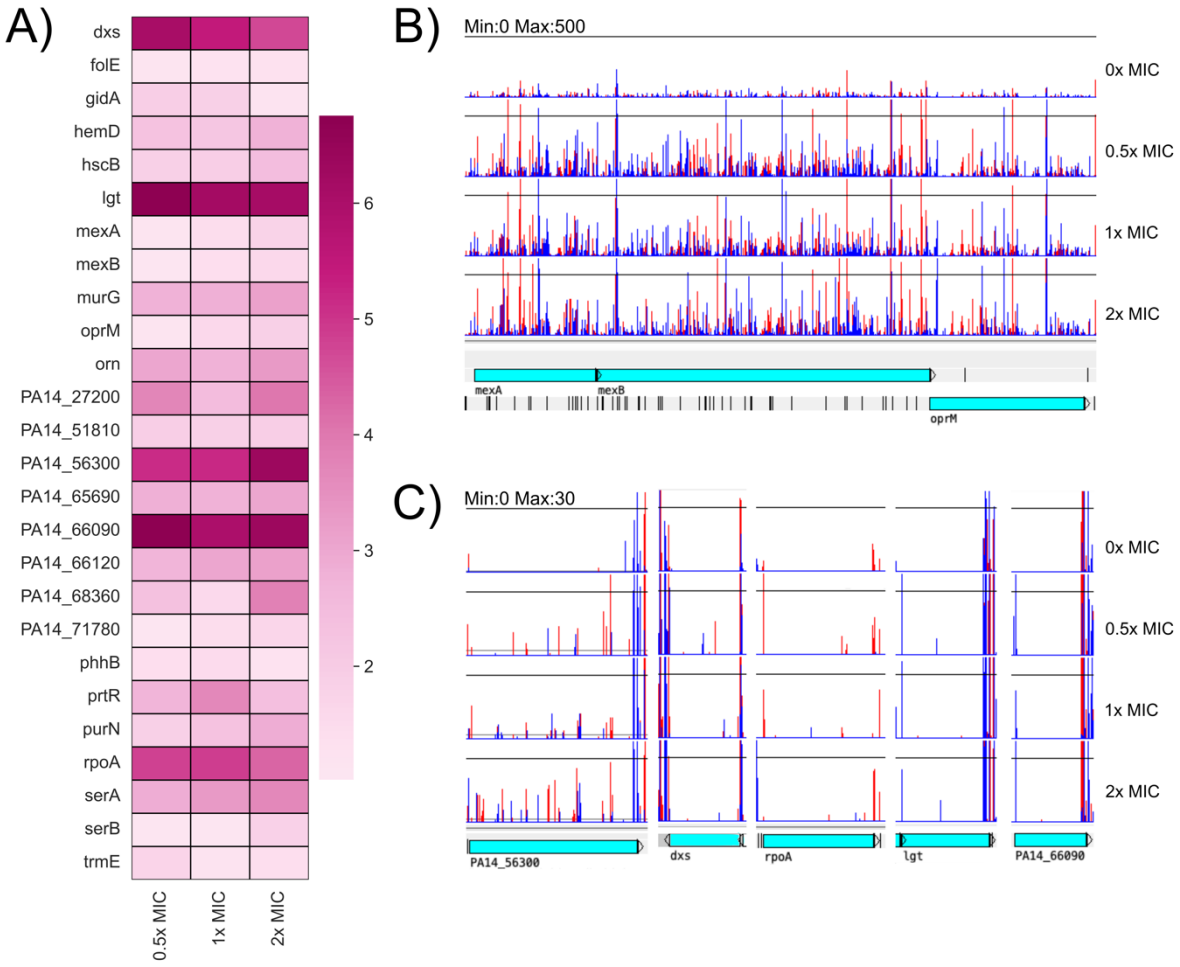


Figure 47: Multiple genes showed increased numbers of transposon insertions. A) Heatmap displaying the Log_2-fold change in transposon insertions for PA14 Δ*bioB-H* in presence of metformin. B-C) TraDIS reads representing transposon insertions within the non-essential genes *mexA*, *mexB*, *and oprM* (B) and the essential genes PA14_56300, *dxs*, *rpoA*, *lgt* and PA14_66090 (C). Colours represent direction of insertion with red representing forward and blue representing reverse. The height of the lines indicate the number of insertions within a given position. Max value indicates the maximum height of insertions contained within the frame. Max heights are 500 (B) and 30 (C).

5.2.11 Genes with decreased numbers of transposon insertions.

Next, genes with reduced numbers of transposon insertions were analysed. Genes with log_2 fold change of ≤ -1 in all metformin concentrations, which were not essential in the base condition were analysed. This included just seven genes, see Figure 48A. The gene with the highest average LogFC across the three conditions encoded for the heat shock response protein FtsH (Basta et al., 2020). Furthermore, a clear reduction in transposon density was observed across the whole *ftsH* gene for the metformin containing conditions (Fig. 48B). This suggests that FtsH function had increased importance in the presence of metformin. Another gene with increased importance was *nfxB,* this corresponds well with the increased insertions found within the MexAB-oprM encoding genes. This is due to *nfxB* function as a repressor of the MexCD-OprJ efflux pump (Purssell and Poole, 2013). Further to this, another gene involved in the repression of efflux pumps, *mvfR* showed reduced insertions, with the highest reduction seen within the 2x MIC metformin condition (Fig. 48A/B). MvfR is a repressor of quorum sensing and also directly represses another gene from this analysis *pqsD*. Finally, three genes associated with thiamine metabolism had decreased numbers of insertions, these were *thiS*, *thiD*, and *thiG* (Jurgenson et al., 2009).

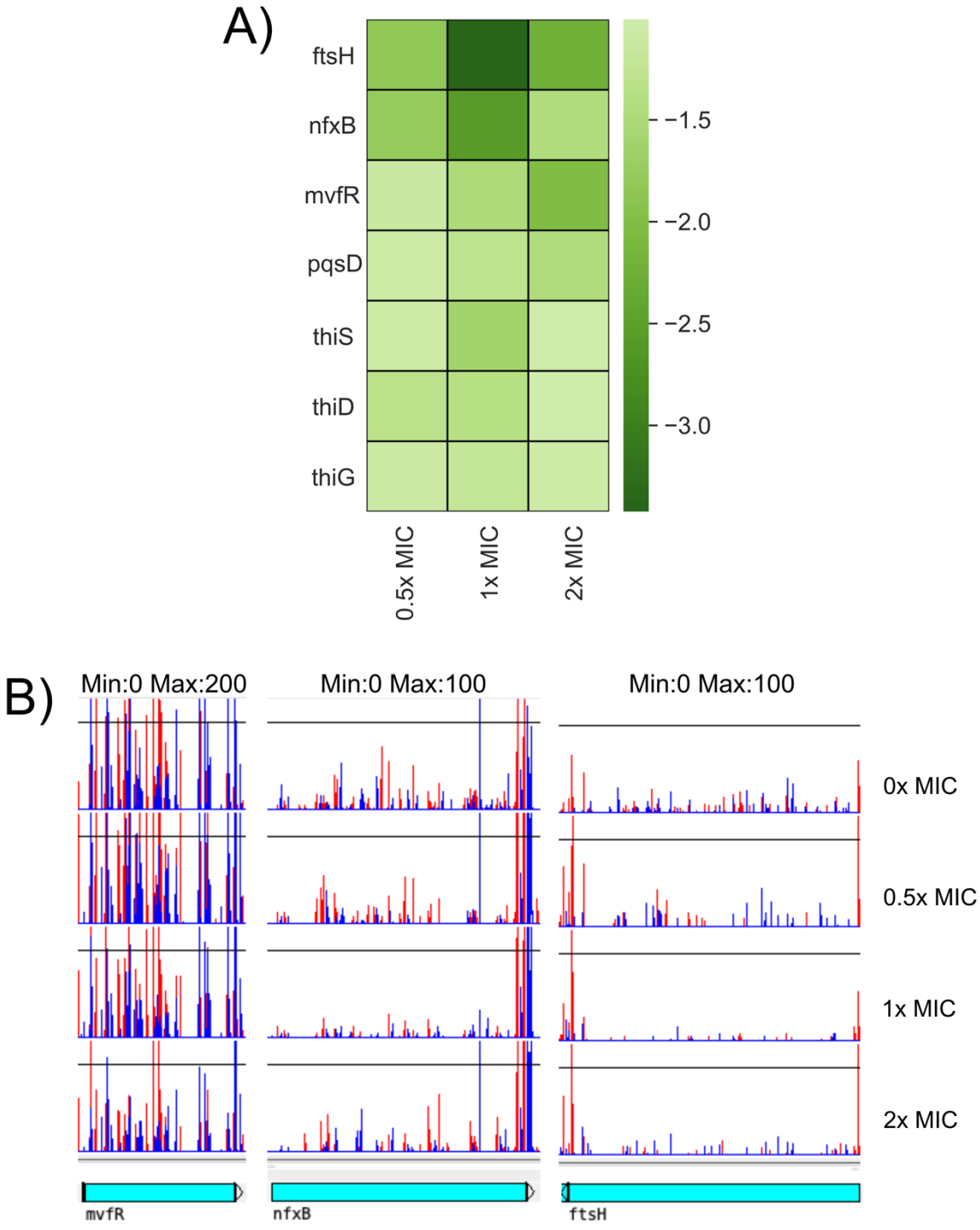


Figure 48: Multiple genes showed decreased numbers of transposon insertions across all metformin containing conditions. A) Heatmap displaying the Log_2-fold change in

transposon insertions for PA14 Δ*bioB-H* in presence of metformin. Only genes with LogFC values ≤ -1 in all concentrations are included. B) TraDIS reads representing transposon insertions within the genes *mvfR*, *nfxB*, and *ftsH*. Colours represent direction of insertion with red representing forward and blue representing reverse. The height of the lines indicate the number of insertions within a given position. Max value indicates the maximum height of insertions contained within the frame. Max insertion height is 200 for *mvfR* and 100 for *nfxB* and *ftsH*.

Since very few genes were reduced in all concentrations of metformin, genes which showed reductions in insertions in the two highest concentrations of metformin were also analysed. For this analysis, the majority of the genes (4/7) were involved in quorum sensing (see Fig. 49A). Of these genes, the highest LogFC value was seen for lasR, a receptor in HSL based quorum sensing which activates virulence genes (Fig. 49A/B). Further to this, the *pqs* genes constitute an alternative quorum sensing pathway. The *pqs* genes, *pqsB*/*C*/*D,* had an average LogFC of -1.14 and -1.36 for 1x MIC and 2x MIC, respectively. This is mirrored within Figure 49B, where both 1x MIC and 2x MIC show a reduction in transposon insertions, with the reduction being clearer for 2x MIC. Two of the genes in this analysis, PA14_51310 and PA14_60650 were hypothetical genes, with little information on their function. Finally, *pykA* showed reduced insertions, with LogFC values of -1.15 and -1.04, meaning the number of insertions was roughly halved for both 1x and 2x MIC (Fig. 49A/B).

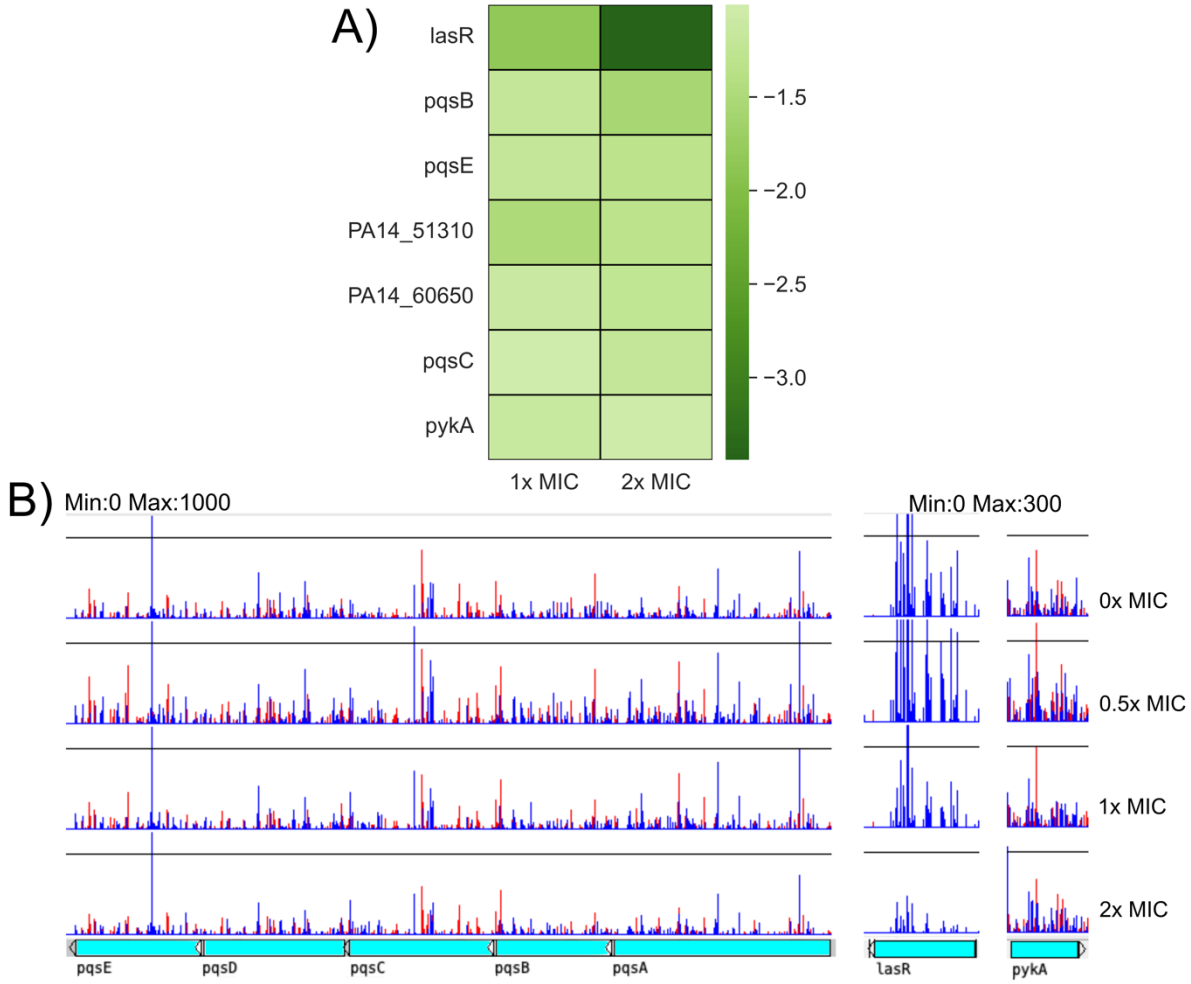


Figure 49: Seven genes showed decreased numbers of transposon insertions within the 1x and 2x MIC conditions. A) Heatmap displaying the Log_2-fold change in transposon insertions for PA14 Δ*bioB-H* in presence of metformin. Only genes with LogFC values ≤ -1 in 1x MIC and 2x MIC are included. B) TraDIS reads representing transposon insertions within the genes *pqsA/B/C/D/E*, *lasR*, and *pykA*. Colours represent direction of insertion with red representing forward and blue representing reverse. The height of the lines indicate the number of insertions within a given position. Max value indicates the maximum height of insertions contained within the frame. Max insertion height is 1000 for *pqs* genes and 300 for *lasR* and *pykA*.

5.3 Discussion

This study found that the biotin synthesis pathway is integral to metformin resistance in *P. aeruginosa*. This was confirmed via the production of clean knockouts of the biotin synthesis pathway and the complementation of resistance with exogenous biotin supply. Although the presence of biotin recovered the wildtype phenotype, the quantity of biotin added (400 nM) was equivalent to around 97 ng/mL and the human levels of biotin are between 0.1-1.2 ng/mL (Carfrae et al., 2020). The constitutive levels of biotin in humans would therefore not recover this phenotype and make therapeutics unfeasible. Not only is this a significant discovery from the PA14 chemical genomic screen in the context of drug development, but it also validates the ability of the screen to accurately reveal biologically relevant phenotypes.

One interesting aspect of this phenotype is that it is not retained amongst Enterobacteriaceae species. One reason proposed for this is the rearrangement of the biotin synthesis operon to encode *bioH* within the operon. As previously described BioH performs a critical step in biotin synthesis, taking intermediates from the fatty acid biosynthesis pathway and syphoning them into the biotin synthesis pathway. A previous study by Cao et al. showed that when *bioH* was operon encoded expression was increased compared to free-standing *bioH* (Cao et al., 2017). The reason for this may be to increase biotin production in order to fulfil the increased requirement for biotin within *P. aeruginosa* (Carfrae et al., 2020). It is probable that this increased requirement is due to the multiple biotin-dependent carboxylases encoded by *P. aeruginosa* (Tong, 2013). This would make sense as with more enzymes dependent on biotin, it needs to be more readily available. This hypothesis is supported by a recent

study which used a streptavidin blot assay to isolate biotinylated proteins in *E. coli* and *P. aeruginosa*. Just one band was achieved for *E. coli*, relating to its singular biotin-dependent carboxylase AccB (Shi et al., 2023; Tong, 2013). However multiple bands were revealed for *P. aeruginosa,* three of which related to known carboxylases pyruvate carboxylase, methyl crotonyl-CoA carboxylase and geranyl-CoA carboxylase (Shi et al., 2023). Shi et al. suggested that the increased number of biotin-dependent carboxylases may aid in *P. aeruginosa* adaptability to multiple harsh environments.

