



**Integrating Imaging and Omics with Machine Learning
Approaches to Infer Antimicrobial Susceptibility in
Gram-negative Bacteria**

Tran Tuan-Anh

**Linacre College
Clinical Medicine**

**A thesis submitted to the degree of Doctor of Philosophy
January 2023**

Integrating Imaging and Omics with Machine Learning Approaches to Infer Antimicrobial Susceptibility in Gram-negative Bacteria

Tran Tuan-Anh, Linacre College. DPhil thesis in Clinical Medicine, Hillary Term 2023.

Abstract

Antimicrobial resistance is a major global health issue, which threatens the efficacy of antimicrobial treatments. Many efforts have been made to study the genetic mechanisms of resistance and adaptation of bacteria under antimicrobial treatments. However, the impacts of these genetic factors on the bacterial adaptation have not been well understood. In this thesis, I holistically investigated the physiological and cytological adaptation of Gram-negative bacteria upon exposure to different antimicrobials. Particularly, proteomic and cellomic profiles along with genomic data of these bacteria were acquired, which provides insights into the molecular mechanism of the adaptation. I found a high prevalence of resistant pathogenic *Escherichia coli* isolated from children with sepsis or diarrhoeal disease. Although antimicrobial exposure generally reduced the growth bacteria, azithromycin accelerated the bacterial growth at sub-inhibitory concentrations. Additionally, tetracycline-resistant isolates were found to possess a fitness advantage over the susceptible isolates. Furthermore, a variety of machine learning models were trained on the proteomic and cellomic profiles, resulting in high accuracy of resistance identification. The models also found strong association of resistance to different antimicrobials with the proteomic and cellomic profiles. The machine learning models suggested that exposure to ciprofloxacin exerted a significant perturbation on bacterial membrane and the expression of various transmembrane proteins, which generates novel hypotheses for further experiments to confirm. Alternatively, I utilised a high-content imaging platform to observe the cytological adaptation of *Salmonella enterica* serovar Typhimurium upon 24-hour of ciprofloxacin

exposure. Strong association of the adaptation with the exposure time and the antimicrobial concentrations was found. I also found that resistant isolates behaved differently compared to the susceptible isolate even without exposure to the antimicrobial. In conclusion, further holistic investigations of antimicrobial resistance should be conducted with the aids of machine learning approaches to provide in-depth understanding on the mechanism of bacterial adaptation, contributing to the fight against antimicrobial resistance.

Acknowledgement

I would like to express my deepest gratitude towards my supervisor, Prof. Stephen Baker, for his patience and endless support during my DPhil journey. Thank you for being there for me during the hard times, for your advice, and your guidance as a mentor, a friend, and a brother. I am also grateful to know your wife, Ms. Tran Vu Thieu Nga. She always stood by my side and supported me in a way that a sister would do for her brother. You are my family.

I wish to send my gratitude and my warmest love to my close friends, Ms. Nguyen Ngoc Minh Chau, Ms. Nguyen Thi Nguyen To, Mr. Pham Duc Trung, Ms. Nguyen Hoang Thu Trang. Our friendship is an asset that I will value for the rest of my life.

I would like to thank my colleagues at OUCRU, Ms. Ha Thanh Tuyen, Ms. Ho Ngoc Dan Thanh, Mr. Le Minh Triet, and Ms. Tran Hanh Triet, for supporting me during the very first stage of my DPhil. I also want to express my appreciation to all my colleagues at the Baker lab, Department of Medicine, University of Cambridge. Special thanks to Ms. Claire Cormie, Ms. Sally Forest, Ms. Sophie Palmer, and Ms. Tamsyn Robert for facilitating my research at the lab.

I want to give a special thanks to my co-supervisors and my mentor, Dr. Pham Thanh Duy, Dr. Hannah Clapham, Dr. Josefin Bartholdson-Scott, and Dr. Chung The Hao.

Finally, I would like to express my love and gratitude to my family, especially my mom, for your unconditional and endless love for me.

Declaration

I, Tran Tuan-Anh, declare that the majority of the work presented in this thesis is on my own and was guided by Professor Stephen Baker at the Oxford University Clinical Research Unit, Vietnam (OUCRU) and the Department of Medicine, University of Cambridge, UK. The details of my contributions were stated in each chapter.

This thesis has not been submitted for a degree or other qualification of this or any other university.

Publications during my DPhil candidature

Zellmer, C., **Tran, T. A.** & Sridhar, S. Seeing the bigger picture. Nat. Rev. Microbiol. 19, 745 (2021). (Chapter 1)

Tuan-Anh, T. et al. Pathogenic *Escherichia coli* Possess Elevated Growth Rates under Exposure to Sub-Inhibitory Concentrations of Azithromycin. Antibiotics 9, 735 (2020). (Chapter 3)

Tuan-Anh, T. et al., Application of machine learning to high content imaging to infer antimicrobial resistance phenotypes in *Salmonella* Typhimurium. (in progress, chapter 4)

Anh, N. T. et al. Viral Metagenomic Analysis of Cerebrospinal Fluid from Patients with Acute Central Nervous System Infections of Unknown Origin, Vietnam. Emerg. Infect. Dis. 27, 205 (2021).

Nguyen, T. N. T. et al. Emerging carbapenem-resistant *Klebsiella pneumoniae* sequence type 16 causing multiple outbreaks in a tertiary hospital in southern Vietnam. Microb. genomics 7, (2021).

Chung The, H. et al. Exploring the genomic diversity and antimicrobial susceptibility of *Bifidobacterium pseudocatenulatum* in a Vietnamese population. Microbiol. Spectr. 9, e00526-21 (2021).

Tuan-Anh, T. & Zaleśny, R. Predictions of High-Order Electric Properties of Molecules: Can We Benefit from Machine Learning? *ACS omega* 5, 5318–5325 (2020).

The, H. C. et al. Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children. *Gut Microbes* 0976, (2017).

Table of Contents

Abstract.....	i
Acknowledgement	iii
Declaration.....	iv
Publications during my DPhil candidature	v
Table of Contents	vii
List of figures	xi
List of tables.....	xiv
Abbreviations	xv
Chapter 1. Introduction and background.....	1
1.1. A brief history of antimicrobials	1
1.2. Antimicrobial resistance.....	10
1.3. Antimicrobial modes of action and resistance mechanisms.....	12
1.3.1. Mechanisms of resistance to (fluoro)quinolones.....	14
1.3.2. Mechanisms of resistance to macrolides	16
1.3.3. Mechanisms of resistance to tetracyclines	18
1.4. The Enterobacteriaceae family	19
1.4.1. <i>Escherichia coli</i>	20
1.4.2. The <i>Salmonella</i> genus and <i>Salmonella</i> Typhimurium.....	21
1.5. Methods to study antimicrobial resistance	24
1.5.1. Culture-based AST	27
1.5.2. Molecular techniques and genomics analysis.....	27
1.5.3. Proteomics analysis with MALDI-TOF MS	30
1.5.4. Cellomic analysis with high-content imaging	32
1.6. Machine learning in biology	42
1.7. Research aims and questions.....	44
Chapter 2. Methods.....	45

2.1. Collections of bacterial isolates.....	45
2.2. Whole genome sequencing.....	46
2.3. Data analysis	47
2.3.1. Genomic analysis.....	47
2.3.2. Statistical analysis and data visualisation.....	48
2.3.3. Machine learning modelling.....	49
2.3.4. Model evaluation	50
2.3.5. Interpretation of machine learning models.....	52
2.3.6. Distance ratio.....	54
2.4. Chapter 3	54
2.4.1. Growth dynamics.....	54
2.4.2. Bright-field imaging	56
2.5. Chapter 4	57
2.5.1. <i>gyrA</i> spontaneous mutants	57
2.5.2. Time kill curves	57
2.5.3. High-content imaging.....	58
2.5.4. Imaging analysis	60
2.6. Chapter 5 and Chapter 6.....	61
2.6.1. Intact protein analysis with MALDI-TOF MS	61
2.6.2. High-content imaging with Opera Phenix	65
2.6.3. Integrating imaging and omics to infer antimicrobial susceptibility: method development	67
2.6.4. Molecular modelling and gene network analysis	69
Chapter 3. Genomic, physiological and morphological characterisation of pathogenic <i>Escherichia coli</i> to study antimicrobial resistance	70
3.1. Introduction	70
3.2. Resistome and morphology associated with AMR	71
3.3. Assessing the growth dynamics of <i>E. coli</i> under exposure to antimicrobials	77
3.3.1. Effects of azithromycin and tetracycline resistance on growth dynamics of <i>E. coli</i>	84
3.3.2. Elevation of growth rate in <i>ermB</i> ⁻ <i>E. coli</i> at sub-MIC concentrations of azithromycin.....	87
3.4. Discussion	89

Chapter 4. Application of machine learning with high-content imaging to infer antimicrobial resistance phenotypes in <i>Salmonella</i> Typhimurium	93
4.1. Introduction	93
4.2. <i>S. Typhimurium</i> isolates display variable growth and morphological characteristics following ciprofloxacin exposure	94
4.3. Development of a random forest classifier to discriminate between isolates and between growth conditions	100
4.4. <i>S. Typhimurium</i> SL1344 and its <i>gyrA</i> mutant display distinct differences in morphological features	106
4.5. Ciprofloxacin susceptible and resistant <i>S. Typhimurium</i> isolates can be distinguished by morphological features	109
4.6. Random forest classifier can distinguish between ciprofloxacin susceptible and non-susceptible isolates without ciprofloxacin exposure	112
4.7. Discussion	116
Chapter 5. Integrating Imaging and Omics with Machine Learning Approaches to Infer Antimicrobial Susceptibility in Gram-negative Bacteria: Method Development	120
5.1. Introduction	120
5.2. Intact protein analysis with MALDI-TOF MS	124
5.2.1. Assessing sample preparation methods	124
5.2.2. Assessing matrix solution and deposit methods	126
5.3. High-content imaging	128
5.3.1. Fluorescent staining	128
5.3.2. Image processing	135
5.4. Integrating imaging with omics	139
Chapter 6. Integrating Imaging and Omics with Machine Learning Approaches to Infer Antimicrobial Susceptibility in Gram-negative Bacteria: Results	147
6.1. Introduction	147
6.2. Phylogeny and resistome analysis	147
6.3. Machine learning to explore the association of proteomics and cellomics with AMR	152

6.4. Proteomic and cellomic analysis with machine learning reveals modifications in protein expression and plasma membrane.....	159
6.5. Discussion	167
Chapter 7. Outlook	171
References	174
Appendix	190

List of figures

Figure 1-1. Chemical structures of historical-milestone antimicrobial.	3
Figure 1-2. Introduction of new antimicrobial classes to clinical practice and first resistance report.	9
Figure 1-3. Antimicrobial modes of action and resistant mechanisms in Gram-negative bacteria.....	13
Figure 1-4. Nomenclature of <i>Salmonella</i> genus	23
Figure 1-5. Overview of AST methods.....	26
Figure 1-6. NGS chemistry overview	29
Figure 1-7. Configuration of a MALDI-TOF MS.	31
Figure 1-8. Overview of microscopic techniques.	34
Figure 1-9. Bacterial cells under examination with various types of microscopes.	36
Figure 2-1. Plate layout for HCI and intact protein analysis of pathogenic <i>E. coli</i> in one batch of experiment	68
Figure 3-1. Bacterial morphology associated with AMR and resistome.	75
Figure 3-2. Three replicates of growth dynamics monitoring	78
Figure 3-3. Comparison between performance of Gompertz's, Baranyi's and three-phase linear model.	80
Figure 3-4. The dynamics of the growth features	82
Figure 3-5. A comparison between the growth features for susceptible isolates and those for resistant isolates.	83

Figure 3-6. A comparison between growth features of resistant and susceptible isolates under exposure to azithromycin and tetracycline	85
Figure 3-7. Comparison between growth parameters estimated from Gompertz's, Baranyi's and three- phase linear model.	86
Figure 3-8. The impact of <i>ermB</i> gene on growth dynamics of azithromycin-resistant <i>E. coli</i>	88
Figure 4-1. <i>S. Typhimurium</i> responses to ciprofloxacin treatment	96
Figure 4-2. Morphological features associated with ciprofloxacin treatments.....	101
Figure 4-3. Random forest model to interpret the association between morphology and antimicrobial exposure.....	103
Figure 4-4. Morphologies of (a) D23580 and (b) VNS20081 over the 24-h exposure to CIP at different concentrations	105
Figure 4-5. Morphological distinction between SL1344 and SL1344gyrA	108
Figure 4-6. Distinguish resistant and susceptible isolates at 0xMIC-22h.....	111
Figure 4-7. Machine learning classifiers to identify resistant isolates at 0xMIC-22h	114
Figure 5-1. Integrating intact protein analysis by MALDI-TOF with HCl by Opera Phenix	123
Figure 5-2. Comparing sample preparation methods.....	125
Figure 5-3. Evaluating different combinations of matrix solutions and target plate coatings.	127
Figure 5-4. Fluorescent staining and cell immobilisation method.....	129
Figure 5-5. Comparing the proposed 'sandwich' method and Sridha's method for fluorescent staining and cell immobilisation	130
Figure 5-6. Evaluating the 'sandwich' method.....	133

Figure 5-7. Image analysis workflow	136
Figure 5-8. Compare the proposed image analysis workflow with Sridha's workflow	138
Figure 5-9. Resistant <i>E. coli</i> segregated from susceptible ones with increased exposure time to different antimicrobials.....	140
Figure 5-10. Standardisation of controls.....	142
Figure 5-11. Effects of batch-correction.....	144
Figure 6-1. Phylogenetic structure of the 70 <i>E. coli</i> isolates with respective sequence type, source of isolation, susceptibility, and resistome profiles.	148
Figure 6-2. Bloodborne and stool isolates behaved differently under exposure to CIP	151
Figure 6-3. Performance of various machine learning models to predict AMR.....	156
Figure 6-4. Performance of various machine learning models to predict MICs.....	158
Figure 6-5. Interpretation of the random forest model to predict resistance to CIP	162
Figure 6-6. Gene network analysis	165