Contrary to this hypothesis was the lack of phenotype displayed in *Acinetobacter baumannii* within this study. Like *P. aeruginosa,* Carfrae showed that *Acinetobacter baumannii* had an increased need for biotin. Furthermore, according to Cao et al. *Acinetobacter* encoded *bioH* within the biotin synthesis operon. This led to the hypothesis that biotin may mediate resistance to metformin in *A. baumannii*, but this was not the case. Despite the claims by Cao et al., when searching BioCyc for the operon structure of *A. baumannii* AB5075, the strain used within this study, the operon showed a *bioAFCD* structure, without *bioB* or *bioH* within the operon. This structure is more similar to that of *Enterobacteriaceae* and therefore may explain the difference in phenotypic response.

Since *P. aeruginosa* encodes multiple biotin-dependent carboxylases, this study aimed to evaluate if any single carboxylase was responsible for the importance of biotin in the presence of metformin. However, only one biotin-dependent carboxylase, GCC, was found to have a change in phenotype in the presence of metformin, and this was likely a false positive. This is due to the phenotype not being retained in other metformin concentrations or in an alternative transposon mutant within the gene. Further to this,

supplementation with acetoacetate did not increase the MIC of PA14 Δ*bioB-H* in metformin. Acetoacetate is the end product of the terpene degradation pathway in which GCC is involved. Oxaloacetate, however, did increase resistance to metformin. Oxaloacetate is the product of biotin-dependent carboxylase pyruvate carboxylase. PC has an important anapleurotic role since many TCA intermediates become depleted in pathways like gluconeogenesis and fatty acid biosynthesis (Lai et al., 2006; Tong, 2013). Therefore, it may be that PC is particularly important for resistance against metformin. In humans, metformin is proposed to target mitochondrial respiratory complex I which plays a prominent role in oxidative phosporylation for the generation of ATP. If the homologue of this complex in *P. aeruginosa* is also targeted by metformin, then generation of ATP may need to be formed in an alternative pathway. The TCA cycle can produce one molecule of ATP per cycle, with the conversion of succinyl-CoA to succinate (Kim and Gadd, 2019). As such, supplying oxaloacetate which combines with acetyl-CoA at the beginning of the TCA cycle, may allow for the replenishment of the TCA cycle for ATP production. Despite oxaloacetate being the only supplement to increase resistance to metformin in PA14 Δ*bioB-H*, it is likely that all of the biotin-dependent carboxylases play a role. It may be that the quantities of supplement supplied were insufficient to restore the phenotypes, and therefore should be repeated with varying concentrations of supplement and perhaps in combination with each other.

Understanding how biotin mediates resistance to metformin would be easier if the target of metformin was elucidated in *P. aeruginosa*. However, this is currently unknown, and even the target in humans in still under heavy debate. Though some recent evidence has shown that biguanides bind to the interface between the ND1 and

NDUFS7 domains of mitochondrial respiratory complex I (Bridges et al., 2023). These are equivalent to the domains encoded by *nuoH* and *nuoB* within *P. aeruginosa*, respectively (Rhein et al., 2016; Sinha et al., 2009). A commonly utilised method for finding the target of drugs, is the development of suppressor mutants, since mutations often occur within the genes which encode the targets (Abrahams et al., 2012; Soto et al., 2018). This was attempted in the current study to elucidate the target of metformin, via exposing PA14 Δ*bioB-H* to increasing concentrations of metformin. However, no spontaneous resistant isolates were gained. Although this was not beneficial to the elucidation of the target of metformin, it does increase the appeal of repurposing metformin in combination with a biotin synthesis inhibitor. Given the current antimicrobial resistance crisis, it is vital that antimicrobial resistance to new antimicrobial therapies is not easily obtained. Given the difficulty of gaining resistant colonies within a biotin synthesis deficient PA14 *in vitro*, it suggests that spontaneous resistance in the clinic would also be difficult. However, this may be due to loss of the entire biotin synthesis operon within the PA14 Δ*bioB-H* strain, and perhaps resistance to a biotin synthesis inhibitor would be more easily obtained.

Due to the failure of the suppressor mutant screen, a TraDIS library was produced with the intention to induce synthetic suppressor mutations. The libraries were succesfully produced, with dense insertions roughly every 13 base pairs. This is comparable to other TraDIS libraries that have previously been produced, including a *Salmonella* Typhi library which also had insertions every 13 bp (Langridge et al., 2009). The libraries were grown in the presence of metformin to discover gene disruptions which gave a fitness advantage. Genes with increased insertions suggest that loss of the gene is advantageous. This is because the number of insertions is linked to how well

a mutant can grow in the given condition. The better the mutant grows the higher the number of clones within the culture. Therefore, when the transposon junctions are amplified there will be a higher abundance of junctions isolated from these clones, leading to increased insertion numbers within the given gene. Since genes with increased numbers of insertions would be most akin to suppressor phenotypes, genes which became non-essential or had increased numbers of transposon insertions were analysed first.

One of the main pathways that became no longer essential in the presence of metformin were the purine and pyrimidine metabolism pathways. This may be due to these pathways being energetically expensive, requiring multiple molecules of ATP for de novo synthesis (Moffatt and Ashihara, 2002). Both of the these pathways also have salvage pathways which consume less ATP during energy limitation. Notably, one of the genes involved in pyrimidine salavage, *thiD*, had reduced numbers of transposon insertions, suggesting increased importance in the presence of metformin. This is further supported by the disruption of other energy consuming pathways also showing loss of essentiality in the presence of metformin. For example multiple genes from the arginine biosytheisis pathway were NLE genes, including *argA*, *argB*, *argC*, *argF*, *argG*, and *argH*. Further to this, the enzymes CarA and CarB which are responsible for producing precursers to both arginine and pyrimidine biosynthesis, were NLE genes (Barrientos-Moreno et al., 2019; Butcher et al., 2016; Zhuo et al., 2015).

The reduced requirement for purine synthesis may also be attributed to induction of the stringent response. As described in 4.1.2.2 the stringent response results in the down-regulation of purine biosynthesis. With purine biosynthesis reduction, the

intermediate PRPP can be used within the histidine and tryptophan biosynthesis pathways. Intriguingly, multiple histidine biosynthesis genes showed decreased numbers of transposon insertions within the 0.5x MIC metformin condition compared to no metformin, see Supplementary Table 6. Since decreased insertions were only observed in a single condition, these genes were not displayed within the analysis in 5.2.10 or 5.2.11. Some further evidence towards this hypothesis is the conditional essentiality of SpoT, an initiator of the stringent response, in PA14 ∆*bioB-H* when grown in minimal media. In opposition to the stringent response hypothesis is the loss of essentiality of *dksA*. DksA is heavily involved within the stringent response and is repsonsible for regulating multiple pathways including amino acid biosynthesis, secondary metabolite synthesis, the TCA cycle, glycogen biosythesis and pyrimidine biosynthesis (Cohen et al., 2022). However, in *P. aeruginosa* there are two redundant copies of *dksA*, which are largely interchangable. Therefore, it is possible that although this one copy of *dksA* becomes non-essential the other remains important, and is able to stimulate aspects of the stringent response. Since the stringent response often allows for survival in nutrient limiting conditions, ensuring its activation may be important for PA14 ∆*bioB-H* survival in metformin. This suggests that the combination of biotin synthesis loss and metformin is causing a severe disruption in energy production and nutrient availablity. Potentially indicating one hypothesis for the mechanism of metformin killing with the loss of biotin synthesis. This hypothesis of energy limitation is supported by the decreased insertions found within the heat-shock protease encoding gene *ftsH*. (Basta et al., 2020). FtsH has been implicated in providing a fitness advantage during energy-limited growth arrest. In *P. aeruginosa*,

this is facilitated by ensuring protein quality control is maintained during energy limitation (Basta et al., 2020).

A further result from the TraDIS screen, which suggested that metformin causes energy limitation, was the acquisition of increased insertions in the MexAB-OprM efflux pump genes. Although efflux pumps are often overexpressed in response to antibiotics, they are energetically expensive to operate. Therefore, if metformin limits energy generation, it would make sense that losing this energetically costly complex would provide a fitness advantage. This hypothesis is supported by a recent paper which investigated the combination of metformin with tetracycline (Liu et al., 2020). This paper showed that metformin exhibited synergy with tetracycline by disrupting the outer membrane and the proton motive force (PMF) in *E. coli* (Liu et al., 2020). A decrease in intracellular ATP production was also described, likely due to the disruption of the PMF which is the major driver of ATP synthesis (Liu et al., 2020). Given the disruption of the PMF, which RND efflux pumps rely on for function, wasting resources on efflux pump expression could be potentially disadvantageous.

A further paper, which investigated the fitness defects associated with MexEF-OprN expression, found that overexpression could result in increased oxygen consumption and mild disruption of ATP synthesis (Olivares et al., 2014). It was hypothesised that, since overexpression would result in increased proton movement down the protein gradient, this would perturb the PMF and thus disrupt ATP synthase activity, which also relies on the PMF. As a result oxygen would be increased to compensate. They also showed that although in a favourable environment this compensation did not result in a fitness cost, in unfavourable conditions overexpression led to a fitness

disadvantage (Olivares et al., 2014). In relation to this, it is notable that the *rpoA* gene was included in the top five genes with increased insertions in all concentrations of metformin. A recent study showed that RpoA played a role in the expression of the MexEF-OprN efflux pump. It was discovered that a point mutation (T262A) in RpoA resulted in decreased contact between RNA polymerase and the promoter of the *mexEF-oprN* operon, leading to reduced expression (Cai et al., 2023). Therefore, the mutants with insertions in *rpoA* may have had a fitness advantage due to decreased expression of MexEF-OprN. However, since RpoA is a major component of RNA polymerase and is essential for growth, it is unusual that transposon insertions were gained within this gene. Although many of the transposon insertions were acquired towards the end of the gene, where disruption is unlikely, insertions were also acquired towards the centre of the gene. Since the presence of metformin appears to affect energy generation, and results in a slow growth phenotype, this may have allowed for the acquisition of these transposon insertions. This is likely due to reduced RNA polymerase activity during slow growth or growth arrest, leading to increased tolerance to mutation (Bergkessel, 2021). Therefore, a more likely hypothesis is that transposon insertion acquisition in *rpoA* is not related to MexEF-OprN expression, but is instead a side effect of energy limitation caused by metformin.

A further efflux pump associated gene found within the TraDIS screen was *nfxB*. Decreased insertions were found in *nfxB*, suggesting its function was important in the presence of metformin. NfxB is a repressor of the MexCD-OprJ efflux pump, and loss of function mutations have been shown to increase MexCD-OprJ expression (Purssell and Poole, 2013; Stickland et al., 2010). Stickland et al. hypothesised that the fitness defects associated with inappropriate efflux pump overexpression could be due to the

high level of oxidative phosphorylation that would be required to generate the PMF needed for pump function. This increased level of oxidative phosphorylation could then result in the formation of free-radicals such as super-oxide, which lead to protein damage (Stickland et al., 2010). Given that metformin has been shown to disrupt the PMF, expression of efflux pumps could put extreme strain on the oxidative phosphorylation pathway in order to compensate, which would potentially result in a severe fitness cost. This could be amplified by metformin potentially targeting oxidative phosphorylation. However, this depends on if metformin binds to the bacterial homologue of its proposed within-human target.

Additional to the evidence presented by Liu et al. that metformin disrupts the proton motive force, the current study showed that loss of the NAD biosynthesis genes *nadABC* resulted in a slight fitness advantage. If metformin is targeting the *nuo* encoded NADH dehydrogenase NDH-1, there could be a disruption of the NAD+/NADH balance within the cell. This is because NDH-1 converts NADH to NAD+ in order to shuttle electrons into the oxidative phosphorylation pathway for energy production (Hreha et al., 2021). With loss of activity this could lead to increased levels of NADH within the cell, which can be toxic in excess. This NADH induced toxicity is the result of increased membrane potential and resultant ROS species production (Arce-Rodríguez et al., 2022). Therefore loss of NAD biosynthesis could be beneficial to reduce production of excess NADH.

A further reason for the reduced importance of efflux pumps could be that efflux pump overexpression causes impairment of quorum sensing in *P. aeruginosa*. Alcalde-Rico et al. showed that MexAB-OprM overexpressing strains display a severe deficit in *pqs*

gene expression (Alcalde-Rico et al., 2020). Multiple genes within the *pqs* operon were shown to have increased importance in the presence of metformin within the TraDIS screen. Some studies have shown that the pqs response is important in the presence of antibiotics or stress, since it can be used as a warning signal amongst bacterial communities (Bru et al., 2019; Häussler and Becker, 2008). Considering metformin is known to be a quorum sensing inhibitor, this may also explain the increased importance of the *pqs* genes within the TraDIS screen (Hegazy et al., 2020).

One set of genes, which had increased insertions, appear to be related to compensation for the loss of biotin synthesis. *Epd*, *dxs*, *pdxA, pdxH* and *pdxJ* all play a role in the formation of vitamin B6 vitamers including PLP (Denise et al., 2023). BioA and BioB are PLP dependent proteins, and rely on the co-factor for function (Mann and Ploux, 2011). Therefore, it is likely that loss of these proteins is beneficial with the deletion of the biotin synthesis genes within PA14 ∆*bioB-H.* These increased insertions may reduce excess PLP. Further to this, three genes associated with thiamine biosynthesis had decreased insertions. During the thiamine biosynthesis pathway PLP is utilised (Jurgenson et al., 2009). Therefore, it is possible that these genes are also important with the loss of biotin synthesis for reducing excess PLP. Excess PLP is known to cause toxcity in both eukaryotes and prokaryotes. In mammals, toxicity is induced via the inhibition of DNA metabolism and central carbon metabolism (Richts et al., 2019). Further to this, excess PLP has been shown to inhibit acetyl-CoA carboxylase, which has a vital role in fatty acid biosynthesis (Lee et al., 2005). Another study, focusing on the toxicity of excess PLP in *E. coli,* showed that amino acid pools are affected, resulting in growth defects (Sugimoto et al., 2017). Further to this, excess PLP has been shown to inactivate the *E. coli* initiation factor 3 (IF3). IF3 ensures

precise and appropriate initiation of translation and is vital for cell growth, therefore excess PLP can be highly damaging (J et al., 2022; Li et al., 2020; Richts et al., 2019). This toxicity based disruption of multiple growth and metabolism related pathways, may explain the increased insertions within PLP synthesis genes in the presence of metformin. This would be especially true if metformin is also disrupting energy metabolism and thus the ability for cells to grow.

In conclusion, the combination of these effects suggests that metformin disrupts energy generation, meaning *P. aeruginosa* must adapt to alternative forms of energy generation. In a biotin producing cell, this disruption of energetics may not be sufficient to induce cell death. However, since biotin-dependent carboxylases are involved in many different aspects of cell viability and metabolism, loss of biotin may result in *P. aeruginosa* no longer being able to compensate. For example, if metformin disrupts oxidative phosphorylation and the TCA cycle is required for energy generation, pyruvate carboxylase plays a critical anaplerotic role. Some evidence for this was displayed in the MIC of PA14 Δ*bioB-H* to metformin being increased with the addition of oxaloacetate. Further to this, MCC and GCC are involved in pathways that produce acetyl-CoA which can be directly used within the TCA cycle in combination with oxaloacetate. Despite this, although multiple genes showed changes in transposon insertions with the addition of metformin, it is still not entirely clear what the mechanism of action of metformin truly is. Further to this, the mechanism of how biotin mediates this resistance has not been elucidated. Therefore further research would be required within this field before a co-therapy could be produced.

CHAPTER 6

FINAL DISCUSSION

6.1 General summary

The aim of this thesis was to understand the lifestyle of *P. aeruginosa* via chemical genomic screening methods. Despite the availability of the screening data for PAO1 and PA14 at the beginning of this research project, the current methods for analysis and quality control were limited. This meant there was a requirement to develop computational scripts to perform this analysis, since early software had become deprecated. The current landscape of the literature showed that multiple groups were producing their own scripts for analysis, leading to inconsistencies in analysis. Therefore, it was decided that a major aim of this project would be to produce and release an easy-to-use software for chemical genomic screen data analysis. The goal being to not only make chemical genomics more accessible, but to make the analysis more consistent across the field. As a result of this, ChemGAPP was produced, with three sub-packages for the analysis of various screening types. As described in Chapter 3, each sub-package was tested against qualified real-world data and accurately revealed biologically relevant phenotypes. This provided evidence for its effectiveness and justification for its release for use in the wider scientific field.

Following the production of ChemGAPP, the aim of Chapter 4 was to analyse both the PAO1 and PA14 phenotypes and elucidate the differences in conditional responses. The chemical genomic screen within this study was unique within the field as it was the first large-scale chemical genomic screen of *P. aeruginosa.* In addition to this, it was also the first to screen two strains of a pathogenic bacteria in parallel. The results of the comparative analysis showed that although genetically PAO1 and PA14 are very similar, that their phenotypic responses to conditional stresses differ extensively. This

has important implications when making assumptions about one strain based on the other, as conclusions may not hold true. Further to this, it suggests that perhaps clinically relevant strains of *P. aeruginosa* may also differ significantly from each other phenotypically. Considering many of the anticorrelated genes were found in conjunction with antibiotic treatment, it is vital to consider that various strains may respond very differently to therapies. Equally, with the development of new antimicrobials for *P. aeruginosa* this must also be considered.