List of tables

Table 1-1. Overview on applications of HCI to study antimicrobials and resistant microorganisms.....	39
Table 2-1. Photophysical properties of fluorescence dyes/stains used in chapter 4, 5, and 6 .	60
Table 3-1. MICs and antimicrobial resistance of ATCC 25922 and 27 clinical isolates used in this study	72
Table 4-1. Resistome analysis of the isolates in relation to fluoroquinolone susceptibility....	98

Abbreviations

λ	lag time (h)
μ	maximum growth rate (h^{-1})
A	maximum cell density
AMR	antimicrobial resistance
AST	antimicrobial susceptibility testing
AUC	area under the curve
AZI	azithromycin
BSA	bovine serum albumin
CAN	acetonitrile
CHCA	α -Cyano-4-hydrocinnamic acid
CIP	ciprofloxacin
CLSI	Clinical and Laboratory Standards Institute
CRO	ceftriaxone
DAPI	4',6-diamidino-2-phenylindole
DIC	differential-interference-contrast
<i>E. coli</i>	<i>Escherichia coli</i>
CN	gentamicin
HCI	high-content imaging
ICE	individual conditional expectation
KNN	k-neighbour nearest

LB	Luria-Bertani
LMICs	low- and middle-income countries
LOESS	Locally estimated scatterplot smoothing
MAE	mean absolute error
MALDI-TOF MS	Matrix-Assisted Laser Desorption/Ionisation – time-of-flight mass spectrometry
MATE	multi drugs and toxic compound extrusion
MC	MacConkey
MDR	multi-drug resistant
MFS	major facilitator superfamily
MH	Mueller-Hinton
MHB2	cation-adjusted MH broth
MIC	minimum inhibitory concentration
MoA	mode of action
MSE	mean square error
NGS	next generation sequencing
PCR	polymerase chain reaction
PCA	principle components analysis
PCoA	principle coordinates analysis
PDP	partial dependence plot
PFA	paraformaldehyde

PMQR	plasmid-mediated quinolones resistance
QRDR	quinolone resistance determining region
R^2	correlation of determination
ρ	correlation coefficient
RND	resistance-nodulation-division
RPPs	ribosomal protection protein
<i>S. Typhimurium</i>	<i>Salmonella enterica</i> serovar Typhimurium
SA	sinapinic acid
SHAP	Shapley additive explanations
ST	sequence type
SVM	support vector machine
TE	tetracycline
TFA	trifluoroacetic acid
TIRF	total internal reflection fluorescent
WGS	whole-genome sequencing

Chapter 1. Introduction and background

All the work in this chapter was performed by me. A part of this was published in collaboration with Caroline Zellmer and Dr. Sushmita Shridha as the following.

Zellmer, C., **Tran, T. A.** & Sridhar, S. Seeing the bigger picture. Nat. Rev. Microbiol. 19, 745 (2021).

1.1. A brief history of antimicrobials

Antibiotic is “a compound made by a microbe to destroy other microbes”¹. To a broader meaning, antimicrobials are defined as a special group of chemotherapeutic agents used to treat diseases caused by microbes². Thus, antibiotics are, by definition, included in antimicrobials. Resistance is the ability of a microorganism to survive under exposure to antimicrobials. Multi-drug resistance (MDR) is defined as resistance to at least three different classes of antimicrobial³.

Although the discovery of penicillin initiated the antimicrobial era, the drug was not the first antimicrobial to be discovered. Indeed, ‘antibiose’ (the origin of the term ‘antibiotic’ [anti-life]) was first used by Paul Vuillemin in 1890, as an antonym to symbiosis to describe antagonism among microorganisms⁴. Three years later, in 1893, the first true antibiotic (mycophenolic, Figure 1-1, compound 1) was isolated from *Penicillium glaucum* (now *P. brevicompactum*) by the Italian physician and microbiologist Bartolomeo Gosio⁵. In 1907, the first antimicrobial used for human health - arsphenamine (Salvarsan, Figure 1-1, compound 2), was synthesised in the laboratories of Paul Ehrlich, who latterly won the 1908 Nobel Prize in Physiology or Medicine. However, the first mainstream antibiotic was

penicillin⁶. Chemical structures of other historical-milestone antimicrobials were described in Figure 1-1.

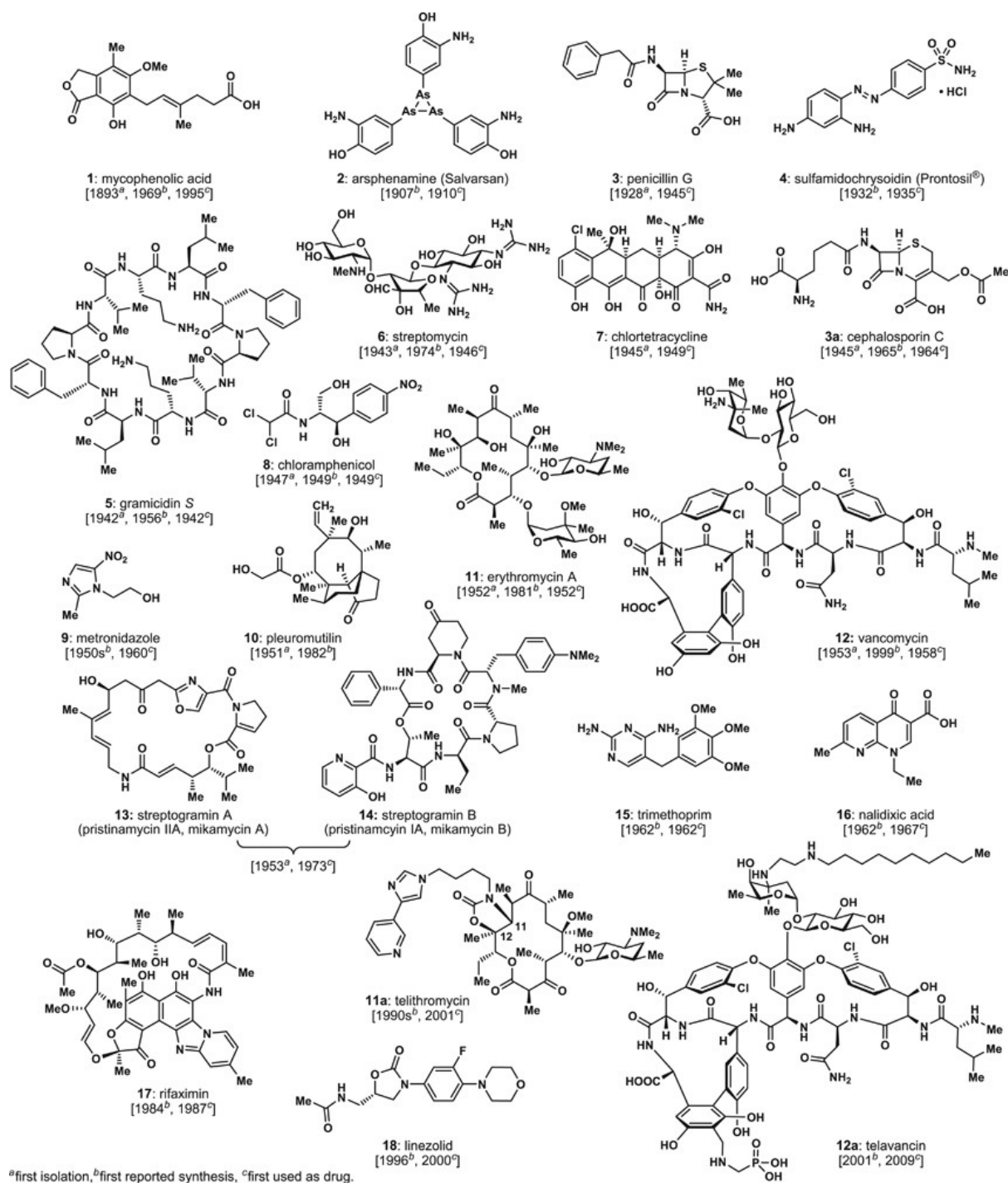


Figure 1-1. Chemical structures of historical-milestone antimicrobial.

For every antimicrobial, a chemical structure was provided with the years of first isolation^a, first reported synthesis^b, and first used as drug^c. Adopted from Nicolaou et al.⁶

In September 1928, returning to his laboratory from a summer vacation, Alexander Fleming observed a plate of *Staphylococcus aureus* (*S. aureus*), left on the bench, contaminated with the fungus *Penicillium notatum* (now *P. chrysogenum*). Further investigating the mould, he discovered that the fungus excreted a substance that killed the bacteria⁶. In March 1929, he named the substance penicillin⁷. Early 1941, three scientists at the University of Oxford, Howard Walter Florey, Ernst Boris Chain, and Norman Heatley, isolated penicillin from fungus and successfully tested it on a mouse model⁶. During World War II, an Anglo-American project was launched to produce sufficient amounts of penicillin to prevent death caused by wound infections amongst the allied forces. In 1945, using X-ray crystallography, Dorothy Crowfoot Hodgkin solved the structure of penicillin⁸ (Figure 1-1, compound 3), enabling the development of semi-synthetic derivatives of the drug⁹. In 1945, Alexander Fleming, Ernst Boris Chain and Howard Walter Florey were ultimately jointly awarded the Nobel Prize in Physiology or Medicine for the discovery and characterisation of penicillin.

Soon after penicillin became mainstream in World War II, Albert Schatz and Selman Waksman discovered streptomycin (Figure 1-1, compound 6), the first aminoglycoside, which was isolated from *Streptomyces griseus*^{10,11}. The drug was proven to be effective against many Gram-negative bacteria as well as *Mycobacterium tuberculosis*. Streptomycin was first used to successfully treat tuberculosis in 1944¹²; Waksman was awarded a Nobel Prize in Physiology or Medicine in 1952. Subsequently, other aminoglycosides have been discovered and developed from *Streptomyces* spp. and *Micromonospora* spp. Antimicrobial agents isolated from the *Streptomyces* group are known as “mycin” aminoglycosides, e.g., neomycin, kanamycin, and tobramycin, whereas the agents isolated from *Micromonospora* group are known as the “micin” aminoglycosides, e.g., gentamicin and sisomicin¹.

In the same year that Alexander Fleming won the Nobel Prize, the first member of the cephalosporin family (a subclass of β -lactams) was first discovered from *Cephalosporium acremonium* (now *Acremonium chrysogenum*) by Giuseppe Brotsu¹³. Unable to attract resources from the Italian pharmaceutical industry, Brotsu decided to send his fungal cultures to the University of Oxford. At the Sir William Duncan School of Pathology, Edward P. Abraham and Guy Newton succeeded in isolating the compound from Brotsu's fungus and named it cephalosporin C (Figure 1-1, compound 3a) in 1955¹⁴. Although cephalosporin C was not approved for clinical use, its structure inspired the development of similar analogues thereafter. As of today, four generations of cephalosporin have been synthesised and introduced to the market, which have broad (extended spectrum) activity against Gram-negative bacteria, but are less effective against Gram-positive microbes¹⁵. The third-generation agents are widely used to treat both hospital-acquired and community infections. Specifically, the third-generation cephalosporins, e.g., ceftriaxone, are well known to penetrate the blood-brain barrier, making them useful for treating meningitis caused by pneumococci, meningococci, *Haemophilus influenzae* (*H. influenzae*), *Escherichia coli* (*E. coli*), *Klebsiella*, and penicillin-resistant *Neisseria gonorrhoeae* (*N. gonorrhoeae*)¹⁵.

In 1945, the American plant physiologist Benjamin M. Duggar isolated chlortetracycline (Figure 1-1, compound 7), the first member of the tetracycline family, from *Streptomyces aureofaciens* (*Strep. aureofaciens*)¹⁶. Soon after, Alexander C. Finlay and his co-workers, at Pfizer (New York, NY, USA), isolated several other members of the tetracycline family, including oxytetracycline (from *Strep. rimous*), demethylchlortetracycline (from *Strep. aureofaciens*) and tetracycline (TE, from *Strep. aureofaciens*)⁶. Other tetracyclines were produced using semisynthetic approaches, which include methacycline, doxycycline, and minocycline. With no major adverse side effects, tetracyclines are commonly prescribed as

broad-spectrum antimicrobials to treat both Gram-negative and Gram-positive bacterial infections in both human and animals. Indeed, these agents have widely been used in veterinary medicine and growth promotion in farmed animals, as tetracyclines are among the cheapest antimicrobials available. The USA licensed chlortetracycline and oxytetracycline as growth promoters for animals^{17,18}; therefore, the country became the biggest consumer of tetracycline antimicrobials worldwide with approximately 3,200 tons per year¹⁹.

In 1950, five years after the discovery of chlortetracycline, pikromycin, the first member of the macrolide family, was isolated from a *Streptomyces* strain²⁰. The agent was named after the ancient Greek word pikro, meaning bitter, due to its distinctive taste. Two years later, Eli Lilly (Indianapolis, IN, USA) isolated erythromycin (Figure 1-1, compound 11) from *Saccharopolyspora erythraea*. The bacterium was found in a soil sample imported from Philippines to the USA. The macrolides are classified based on the size of their macrocyclic lactone ring, including 12-membered ring (e.g., methymycin) 14-membered ring (e.g., pikromycin, erythromycin, clarithromycin and flurithromycin), 15-membered ring (e.g., azithromycin), and 16-membered ring (e.g., carbomycin, niddamycin, and tylosin). Additionally, the 15-membered macrolide, azithromycin (AZI), also displayed anti-inflammatory properties, which have recently been tested to treat viral infections²¹.