The outcome of the screen also raised questions considering the suitability of certain organisms for screening and the phenotypes measured. For both PAO1 and PA14 swarming motility was an issue and may have resulted in artificially significant phenotypes. As such, in future screens comparing multiple phenotypes could provide a better understanding of the lifestyle of both organisms individually, and how they compare. Further to this, if using colony size as the phenotype measure, it is important to consider the motility of an organism and perhaps avoid conditions known to induce this response. Despite this, both screens had high quality scores following benchmarking and successfully produced ~20,000 phenotypes for PAO1 and ~10,000 phenotypes for PA14. Two of these phenotypes were more closely analysed within the study. First, the study revealed that the recently discovered small alarmone hydrolase *Pa*SAH was important for sulfonamide resistance within PA14 but not PAO1. The proposed hypothesis for the killing in the absence of *Pa*SAH, was the induction of GTP and/or purine limitation due to inappropriate activation of the stringent response.

Secondly, biotin was shown to mediate metformin resistance in both PAO1 and PA14. Due to the clinical applications of the second phenotype, Chapter 5 was dedicated to

understanding both the mechanism of action of metformin and the mechanism of resistance of biotin. The major hypothesis put forward by the study was that metformin is targeting oxidative phosphorylation and disrupting the proton motive force. As a result of this, energy limitation is induced, and alternative forms of energy generation are required. Biotin and its dependent carboxylases are involved in multiple aspects of metabolism, with many being directly or indirectly involved in the production of intermediates within the TCA cycle. Therefore, in the presence of biotin, this study proposes that *P. aeruginosa* is able to compensate for this energy disruption. However, with loss of biotin synthesis, *P. aeruginosa* cannot shift to alternative energy production leading to cell death. Despite these hypotheses being based on the work produced within this study, further work would be required to validate them.

6.2 Future work

The chemical genomic screen revealed a promising potential co-therapy for *P. aeruginosa* via combining metformin with a biotin synthesis inhibitor. However, for viability within the clinic, it is important to understand how biotin mediates resistance and how metformin inhibits growth. The current study used TraDIS as a means to discover these mechanisms. However, there is a limitation with this method. TraDIS typically induces gene disruption, which generally produces loss-of function mutants. When attempting to artificially induce suppressors, often over-expression of genes is more informative than loss of function. This is because by introducing random gene over-expression, the target of metformin could be increased leading to a dosage effect

which rescues the phenotype. Therefore, in future work it would be beneficial to produce an overexpression library within PA14 Δ*bioB-H*. Other studies have used overexpression libraries to understand the mechanism of drugs, using techniques such as TraDIS-Xpress and Dub-seq (Mutalik et al., 2019; Yasir et al., 2020). Finally, if a co-therapy is to become viable, first a biotin synthesis inhibitor must be elucidated for *P. aeruginosa* that has synergy with metformin. A BioA inhibitor, MAC13772 has been previously described against *E. coli* BioA (Carfrae et al., 2020). However, it does not have activity against *P. aeruginosa* BioA, which shares only 41.2% protein sequence identity (Carfrae et al., 2020). Therefore, future work would involve the screening of BioA inhibitors for *P. aeruginosa* BioA, before performing checkerboard assays for synergy with metformin.

6.3 Final remarks

This study successfully produced a comprehensive chemical genomic screen data analysis software, ChemGAPP, with broad application amongst the scientific community. ChemGAPP effectively produced reliable fitness values which revealed multiple biologically relevant phenotypes. Two of which had significant clinical relevance showing resistance methods to current antibiotics and potential new antibiotic treatments. With increasing antimicrobial resistance threatening the treatment of *P. aeruginosa* infections it is vital that mechanisms behind these resistances are understood and that new therapies can be developed. This study introduced a potential combination therapy with great promise, due to metformin's

clinically proven safety and biotin synthesis absence within humans. As such, this doctoral research project reveals the power of chemical genomic screening techniques for providing an in depth understanding of bacterial phenotypes on a genome wide level.

REFERENCES

Abrahams, K.A., Cox, J.A.G., Spivey, V.L., et al. (2012) Identification of Novel Imidazo[1,2-a]pyridine Inhibitors Targeting M. tuberculosis QcrB. *PLOS ONE*, 7 (12): e52951. doi:10.1371/journal.pone.0052951.

Aeschlimann, J.R. (2003) The Role of Multidrug Efflux Pumps in the Antibiotic Resistance of Pseudomonas aeruginosa and Other Gram-Negative Bacteria. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 23 (7): 916–924. doi:10.1592/phco.23.7.916.32722.

Aguilar, J.A., Díaz-Pérez, C., Díaz-Pérez, A.L., et al. (2008) Substrate Specificity of the 3-Methylcrotonyl Coenzyme A (CoA) and Geranyl-CoA Carboxylases from Pseudomonas aeruginosa. *Journal of Bacteriology*, 190 (14): 4888–4893. doi:10.1128/JB.00454-08.

Alcalde-Rico, M., Olivares-Pacheco, J., Halliday, N., et al. (2020) The impaired quorum sensing response of Pseudomonas aeruginosa MexAB-OprM efflux pump overexpressing mutants is not due to non-physiological efflux of 3-oxo-C12-HSL. *Environmental Microbiology*, 22 (12): 5167–5188. doi:10.1111/1462-2920.15177.

Aldred, K.J., Kerns, R.J. and Osheroff, N. (2014) Mechanism of Quinolone Action and Resistance. *Biochemistry*, 53 (10): 1565–1574. doi:10.1021/bi5000564.

Alvarez-Ortega, C., Wiegand, I., Olivares, J., et al. (2010) Genetic Determinants Involved in the Susceptibility of Pseudomonas aeruginosa to β-Lactam Antibiotics. *Antimicrobial Agents and Chemotherapy*, 54 (10): 4159–4167. doi:10.1128/AAC.00257-10.

Arabameri, N., Heshmatipour, Z., Eftekhar Ardebili, S., et al. (2021) The role of Gene Mutations (gyrA, parC) in Resistance to Ciprofloxacin in Clinical Isolates of Pseudomonas Aeruginosa. *Iranian Journal of Pathology*, 16 (4): 426–432. doi:10.30699/IJP.2021.520570.2542.

Arce-Rodríguez, A., Pankratz, D., Preusse, M., et al. (2022) Dual Effect: High NADH Levels Contribute to Efflux-Mediated Antibiotic Resistance but Drive Lethality Mediated by Reactive Oxygen Species. *mBio*, 13 (1): e02434-21. doi:10.1128/mbio.02434-21.

Asokan, G.V., Ramadhan, T., Ahmed, E., et al. (2019) WHO Global Priority Pathogens List: A Bibliometric Analysis of Medline-PubMed for Knowledge Mobilization to Infection Prevention and Control Practices in Bahrain. *Oman Medical Journal*, 34 (3): 184–193. doi:10.5001/omj.2019.37.

Ayers, M., Sampaleanu, L.M., Tammam, S., et al. (2009) PilM/N/O/P Proteins Form an Inner Membrane Complex That Affects the Stability of the Pseudomonas aeruginosa Type IV Pilus Secretin. *Journal of Molecular Biology*, 394 (1): 128–142. doi:10.1016/j.jmb.2009.09.034.

Baba, T., Ara, T., Hasegawa, M., et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2 (1): 2006.0008. doi:10.1038/msb4100050.

Bange, G., Brodersen, D.E., Liuzzi, A., et al. (2021) Two P or Not Two P: Understanding Regulation by the Bacterial Second Messengers (p)ppGpp. *Annual Review of Microbiology*, 75 (1): 383–406. doi:10.1146/annurev-micro-042621-122343.

Banzhaf, M., Yau, H.C., Verheul, J., et al. (2020) Outer membrane lipoprotein NlpI scaffolds peptidoglycan hydrolases within multi-enzyme complexes in Escherichia coli. *The EMBO Journal*, 39 (5): e102246. doi:10.15252/embj.2019102246.

Barquist, L., Mayho, M., Cummins, C., et al. (2016) The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics (Oxford, England)*, 32 (7): 1109–1111. doi:10.1093/bioinformatics/btw022.

Barrientos-Moreno, L., Molina-Henares, M.A., Pastor-García, M., et al. (2019) Arginine Biosynthesis Modulates Pyoverdine Production and Release in Pseudomonas putida as Part of the Mechanism of Adaptation to Oxidative Stress. *Journal of Bacteriology*, 201 (22): 10.1128/jb.00454-19. doi:10.1128/jb.00454-19.

Baryshnikova, A., Costanzo, M., Kim, Y., et al. (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods*, 7 (12): 1017–1024. doi:10.1038/nmeth.1534.

Basta, D.W., Angeles-Albores, D., Spero, M.A., et al. (2020) Heat-shock proteases promote survival of Pseudomonas aeruginosa during growth arrest. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (8): 4358–4367. doi:10.1073/pnas.1912082117.

Berg, J.M., Tymoczko, J.L. and Stryer, L. (2002) Summary. *Biochemistry. 5th edition*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK22529/ (Accessed: 27 March 2020).

Bergkessel, M. (2021) Bacterial transcription during growth arrest. *Transcription*, 12 (4): 232–249. doi:10.1080/21541264.2021.1968761.

Beyer, P. and Paulin, S. (2020) The Antibacterial Research and Development Pipeline Needs Urgent Solutions. *ACS Infectious Diseases*, 6 (6): 1289–1291. doi:10.1021/acsinfecdis.0c00044.

Bhagirath, A.Y., Li, Y., Somayajula, D., et al. (2016) Cystic fibrosis lung environment and Pseudomonas aeruginosa infection. *BMC Pulmonary Medicine*, 16. doi:10.1186/s12890-016-0339-5.

Blanco, P., Hernando-Amado, S., Reales-Calderon, J.A., et al. (2016) Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance Determinants. *Microorganisms*, 4 (1): 14. doi:10.3390/microorganisms4010014.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30 (15): 2114–2120. doi:10.1093/bioinformatics/btu170.

Botelho, J., Grosso, F. and Peixe, L. (2019) Antibiotic resistance in Pseudomonas aeruginosa – Mechanisms, epidemiology and evolution. *Drug Resistance Updates*, 44: 100640. doi:10.1016/j.drup.2019.07.002.

Bridges, H.R., Blaza, J.N., Yin, Z., et al. (2023) Structural basis of mammalian respiratory complex I inhibition by medicinal biguanides. *Science*, 379 (6630): 351–357. doi:10.1126/science.ade3332.

Bru, J.-L., Rawson, B., Trinh, C., et al. (2019) PQS Produced by the Pseudomonas aeruginosa Stress Response Repels Swarms Away from Bacteriophage and Antibiotics. *Journal of Bacteriology*, 201 (23): e00383-19. doi:10.1128/JB.00383-19.

Bruchmann, S., Dötsch, A., Nouri, B., et al. (2013) Quantitative Contributions of Target Alteration and Decreased Drug Accumulation to Pseudomonas aeruginosa Fluoroquinolone Resistance. *Antimicrobial Agents and Chemotherapy*, 57 (3): 1361–1368. doi:10.1128/aac.01581-12.

Brugha, R.E. and Davies, J.C. (2011) Pseudomonas aeruginosa in cystic fibrosis: Pathogenesis and new treatments. *British Journal of Hospital Medicine*, 72 (11): 614–619. doi:10.12968/hmed.2011.72.11.614.

Bryant, J.A., Staunton, K.A., Doherty, H.M., et al. (2023) *Bam complex associated proteins in Escherichia coli are functionally linked to peptidoglycan biosynthesis, membrane fluidity and DNA replication*. p. 2023.07.05.547807. doi:10.1101/2023.07.05.547807.

Bullard, K.M. and Dunn, D.L. (2001) *Bloodstream and intravascular catheter infections*. Zuckschwerdt. Available at: https://www.ncbi.nlm.nih.gov/books/NBK7008/ (Downloaded: 6 April 2020).

Butcher, B.G., Chakravarthy, S., D'Amico, K., et al. (2016) Disruption of the carA gene in Pseudomonas syringae results in reduced fitness and alters motility. *BMC Microbiology*, 16 (1): 194. doi:10.1186/s12866-016-0819-z.

Cai, W., Liao, H., Lu, M., et al. (2023) *Discovery of RNA polymerase α-subunit protein as a novel quorum sensing reprograming factor in Pseudomonas aeruginosa*. p. 2022.08.01.502315. doi:10.1101/2022.08.01.502315.

Cao, X., Zhu, L., Hu, Z., et al. (2017) Expression and Activity of the BioH Esterase of Biotin Synthesis is Independent of Genome Context. *Scientific Reports*, 7: 2141. doi:10.1038/s41598-017-01490-0.

Carfrae, L.A., MacNair, C.R., Brown, C.M., et al. (2020) Mimicking the human environment in mice reveals that inhibiting biotin biosynthesis is effective against

antibiotic-resistant pathogens. *Nature Microbiology*, 5 (1): 93–101. doi:10.1038/s41564-019-0595-2.

Carter, M.Q., Chen, J. and Lory, S. (2010) The Pseudomonas aeruginosa Pathogenicity Island PAPI-1 Is Transferred via a Novel Type IV Pilus. *Journal of Bacteriology*, 192 (13): 3249–3258. doi:10.1128/JB.00041-10.

Chowdhury, N., Suhani, S., Purkaystha, A., et al. (2019) Identification of AcrAB-TolC Efflux Pump Genes and Detection of Mutation in Efflux Repressor AcrR from Omeprazole Responsive Multidrug-Resistant Escherichia coli Isolates Causing Urinary Tract Infections. *Microbiology Insights*, 12: 1178636119889629. doi:10.1177/1178636119889629.

Cohen, H., Adani, B., Cohen, E., et al. (2022) The ancestral stringent response potentiator, DksA has been adapted throughout Salmonella evolution to orchestrate the expression of metabolic, motility, and virulence pathways. *Gut Microbes*, 14 (1): 1997294. doi:10.1080/19490976.2021.1997294.

Cole, S.J., Records, A.R., Orr, M.W., et al. (2014) Catheter-Associated Urinary Tract Infection by Pseudomonas aeruginosa Is Mediated by Exopolysaccharide-Independent Biofilms. *Infection and Immunity*, 82 (5): 2048–2058. doi:10.1128/IAI.01652-14.

Collins, S.R. (2006) *EMAP toolbox for MATLAB*. Available at: https://sourceforge.net/projects/emap-toolbox/ (Accessed: 14 October 2021).

Collins, S.R., Schuldiner, M., Krogan, N.J., et al. (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biology*, 7 (7): R63. doi:10.1186/gb-2006-7-7-r63.

Colvin, K.M., Irie, Y., Tart, C.S., et al. (2012) The Pel and Psl polysaccharides provide Pseudomonas aeruginosa structural redundancy within the biofilm matrix. *Environmental Microbiology*, 14 (8): 1913–1928. doi:10.1111/j.1462-2920.2011.02657.x.

Costa Brandão Cruz, D., Lima Santana, L., Siqueira Guedes, A., et al. (2019) Different Ways of Doing the Same: Variations in the Two Last Steps of the Purine Biosynthetic Pathway in Prokaryotes. *Genome Biology and Evolution*, 11 (4): 1235–1249. doi:10.1093/gbe/evz035.

Cui, G., Zhang, Y., Xu, X., et al. (2022) PmiR senses 2-methylisocitrate levels to regulate bacterial virulence in Pseudomonas aeruginosa. *Science Advances*. doi:10.1126/sciadv.add4220.

De Kievit, T.R. (2009) Quorum sensing in Pseudomonas aeruginosa biofilms. *Environmental Microbiology*, 11 (2): 279–288. doi:10.1111/j.1462-2920.2008.01792.x.

De Kievit, T.R., Parkins, M.D., Gillis, R.J., et al. (2001) Multidrug Efflux Pumps: Expression Patterns and Contribution to Antibiotic Resistance in Pseudomonas aeruginosa Biofilms. *Antimicrobial Agents and Chemotherapy*, 45 (6): 1761–1770. doi:10.1128/aac.45.6.1761-1770.2001.

Denise, R., Babor, J., Gerlt, J.A., et al. (2023) Pyridoxal 5'-phosphate synthesis and salvage in Bacteria and Archaea: predicting pathway variant distributions and holes. *Microbial Genomics*, 9 (2). doi:10.1099/mgen.0.000926.

Dettman, J.R. and Kassen, R. (2021) Evolutionary Genomics of Niche-Specific Adaptation to the Cystic Fibrosis Lung in Pseudomonas aeruginosa. *Molecular Biology and Evolution*, 38 (2): 663–675. doi:10.1093/molbev/msaa226.

Diao, J., Komura, R., Sano, T., et al. (2021) Inhibition of Escherichia coli Lipoprotein Diacylglyceryl Transferase Is Insensitive to Resistance Caused by Deletion of Braun's Lipoprotein. *Journal of Bacteriology*, 203 (13): e0014921. doi:10.1128/JB.00149-21.