After the discovery and development of macrolides in the 1950s, the 1960s witnessed the birth of one of the most important classes of antimicrobial, the quinolones. After investigating an impurity within the chemical manufacture of chloroquine (an antimalarial agent), in 1962, George Y. Leshner and colleagues (at Sterling-Winthrop laboratories, Rochester, NY, USA) discovered nalidixic acid (Figure 1-1, compound 16), the first member of the class quinolone. However, a major ocular toxicity was reported soon after the agent the

approval for clinical use²². For this reason, the quinolones did not get as much attention until the development of fluoroquinolones in the 1950s and in the 1960s. Compared to its ascendant, the fluoroquinolones possess extended spectrum against both Gram-negative and Gram-positive microbes and possess superior pharmacokinetics with less severe side effects. In two decades, from a relatively small sale, the (fluoro)quinolones have become major class of antimicrobials with a global market of \$3.04 billion in 1997²³.

Referring back to Waksman's work, by systematically investigating microbes to isolate antibiotics, he initiated the Golden Age of antibiotic discovery from the 1940s to the 1960s²⁴. During this period, new classes of antibiotics were discovered and developed nearly every year to treat bacteria resistant to previous classes²⁴. However, over the last 40 years, no new class of antimicrobial has been developed and introduced to the market (Figure 1-2). Indeed, harvesting all 'low-hanging-fruits' largely exhausted the antimicrobial discovery pipeline²⁴, leading to greater difficulty to discover novel targets for antimicrobial development. Along with this, antimicrobial development faces certain obstacles²⁵ compared to other drug development.

There are three main obstacles. i) Resistance against developing antimicrobials emerges rapidly²⁵. Since the majority of the previous-generation antimicrobials were complex natural or semi-synthesised products with multiple binding targets, they are less likely to select for resistance²⁵. On the other hand, newly developed antimicrobials are often small molecules directly binding to protein targets, which had them selected for resistance rapidly^{26,27}. Moreover, for that reason, new generations of antimicrobials are often reserved to treat bacteria resistant to previous antimicrobials, which limits the sales and profits of drug developers. ii) Novel antimicrobial targets exhibit an undesirably narrow spectrum, e.g., only

effective against specific genera, species or even organisms²⁵. Compared to broad-spectrum encounters, narrow-spectrum antimicrobials are considered more beneficial for patients as they display less disturbing impacts on human microbiome. However, that advantage also holds back the sales of narrow-spectrum antimicrobials since the drugs cannot be used to treat a wide range of infection. For example, fidaxomicin, a recently approved antimicrobial, is only effective against *Clostridium difficile*, which had the drug struggled to compete in the market with older broad-spectrum antimicrobial, such as vancomycin²⁵. iii) Marketed antimicrobials are required to possess high selectivity for bacteria to minimise toxicity on mammalian cells²⁵. That requirement remains a challenge for antimicrobial development, given that many antimicrobials have been discontinued due to reported adversarial effects, including the classic nalidixic acid²² and three antimicrobials, telithromycin²⁸, temafloxacin²⁹, and trovafloxacin³⁰, developed in the 1990s and early 2000s.

Therefore, for those reasons, the cost of developing a new antimicrobial exceeded \$3.12 billion in 2013³¹, having the pharmaceutical and agrochemical companies shut down their antimicrobial discovery programmes. In 2015, Merck closed its antimicrobial arm of Cubist, a subsidiary of the company²⁴. Later in 2017, Sanofi ended its collaborative programme, with Warp Drive Bio, developing cancer therapies and antimicrobial using genomic mining²⁴. Altogether, this loss in expertise has limited the options of antimicrobial to treat infections with AMR and life-threatening MDR bacteria.

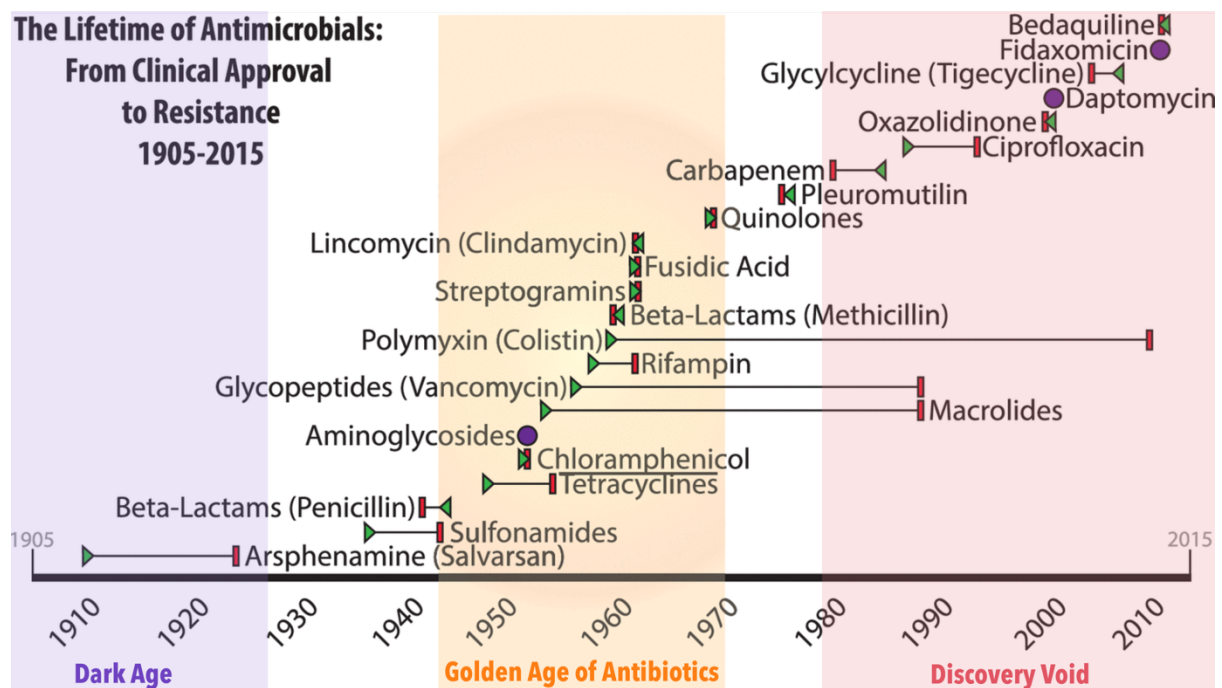


Figure 1-2. Introduction of new antimicrobial classes to clinical practice and first resistance report.

Green triangles represent the years of first clinical use. Red rectangles indicate first resistant identification. Black lines connecting green triangles and red rectangles display resistance-free lifespan of each antimicrobial class. Adopted from Spagnolo et al.³².

1.2. Antimicrobial resistance

Antibiotic synthesis pathways have played an important role in microbial evolution for millions of years^{33,34}. These natural products are hypothesised to act as chemical weapons to kill competitors as well as signalling molecules, interacting with other microbes or eukaryotic hosts^{35–37}. To counteract the impacts of antibiotics, bacterial communities have co-evolved and incorporated antimicrobial resistance (AMR) factors into their genomes³⁸. These factors are called antimicrobial resistance genes (ARGs) and all ARGs in both pathogenic and non-pathogenic bacteria are defined as “resistome”. In fact, metagenomic analysis of ancient DNA concluded that ARGs existed long before the wide use of antimicrobials³⁹. Reinforcing this conclusion, by retrospectively investigating the bacterial collection curated by E. G. D. Murray from 1927 to 1954, a genomic study found that ARGs were circulating within bacterial communities prior the discovery of penicillin^{40–42}.

Resistance to penicillin was identified within a few years before approval of the drug for clinical use⁴³. Two years after its first clinical use to treat typhoid fever, a *Salmonella* Typhi isolate resistant to chloramphenicol was first identified in 1950⁴⁴. Soon thereafter, more resistant microbes were reported within a few years post clinical approvals of new antimicrobial classes, e.g., streptogramins, fusidic acid, and quinolones (Figure 1-2). For aminoglycosides, oxazolidinone, and daptomycin, resistance was reported in the same year of market introduction (Figure 1-2). Spagnolo and his colleagues suggested that human abuse and misuse of antimicrobials was one of the major contributions to the wide spread of AMR³². The father of penicillin, Alexander Fleming, spoke to the New York Times about potential threats, and warned the public about antimicrobial abuses and resistance⁴⁵.

In fact, the more we use antimicrobials, the more resistance is facilitated ^{46–48}. For instance, a 50-year retrospective study in China found that from the 1970s to 2010s the prevalence of ampicillin-resistant *E. coli* almost tripled from 24.39% to 72.86%, TE resistance rose from 48.67% to 79.78%, and resistance to cephalosporins soared from 1.22% to 52.24%⁴⁹. In Africa, the incidence rate of ciprofloxacin-resistant *Enterobacteriaceae* spp. escalated from about 20% to more than 60% in 14 years (2001-2015)⁵⁰. Moreover, from 2016 to 2020, the proportion of *E. coli* resistant to carbapenems has doubled in Europe. Note that carbapenems are one of the last resorts to treat MDR bacteria⁵¹. In Norway, the prevalence of MDR in *S. Typhimurium* increased 84% from 1998 to 2000⁵², whilst in the US, the MDR prevalence increased from 0.6% (1979) to 34% (1996)⁵³. Furthermore, from 2003 to 2006, the prevalence of MDR *S. Typhimurium* isolated in the UK increased 32%⁵⁴. In China, the proportion of *S. Typhimurium* carrying beta-lactam resistance genes raised from 25.93% in 2006-2009 to 68.71% in 2014-2017⁵⁵. Altogether, humanity has witnessed the colossally wide spread of antimicrobial resistance across the globe in recent decades⁵⁶.

Indisputably, the soaring trend of antimicrobial resistance is unignorable. But what makes antimicrobial resistance such a concerning and problematic subject for the World Health Organisation (WHO) to name it as one of the major threats for global health?⁵⁷ The answer lies in the looming storm. A systematic review published in 2019 estimated that approximately 1.27 million deaths were attributable to AMR infections worldwide⁵⁸. Particularly, there were 33,000 deaths caused by resistant bacteria in Europe⁵⁹ and 23,000 deaths in the USA⁶⁰. According to the WHO, the hospitalisation costs for patients infected with AMR bacteria are significantly higher than that for patients infected with susceptible organisms⁶¹. The economic cost of AMR infections in Europe was estimated to be around 1.1 billion Euros⁵¹. In a broader view, without efficient intervention, by 2050, AMR was

predicted to directly or associatively cause 10 million deaths worldwide and will cost the global economy £55 trillion to treat infection with AMR pathogens⁶².

1.3. Antimicrobial modes of action and resistance mechanisms

Over millions of years of evolution, certain microorganisms, including bacteria, have developed antibiotics to inhibit and kill other bacteria based on five major modes of action (MoA). These are, (A) inhibition of cell wall synthesis (e.g. β -lactams), (B) inhibition protein synthesis (e.g. macrolides and tetracyclines), (C) metabolic antagonist action (e.g. trimethoprim), (D) inhibition of nucleic acid synthesis (e.g. fluoroquinolones), and (E) impairment of cytoplasmic membrane (e.g. daptomycin) (Figure 1-3). Synthesis and semi-synthesis of further commercial antimicrobials were inspired by those five MoA. To counteract the effects of antimicrobials, bacteria have consequently evolved four broad resistance mechanisms: these are (1) enzymic inactivation of the antimicrobial, (2) the upregulation of efflux pumps, (3) mutations of or downregulation of porins, and (4) alteration of the antimicrobial target (Figure 1-3).

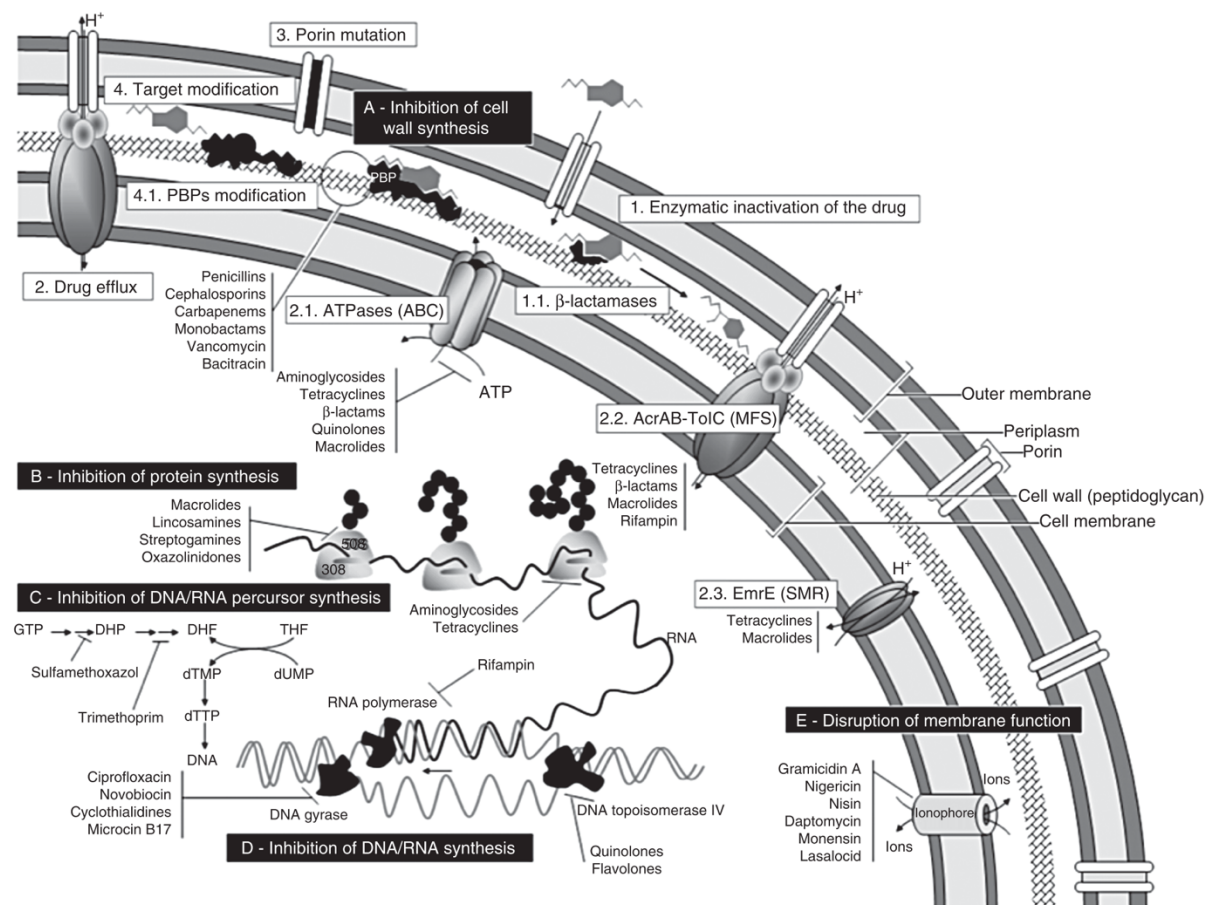


Figure 1-3. Antimicrobial modes of action and resistant mechanisms in Gram-negative bacteria

The figure illustrated a part of a bacterial cell, including membrane, cell wall, and cytoplasm. Black boxes represented modes of action and white boxes showed resistance mechanisms.

Adopted from Fernandes et al.⁶³

1.3.1. Mechanisms of resistance to (fluoro)quinolones

1.3.1.1. Alteration of antimicrobial targets

Gram-negative bacteria acquire high-level of resistance to (fluoro)quinolones by the accumulation of mutations within the quinolone resistance determining regions (QRDRs) in the genes encoding the type II topoisomerases (GyrA, GyrB, ParC, ParE)⁶⁴. Mutations in QRDRs alter the structures of antimicrobial targets, decreasing the binding affinity of the drugs, and thus confer resistance to (fluoro)quinolones^{65,66}. Mutations in *gyrA* and *parC* are more commonly found, compared to *gyrB* and *parE*⁶⁷. For *E. coli*, the Ser83 mutation site in *gyrA* was most frequently found, followed by Asp87 on the same target protein⁶⁵. Bacteria acquire higher levels of (fluoro)quinolone resistance with increased higher number of mutations on the QRDRs. Typically, a single mutation, on DNA gyrase or topoisomerase IV, confers eight- to 16-fold increase in resistance level, compared to the wild-type^{65,67,68}. Meanwhile, mutation in both targets results in greater resistance levels (10-100 times)^{69,70}.

1.3.1.2. Reduced antimicrobial accumulation

As a hydrophilic compound, the (fluoro)quinolones require the aid of porin channels to enter the intracellular environment of Gram-negative bacteria^{64-67,71,72}. Certain mutations can result in either loss or downregulation of porin channels, such those occurring in genes encoding OmpA, OmpC, OmpD, and OmpF. The loss and downregulation can lead to a reduction in intracellular concentration of (fluoro)quinolones and consequently confer resistance. OmpX has been known as an inhibitor of porin channels. Thus, overexpression of OmpX could downregulate OmpC, OmpD, OmpF, LamB, and the Txs porins^{67,71-73}. The downregulation confers resistance to a wide range of antimicrobials including (fluoro)quinolones. Besides porin channels, (fluoro)quinolones can also intrude bacterial cells via the lipid-mediated

pathway. Hence, mutations leading to modification of lipopolysaccharide (LPS) structures increases the resistance level to (fluoro)quinolones⁶⁴.

In concert with porin downregulation and LPS modification, overexpression of efflux pumps enhances the resistance level to (fluoro)quinolone and other antimicrobials^{64,66,67}. For *E. coli*, the resistance-nodulation-division (RND) family is the primary involved efflux pumps, contributing to resistance against (fluoro)quinolones. A typical structure of RND pumps, includes an inner membrane pump protein, an outer membrane channel protein, and a periplasmic-membrane fusion protein (e.g. AcrAD-TolC)^{67,68,72}. Additionally, a few members of MFS (Major Facilitator Superfamily, e.g. EmrAB-TolC and MdfA) and MATE (Multi Drugs and Toxic Compound Extrusion, e.g. NorE and YdhE) also contribute to (fluoro)quinolone resistance^{67,68,72}. Compared to mutations on target proteins (e.g. GyrA and ParC), mutations leading to downregulation of porin channels and overexpression of efflux pumps confer lower resistance level (typically four- to eight-fold)⁶⁴.

1.3.1.3. The Plasmid-mediated quinolones resistance (PMQR) genes

The primary PMQR mechanism is protecting the type II topoisomerases from (fluoro)quinolone inhibition with Qnr proteins⁶⁴. The secondary PMQR mechanism is modifying (fluoro)quinolones and thus, decreasing the antimicrobial activities. Mutations on plasmid-encoded aminoglycoside-modifying acetyltransferase AAC(6')-Ib-cr confer resistance by acylation of the unsubstituted nitrogen of the C7 piperazine on (fluoro)quinolone^{64,68}. The third mechanism of PMQR is to extrude (fluoro)quinolone from bacterial cytoplasm with efflux pumps QepA and OqxAB encoded on plasmids. QepA belongs to the MFS family which actively pumps out hydrophilic fluoroquinolones like

ciprofloxacin (CIP) and norfloxacin. OqxAB is a part of the RND family and works on a wide range of antimicrobials beyond (fluoro)quinolone, including tetracyclines, chloramphenicols, and trimethoprim^{64,74}. Generally, PMQR genes alone confer a low level of resistance to (fluoro)quinolones, which generally does not reach clinical breakpoints⁶⁴.

1.3.2. Mechanisms of resistance to macrolides

1.3.2.1. Alteration of antimicrobial targets

The ribosome is a macromolecular biological complex, comprising of 65% rRNA and 35% ribosomal proteins⁷⁵. Therefore, mutations in genes encoding both the rRNA and ribosomal proteins can confer resistance to macrolides. Studies on role of target alteration in development of macrolide resistance have been mainly facilitated by mutagenesis or spontaneous mutants *in vitro*⁷⁶. Base substitutions in the 23S rRNA (encoded on *rrn* operon) have been reported to increase resistance to macrolides⁷⁷. Since bacteria possess a varied number of *rrn* copies on their chromosome, the effect of substitution on 23S rRNA is highly species dependent⁷⁶. For the Enterobacteriaceae, the species have up to eight *rrn* loci^{78,79}. Thus, the mechanism is less relevant to the development of macrolide resistance in those species⁷⁶. For alteration of ribosomal proteins, mutations on L4 (encoded by *rplD* gene) and L22 (encoded by *rplV* gene) have been shown to contribute to macrolide resistance⁷⁶. The mutations have been hypothesised to change the structure of ribosome complex and thus, to complicate the access of macrolides to their binding site. Alteration of amino acid at positions Q62, K63, G64, or G66 on L4, as well as K90 on L22 have mainly been found to reduce the binding of macrolides to ribosome^{80,81}. Insertion and deletion on both L4 and L22 have been reported to be involved in macrolides resistance. In *E. coli*, four insertion sites on L4 and L22, eight deletion sites on L4, and five deletion sites on L22 were described to confer resistance to macrolides⁷⁶.

Methylation of the post-transcriptional 23S rRNA can also confer resistance to macrolides. In the presence of methyltransferase, methyl groups are covalently added to the 23S rRNA, preventing the macrolides from interacting with the ribosome⁷⁶. Among the methyltransferases, adenine N-6-methyltransferase (encoded in *erm* genes) are the most extensively studied. The N-6-methyltransferase confer resistance to macrolides by adding one or two methyl groups to an adenine of post-transcribed 23S rRNA, mainly at position A2058⁸². Within the Enterobacteriaceae, at least six variants of the *erm* gene, including *erm(B)*, *erm(C)*, *erm(D)*, *erm(E)*, *erm(F)*, and *erm(42)*, have been described on the bacterial chromosome^{83,84}. These *erm* genes have also been found in mobile elements. In *E. coli*, *erm(B)* and *erm(C)* are encoded on pTN48 plasmid along with another macrolide resistance determinant, the *mphA* gene⁸⁵.

1.3.2.2. Reduced antimicrobial accumulation

Reduced antimicrobial accumulation can also induce resistance to macrolides, in a manner comparable to the (fluoro)quinolones and can be mediated by extrusion of the antimicrobials with chromosomal efflux pumps⁷⁶. However, due to specific membrane permeation, basal efflux pump activities, and natural binding to ribosome, the effects of efflux pumps on macrolide resistance can only be observed indirectly⁸⁶. For the Enterobacteriaceae, the AcrAB-TolC, a RND efflux pump, is the most well studied transport systems. Previous studies on *E. coli* have shown that inhibition of RND pumps by PA β N resulted in four- to 16-fold decrease in MIC⁸⁷. Another study on *S. Typhimurium* discovered that inactivation of *arcB* gene led to 64-fold decrease in MIC against erythromycin⁸⁷. In accordance with this finding, Wehmier and his colleagues observed that disruption of *arcB* reduced MIC against erythromycin, clarithromycin, and azithromycin by 128-fold⁸⁸. Point mutations on genes

encoding for efflux pump components can also disrupt the activity of the pump and result in a reduction in the MIC⁷⁶.

In addition to the chromosome, efflux pump genes can also be carried on mobile elements, such as plasmids. To date, five transferable efflux pumps have been reported in the Enterobacteriaceae. Of those, *msr(A)*, *msr(D)*, and *msr(E)* belong to the ABC family, while *mef(A)* and *mef(B)* belong to the MF family^{76,83,89}. At present, *mef(A)* is the most identified transferable efflux-pumps-gene that confers resistance to macrolides⁹⁰. The *mef(B)* gene, which is 30% identical to *mef(A)*, has only been found in *E. coli* isolated from porcine samples⁹¹. Regarding *Msr(D)*, the efflux pump itself does not extrude macrolides molecules, but expresses enhanced pump out activity in the presence of *Mef(A)*⁸⁹. Note that the *msr(D)* and *mef(A)* gene are regulated under by the same operon. Like *msr(D)* and *mef(A)*, the *msr(E)* gene is associated with *mph(E)* as part of a common operon⁹².

1.3.3. Mechanisms of resistance to tetracyclines

1.3.3.1. Alteration of antimicrobial targets

Comparable to macrolide resistance, resistance to tetracyclines can be conferred through point mutations in genes coding the rRNA. This type of resistance is usually found in bacteria with low rRNA gene copy numbers, e.g. *Propionibacterium acnes* (2-3 copies), *Helicobacter pylori* (1-2 copies), and *Mycoplasma bovis* (1-2 copies)⁹³. In contrast, since genes coding for ribosomal proteins are in single copy, mutations in these genes can confer resistance to tetracyclines. For example, *in vitro* studies found that mutations on *rpsJ* can confer resistance in *E. coli*⁹⁴ and *K. pneumoniae*⁹⁵. The gene *rpsJ* encodes for the 30S ribosomal subunit protein S10.

1.3.3.2. Target protection and enzymatic inactivation

The tetracycline ribosomal protection proteins (RPPs) can catalyse the GTP-dependent release of tetracycline from ribosome^{96,97}. Tet(O) and Tet(M) are the two most well studied RPPs, which are commonly found in both Gram-positive and Gram-negative bacteria⁹⁸. Plasmids carrying *tet(X)*, encoding for flavin-dependent monooxygenase, can also confer tetracycline resistance. This gene was originally identified in *Bacteroides*, but has become widespread to Enterobacteriaceae due to the conjugative nature of the carrying plasmids⁹³.

1.3.3.3. Reduced antimicrobial accumulation

Tetracycline-specific efflux pumps extrude intracellular tetracyclines to extracellular environment. These pumps mostly belong to the MFS and can be classified into seven groups⁹³. While *tet(A)* and *tet(B)* of group 1 are commonly found in Gram-negative bacteria, group 2 pumps including *tet(K)* and *tet(L)* are mostly found in Gram-positive bacteria⁹³.

1.4. The Enterobacteriaceae family

Rahn, in his publication in 1936, proposed a classification system to group bacteria species of similar phenotypic traits⁹⁹. Afterwards, the family Enterobacteriaceae was later coined to host a single genus (*Enterobacter*) of 112 species^{100,101}. Since then, the classification has been modified several times and more species have been formally included to the family. As of June 2014, the family Enterobacteriaceae consists of 23 genera and more than 210 species¹⁰² which can be found in various habitats including water, soil vegetation, and warm-blood animals. All members of the family are Gram-negative, nonsporulating, rod-shape (bacilli), facultative aerobic, either motile or nonmotile, oxidase-negative (except *Plesiomonas*

shigelloides), and catalase-positive. They utilise glucose as a carbon source via fermentation process and reduce nitrates to nitrites during respiration (except *Photorhabdus* and *Xenorhabdus*)¹⁰².

Through the history of microbiology, the Enterobacteriaceae members have been considered as amongst the most important families in the microbial world with a great impact on veterinaries, public health, and the global community^{100,103,104}. While many Enterobacteriaceae species are non-pathogenic as a part of normal human gut microbiota (commensal microorganisms), certain species of the family are pathogenic and can cause both intestinal and extraintestinal diseases. Of these species, *K. pneumoniae* accounted for 6.2% isolates from adults with community-acquired pneumonia in Asia¹⁰⁵ and *Klebsiella* spp. accounted for 15.1% isolates from blood samples in Vietnam¹⁰⁶. In Africa, approximately 11.7% isolates from urine samples of patient with urinary tract infection were *Proteus* spp.¹⁰⁷. *Salmonella* spp. was detected in 115 of 657 stool samples (17.5%) from children with gastroenteritis in Taiwan¹⁰⁸.

1.4.1. *Escherichia coli*

E. coli is a member of the family Enterobacteriaceae, and a part of normal gut microbiota of humans along with other common species including *Lactobacillus*, *Clostridium*, *Bacteroides*, *Bifidobacterium*, *Fusobacteria*, and *Ruminococcus*, etc.¹⁰⁹. *E.coli* displays all characteristics of a typical Enterobacteriaceae bacterium and possesses specific bacteriological features, including positive ortho-nitrophenyl- β -D-galactoside reaction, positive methyl-red reaction, negative Voges-Proskauer reaction, negative phenylalanine-deaminase reaction, negative Simmons' reaction, indole production, hydrogen sulphide production¹¹⁰. While commensal *E. coli* are considered harmless, other variants are pathogens and can be broadly classified into two main groups: (1) gastrointestinal pathogens including enteropathogenic *E. coli* (EPEC),

enterotoxigenic *E. coli* (ETEC), enterohemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) and (2) extraintestinal pathogens including uropathogenic *E. coli* (UPEC), neonatal meningitis *E. coli* (NMEC), adherent invasive *E. coli* (AIEC), extra-intestinal *E. coli* (ExPEC)¹¹¹.