Dixon, S.J., Costanzo, M., Baryshnikova, A., et al. (2009) Systematic Mapping of Genetic Interaction Networks. *Annual Review of Genetics*, 43 (1): 601–625. doi:10.1146/annurev.genet.39.073003.114751.

Doherty, H.M., Kritikos, G., Galardini, M., et al. (2023) ChemGAPP: a tool for chemical genomics analysis and phenotypic profiling. *Bioinformatics*, 39 (4): btad171. doi:10.1093/bioinformatics/btad171.

Donlan, R.M. (2002) Biofilms: microbial life on surfaces. *Emerging Infectious Diseases*, 8 (9): 881–890. doi:10.3201/eid0809.020063.

Dötsch, A., Becker, T., Pommerenke, C., et al. (2009) Genomewide Identification of Genetic Determinants of Antimicrobial Drug Resistance in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 53 (6): 2522–2531. doi:10.1128/AAC.00035-09.

Dreier, J. and Ruggerone, P. (2015) Interaction of antibacterial compounds with RND efflux pumps in Pseudomonas aeruginosa. *Frontiers in Microbiology*, 6: 660. doi:10.3389/fmicb.2015.00660.

Driscoll, P., Lecky, F. and Crosby, M. (2000) Article 4. An introduction to estimation—1. Starting from Z. *Emergency Medicine Journal*, 17 (6): 409–415. doi:10.1136/emj.17.6.409.

Elamin, A.A., Steinicke, S., Oehlmann, W., et al. (2017) Novel drug targets in cell wall biosynthesis exploited by gene disruption in Pseudomonas aeruginosa. *PLoS ONE*, 12 (10): e0186801. doi:10.1371/journal.pone.0186801.

Elkins, C.A. and Nikaido, H. (2002) Substrate Specificity of the RND-Type Multidrug Efflux Pumps AcrB and AcrD of Escherichia coli Is Determined Predominately by Two

Large Periplasmic Loops. *Journal of Bacteriology*, 184 (23): 6490–6498. doi:10.1128/JB.184.23.6490-6498.2002.

Fernández-Villa, D., Aguilar, M.R. and Rojo, L. (2019) Folic Acid Antagonists: Antimicrobial and Immunomodulating Mechanisms and Applications. *International Journal of Molecular Sciences*, 20 (20): 4996. doi:10.3390/ijms20204996.

Fetar, H., Gilmour, C., Klinoski, R., et al. (2011) mexEF-oprN Multidrug Efflux Operon of Pseudomonas aeruginosa: Regulation by the MexT Activator in Response to Nitrosative Stress and Chloramphenicol. *Antimicrobial Agents and Chemotherapy*, 55 (2): 508–514. doi:10.1128/aac.00830-10.

Fleming, D., Niese, B., Redman, W., et al. (2022) Contribution of Pseudomonas aeruginosa Exopolysaccharides Pel and Psl to Wound Infections. *Frontiers in Cellular and Infection Microbiology*, 12. Available at: https://www.frontiersin.org/article/10.3389/fcimb.2022.835754 (Accessed: 11 May 2022).

Fraud, S., Campigotto, A.J., Chen, Z., et al. (2008) MexCD-OprJ Multidrug Efflux System of Pseudomonas aeruginosa: Involvement in Chlorhexidine Resistance and Induction by Membrane-Damaging Agents Dependent upon the AlgU Stress Response Sigma Factor. *Antimicrobial Agents and Chemotherapy*, 52 (12): 4478–4482. doi:10.1128/aac.01072-08.

French, S., Mangat, C., Bharat, A., et al. (2016) A robust platform for chemical genomics in bacterial systems. *Molecular Biology of the Cell*, 27 (6): 1015–1025. doi:10.1091/mbc.E15-08-0573.

Friedrich, T. and Scheide, D. (2000) The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases11Dedicated to Prof. Dr. Hanns Weiss on the occasion of his 60th birthday. *FEBS Letters*, 479 (1): 1–5. doi:10.1016/S0014-5793(00)01867-6.

Gahlot, R., Nigam, C., Kumar, V., et al. (2014) Catheter-related bloodstream infections. *International Journal of Critical Illness and Injury Science*, 4 (2): 162–167. doi:10.4103/2229-5151.134184.

Galardini, M., Busby, B.P., Vieitez, C., et al. (2019) The impact of the genetic background on gene deletion phenotypes in Saccharomyces cerevisiae. *Molecular Systems Biology*, 15 (12): e8831. doi:10.15252/msb.20198831.

Galardini, M., Koumoutsi, A., Herrera-Dominguez, L., et al. (2017) Phenotype inference in an Escherichia coli strain panel. *eLife*, 6: e31035. doi:10.7554/eLife.31035.

Gallagher, L.A., Ramage, E., Weiss, E.J., et al. (2015) Resources for Genetic and Genomic Analysis of Emerging Pathogen Acinetobacter baumannii. *Journal of Bacteriology*, 197 (12): 2027–2035. doi:10.1128/JB.00131-15.

Gellatly, S.L. and Hancock, R.E.W. (2013) Pseudomonas aeruginosa : new insights into pathogenesis and host defenses. *Pathogens and Disease*, 67 (3): 159–173. doi:10.1111/2049-632X.12033.

Ghafoor, A., Hay, I.D. and Rehm, B.H.A. (2011) *Role of Exopolysaccharides in Pseudomonas aeruginosa Biofilm Formation and Architecture | Applied and Environmental Microbiology*. Available at: https://journals.asm.org/doi/10.1128/AEM.00637-11 (Accessed: 21 April 2022).

Gomez, M.J. and Neyfakh, A.A. (2006) Genes Involved in Intrinsic Antibiotic Resistance of Acinetobacter baylyi. *Antimicrobial Agents and Chemotherapy*, 50 (11): 3562–3567. doi:10.1128/AAC.00579-06.

Grace, A., Sahu, R., Owen, D.R., et al. (2022) Pseudomonas aeruginosa reference strains PAO1 and PA14: A genomic, phenotypic, and therapeutic review. *Frontiers in Microbiology*, 13. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1023523 (Accessed: 22 May 2023).

Green, J.M. and Matthews, R.G. (2007) Folate Biosynthesis, Reduction, and Polyglutamylation and the Interconversion of Folate Derivatives. *EcoSal Plus*, 2 (2): 10.1128/ecosalplus.3.6.3.6. doi:10.1128/ecosalplus.3.6.3.6.

Guenezan, J., Drugeon, B., Marjanovic, N., et al. (2018) Treatment of central line-associated bloodstream infections. *Critical Care*, 22 (1): 303, s13054-018-2249–9. doi:10.1186/s13054-018-2249-9.

Guilbaud, M., Bruzaud, J., Bouffartigues, E., et al. (2017) Proteomic Response of Pseudomonas aeruginosa PAO1 Adhering to Solid Surfaces. *Frontiers in Microbiology*, 8: 1465. doi:10.3389/fmicb.2017.01465.

Hannon, G.J. (2010) *FASTX-Toolkit*. Available at: http://hannonlab.cshl.edu/fastx_toolkit/index.html (Accessed: 17 February 2023).

Harrison, E.M., Carter, M.E.K., Luck, S., et al. (2010) Pathogenicity Islands PAPI-1 and PAPI-2 Contribute Individually and Synergistically to the Virulence of Pseudomonas aeruginosa Strain PA14. *Infection and Immunity*, 78 (4): 1437–1446. doi:10.1128/iai.00621-09.

Hauser, A.R., Jain, M., Bar-Meir, M., et al. (2011) Clinical Significance of Microbial Infection and Adaptation in Cystic Fibrosis. *Clinical Microbiology Reviews*, 24 (1): 29–70. doi:10.1128/CMR.00036-10.

Häussler, S. and Becker, T. (2008) The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in Pseudomonas aeruginosa Populations. *PLOS Pathogens*, 4 (9): e1000166. doi:10.1371/journal.ppat.1000166.

He, J., Baldini, R.L., Déziel, E., et al. (2004) The broad host range pathogen Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (8): 2530. doi:10.1073/pnas.0304622101.

Hegazy, W.A.H., Khayat, M.T., Ibrahim, T.S., et al. (2020) Repurposing Anti-diabetic Drugs to Cripple Quorum Sensing in Pseudomonas aeruginosa. *Microorganisms*, 8 (9): 1285. doi:10.3390/microorganisms8091285.

Heywood, A. and Lamont, I.L. (2020) Cell envelope proteases and peptidases of Pseudomonas aeruginosa: multiple roles, multiple mechanisms. *FEMS Microbiology Reviews*, 44 (6): 857–873. doi:10.1093/femsre/fuaa036.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., et al. (2008) The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. *Science*, 320 (5874): 362–365. doi:10.1126/science.1150021.

Høiby, N., Ciofu, O. and Bjarnsholt, T. (2010) Pseudomonas aeruginosa biofilms in cystic fibrosis. *Future Microbiology*, 5 (11): 1663–1674. doi:10.2217/fmb.10.125.

Holloway, B.W. (1955) Genetic Recombination in Pseudomonas aeruginosa. *Microbiology*, 13 (3): 572–581. doi:10.1099/00221287-13-3-572.

Horna, G., López, M., Guerra, H., et al. (2018) Interplay between MexAB-OprM and MexEF-OprN in clinical isolates of Pseudomonas aeruginosa. *Scientific Reports*, 8: 16463. doi:10.1038/s41598-018-34694-z.

Housseini B Issa, K., Phan, G. and Broutin, I. (2018) Functional Mechanism of the Efflux Pumps Transcription Regulators From Pseudomonas aeruginosa Based on 3D Structures. *Frontiers in Molecular Biosciences*, 5. Available at: https://www.frontiersin.org/article/10.3389/fmolb.2018.00057 (Accessed: 15 May 2022).

Hreha, T.N., Foreman, S., Duran-Pinedo, A., et al. (2021) The three NADH dehydrogenases of Pseudomonas aeruginosa: Their roles in energy metabolism and links to virulence. *PLoS ONE*, 16 (2): e0244142. doi:10.1371/journal.pone.0244142.

Hsu, D.I., Okamoto, M.P., Murthy, R., et al. (2005) Fluoroquinolone-resistant Pseudomonas aeruginosa: risk factors for acquisition and impact on outcomes. *Journal of Antimicrobial Chemotherapy*, 55 (4): 535–541. doi:10.1093/jac/dki026.

Hu, Z. and Cronan, J.E. (2020) The primary step of biotin synthesis in mycobacteria. *Proceedings of the National Academy of Sciences*, 117 (38): 23794–23801. doi:10.1073/pnas.2010189117.

Huang, W. and Wilks, A. (2017) A rapid seamless method for gene knockout in Pseudomonas aeruginosa. *BMC Microbiology*, 17 (1): 199. doi:10.1186/s12866-017-1112-5.

Ibrahim, D., Jabbour, J.-F. and Kanj, S.S. (2020) Current choices of antibiotic treatment for Pseudomonas aeruginosa infections. *Current Opinion in Infectious Diseases*, 33 (6): 464–473. doi:10.1097/QCO.0000000000000677.

Irving, S.E., Choudhury, N.R. and Corrigan, R.M. (2021) The stringent response and physiological roles of (pp)pGpp in bacteria. *Nature Reviews. Microbiology*, 19 (4): 256–271. doi:10.1038/s41579-020-00470-y.

J, S., Rk, M., Sa, A., et al. (2022) The initiation factor 3 (IF3) residues interacting with initiator tRNA elbow modulate the fidelity of translation initiation and growth fitness in Escherichia coli. *Nucleic acids research*, 50 (20). doi:10.1093/nar/gkac1053.

Jacobs, M.A., Alwood, A., Thaipisuttikul, I., et al. (2003) Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, 100 (24): 14339–14344. doi:10.1073/pnas.2036282100.

Jennings, L.K., Storek, K.M., Ledvina, H.E., et al. (2015) Pel is a cationic exopolysaccharide that cross-links extracellular DNA in the Pseudomonas aeruginosa biofilm matrix. *Proceedings of the National Academy of Sciences*, 112 (36): 11353–11358. doi:10.1073/pnas.1503058112.

Jolivet, P. and Foley, J.W. (2020) *SPRI bead mix*. Available at: https://www.protocols.io/view/spri-bead-mix-bnz4mf8w (Accessed: 10 February 2023).

Jurado-Martín, I., Sainz-Mejías, M. and McClean, S. (2021) Pseudomonas aeruginosa: An Audacious Pathogen with an Adaptable Arsenal of Virulence Factors. *International Journal of Molecular Sciences*, 22 (6): 3128. doi:10.3390/ijms22063128.

Jurgenson, C.T., Begley, T.P. and Ealick, S.E. (2009) The Structural and Biochemical Foundations of Thiamin Biosynthesis. *Annual Review of Biochemistry*, 78 (1): 569–603. doi:10.1146/annurev.biochem.78.072407.102340.

Kanamori, T., Kanou, N., Atomi, H., et al. (2004) Enzymatic Characterization of a Prokaryotic Urea Carboxylase. *Journal of Bacteriology*, 186 (9): 2532–2539. doi:10.1128/JB.186.9.2532-2539.2004.

Keil, C., Leach, R.W., Faizaan, S.M., et al. (2018) *Treeview 3.0 (beta 1) - Visualization and analysis of large data matrices*. doi:10.5281/zenodo.1303402.

Kim, B.H. and Gadd, G.M. (eds.) (2019) “Tricarboxylic acid (TCA) cycle, electron transport and oxidative phosphorylation.” In *Prokaryotic Metabolism and Physiology*. 2nd ed. Cambridge: Cambridge University Press. pp. 80–114. doi:10.1017/9781316761625.005.

King, J.D., Kocíncová, D., Westman, E.L., et al. (2009) Review: Lipopolysaccharide biosynthesis in Pseudomonas aeruginosa. *Innate Immunity*, 15 (5): 261–312. doi:10.1177/1753425909106436.

Kirisits, M.J., Prost, L., Starkey, M., et al. (2005) Characterization of Colony Morphology Variants Isolated from Pseudomonas aeruginosa Biofilms. *Applied and Environmental Microbiology*, 71 (8): 4809–4821. doi:10.1128/AEM.71.8.4809-4821.2005.

Klemm, C., Howell, R.S.M. and Thorpe, P.H. (2022) ScreenGarden: a shinyR application for fast and easy analysis of plate-based high-throughput screens. *BMC Bioinformatics*, 23 (1): 60. doi:10.1186/s12859-022-04586-1.

Klockgether, J., Cramer, N., Wiehlmann, L., et al. (2011) Pseudomonas aeruginosa Genomic Structure and Diversity. *Frontiers in Microbiology*, 2. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2011.00150 (Accessed: 22 May 2023).

Koch, C. (2002) Early infection and progression of cystic fibrosis lung disease. *Pediatric Pulmonology*, 34 (3): 232–236. doi:10.1002/ppul.10135.

Köhler, T., Curty, L.K., Barja, F., et al. (2000) Swarming of Pseudomonas aeruginosa Is Dependent on Cell-to-Cell Signaling and Requires Flagella and Pili. *Journal of Bacteriology*, 182 (21): 5990–5996.

Koo, B.-M., Kritikos, G., Farelli, J.D., et al. (2017) Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Systems*, 4 (3): 291-305.e7. doi:10.1016/j.cels.2016.12.013.

Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., et al. (2017) A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nature Microbiology*, 2 (5): 17014. doi:10.1038/nmicrobiol.2017.14.

Krivoruchko, A., Zhang, Y., Siewers, V., et al. (2015) Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic Engineering*, 28: 28–42. doi:10.1016/j.ymben.2014.11.009.

Lai, H., Kraszewski, J.L., Purwantini, E., et al. (2006) Identification of Pyruvate Carboxylase Genes in Pseudomonas aeruginosa PAO1 and Development of a P. aeruginosa-Based Overexpression System for α4- and α4β4-Type Pyruvate Carboxylases. *Applied and Environmental Microbiology*, 72 (12): 7785–7792. doi:10.1128/AEM.01564-06.

Langridge, G.C., Phan, M.-D., Turner, D.J., et al. (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, 19 (12): 2308–2316. doi:10.1101/gr.097097.109.

Lau, G.W., Hassett, D.J., Ran, H., et al. (2004) The role of pyocyanin in Pseudomonas aeruginosa infection. *Trends in Molecular Medicine*, 10 (12): 599–606. doi:10.1016/j.molmed.2004.10.002.

Lee, D.G., Urbach, J.M., Wu, G., et al. (2006) Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. *Genome Biology*, 7 (10): R90. doi:10.1186/gb-2006-7-10-r90.

Lee, J., Wu, J., Deng, Y., et al. (2013) A cell-cell communication signal integrates quorum sensing and stress response. *Nature Chemical Biology*, 9 (5): 339–343. doi:10.1038/nchembio.1225.

Lee, K. and Yoon, S.S. (2017) Pseudomonas aeruginosa Biofilm, a Programmed Bacterial Life for Fitness. *Journal of Microbiology and Biotechnology*, 27 (6): 1053–1064. doi:10.4014/jmb.1611.11056.

Lee, W.M., Elliott, J.E. and Brownsey, R.W. (2005) Inhibition of Acetyl-CoA Carboxylase Isoforms by Pyridoxal Phosphate *. *Journal of Biological Chemistry*, 280 (51): 41835–41843. doi:10.1074/jbc.M510728200.