Epidemiologically, *E. coli* was found to be the most prevalent pathogen associated with bacteraemia in Asia (43.8%)¹⁰⁶ and urinary tract infections in Africa (41.9%)¹⁰⁷.

Diarrheagenic *E. coli* was responsible for 45.2% diarrheal cases in children under five years for which ETEC, EPEC, and EAEC accounted 14.6%, 12.8%, and 24.7%, respectively¹¹².

Particularly, ETEC caused 222.6 million diarrheal cases and 51,200 deaths globally in 2016¹¹³. The high prevalence and burden of *E. coli* has made it one of the most important pathogens belonging to the family Enterobacteriaceae and most well studied microbial species.

1.4.2. The *Salmonella* genus and *Salmonella* Typhimurium

The genus *Salmonella* was named after D.E. Salmon, an American bacteriologist, who isolated *Salmonella enterica* serotype Choleraesuis from pigs with hog cholera¹⁰⁰. The nomenclature of *Salmonella* has been a heated debate and has been updated several times through the ~100-year history of the genus. Currently, the genus comprises of two species, namely *Salmonella bongori* (*S. bongori*) and *Salmonella enterica* (*S. enterica*) (Figure 1-4). Of these species, *Salmonella enterica* has been classified into six subspecies falling into approximately 2,500 serotypes according to the Kauffmann-White scheme. These subspecies include (I) *S. enterica* subsp. *enterica*, (II) *S. enterica* subsp. *salamae*, (IIIa), *S.*

enterica subsp. *arizonae*, (IIIb) *S. enterica* subsp. *diarizonae*, (IV) *S.*

enterica subsp. *houtenae*, and (VI) *S. enterica* subsp. *indica*.

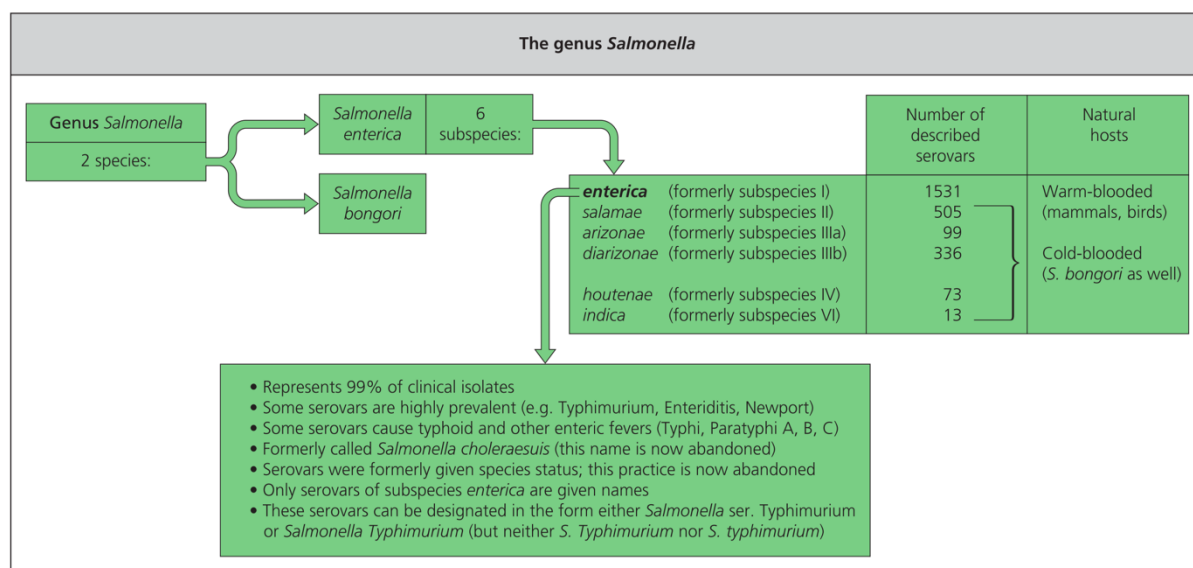


Figure 1-4. Nomenclature of *Salmonella* genus

The diagram described the nomenclature of *Salmonella* genus, which hosts two species of *S. enterica* and *S. bongori*. The *S. enterica* was sub-classified into six subspecies, of which *S. enterica* subsp. *enterica* belongs to more than 1,500 serovars.

As a genus in the family Enterobacteriaceae, *Salmonella* species exhibit all metabolism and physiological characteristics of the family, but also display unique properties which distinguish *Salmonella* from other genera belonging to the family. *Salmonella* spp. are catalase positive, lysine and ornithine decarboxylase positive, oxidase positive, indole negative, tryptophan and phenylalanine deamination negative, methyl red positive, Voges-Proskauer negative, and urease negative. They produce acid (and gas) through fermentation of glucose, mannitol, and sorbitol. The bacteria are able to ferment sucrose, but rarely adonitol and lactose and produce H₂S when grown on triple iron sugar agar.¹¹⁴

Salmonella enterica serovar Typhimurium (*S. Typhimurium*) is a food-borne bacterium found worldwide ^{115–118} that can cause gastroenteritis, diarrhoea, and in certain settings, invasive disease. Together with *Salmonella* serovar Enteritidis, *S. Typhimurium* accounted for approximately 35% to 40% of culture-confirmed infection, making them the two most common serovars in the USA¹¹⁹. In low- and middle-income countries (LMICs), *Salmonella* infection, especially *S. Typhimurium*, is still a high burden due to low standard of environmental hygiene and food safety. In 2019, a study reported that 21 of 50 samples of soft drink in Jakarta, Indonesia were contaminated with *S. Typhimurium*¹²⁰. In another study, 29% and 16% of 900 samples from swine farms in North Nigeria were contaminated with *Salmonella* Agama and *S. Typhimurium*, respectively¹²¹.

1.5. Methods to study antimicrobial resistance

Given the emergence and widespread of resistant bacteria, antimicrobial susceptibility testing (AST) plays a key role in antimicrobial stewardship and AMR management. AST not only helps to preserve the currently available antimicrobials, but also provides vital information to

physicians for the optimal prescription of antimicrobials. Based on the target of observation or detection, AST was classified into two main groups of (1) phenotypic AST and (2) AMR detection¹²² (Figure 1-5). In phenotypic AST, resistance is identified by observation of phenotypic changes of bacteria under exposure to different antimicrobials¹²². These changes can be measured as inhibitory zones, bacterial suspension density, morphology, proteomics, metabolomics, etc. Regarding AMR detection, resistance is determined by the presence of specific ARGs, which can be detected by either molecular techniques or by genomic sequencing¹²².

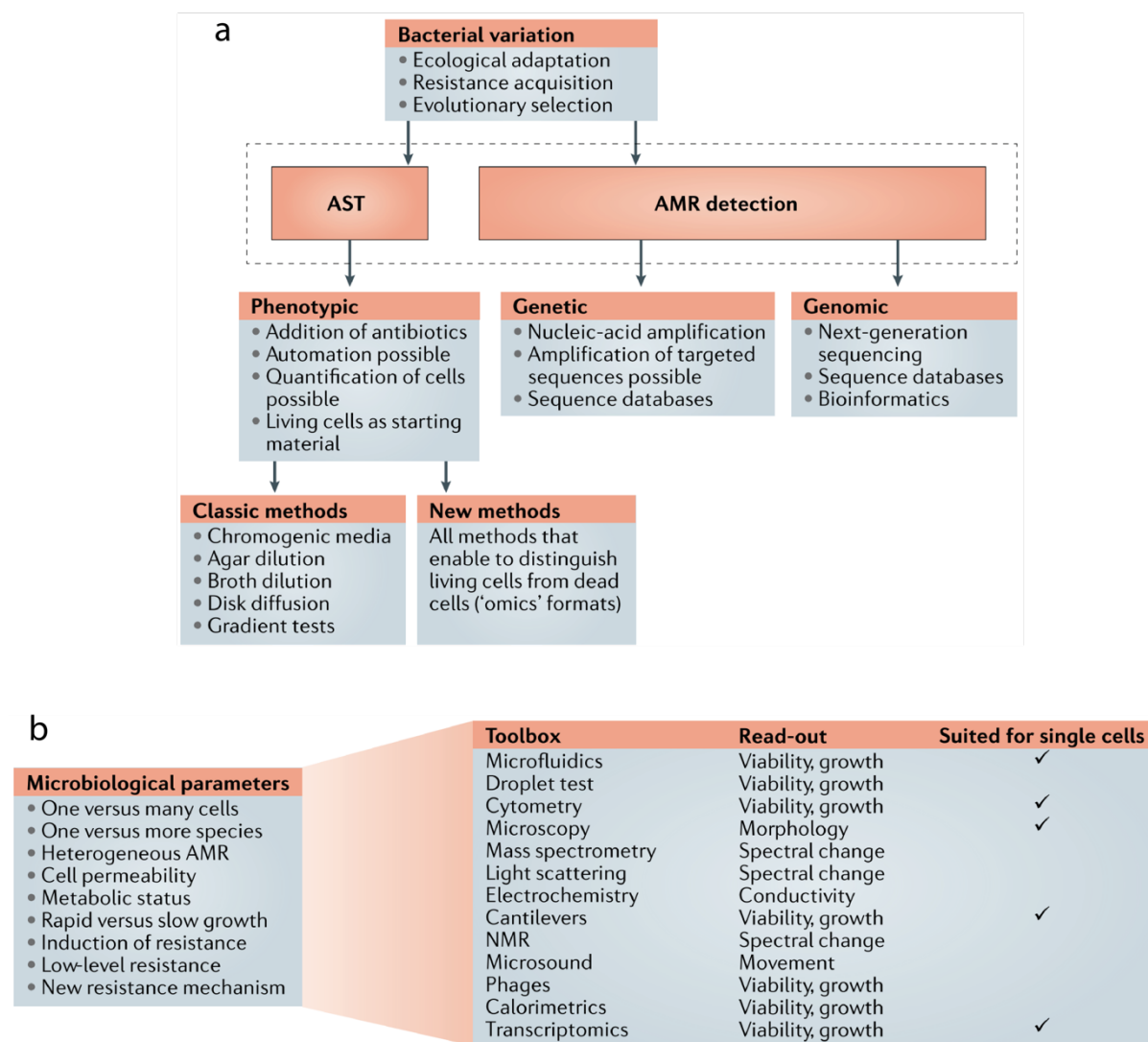


Figure 1-5. Overview of AST methods

(a) General classification of AST methods. (b) Emergence innovative technologies for phenotypic AST with corresponding characteristics. Adopted from van Belkum et al.¹²².

1.5.1. Culture-based AST

Roughly three decades after the first detection of resistant bacteria, agar dilution test was introduced in the 1950s as the pioneering method for susceptibility testing¹²². Latterly, disk diffusion, broth dilution, and VITEK instrument were introduced in the 1960s, 1970s, and 1980s sequentially¹²². The mechanism of these culture-based AST is to observe visual growth of bacteria in the presence of different antimicrobials at specific concentrations. Of these classic AST methods, disk diffusion, gradient diffusion and broth dilution are considered as the gold-standards for detection of resistant organisms, which were released by FDA for *in vitro* diagnostics¹²². Compared to other methods, culture-based AST is low cost, robust and does not require any prior knowledge about resistance mechanisms. However, these methods take long time to generate results (usually 16-24h)¹²³, require high standard laboratories, depend heavily on experience of performers, and are unsuitable for high-throughput. Therefore, innovative technologies are in sustainable demand to overcome the disadvantages of these classical methods, providing a more effective solution for development of novel antimicrobial susceptibility test.

1.5.2. Molecular techniques and genomics analysis

The invention of polymerase chain reaction (PCR) by Kary Mullis gave birth to AMR detection techniques using conventional PCR, quantitative PCR (qPCR), or quantitative real-time PCR (RT-PCR). The idea behind these methods is to detect the presence of AMR factors, that were previously known as resistance determinants. Compared to classical AST methods, ARG detection techniques requires >6h to generate results¹²³. However, the detection of all AMR genes in bacterial resistome is costly, labour intensive and time consuming.

For analysis of the entire resistome, whole-genome sequencing (WGS) is a preferable method. Thanks to the advances in technology, especially the next generation sequencing (NGS) platforms, WGS of bacteria has become more affordable, which enables detection of all resistance factors within genomes of multiple bacterial isolates simultaneously¹⁰⁶. Additionally, the technology has been used to study population structure and evolution of pathogenic bacteria, including *E. coli*^{124,125}, *S. Typhimurium*^{126,127}, *K. pneumoniae*¹²⁸. Furthermore, functional genomics and metagenomics can be employed to discover novel mechanisms of antimicrobial resistance in bacteria. However, compared to other phenotypic ASTs and nucleic-acid-detection methods, genome analysis is more expensive and labour-intensive, which is not suitable to be integrated into routine clinical diagnostics.

A typical Illumina NGS study includes four steps: library preparation, cluster generation, sequencing, and data analysis (Figure 1-6)¹²⁹. In the library preparation step, transposases are added to simultaneously and randomly fragment the input DNA, following by tagging these fragments with adaptors and indexes (Figure 1-6A). The cluster generation is the process in which each tagged DNA fragment is isothermally amplified through a flow cell coated with forward and reverse oligos (Figure 1-6B). During sequencing step, in every cycle, one of four types of fluorescently tagged nucleotides (A, T, G, C) is added to the growing chain complementarily to the fragment templates and emits light signal (Figure 1-6C). Finally, the read data undergoes a data analysis process, depending on the proposed research questions (Figure 1-6D).