Leighton, T.L., Mok, M.C., Junop, M.S., et al. (2018) Conserved, unstructured regions in Pseudomonas aeruginosa PilO are important for type IVa pilus function. *Scientific Reports*, 8 (1): 2600. doi:10.1038/s41598-018-20925-w.

Li, H., Luo, Y.-F., Williams, B.J., et al. (2012) Structure and function of OprD protein in Pseudomonas aeruginosa: From antibiotic resistance to novel therapies. *International Journal of Medical Microbiology*, 302 (2): 63–68. doi:10.1016/j.ijmm.2011.10.001.

Li, L., Palmer, S.O., Gomez, E.A., et al. (2020) 1H, 13C and 15N resonance assignments of translation initiation factor 3 from Pseudomonas aeruginosa. *Biomolecular NMR Assignments*, 14 (1): 93–97. doi:10.1007/s12104-020-09926-x.

Li, Y., Liu, X., Tang, K., et al. (2019) Excisionase in Pf filamentous prophage controls lysis-lysogeny decision-making in Pseudomonas aeruginosa. *Molecular Microbiology*, 111 (2): 495–513. doi:10.1111/mmi.14170.

Liberati, N.T., Urbach, J.M., Miyata, S., et al. (2006) An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Sciences*, 103 (8): 2833–2838. doi:10.1073/pnas.0511100103.

Lin, S. and Cronan, J.E. (2011) Closing in on complete pathways of biotin biosynthesis. *Molecular BioSystems*, 7 (6): 1811–1821. doi:10.1039/C1MB05022B.

Liu, Y., Jia, Y., Yang, K., et al. (2020) Metformin Restores Tetracyclines Susceptibility against Multidrug Resistant Bacteria. *Advanced Science*, 7 (12): 1902227. doi:10.1002/advs.201902227.

Llanes, C., Köhler, T., Patry, I., et al. (2011) Role of the MexEF-OprN Efflux System in Low-Level Resistance of Pseudomonas aeruginosa to Ciprofloxacin. *Antimicrobial Agents and Chemotherapy*, 55 (12): 5676–5684. doi:10.1128/aac.00101-11.

López, M., Rojo-Bezares, B., Chichón, G., et al. (2022) Resistance to Fluoroquinolones in Pseudomonas aeruginosa from Human, Animal, Food and Environmental Origin: The Role of CrpP and Mobilizable ICEs. *Antibiotics*, 11 (9): 1271. doi:10.3390/antibiotics11091271.

Lucock, M. (2000) Folic Acid: Nutritional Biochemistry, Molecular Biology, and Role in Disease Processes. *Molecular Genetics and Metabolism*, 71 (1): 121–138. doi:10.1006/mgme.2000.3027.

Lun, S., Miranda, D., Kubler, A., et al. (2014) Synthetic Lethality Reveals Mechanisms of Mycobacterium tuberculosis Resistance to β-Lactams. *mBio*, 5 (5): e01767-14. doi:10.1128/mBio.01767-14.

Ma, L., Conover, M., Lu, H., et al. (2009) Assembly and Development of the Pseudomonas aeruginosa Biofilm Matrix Ausubel, F.M. (ed.). *PLoS Pathogens*, 5 (3): e1000354. doi:10.1371/journal.ppat.1000354.

van der Maaten, L. and Hinton, G. (2008) Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9: 2579–2605.

Madiraju, A.K., Erion, D.M., Rahimi, Y., et al. (2014) Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. *Nature*, 510 (7506): 542–546. doi:10.1038/nature13270.

Mani, R., St.Onge, R.P., Hartman, J.L., et al. (2008) Defining genetic interaction. *Proceedings of the National Academy of Sciences*, 105 (9): 3461–3466. doi:10.1073/pnas.0712255105.

Mann, S. and Ploux, O. (2011) Pyridoxal-5′-phosphate-dependent enzymes involved in biotin biosynthesis: Structure, reaction mechanism and inhibition. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814 (11): 1459–1466. doi:10.1016/j.bbapap.2010.12.004.

Masters, P.A., O'Bryan, T.A., Zurlo, J., et al. (2003) Trimethoprim-Sulfamethoxazole Revisited. *Archives of Internal Medicine*, 163 (4): 402–410. doi:10.1001/archinte.163.4.402.

Masuda, N., Sakagawa, E., Ohya, S., et al. (2000) Substrate Specificities of MexAB-OprM, MexCD-OprJ, and MexXY-OprM Efflux Pumps in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 44 (12): 3322–3327.

Mathee, K. (2018) Forensic investigation into the origin of Pseudomonas aeruginosa PA14 — old but not lost. *Journal of Medical Microbiology*, 67 (8): 1019–1021. doi:10.1099/jmm.0.000778.

Meachery, G., Soyza, A.D., Nicholson, A., et al. (2008) Outcomes of lung transplantation for cystic fibrosis in a large UK cohort. *Thorax*, 63 (8): 725–731. doi:10.1136/thx.2007.092056.

Miethke, M., Pieroni, M., Weber, T., et al. (2021) Towards the sustainable discovery and development of new antibiotics. *Nature Reviews Chemistry*, 5 (10): 726–749. doi:10.1038/s41570-021-00313-1.

Mikkelsen, H., McMullan, R. and Filloux, A. (2011) The Pseudomonas aeruginosa Reference Strain PA14 Displays Increased Virulence Due to a Mutation in ladS. *PLoS ONE*, 6 (12): e29113. doi:10.1371/journal.pone.0029113.

Miller, M.B. and Bassler, B.L. (2001) Quorum Sensing in Bacteria. *Annual Review of Microbiology*, 55 (1): 165–199. doi:10.1146/annurev.micro.55.1.165.

Minato, Y., Dawadi, S., Kordus, S.L., et al. (2018) Mutual potentiation drives synergy between trimethoprim and sulfamethoxazole. *Nature Communications*, 9: 1003. doi:10.1038/s41467-018-03447-x.

Mittal, R., Aggarwal, S., Sharma, S., et al. (2009) Urinary tract infections caused by Pseudomonas aeruginosa: A minireview. *Journal of Infection and Public Health*, 2 (3): 101–111. doi:10.1016/j.jiph.2009.08.003.

Moffatt, B.A. and Ashihara, H. (2002) Purine and Pyrimidine Nucleotide Synthesis and Metabolism. *The Arabidopsis Book / American Society of Plant Biologists*, 1: e0018. doi:10.1199/tab.0018.

Mooij, M.J., Drenkard, E., Llamas, M.A., et al. (2007) Characterization of the integrated filamentous phage Pf5 and its involvement in small-colony formation. *Microbiology*, 153 (6): 1790–1798. doi:10.1099/mic.0.2006/003533-0.

Moradali, M.F., Ghods, S. and Rehm, B.H.A. (2017) Pseudomonas aeruginosa Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. *Frontiers in Cellular and Infection Microbiology*, 7: 39. doi:10.3389/fcimb.2017.00039.

Mráček, T., Drahota, Z. and Houštěk, J. (2013) The function and the role of the mitochondrial glycerol-3-phosphate dehydrogenase in mammalian tissues. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1827 (3): 401–410. doi:10.1016/j.bbabio.2012.11.014.

Munita, J.M. and Arias, C.A. (2016) “Mechanisms of Antibiotic Resistance.” In *Virulence Mechanisms of Bacterial Pathogens*. John Wiley & Sons, Ltd. pp. 481–511. doi:10.1128/9781555819286.ch17.

Murray, T.S. and Kazmierczak, B.I. (2008) Pseudomonas aeruginosa Exhibits Sliding Motility in the Absence of Type IV Pili and Flagella. *Journal of Bacteriology*, 190 (8): 2700–2708. doi:10.1128/jb.01620-07.

Mutalik, V.K., Novichkov, P.S., Price, M.N., et al. (2019) Dual-barcoded shotgun expression library sequencing for high-throughput characterization of functional traits in bacteria. *Nature Communications*, 10 (1): 308. doi:10.1038/s41467-018-08177-8.

Nichols, R.J., Sen, S., Choo, Y.J., et al. (2011) Phenotypic landscape of a bacterial cell. *Cell*, 144 (1): 143–156. doi:10.1016/j.cell.2010.11.052.

Noda, S., Takezawa, Y., Mizutani, T., et al. (2006) Alterations of cellular physiology in Escherichia coli in response to oxidative phosphorylation impaired by defective F1-ATPase. *Journal of Bacteriology*, 188 (19): 6869–6876. doi:10.1128/JB.00452-06.

Nouri, R., Ahangarzadeh Rezaee, M., Hasani, A., et al. (2016) The role of gyrA and parC mutations in fluoroquinolones-resistant Pseudomonas aeruginosa isolates from Iran. *Brazilian Journal of Microbiology*, 47 (4): 925–930. doi:10.1016/j.bjm.2016.07.016.

Olivares, J., Álvarez-Ortega, C. and Martinez, J.L. (2014) Metabolic Compensation of Fitness Costs Associated with Overexpression of the Multidrug Efflux Pump MexEF-OprN in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, 58 (7): 3904–3913. doi:10.1128/AAC.00121-14.

Overhage, J., Bains, M., Brazas, M.D., et al. (2008) Swarming of Pseudomonas aeruginosa Is a Complex Adaptation Leading to Increased Production of Virulence Factors and Antibiotic Resistance. *Journal of Bacteriology*, 190 (8): 2671–2679. doi:10.1128/JB.01659-07.

Overhage, J., Lewenza, S., Marr, A.K., et al. (2007) Identification of Genes Involved in Swarming Motility Using a Pseudomonas aeruginosa PAO1 Mini-Tn5-lux Mutant Library. *Journal of Bacteriology*, 189 (5): 2164–2169. doi:10.1128/JB.01623-06.

Page, M.G.P. (2020) A Unified Numbering Scheme for Class C β-Lactamases. *Antimicrobial Agents and Chemotherapy*, 64 (3): e02247-19. doi:10.1128/AAC.02247-19.

Pailler, J., Aucher, W., Pires, M., et al. (2012) Phosphatidylglycerol::Prolipoprotein Diacylglyceryl Transferase (Lgt) of Escherichia coli Has Seven Transmembrane Segments, and Its Essential Residues Are Embedded in the Membrane. *Journal of Bacteriology*, 194 (9): 2142–2151. doi:10.1128/JB.06641-11.

Pang, Z., Raudonis, R., Glick, B.R., et al. (2019) Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies. *Biotechnology Advances*, 37 (1): 177–192. doi:10.1016/j.biotechadv.2018.11.013.

Parkins, M.D., Somayaji, R. and Waters, V.J. (2018) Epidemiology, Biology, and Impact of Clonal *Pseudomonas aeruginosa* Infections in Cystic Fibrosis. *Clinical Microbiology Reviews*, 31 (4): e00019-18, /cmr/31/4/e00019-18.atom. doi:10.1128/CMR.00019-18.

Parsons, A.B., Lopez, A., Givoni, I.E., et al. (2006) Exploring the Mode-of-Action of Bioactive Compounds by Chemical-Genetic Profiling in Yeast. *Cell*, 126 (3): 611–625. doi:10.1016/j.cell.2006.06.040.

Parsons, J.B. and Rock, C.O. (2011) Is Bacterial Fatty Acid Synthesis a Valid Target for Antibacterial Drug Discovery? *Current opinion in microbiology*, 14 (5): 544–549. doi:10.1016/j.mib.2011.07.029.

Pendleton, J.N., Gorman, S.P. and Gilmore, B.F. (2013) Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-infective Therapy*, 11 (3): 297–308. doi:10.1586/eri.13.12.

Polyak, S.W., Abell, A.D., Wilce, M.C.J., et al. (2012) Structure, function and selective inhibition of bacterial acetyl-coa carboxylase. *Applied Microbiology and Biotechnology*, 93 (3): 983–992. doi:10.1007/s00253-011-3796-z.

Poulsen, B.E., Yang, R., Clatworthy, A.E., et al. (2019) Defining the core essential genome of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (20): 10072–10080. doi:10.1073/pnas.1900570116.

Purssell, A. and Poole, K. (2013) Functional characterization of the NfxB repressor of the mexCD–oprJ multidrug efflux operon of Pseudomonas aeruginosa. *Microbiology*, 159 (Pt_10): 2058–2073. doi:10.1099/mic.0.069286-0.

Rasamiravaka, T., Labtani, Q., Duez, P., et al. (2015) The Formation of Biofilms by *Pseudomonas aeruginosa* : A Review of the Natural and Synthetic Compounds Interfering with Control Mechanisms. *BioMed Research International*, 2015: 1–17. doi:10.1155/2015/759348.

Rashid, M.H. and Kornberg, A. (2000) Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (9): 4885–4890.

Rhein, V.F., Carroll, J., Ding, S., et al. (2016) NDUFAF5 Hydroxylates NDUFS7 at an Early Stage in the Assembly of Human Complex I. *The Journal of Biological Chemistry*, 291 (28): 14851–14860. doi:10.1074/jbc.M116.734970.

Rice, S.A., Tan, C.H., Mikkelsen, P.J., et al. (2009) The biofilm life cycle and virulence of Pseudomonas aeruginosa are dependent on a filamentous prophage. *The ISME Journal*, 3 (3): 271–282. doi:10.1038/ismej.2008.109.

Richts, B., Rosenberg, J. and Commichau, F.M. (2019) A Survey of Pyridoxal 5′-Phosphate-Dependent Proteins in the Gram-Positive Model Bacterium Bacillus subtilis. *Frontiers in Molecular Biosciences*, 6: 32. doi:10.3389/fmolb.2019.00032.

Rossi, E., Falcone, M., Molin, S., et al. (2018) High-resolution in situ transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-phenotypes in cystic fibrosis lungs. *Nature Communications*, 9 (1): 3459. doi:10.1038/s41467-018-05944-5.

Rousselet, G.A. and Pernet, C.R. (2012) Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience*, 6: 119. doi:10.3389/fnhum.2012.00119.

Ryan, C.J., Roguev, A., Patrick, K., et al. (2012) Hierarchical modularity and the evolution of genetic interactomes across species. *Molecular Cell*, 46 (5): 691–704. doi:10.1016/j.molcel.2012.05.028.

Salaemae, W., Azhar, A., Booker, G.W., et al. (2011) Biotin biosynthesis in Mycobacterium tuberculosis: physiology, biochemistry and molecular intervention. *Protein & Cell*, 2 (9): 691–695. doi:10.1007/s13238-011-1100-8.

Sana, T.G., Lomas, R., Gimenez, M.R., et al. (2019) Differential Modulation of Quorum Sensing Signaling through QslA in Pseudomonas aeruginosa Strains PAO1 and PA14. *Journal of Bacteriology*, 201 (21): 10.1128/jb.00362-19. doi:10.1128/jb.00362-19.

Schein, C.H. (2020) Repurposing approved drugs on the pathway to novel therapies. *Medicinal Research Reviews*, 40 (2): 586–605. doi:10.1002/med.21627.

Schuldiner, M., Collins, S.R., Thompson, N.J., et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell*, 123 (3): 507–519. doi:10.1016/j.cell.2005.08.031.

Schuster, M., Hawkins, A.C., Harwood, C.S., et al. (2004) The Pseudomonas aeruginosa RpoS regulon and its relationship to quorum sensing. *Molecular Microbiology*, 51 (4): 973–985. doi:10.1046/j.1365-2958.2003.03886.x.

Sharma, L.K., Lu, J. and Bai, Y. (2009) Mitochondrial Respiratory Complex I: Structure, Function and Implication in Human Diseases. *Current medicinal chemistry*, 16 (10): 1266–1277.

Shi, Y., Cao, Q., Sun, J., et al. (2023) The opportunistic pathogen Pseudomonas aeruginosa exploits bacterial biotin synthesis pathway to benefit its infectivity. *PLOS Pathogens*, 19 (1): e1011110. doi:10.1371/journal.ppat.1011110.

Shiver, A.L., Osadnik, H., Kritikos, G., et al. (2016) A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S Hughes, D. (ed.). *PLOS Genetics*, 12 (6): e1006124. doi:10.1371/journal.pgen.1006124.

Shiver, A.L., Osadnik, H., Peters, J.M., et al. (2021) Chemical-genetic interrogation of RNA polymerase mutants reveals structure-function relationships and physiological tradeoffs. *Molecular Cell*, 81 (10): 2201-2215.e9. doi:10.1016/j.molcel.2021.04.027.

Shuman, J., Giles, T.X., Carroll, L., et al. (2018) Transcriptome analysis of a Pseudomonas aeruginosasn-glycerol-3-phosphate dehydrogenase mutant reveals a disruption in bioenergetics. *Microbiology,* 164 (4): 551–562. doi:10.1099/mic.0.000646.

Silby, M.W., Winstanley, C., Godfrey, S.A.C., et al. (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiology Reviews*, 35 (4): 652–680. doi:10.1111/j.1574-6976.2011.00269.x.

Sinha, P.K., Torres-Bacete, J., Nakamaru-Ogiso, E., et al. (2009) Critical Roles of Subunit NuoH (ND1) in the Assembly of Peripheral Subunits with the Membrane Domain of Escherichia coli NDH-1. *The Journal of Biological Chemistry*, 284 (15): 9814–9823. doi:10.1074/jbc.M809468200.

Soberón-Chávez, G., Lépine, F. and Déziel, E. (2005) Production of rhamnolipids by Pseudomonas aeruginosa. *Applied Microbiology and Biotechnology*, 68 (6): 718–725. doi:10.1007/s00253-005-0150-3.

Soto, R., Perez-Herran, E., Rodriguez, B., et al. (2018) Identification and characterization of aspartyl-tRNA synthetase inhibitors against Mycobacterium tuberculosis by an integrated whole-cell target-based approach. *Scientific Reports*, 8 (1): 12664. doi:10.1038/s41598-018-31157-3.

Stapper, A.P., Narasimhan, G., Ohman, D.E., et al. (2004) Alginate production affects Pseudomonas aeruginosa biofilm development and architecture, but is not essential for biofilm formation. *Journal of Medical Microbiology*, 53 (7): 679–690. doi:10.1099/jmm.0.45539-0.

Steinchen, W., Ahmad, S., Valentini, M., et al. (2021) Dual role of a (p)ppGpp- and (p)ppApp-degrading enzyme in biofilm formation and interbacterial antagonism. *Molecular Microbiology*, 115 (6): 1339–1356. doi:10.1111/mmi.14684.

Stickland, H.G., Davenport, P.W., Lilley, K.S., et al. (2010) Mutation of nfxB Causes Global Changes in the Physiology and Metabolism of Pseudomonas aeruginosa. *Journal of Proteome Research*, 9 (6): 2957–2967. doi:10.1021/pr9011415.

Stover, C.K., Pham, X.Q., Erwin, A.L., et al. (2000) Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature*, 406 (6799): 959–964. doi:10.1038/35023079.

Sugimoto, R., Saito, N., Shimada, T., et al. (2017) Identification of YbhA as the pyridoxal 5'-phosphate (PLP) phosphatase in *Escherichia coli*: Importance of PLP homeostasis on the bacterial growth. *The Journal of General and Applied Microbiology*, 63 (6): 362–368. doi:10.2323/jgam.2017.02.008.

Thacharodi, A. and Lamont, I.L. (2022) Aminoglycoside-Modifying Enzymes Are Sufficient to Make Pseudomonas aeruginosa Clinically Resistant to Key Antibiotics. *Antibiotics*, 11 (7): 884. doi:10.3390/antibiotics11070884.

Thi, M.T.T., Wibowo, D. and Rehm, B.H.A. (2020) Pseudomonas aeruginosa Biofilms. *International Journal of Molecular Sciences*, 21 (22): 8671. doi:10.3390/ijms21228671.

Tong, A.H., Evangelista, M., Parsons, A.B., et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.)*, 294 (5550): 2364–2368. doi:10.1126/science.1065810.

Tong, L. (2013) Structure and function of biotin-dependent carboxylases. *Cellular and molecular life sciences : CMLS*, 70 (5): 863. doi:10.1007/s00018-012-1096-0.

Toutain, C.M., Zegans, M.E. and O'Toole, G.A. (2005) Evidence for Two Flagellar Stators and Their Role in the Motility of Pseudomonas aeruginosa. *Journal of Bacteriology*, 187 (2): 771–777. doi:10.1128/JB.187.2.771-777.2005.

Turcios, N.L. (2020) Cystic Fibrosis Lung Disease: An Overview. *Respiratory Care*, 65 (2): 233–251. doi:10.4187/respcare.06697.

Ude, J., Tripathi, V., Buyck, J.M., et al. (2021) Outer membrane permeability: Antimicrobials and diverse nutrients bypass porins in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, 118 (31): e2107644118. doi:10.1073/pnas.2107644118.

Ulrich, M., Worlitzsch, D., Viglio, S., et al. (2010) Alveolar Inflammation in Cystic Fibrosis. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society*, 9 (3): 217–227. doi:10.1016/j.jcf.2010.03.001.

Van Laar, T.A., Esani, S., Birges, T.J., et al. (2018) Pseudomonas aeruginosa gshA Mutant Is Defective in Biofilm Formation, Swarming, and Pyocyanin Production. *mSphere*, 3 (2): e00155-18. doi:10.1128/mSphere.00155-18.

Vélez-Cruz, R. and Osheroff, N. (2004) "DNA Topoisomerases: Type II." In Lennarz, W.J. and Lane, M.D. (eds.) *Encyclopedia of Biological Chemistry*. New York: Elsevier. pp. 806–811. doi:10.1016/B0-12-443710-9/00680-3.

Viollet, B., Guigas, B., Sanz Garcia, N., et al. (2012) Cellular and molecular mechanisms of metformin: an overview. *Clinical Science (London, England : 1979)*, 122 (6): 253–270. doi:10.1042/CS20110386.

Wagih, O. and Parts, L. (2014) gitter: A Robust and Accurate Method for Quantification of Colony Sizes From Plate Images. *G3: Genes|Genomes|Genetics*, 4 (3): 547. doi:10.1534/g3.113.009431.

Wang, K., Meng, X. and Guo, Z. (2021) Elastin Structure, Synthesis, Regulatory Mechanism and Relationship With Cardiovascular Diseases. *Frontiers in Cell and Developmental Biology*, 9. Available at: https://www.frontiersin.org/article/10.3389/fcell.2021.596702 (Accessed: 12 April 2022).

Wang, Y.-T., Lee, M.-F. and Peng, C.-F. (2014) Mutations in the quinolone resistance-determining regions associated with ciprofloxacin resistance in Pseudomonas

aeruginosa isolates from Southern Taiwan. *Biomarkers and Genomic Medicine*, 6 (2): 79–83. doi:10.1016/j.bgm.2014.03.003.

Webb, J.S., Lau, M. and Kjelleberg, S. (2004) Bacteriophage and Phenotypic Variation in Pseudomonas aeruginosa Biofilm Development. *Journal of Bacteriology*, 186 (23): 8066–8073. doi:10.1128/jb.186.23.8066-8073.2004.

Weiner-Lastinger, L.M., Abner, S., Edwards, J.R., et al. (2020) Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: Summary of data reported to the National Healthcare Safety Network, 2015–2017. *Infection Control & Hospital Epidemiology*, 41 (1): 1–18. doi:10.1017/ice.2019.296.

Wu, P.I.-F., Ross, C., Siegele, D.A., et al. (2020) Insights from the reanalysis of high-throughput chemical genomics data for Escherichia coli K-12. *G3: Genes|Genomes|Genetics*, 11 (1): jkaa035. doi:10.1093/g3journal/jkaa035.

Yasir, M., Turner, A.K., Bastkowski, S., et al. (2020) TraDIS-Xpress: a high-resolution whole-genome assay identifies novel mechanisms of triclosan action and resistance. *Genome Research*, 30 (2): 239–249. doi:10.1101/gr.254391.119.

Zhuo, T., Rou, W., Song, X., et al. (2015) Molecular study on the carAB operon reveals that carB gene is required for swimming and biofilm formation in Xanthomonas citri subsp. citri. *BMC Microbiology*, 15 (1): 225. doi:10.1186/s12866-015-0555-9.

Zimlichman, E., Henderson, D., Tamir, O., et al. (2013) Health Care–Associated Infections: A Meta-analysis of Costs and Financial Impact on the US Health Care System. *JAMA Internal Medicine*, 173 (22): 2039–2046. doi:10.1001/jamainternmed.2013.9763.

APPENDICES

Supplementary Table 1: MSR genes for PAO1 and their PseudoCAP functions. Count represents count of phenotypes.

Mutant	Count	Locus	PseudoCAP Function
PW8494__PA4450__murA	78	PA4450	Cell wall / LPS / capsule
PW7517__PA5488__	56	PA5488	Hypothetical
PW4193__PA1818__cadA	53	PA1818	Amino acid biosynthesis and metabolism
PW3907__PA3615__	49	PA3615	Hypothetical
PW9855__PA5261__algR	43	PA5261	Secreted Factors (toxins, enzymes, alginate)
PW1292__PA0171__siaB	40	PA0171	Motility & Attachment
PW3513__PA1382__	39	PA1382	Protein secretion/export apparatus
PW8457__PA4426__	39	PA4426	Hypothetical
PW4692__PA2155__	38	PA2155	Putative enzymes
PW4797__PA2230__	38	PA2230	Hypothetical
PW9474__PA5044__pilM	38	PA5044	Motility & Attachment
PW8674__PA4553__pilX	36	PA4553	Motility & Attachment
PW5039__PA2399__pvdD	35	PA2399	Secreted Factors (toxins, enzymes, alginate)
PW10330__PA5511__mifR	32	PA5511	Two-component regulatory systems
PW1783__PA0427__oprM	32	PA0427	Antibiotic resistance and susceptibility
PW6354__PA3197__	32	PA3197	Hypothetical
PW8676__PA4554__pilY1	31	PA4554	Motility & Attachment
PW1824__PA0451__	30	PA0451	Membrane proteins
PW8622__PA4525__pilA	30	PA4525	Motility & Attachment

Supplementary Table 2: MSR genes for PA14 and their PseudoCAP functions. Count represents count of phenotypes.

Mutant	Count	Locus	PseudoCAP Function
33008__PA14_34540__	73	PA14_34540	Putative enzymes
40470__PA14_00570__	68	PA14_00570	Cell wall / LPS / capsule
31924__PA14_17880__	64	PA14_17880	Carbon compound catabolism
53127__PA14_03010__	45	PA14_03010	Transcriptional regulators
54440__PA14_57170__	44	PA14_57170	Two-component regulatory systems
25862__PA14_16470__	39	PA14_16470	Two-component regulatory systems
29758__PA14_15600__	39	PA14_15600	Hypothetical
32003__PA14_04410__ptsP	39	PA14_04410	Transport of small molecules
23942__PA14_28130__	38	PA14_28130	Transcriptional regulators
43633__PA14_07620__cca	38	PA14_07620	Transcription, RNA processing and degradation
23029__PA14_65250__hisZ	36	PA14_65250	Transport of small molecules
25469__PA14_59470__	36	PA14_59470	Hypothetical
48761__PA14_46110__	32	PA14_46110	Transport of small molecules
53600__PA14_68360__	31	PA14_68360	Fatty acid and phospholipid metabolism

Supplementary Table 3: Ortholog genes for PAO1 and PA14 with Pearson correlation coefficient r values ≥ 0.4.

Gene	PAO1 Locus	PA14 Locus	r value	Function
aatJ	PA1342	PA14_46910	0.401	Transport of small molecules
aceA	PA2634	PA14_30050	0.481	Putative enzymes
acsA	PA0887	PA14_52800	0.480	Carbon compound catabolism
ampC	PA4110	PA14_10790	0.415	Adaptation, Protection
atuE	PA2890	PA14_26690	0.406	Putative enzymes
copA1	PA3920	PA14_13170	0.667	Transport of small molecules
dadA	PA5304	PA14_70040	0.416	Energy metabolism
fpvG	PA2403	PA14_33600	0.478	Membrane proteins
gltD	PA5035	PA14_66560	0.410	Amino acid biosynthesis and metabolism
gshA	PA5203	PA14_68730	0.478	Amino acid biosynthesis and metabolism
ilvE	PA5013	PA14_66260	0.418	Amino acid biosynthesis and metabolism
mexB	PA0426	PA14_05540	0.682	Transport of small molecules
ntrC	PA5125	PA14_67680	0.470	Transcriptional regulators
nuoJ	PA2645	PA14_29900	0.414	Energy metabolism
nuoN	PA2649	PA14_29850	0.424	Energy metabolism
oprM	PA0427	PA14_05550	0.579	Transport of small molecules
pilI	PA0410	PA14_05340	0.402	Motility & Attachment
prpB	PA0796	PA14_53940	0.545	Carbon compound catabolism
prpC	PA0795	PA14_53950	0.450	Carbon compound catabolism
pstC	PA5368	PA14_70850	0.409	Transport of small molecules
thrB	PA5495	PA14_72510	0.616	Amino acid biosynthesis and metabolism
uvrB	PA3138	PA14_23510	0.407	DNA replication, recombination, modification and repair

Supplementary Table 4: Table displaying the anticorrelated genes for PAO1 and PA14. MC and MN indicate minimal media based conditions.

Condition	Gene	PAO1 Locus	PA14 Locus	PAO1 S-score	PA14 S-score
Adult bovine serum 20%	-	PA5151	PA14_68050	-4.74	3.85
Adult bovine serum 20%	-	PA4136	PA14_10470	-2.91	7.14
Adult bovine serum 20%	cbrA	PA4725	PA14_62530	-3.11	3.56
Adult bovine serum 20%	ctpA	PA5134	PA14_67810	-3.84	3.25
Adult bovine serum 20%	gcvT1	PA5215	PA14_68870	-5.14	5.41
Adult bovine serum 20%	gcvT1	PA5215	PA14_68870	-5.14	5.41
Adult bovine serum 20%	gcvT1	PA5215	PA14_68870	-5.14	5.41
Adult bovine serum 20%	gcvT1	PA5215	PA14_68870	-5.14	5.41
Adult bovine serum 20%	mgtE	PA0913	PA14_52460	-2.75	4.03
Adult bovine serum 20%	pheA	PA3166	PA14_23280	-3.41	6.94
Adult bovine serum 20%	uvrC	PA2585	PA14_30660	-3.96	4.07
Adult bovine serum 20%	uvrC	PA2585	PA14_30660	-3.96	4.07
Adult bovine serum 5%	pilS	PA4546	PA14_60250	-3.54	6.10
Adult bovine serum 5%	pilS	PA4546	PA14_60250	-3.54	6.10
Adult bovine serum 5%	pilS	PA4546	PA14_60250	-3.54	6.10
Adult bovine serum 5%	pilS	PA4546	PA14_60250	-3.54	6.10
Ampicillin 100 µg/ml	amgS	PA5199	PA14_68680	-3.26	6.27
Ampicillin 100 µg/ml	amgS	PA5199	PA14_68680	-3.26	6.27
Ampicillin 100 µg/ml	amgS	PA5199	PA14_68680	-3.26	6.27
Ampicillin 150 µg/ml	amgS	PA5199	PA14_68680	-4.17	3.91
Ampicillin 150 µg/ml	amgS	PA5199	PA14_68680	-4.17	3.91
Ampicillin 150 µg/ml	amgS	PA5199	PA14_68680	-4.17	3.91
Ampicillin 150 µg/ml	lpoA	PA4423	PA14_57480	3.68	-3.30
Ampicillin 150 µg/ml	phzC2	PA1901	PA14_09460	7.70	-3.79
Ampicillin 150 µg/ml	phzC2	PA1901	PA14_09460	7.70	-3.79
Azlocillin 2 µg/ml	-	PA1940	PA14_39460	-2.91	3.18
Azlocillin 2 µg/ml	galU	PA2023	PA14_38350	-3.04	3.60
Azlocillin 2 µg/ml	galU	PA2023	PA14_38350	-3.04	3.60
Azlocillin 2 µg/ml	galU	PA2023	PA14_38350	-3.04	3.60
Azlocillin 2 µg/ml	galU	PA2023	PA14_38350	-3.04	3.60
Azlocillin 2 µg/ml	galU	PA2023	PA14_38350	-3.04	3.60
Azlocillin 2 µg/ml	gcvT1	PA5215	PA14_68870	3.04	-2.73
Azlocillin 2 µg/ml	gcvT1	PA5215	PA14_68870	3.04	-2.73

Azlocillin 2 µg/ml	gcvT1	PA5215	PA14_68870	3.04	-2.73
Azlocillin 2 µg/ml	gcvT1	PA5215	PA14_68870	3.04	-2.73
Bipyridil 0.25 mm	-	PA2814	PA14_27740	-3.86	4.56
Bipyridil 0.25 mM	fimU	PA4550	PA14_60280	-3.62	6.56
Bipyridil 0.25 mM	pilG	PA0408	PA14_05320	-3.41	3.67
Bipyridil 0.25 mM	pilG	PA0408	PA14_05320	-3.41	3.67
Bipyridil 0.25 mM	pilG	PA0408	PA14_05320	-3.41	3.67
Bipyridil 0.25 mM	pilG	PA0408	PA14_05320	-3.41	3.67
Bipyridil 0.25 mM	pilG	PA0408	PA14_05320	-3.41	3.67
Bipyridil 0.25 mM	recB	PA4284	PA14_55670	-4.07	4.83
Bipyridil 0.25 mM	recB	PA4284	PA14_55670	-4.07	4.83
Chloramphenicol 12 µg/ml	mexF	PA2494	PA14_32390	3.44	-3.14
Chloramphenicol 12 µg/ml	mexF	PA2494	PA14_32390	3.44	-3.14
Ciprofloxacin 0.025 µg/ml	pilI	PA0410	PA14_05340	4.10	-4.56
Ciprofloxacin 0.025 µg/ml	pilI	PA0410	PA14_05340	4.10	-4.56
Ciprofloxacin 0.025 µg/ml	pilI	PA0410	PA14_05340	4.10	-4.56
Ciprofloxacin 0.025 µg/ml	pilI	PA0410	PA14_05340	4.10	-4.56
Cobalt II chloride hexahydrate 0.5 mM	dadA	PA5304	PA14_70040	-3.07	3.83
Cobalt II chloride hexahydrate 0.5 mM	dadA	PA5304	PA14_70040	-3.07	3.83
Cobalt II chloride hexahydrate 0.5 mM	HA62_01445	PA2897	PA14_26590	-4.00	4.61
Copper II chloride dihydrate 4 mM	-	PA5472	PA14_72220	-5.45	3.27
Copper II chloride dihydrate 4 mM	-	PA4032	PA14_11680	-3.40	3.53
Copper II chloride dihydrate 4 mM	-	PA5551	PA14_73200	3.86	-3.61
Copper II chloride dihydrate 4 mM	-	PA5551	PA14_73200	3.86	-3.61
Copper II chloride dihydrate 4 mM	-	PA0339	PA14_04430	3.08	-4.58
Copper II chloride dihydrate 4 mM	truB	PA4742	PA14_62730	4.20	-4.13
Copper II chloride dihydrate 4 mM	truB	PA4742	PA14_62730	3.11	-4.13
Copper II chloride dihydrate 4 mM	vfr	PA0652	PA14_08370	4.19	-6.45