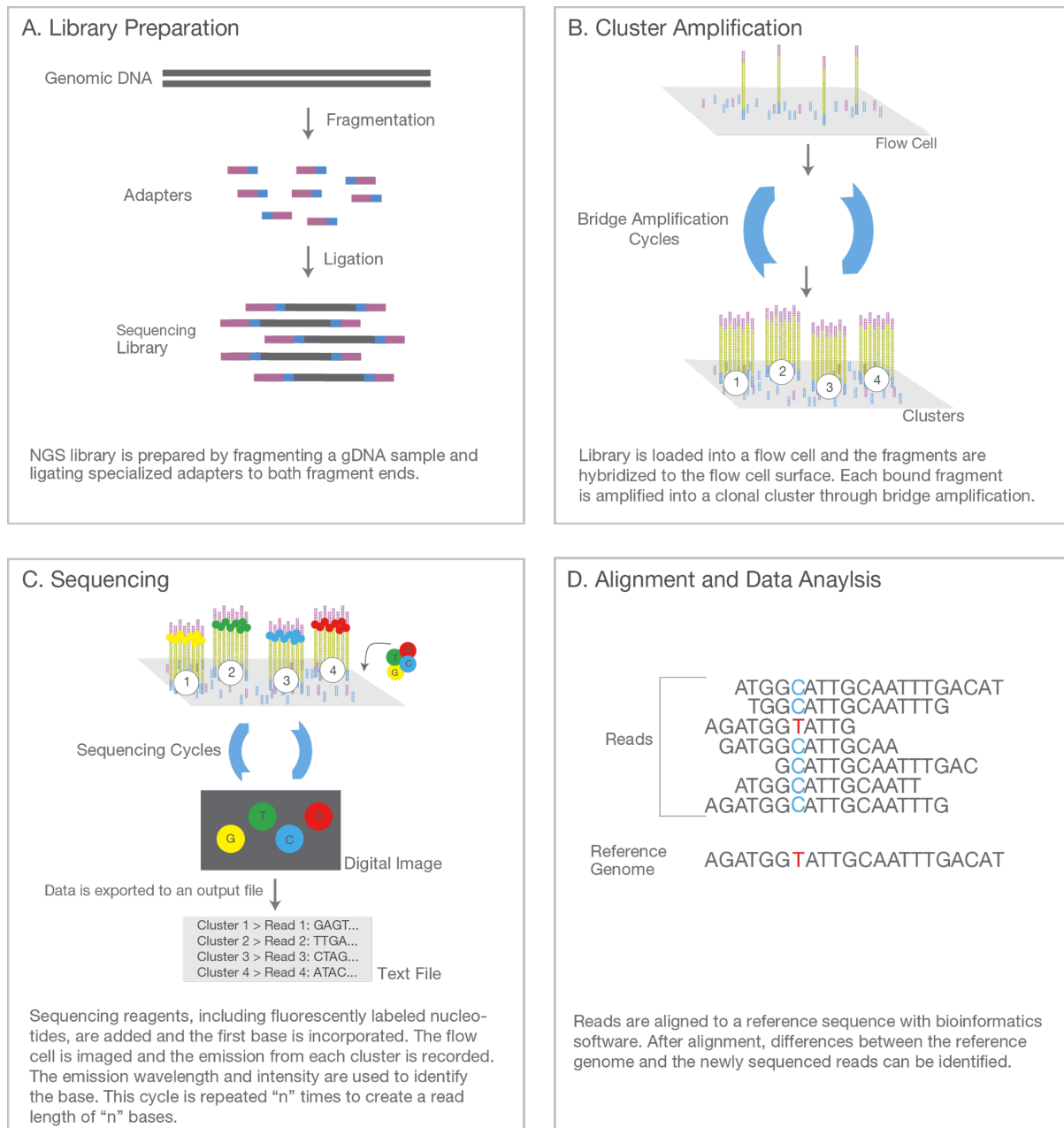


Figure 1-6. NGS chemistry overview

(A) library preparation, (B) cluster generation, (C) sequencing, (D) data analysis¹²⁹.

1.5.3. Proteomics analysis with MALDI-TOF MS

Matrix-Assisted Laser Desorption/Ionisation – time-of-flight mass spectrometry (MALDI-TOF MS) has been widely used in different fields of biomedical research, including analysis of high-molecular-weight molecules of 2kDa-100kDa (proteins¹³⁰, peptides¹³¹, and polymers¹³²), the identification of pathogenic microorganisms¹³³, and screening of biomarkers (especially in AMR detection)¹³⁴, etc. Experiments with MALDI-TOF MS are comparatively straightforward, expeditious and low-cost at <£1 per assay¹²³. Given these advantages, MALDI-TOF MS has been installed and employed in medical laboratories for routine identification of pathogenic microorganisms cultured from biological samples.

In a typical experiment with MALDI-TOF MS, samples are combined with matrix solution and co-crystallised on a metal target plate, which is then loaded into the mass spectrometry. In the MALDI-TOF, a vacuum environment is generated; the target plate is irradiated by a UV laser (usually 337nm). Upon the laser bombardment, both samples and matrix crystals are sublimated and entered the gas phase, releasing ionised particles into the vacuum chamber, e.g. proteins and peptides. The ions are subsequently accelerated in an electrical field generated by extraction electrodes. Under acceleration, the ions fly through the vacuum chamber and hit the detector at the other end of the chamber; time-of-flight for each ion is recorded. During the flight, the ions are separated by their mass, such that lighter ions require less time to reach the detector (mass analyser). Finally, the time-of-flight is used to calculate the mass-to-charge ratio (m/z) for each ionised particle, resulting a mass spectrum of all analytes in a sample¹³⁵. Figure 1-7 described the structure of a MALDI-TOF MS.

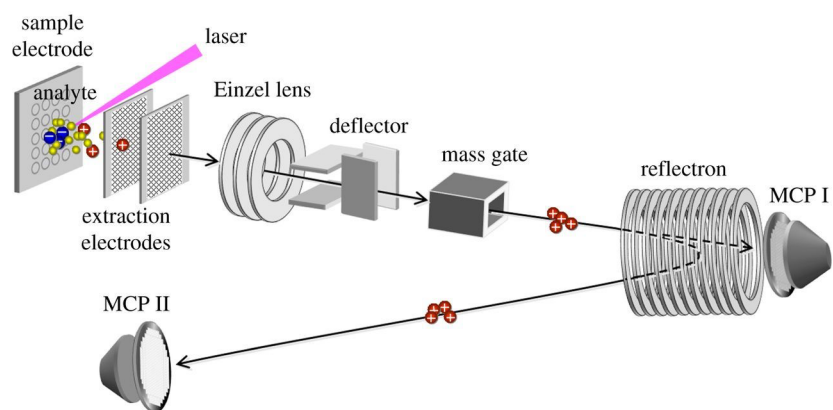


Figure 1-7. Configuration of a MALDI-TOF MS.

The ion source consists of a sample electrode (target plate with laser) and two extraction electrodes. The mass analyser contains an Einzel lens, deflectors, and a mass gate. Ions are detected with the MCP I in the linear mode and with the MCP II in the reflection mode (with a reflectron). Adopted from Wang et al.¹³⁵

Many efforts have been made to adopt MALDI-TOF MS for rapid susceptibility testing. Recently, retrospective mass spectra have been collated and re-analysed by Weis and co-workers in 2022 to predict AMR phenotype of clinical bacterial isolates resulting in considerable predictive performance without prior knowledge on resistance determinants¹³⁶. In another study, MALDI-TOF was able to detect variants of β -lactamases in Gram-negative bacteria within 1 to 2.5 hours¹³⁷. For Gram-positive bacteria, MALDI-TOF was used to detect benzylpenicillin and MDR in *Staphylococcus aureus* with appreciable performance of 97.64% accuracy, 99.93% sensitivity, and 95.04% specificity¹³⁸.

1.5.4. Cellomic analysis with high-content imaging

Based on the beam of matters interacting with samples, microscopy is classified into electron microscopy (electron particles) and light microscopy (photon particles) (Figure 1-8b). Of these techniques, electron microscopy provides the highest resolution (Figure 1-8a), but it is not suitable for high-throughput and high-content imaging. Alternatively, light microscopy produces lower-resolution images (Figure 1-8a), but is low-cost, fast, automatable, and thus suitable for high-throughput and -content imaging.

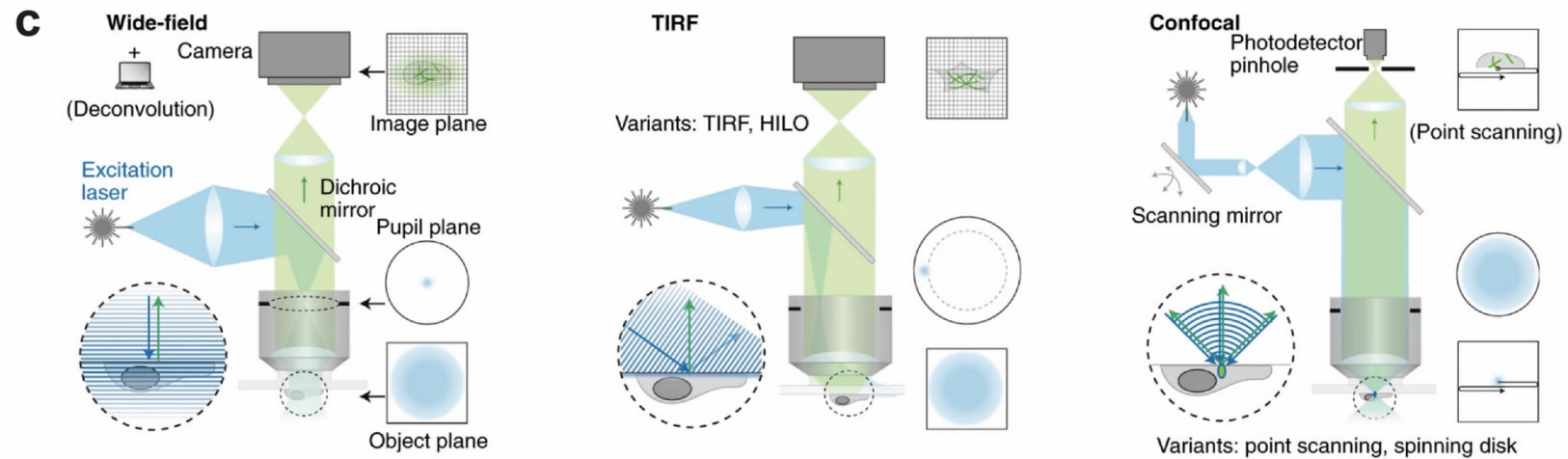
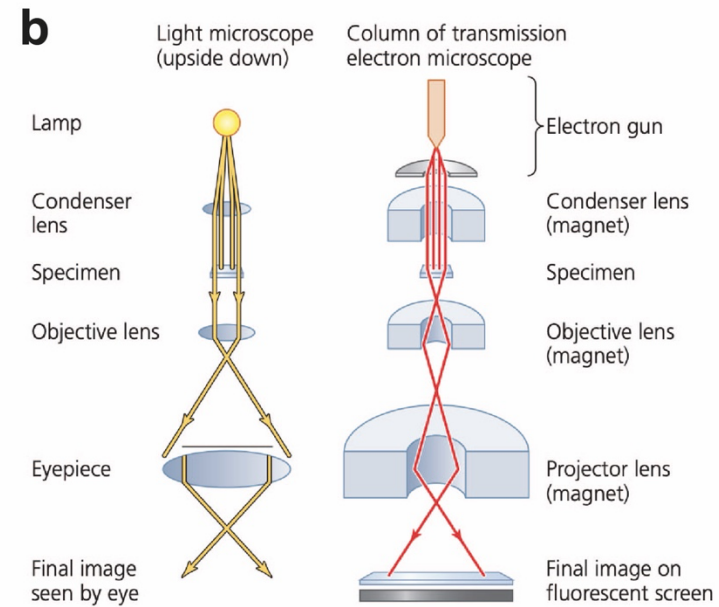
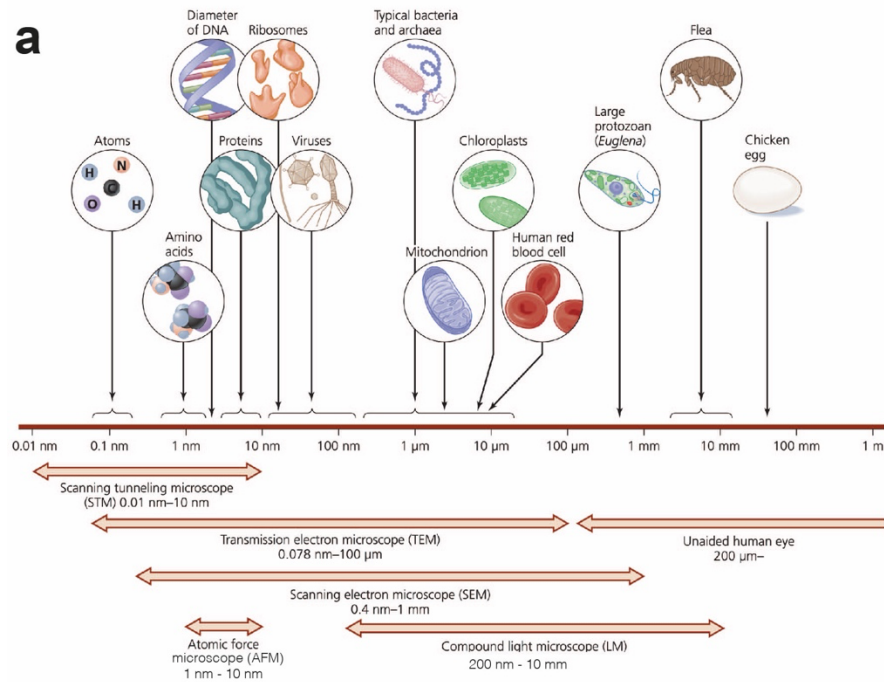


Figure 1-8. Overview of microscopic techniques.

(a) Limits of resolution (and some respective objects within those range) of various types of microscopes. (b) The path of electrons through a transmission electron microscope (right) compared to the path of light (photons) through a light microscope (left). (a) and (b) were adopted from Bauman et al.¹³⁹.(c) Three major light-paths in light microscopy with corresponding optical designs, including widefield (left), TIRF (middle), and confocal (right). Adopted from Schermelleh et al.¹⁴⁰

According to the light-paths, light microscopy is divided into widefield microscopes, total internal reflection fluorescent (TIRF) microscopes, and confocal microscope (Figure 1-8c). The TIRF technique is well known for its applications in cell biology, while the alternative techniques are used across all fields of biological science. Compared to confocal microscopy, widefield microscopy is cheaper, faster, and more user-friendly. The widefield microscopy itself is sub-classified into at least five sub-categories, including brightfield, darkfield, phase-contrast, differential-interference-contrast (DIC), and epifluorescence microscopy. Among widefield microscopes, brightfield microscopy is the simplest technique which has been widely used to examine fixed samples, such as Gram-stained bacteria and histological tissues. For samples that cannot be stained, phase-contrast and DIC microscopy are used as alternatives for both fixed-cell and live-cell imaging.

Additionally, epifluorescence microscopy is ideally used to examine sub-cellular components, e.g. nucleoids, plasma membrane, proteins, and organelles. For samples requiring higher axial resolution, confocal microscopy is used to provide optical sectioning of the samples. This technique is classified into scanning confocal microscopy and spinning-disk confocal microscopy. Compared to spinning-disk microscopy, scanning confocal microscopy produces higher-resolution images. However, spinning-disk confocal microscopy is faster with lower phototoxicity, which makes this technique more ideal for live-cell and high-content imaging. Examples of how these microscopic techniques have been used in bacteriology were shown in Figure 1-9.

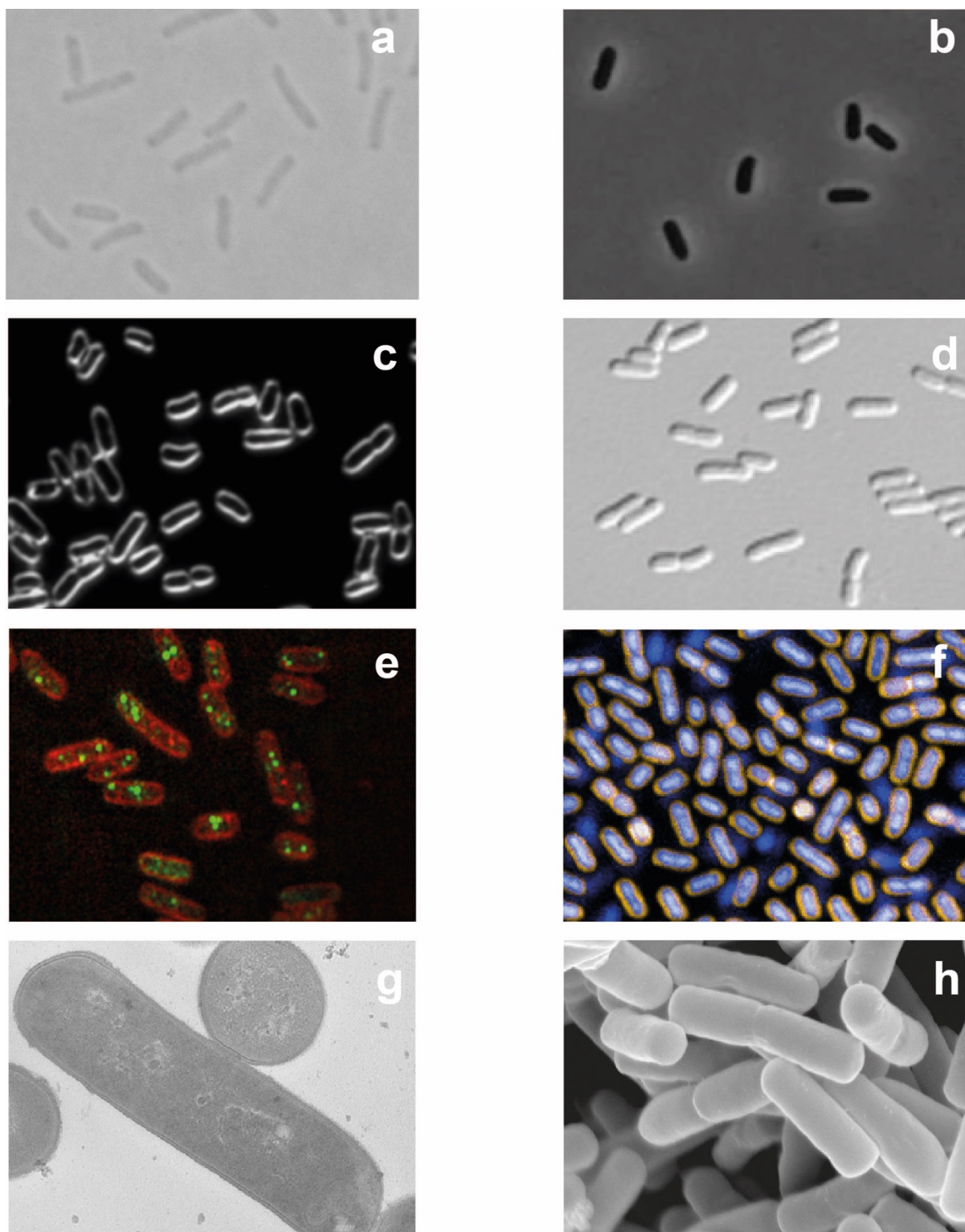


Figure 1-9. Bacterial cells under examination with various types of microscopes.

(a) Brightfield³, (b) Phase-contrast¹⁴¹, (c) Darkfield¹⁴², (d) DIC¹⁴³, (e) Epifluorescence¹⁴⁴, (f) Confocal¹⁴⁵, (g) Transmission electron¹⁴⁶ and (h) Scanning electron¹⁴⁶ microscope.

High-content screening (HCS), high-content analysis (HCA) or high-content imaging (HCI) refers a technique utilising automated microscope to measure a variety of phenotypes in individual cells and heterogeneous responses analyses¹⁴⁷. Unlike high-throughput screening, which measures averaged biological signals over total cells within a microplate well, HCI samples and measures the signal of every single cell of a population within a well¹⁴⁸. Assisted by automated motorized microscope, HCI can sample a large number of individual cells to assess the diversity and the heterogeneity of a microscopic population without compromising measurement representativeness. Additionally, HCI can capture localization and dynamics of fluorescent-labelled subcellular particles within cells, facilitating single-molecule imaging, e.g. a single protein¹⁴⁹. Therefore, HCI has been widely used for library and drug screening in cell biology^{150,151}, especially cancer research¹⁵².

In comparison to eukaryotic cells, the application of HCI in studying prokaryotic organisms, particularly bacterial pathogens, is limited^{145,153,162–165,154–161}. The most common use of HCI with respect to bacteria is to examine the association between antimicrobial mode of action (MoA) and bacterial morphology under exposure to a drug^{145,155,157–160}, allowing qualitative prediction of new compounds' MoA based on dimensional analysis. Additionally, HCI has been employed for AMR identification¹⁵⁶, library screening^{163,165}, antibody screening¹⁵⁴, predicting drug targets¹⁶⁴, activity-guided purification¹⁵⁸, and bacterial enumeration¹⁶². While most studies in the field utilised phase contrast¹⁶⁵ or widefield fluorescent microscopy^{156–164}, only four of them have used confocal high-content microscope imaging^{145,153–155}. Compared to widefield fluorescent microscope, confocal microscope provides optical sectioning with less out-focus signals and better resolution at 180-250nm¹⁶⁶, which allows capturing subtle morphological patterns of bacterial cells within a population. However, the correlation between these subtle patterns and AMR have not been well investigated, which requires

powerful analytic techniques to provide insight into the yet unveiled correlation. More details on applications of HCI in studies of prokaryotic organisms was described as in Table 1-1.

Table 1-1. Overview on applications of HCI to study antimicrobials and resistant microorganisms.

Organism	HCS platform / microscope	Image analysis	Data analysis	Aims	Ref
<i>K. pneumoniae</i> <i>S. Typhimurium</i> <i>S. aureus</i>	Opera Phenix (Perkin Elmer), confocal microscope	Harmony software	Dimensional analysis	Distinguish MoA of antimicrobials	145
<i>S. Typhimurium</i>	Opera Phenix (Perkin Elmer), confocal microscope	Harmony software	Cell length	Morphological responses under exposure to CIP	153
<i>E. coli</i>	Opera Phenix (Perkin Elmer), confocal microscope	Harmony software	Dimensional analysis combined with phylogenetic analysis and statistics	Monoclonal antibody screening	154
<i>A. baumannii</i> <i>E. coli</i> <i>K. pneumoniae</i> <i>P. aeruginosa</i>	Opera QEHS, confocal microscope	Acapella	Dimensional analysis Feature selection and MOA classification using random forest	Predict MoA of antimicrobials for drug discovery	155
<i>S. aureus</i>	Applied Precision Spectris Microscope, widefield fluorescence	Deconvolution with softWoRx and image analysis with FIJI and CellProfiler 2.0	Dimensional analysis combined with phylogenetic analysis	Distinguish methicillin-resistant and -susceptible isolates	156

<i>E. coli</i>	Applied Precision Spectris Microscope, widefield fluorescence	Deconvolution with softWoRx and image analysis with ImageJ	Dimensional analysis and clustering	Distinguish MoA of antimicrobials	157
<i>B. subtilis</i> <i>E. coli</i>	Applied Precision Spectris Microscope, widefield fluorescence	Deconvolution with softWoRx and image analysis with ImageJ	Qualitative image analysis	Distinguish MoA of antimicrobials for activity-guided purification	158
<i>M. tuberculosis</i>	DeltaVision PersonalDV, widefield fluorescence microscope	BaSic and MicrobeJ	Dimensional analysis, machine learning classification, and clustering	Predict MoA of antimicrobials	159
<i>A. baumannii</i>	N/A	FIJI and CellProfiler 3.0	Dimensional analysis, clustering and statistics	Distinguish MoA of antimicrobials	160
<i>S. aureus</i>	Applied Precision DV Elite epifluorescent microscope	Deconvolution, FIJI and Adobe Photoshop CellProfiler 3.0	Dimensional analysis	Differentiate effects of bacteriological and physiological media	161
Complex microbial samples	GE INCell 2200, widefield fluorescence microscope	GE Developer	Machine learning classification	Bacterial enumeration	162

<i>M. tuberculosis</i>	Zeiss Axio Observer Z1	FIJI and MicrobeJ	Machine learning classification and clustering	Library screening	163
<i>E. coli</i>	ImageXpress XLS, widefield fluorescence microscope	M-CNN	Deep learning for image classification	Predict morphology and target proteins	164
<i>E. coli</i>	Nikon Eclips TE200, brightfield imaging	ImageJ	Dimensional analysis	Library screening	165

1.6. Machine learning in biology

Machine learning is a subfield of artificial intelligence that involves the development of algorithms and statistical models that enable computers to learn and make decision based on data, without being explicitly programmed¹⁶⁷. The term ‘machine learning’ was first coined by Arthur Samuel, the father of machine learning, in 1959¹⁶⁸. One of the earliest and simplest examples of machine learning was the perceptron, which was developed by Frank Rosenblatt in the 1950s¹⁶⁹. During the 1960s and the 1970s, more sophisticated machine learning models were developed, including decision tree¹⁷⁰, neural network¹⁷¹, and support vector machine (SVM)¹⁷². Since then, machine learning has been widely used in various fields of science, including biology and biomedical science^{173–179}. From the 1980s to the 1990s, the advance in semiconductor technology facilitated the development of more powerful computers¹⁸⁰ and thus enabled researchers train deeper neural network on large datasets, with multiple hidden layers and more complex architectures. In 2006, Geoffrey Hinton and his team at the University of Toronto first developed a deep network architecture that outperformed conventional machine learning methods in image recognition task¹⁸¹. This work was a breakthrough in machine learning timeline, which opened a new era of applied deep learning in both science and technology.

As a consequence of advances in biological devices and instruments, an enormous amount of data has been produced and shared within the scientific community. Machine learning is an appropriate approach for such large datasets as it can capture complex inter-relationships among multiple variables to yield a higher performance than traditional statistics. In cell biology, machine learning has been utilised for cell imaging analysis tasks, including cell segmentation, cell tracking, and cell classification based on morphology and phenotype¹⁷³. With the development of automated microscopy, more microscopic images have been

produced as the systems are able to acquire images more rapidly and precisely without human supervision. This rich source of data is a fertile land for deep learning and extract biological insight from. Moreover, machine learning, especially deep learning, has been used to explore and analyse the genotype-phenotype complex relationship. Furthermore, machine learning can be used to study human microbiome. For instance, many previous research used machine learning methods to define the relationships between microbiome structure, age, nutrition and disease^{174,175}. Overall, machine learning has been widely used in the field of biological science, from image analysis to genome analysis and microbiome studies. Looking beyond those applications, there are more potentials of machine learning that can be utilised to address broader healthcare challenges at a global scale.

To approach the global challenge of AMR, machine learning has been used extensively to predict and identify resistance, explore resistance mechanisms, design novel antibacterial compounds, and track the spread of AMR¹⁷⁸. A typical example of applied machine learning to tackle AMR is the decision support system. Using electronic medical records, machine learning was trained to predict the risk of AMR infection in hospitalised patients^{176–179}, which provides critical information for optimal administrations of antimicrobial. Machine learning also played as a complement to routine laboratory diagnostics to identify AMR in a timely manner^{182–184}. These rapid susceptibility tests were proven to improve the efficacy and outcome of antimicrobial treatments against bacterial infections^{185,186}. Moreover, the development of deep learning has facilitated genome mining in resistance bacteria to explore new mechanisms of resistance^{187–189}, which creates hypotheses for further laboratory investigation. Overall, machine learning has the potential to play a significant role in tackling the crisis of AMR by providing data-driven insight and decision-making tools to healthcare providers.