Copper II chloride dihydrate 4 mM	vfr	PA0652	PA14_08370	4.19	-6.45
Copper II chloride dihydrate 4 mM	vfr	PA0652	PA14_08370	4.19	-6.45
Copper II chloride dihydrate 4 mM	vfr	PA0652	PA14_08370	4.19	-6.45
Deoxycholate 0.02%	-	PA5145	PA14_67970	-4.21	4.80
Erythromycin 3 µg/ml	fimU	PA4550	PA14_60280	3.32	-4.31
Erythromycin 3 µg/ml	fimU	PA4550	PA14_60280	6.15	-4.31
MC arabinose 24 mM green	argH	PA5263	PA14_69500	-11.35	10.60
MC arabinose 24 mM green	argH	PA5263	PA14_69500	-11.35	10.60
MC arabinose 24 mM green	argH	PA5263	PA14_69500	-11.35	10.60
MC arabinose 24 mM green	ilvI	PA4696	PA14_62160	-5.28	18.47
MC arabinose 24 mM green	ilvI	PA4696	PA14_62160	-5.28	18.47
MC arabinose 24 mM green	ilvI	PA4696	PA14_62160	-5.28	18.47
MC arabinose 24 mM green	ilvI	PA4696	PA14_62160	-5.28	18.47
MC fructose 20 mM	hisC2	PA3165	PA14_23290	-5.54	3.45
MC fructose 20 mM	hisC2	PA3165	PA14_23290	-5.54	3.45
MC fructose 20 mM	hisC2	PA3165	PA14_23290	-5.54	3.45
MC fructose 20 mM	hisC2	PA3165	PA14_23290	-5.54	3.45
MC fructose 20 mM	hisC2	PA3165	PA14_23290	-5.54	3.45
MC fructose 20 mM	hisC2	PA3165	PA14_23290	-5.54	3.45
MC glucose 20 mM	proA	PA4007	PA14_12010	-3.56	3.91
MC glucose 20 mM	proA	PA4007	PA14_12010	-3.56	3.91
MC glucose 20 mM	uvrC	PA2585	PA14_30660	-4.56	7.91
MC glucose 20 mM	uvrC	PA2585	PA14_30660	-4.56	7.91
MC glycerol 40 mM	cysD	PA4443	PA14_57720	-8.07	3.43
MC glycerol 40 mM	cysD	PA4443	PA14_57720	-8.07	3.43
MC glycerol 40 mM	cysD	PA4443	PA14_57720	-8.07	3.43
MC glycerol 40 mM	cysD	PA4443	PA14_57720	-8.07	3.43
MC glycerol 40 mM	cysG	PA2611	PA14_30340	-3.49	24.49
MC glycerol 40 mM	cysG	PA2611	PA14_30340	-3.51	24.49
MC glycerol 40 mM	pyrC	PA0401	PA14_05250	-4.07	3.60
MC glycerol ph8	cysG	PA2611	PA14_30340	-3.43	15.79
MC glycerol ph8	hisZ	PA4939	PA14_65250	-3.55	9.44
MC glycerol ph8	leuC	PA3121	PA14_23750	-4.27	8.68

MC glycerol ph8	leuC	PA3121	PA14_23750	-4.27	8.68
MC glycerol ph8	trpA	PA0035	PA14_00440	-4.10	18.35
MC glycerol ph8	trpA	PA0035	PA14_00440	-4.10	18.35
MC lactose 10 mM	leuC	PA3121	PA14_23750	-3.51	7.54
MC lactose 10 mM	leuC	PA3121	PA14_23750	-3.51	7.54
MC lactose 10 mM	mqoB	PA4640	PA14_61400	-5.30	13.40
MC lactose 10 mM	mqoB	PA4640	PA14_61400	-5.30	13.40
MC lactose 10 mM	proA	PA4007	PA14_12010	-4.00	18.89
MC lactose 10 mM	proA	PA4007	PA14_12010	-4.00	18.89
MC lactose 10 mM	trpA	PA0035	PA14_00440	-3.51	15.48
MC lactose 10 mM	trpA	PA0035	PA14_00440	-3.51	15.48
MC mannitol 20 mM	leuB	PA3118	PA14_23790	-4.50	4.31
MC mannitol 20 mM	leuB	PA3118	PA14_23790	-4.50	4.31
MC mannitol 20 mM	pyrC	PA0401	PA14_05250	-3.98	5.02
MC pyruvate 40 mM	-	PA4683	PA14_61950	-3.84	3.47
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-4.96	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-4.96	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-4.96	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-4.96	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-3.73	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-3.73	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-3.73	3.71
MC pyruvate 40 mM	cysD	PA4443	PA14_57720	-3.73	3.71
MC pyruvate 40 mM	trpB	PA0036	PA14_00450	-5.69	4.80
MC pyruvate 40 mM	trpB	PA0036	PA14_00450	-5.69	4.80
MC succinate 30 mM	argC	PA0662	PA14_08480	-3.78	13.58
MC succinate 30 mM	cysG	PA2611	PA14_30340	-4.46	18.46
MC succinate 30 mM	cysG	PA2611	PA14_30340	-4.49	18.46
MC succinate 30 mM	hisZ	PA4939	PA14_65250	-3.50	5.05
MC succinate 30 mM	hisZ	PA4939	PA14_65250	-3.50	10.32
MC succinate 30 mM	ilvD	PA0353	PA14_04630	-4.73	9.01
MC succinate 30 mM	ilvD	PA0353	PA14_04630	-4.73	9.01
Meropenem 0.6 µg/ml	-	PA5188	PA14_68530	2.83	-7.23
Meropenem 0.6 µg/ml	-	PA4390	PA14_57060	3.25	-5.98
Meropenem 0.6 µg/ml	-	PA3978	PA14_12370	3.36	-3.03
Meropenem 0.6 µg/ml	-	PA2949	PA14_25910	3.20	-2.84
Meropenem 0.6 µg/ml	atuF	PA2891	PA14_26670	3.78	-8.82
Meropenem 0.6 µg/ml	copA2	PA1549	PA14_44440	-3.33	4.33
Meropenem 0.6 µg/ml	cupE5	PA4652	PA14_61540	3.49	-3.30
Meropenem 0.6 µg/ml	cupE5	PA4652	PA14_61540	3.49	-3.30

Meropenem 0.6 µg/ml	cupE5	PA4652	PA14_61540	3.49	-3.30
Meropenem 0.6 µg/ml	cysH	PA1756	PA14_41840	3.59	-5.18
Meropenem 0.6 µg/ml	dusA	PA2795	PA14_27980	3.12	-8.78
Meropenem 0.6 µg/ml	eco	PA2755	PA14_28450	3.43	-11.06
Meropenem 0.6 µg/ml	mexJ	PA3677	PA14_16800	3.02	-9.62
Meropenem 0.6 µg/ml	mexJ	PA3677	PA14_16800	3.02	-9.62
Meropenem 0.6 µg/ml	mexK	PA3676	PA14_16820	2.74	-8.89
Meropenem 0.6 µg/ml	mexK	PA3676	PA14_16820	2.74	-8.89
Meropenem 0.6 µg/ml	mpaR	PA4132	PA14_10530	-4.81	3.20
Meropenem 0.6 µg/ml	muxC	PA2526	PA14_31900	2.90	-6.87
Meropenem 0.6 µg/ml	muxC	PA2526	PA14_31900	2.90	-6.87
Meropenem 0.6 µg/ml	pldB	PA0308	PA14_04030	4.45	-9.15
Meropenem 0.6 µg/ml	pnaB	PA0195.1	PA14_02460	2.96	-11.49
Meropenem 0.6 µg/ml	rsmC	PA4627	PA14_61220	2.72	-10.67
Meropenem 0.6 µg/ml	scpA	PA3198	PA14_22840	2.87	-4.30
Meropenem 0.6 µg/ml	ybiB	PA2609	PA14_30360	3.39	-5.80
Meropenem 0.6 µg/ml	ylbA	PA1140	PA14_49650	3.51	-4.42
Metformin 100 mM	-	PA4523	PA14_58690	4.11	-3.20
Metformin 100 mM	bbsG	PA2550	PA14_31580	4.38	-4.07
Metformin 100 mM	bbsG	PA2550	PA14_31580	4.38	-4.07
Metformin 50 mM	alaA	PA2828	PA14_27500	4.66	-3.40
Metformin 50 mM	alaA	PA2828	PA14_27500	4.66	-3.40
Metformin 50 mM	alaA	PA2828	PA14_27500	4.66	-3.40
Metformin 50 mM	alaA	PA2828	PA14_27500	4.66	-3.40
Metformin 50 mM	fimX	PA4959	PA14_65540	5.75	-3.69
Minocycline 24 µg/ml	-	PA4222	PA14_09320	-7.94	3.99
Minocycline 24 µg/ml	armZ	PA5471	PA14_72210	2.99	-4.95
MN glycine 20 mM	soxA	PA2326	PA14_34540	-6.39	7.39
MN glycine 20 mM	yjjK	PA4595	PA14_60800	-2.95	3.61
MN isoleucine 20 mM	hisH1	PA5142	PA14_67920	-5.71	5.80
MN isoleucine 20 mM	leuC	PA3121	PA14_23750	-3.65	5.54
MN isoleucine 20 mM	leuC	PA3121	PA14_23750	-3.65	5.54
MN L-alanine 20 mM	thiD	PA3975	PA14_12410	3.91	-3.44
MN L-asparagine 10 mM	hisZ	PA4939	PA14_65250	-3.78	4.85
MN L-asparagine 10 mM	pyrD	PA3050	PA14_24640	-3.17	3.51
MN L-aspartic acid 20 mM	gidA	PA5565	PA14_73370	5.26	-3.25
MN L-aspartic acid 20 mM	gidA	PA5565	PA14_73370	5.26	-3.25
MN L-aspartic acid 20 mM	nrtR	PA4916	PA14_64930	4.44	-3.40

MN leucine 20 mM	-	PA3534	PA14_18640	-5.26	4.12
MN L-glutamate 20 mM	aatQ	PA1341	PA14_46920	3.39	-10.56
MN L-glutamate 20 mM	aatQ	PA1341	PA14_46920	3.39	-10.56
MN L-glutamate 20 mM	aatQ	PA1341	PA14_46920	3.39	-10.56
MN L-glutamate 20 mM	ntrB	PA5124	PA14_67670	3.62	-6.75
MN L-glutamate 20 mM	ntrB	PA5124	PA14_67670	3.62	-6.75
MN L-glutamate 20 mM	ntrB	PA5124	PA14_67670	3.62	-6.75
MN L-glutamate 20 mM	pilM	PA5044	PA14_66660	3.38	-4.17
MN L-glutamate 20 mM	pilM	PA5044	PA14_66660	3.38	-4.17
MN L-glutamate 20 mM	pilO	PA5042	PA14_66640	4.62	-3.27
MN L-glutamate 20 mM	pilO	PA5042	PA14_66640	5.57	-3.27
MN L-glutamate 20 mM	pilO	PA5042	PA14_66640	4.62	-3.27
MN L-glutamate 20 mM	pilO	PA5042	PA14_66640	5.57	-3.27
MN L-glutamate 20 mM	pilQ	PA5040	PA14_66620	3.97	-4.50
MN L-glutamate 20 mM	pilQ	PA5040	PA14_66620	3.97	-4.50
MN proline 20 mM	hisZ	PA4939	PA14_65250	-6.23	6.24
MN proline 20 mM	hisZ	PA4939	PA14_65250	-6.23	10.98
MN proline 20 mM	ilvI	PA4696	PA14_62160	-27.07	12.24
MN proline 20 mM	ilvI	PA4696	PA14_62160	-27.07	12.24
MN proline 20 mM	ilvI	PA4696	PA14_62160	-27.07	12.24
MN proline 20 mM	ilvI	PA4696	PA14_62160	-27.07	12.24
MN proline 20 mM	kinB	PA5484	PA14_72390	5.13	-3.85
MN proline 20 mM	kinB	PA5484	PA14_72390	5.13	-3.85
MN proline 20 mM	kinB	PA5484	PA14_72390	5.13	-3.85
MN proline 20 mM	purD	PA4855	PA14_64220	-3.35	4.31
MN serine 20 mM	-	PA4520	PA14_58650	3.81	-9.84
MN serine 20 mM	-	PA1006	PA14_51310	3.43	-3.94
MN serine 20 mM	-	PA1006	PA14_51310	3.43	-3.94
MN serine 20 mM	-	PA4520	PA14_58650	3.81	-9.84
MN serine 20 mM	-	PA4520	PA14_58650	3.81	-9.84
MN serine 20 mM	gltB	PA5036	PA14_66570	3.46	-3.20
MN serine 20 mM	gltB	PA5036	PA14_66570	3.46	-3.20
MN serine 20 mM	gltB	PA5036	PA14_66570	3.46	-3.20
MN serine 20 mM	mqoB	PA4640	PA14_61400	3.74	-3.25
MN serine 20 mM	mqoB	PA4640	PA14_61400	3.74	-3.25
MN serine 20 mM	ntrB	PA5124	PA14_67670	3.21	-3.33
MN serine 20 mM	ntrB	PA5124	PA14_67670	3.21	-3.33
MN serine 20 mM	ntrB	PA5124	PA14_67670	3.21	-3.33
MN serine 20 mM	pycA	PA5436	PA14_71740	3.41	-6.11
MN serine 20 mM	pycA	PA5436	PA14_71740	3.41	-6.11

MN serine 20 mM	pycA	PA5436	PA14_71740	5.83	-6.11
MN serine 20 mM	pycA	PA5436	PA14_71740	5.83	-6.11
MN serine 20 mM	pycA	PA5436	PA14_71740	4.62	-6.11
MN serine 20 mM	pycA	PA5436	PA14_71740	4.62	-6.11
MN serine 20 mM	tdcF	PA5339	PA14_70480	3.69	-6.98
MN serine 20 mM	trpG	PA0649	PA14_08340	4.24	-4.76
MN serine 20 mM	trpG	PA0649	PA14_08340	4.24	-4.76
MN serine 20 mM	trpG	PA0649	PA14_08340	4.24	-4.76
MN serine 20 mM	trpG	PA0649	PA14_08340	4.24	-4.76
NaCl 300 mM	damX	PA5037	PA14_66580	-3.85	3.44
NaCl 300 mM	tpbA	PA3885	PA14_13660	3.48	-7.86
pH 7	fimU	PA4550	PA14_60280	-3.13	4.02
pH 7	fimU	PA4550	PA14_60280	-3.33	4.02
pH 7	HA62_32075	PA5005	PA14_66170	-3.24	4.14
pH 7	HA62_32075	PA5005	PA14_66170	-5.19	4.14
pH 8	nrtR	PA4916	PA14_64930	5.70	-4.13
pH 8	pilX	PA4553	PA14_60300	-4.41	4.50
pH 8	pilX	PA4553	PA14_60300	-4.41	4.50
pH 9	ilvE	PA5013	PA14_66260	6.65	-3.42
pH 9	ilvE	PA5013	PA14_66260	6.65	-3.42
pH 9	ilvE	PA5013	PA14_66260	6.65	-3.42
pH 9	ilvE	PA5013	PA14_66260	6.65	-3.42
pH 9	ilvE	PA5013	PA14_66260	6.65	-3.42
Plumbagin 500 µg/mL	ruvB	PA0967	PA14_51780	-6.42	6.15
Plumbagin 500 µg/mL	ruvB	PA0967	PA14_51780	-6.42	6.15
Rifaximin 2 µg/mL	soxA	PA2326	PA14_34540	-5.93	3.89
Spectinomycin 5 µg/mL	pilK	PA0412	PA14_05380	-4.59	5.97
Spectinomycin 5 µg/mL	pilK	PA0412	PA14_05380	-4.59	5.97
Spectinomycin 5 µg/mL	pilK	PA0412	PA14_05380	-4.59	5.97
Sulfamethoxazole 100 µg/ml	algR	PA5261	PA14_69470	-8.53	4.82
Sulfamethoxazole 100 µg/ml	algR	PA5261	PA14_69470	-8.94	4.82
Sulfamethoxazole 100 µg/ml	algR	PA5261	PA14_69470	-8.53	4.82
Sulfamethoxazole 100 µg/ml	algR	PA5261	PA14_69470	-8.94	4.82
Sulfamethoxazole 100 µg/ml	pilO	PA5042	PA14_66640	-5.50	3.61
Sulfamethoxazole 100 µg/ml	pilO	PA5042	PA14_66640	-3.36	3.61

Sulfamethoxazole 100 µg/ml	pilO	PA5042	PA14_66640	-5.50	3.61
Sulfamethoxazole 100 µg/ml	pilO	PA5042	PA14_66640	-3.36	3.61
Temperature 42 ˚C	folE2	PA1674	PA14_42850	3.40	-5.73
Temperature2 30 ˚C	cpoB	PA0974	PA14_51690	4.62	-3.62
Temperature2 30 ˚C	pcm	PA3624	PA14_17460	-4.03	5.96
Temperature2 rt	cpoB	PA0974	PA14_51690	4.20	-4.14
Temperature2 rt	HA62_32075	PA5005	PA14_66170	-3.80	6.29
Temperature2 rt	HA62_32075	PA5005	PA14_66170	-7.42	6.29
Triclosanirgasan 0.025 µg/mL	-	PA0339	PA14_04430	-3.74	4.40
Triclosanirgasan 0.025 µg/mL	-	PA5145	PA14_67970	-6.25	13.10
Triclosanirgasan 0.025 µg/mL	fimU	PA4550	PA14_60280	-3.60	3.67
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	-	PA3091	PA14_24180	3.70	-4.28
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	algR	PA5261	PA14_69470	-3.38	7.01
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	algR	PA5261	PA14_69470	-3.41	7.01
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	algR	PA5261	PA14_69470	-3.38	7.01
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	algR	PA5261	PA14_69470	-3.41	7.01
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	cbrB	PA4726	PA14_62540	3.97	-3.04
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	cbrB	PA4726	PA14_62540	3.97	-3.04
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	cycH	PA1483	PA14_45280	-3.29	4.24
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	oprG	PA4067	PA14_11270	3.67	-3.18

Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	pcnA	PA4918	PA14_64950	2.97	-3.27
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	pilQ	PA5040	PA14_66620	-3.28	3.87
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	pilQ	PA5040	PA14_66620	-3.28	3.87
Trimethoprim + sulfamethoxazole, 1 µg/ml + 100 µg/ml	ydgK	PA3573	PA14_18090	3.30	-4.28
Tylosin 20	fimU	PA4550	PA14_60280	-4.42	5.34
UV 4	-	PA5145	PA14_67970	-9.99	8.12
UV 4	dsbA	PA5489	PA14_72450	3.98	-3.71
UV 4	dsbA	PA5489	PA14_72450	3.98	-3.71
UV 4	pfpI	PA0355	PA14_04650	2.92	-4.10
UV 4	prhR	PA2051	PA14_37980	3.42	-4.06
UV 4	purF	PA3108	PA14_23920	3.32	-4.71
UV 4	purF	PA3108	PA14_23920	3.32	-4.71
UV 4	purF	PA3108	PA14_23920	3.32	-4.71
Uv 4	quiP	PA1032	PA14_50980	3.51	-4.29
UV 8	-	PA3670	PA14_16890	2.70	-4.83
UV 8	purF	PA3108	PA14_23920	2.76	-4.53
UV 8	purF	PA3108	PA14_23920	2.76	-4.53
UV 8	purF	PA3108	PA14_23920	2.76	-4.53
UV 8	rmcA	PA0575	PA14_07500	3.39	-4.01
UV 8	rmcA	PA0575	PA14_07500	2.68	-4.01
UV 8	roxS	PA4494	PA14_58320	-9.76	3.80
UV 8	roxS	PA4494	PA14_58320	-9.76	3.80
UV 8	roxS	PA4494	PA14_58320	-9.76	3.80
UV 8	uvrD	PA5443	PA14_71870	3.55	-4.81
UV 8	uvrD	PA5443	PA14_71870	3.55	-4.81
UV 8	uvrD	PA5443	PA14_71870	3.55	-4.81

Supplementary Table 5: Genes with increased transposon insertions for the PA14 Δ*bioB-H* library in different concentrations of metformin.

	Log2 Fold Change		
Gene	**0.5x MIC**	**1x MIC**	**2x MIC**
aceE			2.31975836
aceF			1.853757962
argA			2.275165841
argB		1.289897815	2.450441371
argC			2.645294392
argF	1.610995854	1.428533549	3.028153334
argG	1.177254356	1.309263935	2.527965503
argH	1.580629849	1.563227411	3.146380909
aroB			3.40547737
aroC			6.200039611
aspS	4.463059809		4.580657496
atpC		1.63883225	
carA	1.012137115	1.329038705	2.971028236
carB	1.358216568	1.292443817	3.07456994
cc4	1.085631952		1.950118626
cca	1.078385783		1.32995056
ccmA		1.679918853	3.279334177
ccmB			1.984464329
ccmC	4.305249409		4.587230272
ccmF			2.093305878
ccmG			1.830023945
ccmH			1.200575879
clpX	1.737388359	1.41990641	
cupC2		1.375477538	
cycB			1.010609944
cysD			1.307130837
cysN			1.252370289
dapE	2.908626324		
dcd		4.489439822	
dksA		1.188157843	2.370203041
dnaK		1.393097506	1.401192469
dnaX		4.713794187	
dsbA1			1.432692759
dxr		4.398588128	

dxs	6.086890763	5.447250171	4.718680908
epd	2.755511064	3.602313349	4.016969011
fdxA	4.387644679		
ffh		3.946022018	4.611119023
folB	1.761166039	1.488785184	
folE	1.129846911	1.230446367	1.28779749
frr	3.342612139		
ftsW	5.536910447		6.169291883
galU	2.002182386	2.172975131	
gidA	1.912708712	1.880063153	1.211838303
glmM	1.67022322	1.560412458	
glmS			5.145305831
glnA		4.127649589	
glpD			1.469987292
gltA			2.274096354
gmk			3.237360201
gshA			1.087184386
hemA			2.302349409
hemB		4.435181054	
hemD	2.310776333	2.186117419	2.805366633
hemL	2.56599999		
hisC1			1.06802467
hom			1.049598995
hscA		1.66685402	
hscB	1.846081183	1.951207494	2.443678475
ilvC			1.503983479
ilvD			1.319419388
ilvH			2.695831481
ilvI			1.887876559
ipk	4.963666236		
ispA	4.794500871		5.003142319
ispB			2.885420822
lepA			1.137209615
leuC			1.181210094
leuD			1.238629911
lgt	6.753706099	6.182439198	6.170571974
lip22		2.444807598	1.616293572
lis	1.239879193		
lnt		3.946152199	
lpxK	1.750648504	1.566899139	

masA		1.893388864	1.544473521
metR			1.0789454
metX			1.175627259
mexA	1.165034505	1.485417712	1.768441959
mexB	1.12425628	1.41465786	1.633953532
miaA	1.527565113	1.568205285	1.372487559
minC			1.351409137
mraY			3.469340825
murC			5.747010929
murD		5.247421083	
murG	2.770643671	2.827361368	3.156993493
nadA	1.366262239	2.065029605	2.66206935
nadB	2.699122581	2.996997742	3.512860783
nadC	1.128235227	1.196636558	2.19800665
ndk			4.700971536
nusA			1.334129318
oprM	1.158609003	1.602665296	1.906239987
orfE		2.777455201	
orfL	1.24263213		
orfM	1.128711072		
orn	3.018644438	2.74996015	3.296403342
PA14_04860		1.044076972	1.377458301
PA14_07600		1.272234329	1.623059226
PA14_08400			2.931439696
PA14_08470		1.932206647	
PA14_10530			1.466198043
PA14_11860			1.951788724
PA14_14680	7.155119124	5.78906057	
PA14_16710		4.127413916	
PA14_17150		2.521576635	
PA14_17330			1.358174944
PA14_19150			2.909259251
PA14_21810		1.218704606	
PA14_23310			1.11080835
PA14_24870		4.498668824	
PA14_25050		1.011669543	1.399781911
PA14_27180	1.088215565		
PA14_27200	3.731796876	2.494511233	4.054561479
PA14_27940			1.2580958
PA14_29720	3.331148124		

PA14_30930			1.0126058
PA14_35210		1.029175282	2.304934009
PA14_36620	1.019664065		
PA14_40630	1.808963246		1.276093573
PA14_42380	1.15789264		
PA14_44270	4.395816837	5.145418725	
PA14_44370	1.106009935		
PA14_44380	1.512808768	1.537825273	1.482779599
PA14_44430	1.274851608	1.21255782	2.582978508
PA14_44460	1.03148233		1.68244048
PA14_45340			1.886142547
PA14_48900			1.523538071
PA14_51810	1.938305489	1.889826677	1.951594579
PA14_52530	4.980494693	6.389764935	
PA14_56300	5.13835405	5.994693444	6.355340541
PA14_57170			1.194670488
PA14_57470			1.101311867
PA14_57560			1.296461937
PA14_57570			1.31364455
PA14_58300			1.38609762
PA14_58320			1.03083364
PA14_61840			1.584491448
PA14_61890		4.316139961	
PA14_62050	1.141896126		
PA14_63460	1.646693362	1.531065106	
PA14_65690	2.832672505	2.775408156	3.020693233
PA14_66000		3.060915431	
PA14_66090	6.699051565	5.994693444	6.358262167
PA14_66120	2.697146572	3.014696345	3.149531209
PA14_66160	1.080499903		
PA14_66190		1.169977667	1.584866728
PA14_66800		1.104273153	
PA14_68360	2.351951953	1.552405203	3.847014287
PA14_68670	1.656343227		
PA14_70480			1.100930463
PA14_71780	1.105113195	1.449908784	1.665333028
pckA			1.013501904
pdxA	1.303955075	1.725940099	3.131785763
pdxB			1.179572951
pdxJ	2.663134781	3.394949643	4.048463023

pgk	1.230701448	1.468826624	
pgsA		4.300294264	
pheA			2.459501673
phhB	1.420976338	1.485712979	1.229016178
phoU			1.134250897
pilH		1.174411971	1.506198129
ppa			5.394396801
proA			1.704133975
proB			1.610704107
proC			1.693190736
prs			5.14749828
prtR	2.727798337	3.64447193	2.412067888
pstB			1.150139558
purD	1.146233292	1.164247491	2.952545597
purE	2.839232549	2.876092841	4.022891023
purF	1.463025284		2.767176279
purK		1.535147214	2.475864199
purL		1.19465514	2.906047742
purM			3.224637929
purN	1.8535122	2.280217508	2.889864475
pyrB	1.335213555	1.035324804	2.936377336
pyrC			6.121856042
pyrC			2.764771971
pyrD		1.55711677	3.909940986
pyrE	1.592177828	1.139612561	3.412278158
pyrF	2.451979922		4.310776133
rbfA	1.073039143	1.847806727	1.60356008
rimM		4.132585025	
rmlB	5.047027778	5.493671199	
rnc	1.152459085	1.233108651	1.035273968
rnhA			5.435063873
rnt	2.073508188	1.76120397	
rplC		4.623743088	
rplD		4.612257515	
rplS		4.1323331	
rpmC			2.302400742
rpoA	4.834733478	4.869693014	4.273098227
rpsI			5.293695227
sdhA			1.840116655
secF		4.978432093	5.268944364

serA	2.860973658	3.316118221	3.70118419
serB	1.046208144	1.01488991	1.826215463
speD			1.093460099
sucC		1.647813914	2.020067756
tatA			2.335269266
tatC			2.140024615
thrC			1.208287639
tktA		2.566881953	
topA			3.178670313
trmE	1.752777749	1.174245178	1.383913822
trpA			1.740285062
trpB			1.190333026
trpD			1.086639394
trpF			1.372278846
trpG			1.62643477
trpS			4.822292902
truB	1.190685857	1.571163252	1.227551035
tyrB			1.675376642
ubiB	3.512205886		
waaC			2.039034778
waaF			2.04685526
yqgF	1.01164803	1.06547298	

Supplementary Table 6: Genes with decreased transposon insertions for the PA14 Δ*bioB-H* library in different concentrations of metformin.

	Log_2 Fold Change		
Gene	**0.5x MIC**	**1x MIC**	**2x MIC**
cysB		-1.1093561	
cysI	-1.0978205		
fabF1		-1.2067041	
ftsH	-1.8424195	-3.4228233	-2.2304334
gpsA	-1.1019616	-1.3108769	
hfq			-1.1387152
hisA	-1.1208877	-1.0928822	

hisB	-1.5807384		
hisD	-1.4510603		
hisE	-1.4286648		
hisF1		-1.1685794	
hisH	-1.6813494		
hisZ	-1.0574461		
hom	-1.110256		
lasR		-1.8312422	-3.4401861
leuB	-1.0373617		
lysA	-1.1582992	-1.3027113	
metZ	-1.2118855		
mgtE			-1.0616547
moeB		-1.2219432	
mvfR	-1.1250056	-1.4867155	-2.021516
nfxB	-1.7663576	-2.5755185	-1.4693962
nqrD			-1.1142287
nqrE			-1.0001581
omlA	-1.0584443	-1.1247767	
orfN			-1.9850744
PA14_04380	-1.3468172		
PA14_04970	-1.0531241	-1.6225757	-1.0364413
PA14_05070	-1.2504869		
PA14_12180	-1.1070946		
PA14_20850			-1.1727767
PA14_21800			-1.8758616
PA14_24110			-1.982459
PA14_25370			-1.1536681
PA14_26850			-1.1256814
PA14_32420			-1.0438383
PA14_34740			-1.2637619
PA14_38380			-1.0153251
PA14_41200		-1.358514	
PA14_41320	-5.8236497	-5.6046796	-5.5283057
PA14_41680			-1.2421388
PA14_46530			-1.1023695
PA14_47640			-1.17529
PA14_49410			-1.0578485
PA14_51310		-1.504533	-1.2929012
PA14_51660	-4.8829379		
PA14_55620			-3.2714019

PA14_57970		-1.0981444	
PA14_60650		-1.1075031	-1.2483982
pqsA			-1.2647863
pqsB		-1.2078164	-1.5769346
pqsC		-1.0073542	-1.1916248
pqsD	-1.0728433	-1.2653151	-1.4588019
pqsE		-1.209194	-1.3163058
pykA		-1.1499254	-1.0385594
rpmE		-1.1450132	
rsmA			-1.6316525
shaA			-1.0187145
thiD	-1.3208091	-1.3697531	-1.0072963
thiE		-1.0586359	
thiG	-1.0956869	-1.2153471	-1.0491758
tpiA	-1.4146538	-1.5025373	

The lipoprotein DolP affects cell separation in Escherichia coli, but not as an upstream regulator of NlpD

Gabriela Boelter, Jack A Bryant , **Hannah Doherty**, Peter Wotherspoon, Dema Alodaini, Xuyu Ma, Micheal B Alao, Patrick J Moynihan, Danesh Moradigaravand, Monika Glinkowska, Timothy J Knowles, Ian R Henderson, Manuel Banzhaf.

Microbiology (Reading). 2022 May;168(5). DOI: 10.1099/mic.0.001197.

This study focused on characterising the *Escherichia coli* lipoprotein DolP. Previous studies had linked DolP to the bacterial amidase regulator NlpD, suggesting DolP was an upstream regulator of NlpD. Amidases are essential for cell division, being responsible for daughter cell separation. Interactions between DolP and NlpD were tested via pull-down and MST (MicroScale Thermophoresis) assays. However, no interactions were found. Therefore, double knockout mutants of *dolP* and the amidases *amiA, amiB* and *amiC* or the amidase regulators *envC* and *actS* were analysed. Microscopy revealed increased chain lengths for Δ*amiA*Δ*dolP* and Δ*envC*Δ*dolP* mutants, indicating a role in cell division. As such, MST assays were performed between DolP and the amidases and their regulators. MST displayed interactions between DolP and AmiA, AmiC and EnvC, suggesting a role for DolP in daughter cell separation.

During this study, I contributed via the purification of recombinant AmiA, AmiB, AmiC, NlpD and EnvC. Further to this I performed the MST analysis within Figure 1C.

Bam complex associated proteins in Escherichia coli are functionally linked to peptidoglycan biosynthesis, membrane fluidity and DNA replication.

Jack A. Bryant, Kara A. Staunton, **Hannah M. Doherty**, Micheal B. Alao, Xuyu Ma, Joanna Morcinek-Orłowska, Emily C. A. Goodall, Jessica Gray, Mathew Milner, Jeffrey A. Cole, Felicity de Cogan, Timothy J. Knowles, Monika Glinkowska, Danesh Moradigaravand, Ian R. Henderson, Manuel Banzhaf

BioRxiv. July 2023. DOI: https://doi.org/10.1101/2023.07.05.547807.

This study performed a variety of techniques to characterise the beta-barrel assembly machine (Bam) complex and its associated chaperones. The Bam complex is responsible for the transport and insertion of outer membrane proteins (OMPs). Phenotypic analysis via a small-scale chemical genomic screen, a TraDIS screen and a conservation analysis were performed to evaluate the function of the Bam accessory lipoproteins and chaperones. Disruptions in the regulation of DNA replication initiation were revealed in the absence of BamB. The study also connected the Bam complex to LPS, since LPS truncation led to increased membrane fluidity and decreased Bam activity. Further to this, the study showed essentiality of meso-DAP in the absence of BamB, BamC or BamE. Therefore, displaying a connection to peptidoglycan biosynthesis as well. As such, the study concluded that OMP biogenesis was functionally linked to multiple cellular processes.

In this study I performed the data analysis for the phenotypic screen (Fig. 1 and S1) and the genetic interaction screen (Fig. 6D). Further to this I performed the conservation analysis displayed in Figure 7.