1.7. Research aims and questions

Although AMR is a global health challenge, systematic research on how resistant bacteria response to antimicrobial treatment remains scarce. In this thesis, I sought to employ different “omic” methodologies to measuring the bacterial responses from a range of biological perspectives. Specifically, I integrated genomic (WGS) and proteomic (MALDI-TOF) with cellomic (HCI) data to provide a holistic view of bacterial adaptation upon exposure to antimicrobials. Then, I asked whether the powerful machine learning approaches were able to infer antimicrobial susceptibility from the integrated dataset. That was the main research question of this thesis. To address this major question, I segregated my work into differing aims to develop the techniques and methodologies as the thesis evolved. In chapter 3, I aimed to study holistic bacterial responses under exposure to different antimicrobials. This pilot experiment served as a proof-of-concept for further research targets and aimed to ask if measurable responses of bacteria are informative for susceptibility inference. In chapter 4, I performed a temporal examination of how bacteria adapt to antimicrobial exposure using HCI. Since applications of HCI in AMR detection are limited, this chapter asked if longitudinal cellomic data produced by HCI is sufficient for AMR inference. In chapter 5 and chapter 6, I sought to develop a robust, reproducible, and high-throughput experiment to produce multi-omic data (genomics, proteomics, and cellomics). Consequently, I Integrated multi-omic and imaging data with machine learning to infer antimicrobial susceptibility. I aimed to explore and understand the association between multi-omics and antimicrobial resistance profiles of bacteria.

Chapter 2. Methods

2.1. Collections of bacterial isolates

Collection 1. From 2017 to 2018, twenty-seven *E. coli* were isolated from blood samples of children hospitalised with sepsis in Vietnam¹⁹⁰. The study was approved by the Oxford Tropical Research Ethic Committee (OxTREC, reference number 35-16) and the Ethic Committee of Children's Hospital 1 (reference number 73/GCN/BVND1).

Collection 2. From 2014 to 2017, forty-three isolates of *E. coli* were collected from stool samples of Vietnamese children hospitalised with bloody diarrhoea¹⁹¹. Ethical approvals were obtained from the OxTREC (reference number 1058-13) and the Ethic Committee of Hospital for Tropical Disease (reference number 63/HDDD-QD).

Collection 3. Seventeen *S. Typhimurium* isolates used in this DPhil were a part of the Baker lab collection, in collaboration with Institute of Tropical Medicine (Congo), National Institute for Biomedical Research (Congo), International Vaccine Institute (Korea), and Quadram Institute (UK). Details described in (Sridhar, 2020)¹⁹². Those isolates were collected from 2004 to 2014.

All the isolates were stored at -80°C in either Microbank™ vials (Pro-Lab Diagnostics, UK) or in 1.5-mL vials with 1mL of 10-20% Glycerol (Merck, Germany) in Brain Heart Infusion (BHI) broth (Oxoid, UK). When applicable, the isolates were recovered at 37°C overnight, on either Isosensitest agar (Oxoid, UK), Nutrient agar (Oxoid, UK), MacConkey (MC) agar (Merk, Germany or Oxoid, UK), or Lubria-Bertani (LB) agar (Sigma-Aldrich, Germany or Oxoid, UK).

For antimicrobial susceptibility test, Etest strips (bioMérieux, France) were used to obtain MICs (Minimum Inhibitory Concentrations) for each isolate. Briefly, the isolates were recovered overnight on Isosensitest or LB agar. For each isolate, a few colonies were inoculated in 0.45% saline (Bruker Dalton, Germany) to reach an OD₆₀₀ of approximately 0.45-0.55. The bacterial suspensions were spread on Isosensitest or Mueller-Hinton (MH) agar (Merk, Germany or Oxoid, UK). Then, Etest strips were placed on the agar plates, which were subsequently incubated at 37°C for up to 24 hours. Results were read and interpreted according to the manufacturer's manual and CLSI guideline¹⁹³.

2.2. Whole genome sequencing

Bacterial isolates were overnight cultured on either LB, NA, or MC agar plates from frozen stocks. For each isolate, approximately 10 µL of biomass on the plate was collected using a 10µL inoculate loop, resuspended in HPLC grade water, and subjected to DNA extraction using the Wizard® Genomic DNA Purification Kit (Promega, USA). For collection 1 in Chapter 3 and Chapter 6, DNA fragmentation and adapter ligation was performed using Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). For collection 2 in Chapter 6, DNA library was prepared with NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England BioLabs, UK). WGS for collection 1 and 2 was performed on a MiSeq platform, using MiSeq Reagent Kit V2 from Illumina. For collection 3 in Chapter 5, library preparation for Illumina sequencing was undertaken at the Wellcome Sanger Institute using automated systems using the IHTP WGS NEB Ultra II library kit. Libraries were sequenced on an Illumina HiSeq platform (Illumina, San Diego, CA, USA) using standard running protocols.

2.3. Data analysis

2.3.1. Genomic analysis

2.3.1.1. Core-genome and resistome analysis for *E. coli*

Multi-locus sequencing typing (MLST) of *E. coli* was inferred from the whole genome sequencing data using the SRST2 pipeline¹⁹⁴ on the seven housekeeping genes, including *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*. Resistome analysis with AMR gene identification was performed on the SRST2 pipeline with Comprehensive Antimicrobial Resistance Database (CARD)¹⁹⁵. *De novo* assembly of sequencing data was performed using SPAdes¹⁹⁶, gene annotation was completed using Prokka¹⁹⁷ and the pan- and core- genome were deduced using Roary¹⁹⁸ with minimum 99% blastp identity. Finally, a phylogenetic tree was constructed on RAxML¹⁹⁹ using the alignment of the core genome from Roary's output and was visualized using the 'ggtree' package.

2.3.1.2. Genomic mapping for *S. Typhimurium*

Illumina adapter content was removed from the reads using Trimmomatic v.0.33. Reads mapping was undertaken using the WSI bacterial mapping pipeline, which uses bwa, and de novo assembly was performed using Velvet²⁰⁰. For SL1344gyrA and D23580gyrA mutants, Illumina HiSeq reads for the isogenic mutants and parental strains were mapped to the parental reference strain: SL1344 (FQ312003.1) and D23580 (FN424405.1), respectively, using SMALT v0.7.4 to produce a BAM file^{201–203}. Briefly, variant detection was performed as detailed here: SAMtools mpileup v0.1.19 with parameters -d 1000 -DSugBf and bcftools v0.1.19 were used to generate a BCF file of all variant sites. The bcftools variant quality score was set as greater than 50, mapping quality was set as greater than 30, the allele frequency was determined as either 0 for bases called same as the reference or 1 for bases

called as a SNP ($\text{af1} < 0.95$), the majority base call was set to be present in at least 75% of reads mapping at the base ($\text{ratio} < 0.75$), the minimum mapping depth was four reads, a minimum of two of the four had to map to each strand, and strand_bias was set as less than 0.001, map bias less than 0.001, and tail_bias less than 0.001. Bases that did not meet those criteria were called uncertain and removed. A pseudogenome was constructed by substituting the base calls in the BCF file in the reference genome. Recombinant regions in the chromosome such as prophage regions were removed from the alignment and checked using Gubbins v1.4.10. SNP sites were extracted from the alignment using snp-sites and analysed manually. *gyrA* mutations identified by PCR were confirmed. ARIBA v 2.14.6 using the CARD v3.1.3 with default parameters was used on reads data of all isolates to determine AMR genes^{204,205}.

2.3.2. Statistical analysis and data visualisation

All data were analysed using R (version 4.0.5)²⁰⁶ programming language based on toolbox, package and library availability. For dimensional analysis, Principle Coordinate Analysis (PCoA) was performed using ‘vegan’ package²⁰⁷. Heatmap and spider plots were generated with ‘pheatmap’²⁰⁸ and ‘fmsb’²⁰⁹ package respectively. Other plots were created with ‘ggplot2’²¹⁰ and ‘ggsci’²¹¹ package on R. Locally estimated scatterplot smoothing (LOESS) regression was performed with the ‘ggplot2’²¹⁰ package. Pairwise Kruskal-Wallis tests were performed using ‘ggpubr’ package²¹². For feature selection, ‘randomForestSRC’ package²¹³ was utilised to train random forest models with default parameters and 1,000 decision trees. Features were selected and ranked decreasingly by importance index extracted from the random forests.

2.3.3. Machine learning modelling

Machine learning models were employed for classification and regression tasks in chapter 4 and 6. Given the limited amount of experimental data for such tasks, I chose relatively simple traditional machine learning techniques, instead of advanced complex methods, to avoid overfitting the models. Those machine learning methods included Naïve Bayes classifier, KNN (K-neighbour nearest) classifier, SVM, random forest, neural network, CatBoost, xgBoost, glmBoost, and MARS (multivariate adaptive regression spline). Those models have been known for working well on small biological datasets^{138,214–216} with high prediction performance and high level of generalisability. The implementation of those machine learning models was described in the following.

For chapter 4, ‘Statistics and Machine learning’ toolbox on Matlab was used to train Naïve Bayes classifier, KNN classifier, SVM, and random forest. Neural network and CatBoost was trained with ‘Neural network’ toolbox on Matlab and ‘CatBoost’ library²¹⁷ on python, respectively. Nested resampling was performed on either Python or Matlab for hyperparameter optimisation and evaluation of machine learning models. In that, the resampling was performed 1,000 times with a 50/25/25 data partition for training, validation, and test sets. The training sets were used to train the machine learning models and the predictive performance of machine learning models was reported on the test sets. On the validation sets, grid-search algorithm was employed to optimise the hyperparameters of all the machine learning models.

For chapter 6, all machine learning models were constructed using the ‘mlr3’ package on R. The models included neural network, naïve Bayes classifier, KNN, SVM, random forest,

xgBoost, glmBoost and MARS. For nested resampling, data partition was performed 10 times with 30% of data for test sets (outer resampling). On the 70% remaining data, 5-fold validation was executed to train the models and optimise the hyperparameters (inner resampling). The machine learning models were evaluated on the test sets.

2.3.4. Model evaluation

For classification tasks, machine learning models were evaluated using multiple performance metrics, including accuracy, sensitivity, specificity, precision, and F1 score. Accuracy of a model was defined as the following

$$ACC = \frac{TP + TN}{P + N}$$

where P was the number of actual resistant isolates in the data, N was the number of actual susceptible isolates, TP was the number of accurately predicted resistant isolates, and TN was the number of accurately predicted susceptible isolates. Other metrics for performance evaluation, including sensitivity (SEN), specificity (SPE), precision (PRE), F1 score, were described as the following:

$$SEN = \frac{TP}{P}$$

$$SPE = \frac{TN}{P}$$

$$PRE = \frac{TP}{PP}$$

$$F1 = \frac{2PRE \times SEN}{PRE + SEN}$$

where PP was number of predicted resistant isolates. Area under the curve (AUC) of the ROC (receiver operating characteristic) was calculated with ‘perfcurve’ function on Matlab or ‘mlr3’ package on R.

For regression tasks, machine learning models were evaluated with mean squared error (MSE), mean absolute error (MAE), relative absolute error (RAE), correlation coefficient (ρ), and correlation of determination (R^2). Those metrics were also used to evaluate the bacterial growth models. Details of those metrics were described as the following.

$$MSE = \frac{1}{N} \sum_{i=1}^N \left(y^{(i)} - \widehat{y}^{(i)} \right)^2$$

$$MAE = \frac{1}{N} \sum_{i=1}^N \left| y^{(i)} - \widehat{y}^{(i)} \right|$$

$$RAE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y^{(i)} - \widehat{y}^{(i)}}{y^{(i)}} \right|$$

$$\rho = \frac{\mathbb{E}[(Y - \mu_Y)(\hat{Y} - \mu_{\hat{Y}})]}{\sigma_Y \sigma_{\hat{Y}}}$$

$$R^2 = 1 - \frac{\sum_i (y^{(i)} - \widehat{y}^{(i)})^2}{\sum_i (y^{(i)} - \mu_Y)^2}$$

where N was the number of data instances, Y and y were the regression targets, $\hat{Y}$ and $\hat{y}$ were the predicted targets, i was the index for the i th target, $\mathbb{E}(\cdot)$ was the mean operator, μ was mean, and σ is standard deviation.

2.3.5. Interpretation of machine learning models

The high complexity of machine learning models had the technique considered as black boxes since how machine learning associates features to prediction targets is difficult for human to comprehend. That property is preventing machine learning being broadly used in biological science. In chapter 4 and 6, a range of machine learning models were trained on proteomic and cellomic features to predict bacterial phenotypes and characteristics, including resistance, MICs, and source of isolation. To reveal the behaviours of those models, interpretable machine learning techniques were used as the following.

The first step in interpretation of machine learning models is to select the most important features amongst a large number of features. To calculate the feature importance for chapter 4 and 6, a permutation method was used due to its low complexity and low computational cost²¹⁸. For each feature, permutation feature importance measures the change in prediction error of machine learning models after the feature being permuted which disturbs the relationship between the feature and the prediction target. An increase in prediction error indicates that a feature is important, while decreased error suggests a feature is less important and likely to introduce noise to the models. Given the j^{th} feature, the permutation feature importance FI_j was defined as

$$FI_j = e_{perm} - e_{orig}$$

or

$$FI_j = \frac{e_{perm}}{e_{orig}}$$

or

$$FI_j = \frac{e_{perm} - e_{orig}}{e_{orig}}$$

where e_{orig} was the prediction error of models trained on original data and e_{perm} was the error on data with the j^{th} feature permuted. For implementation, ‘randomForestSRC’, ‘mlr3’, and ‘iml’ packages were used on R programming language to calculate the permutation feature importance.

For each important feature, an in-house code was used to generate Partial Dependence Plot (PDP)²¹⁹ and Individual Conditional Expectation (ICE)²²⁰ plot. Considering one important feature, the ICE plot displays one line per datapoint that shows how the prediction for that datapoint changes when that feature varies. In ICE plot, for each datapoint in

$\{(x_S^{(i)}, x_C^{(i)})\}_{i=1}^N$, the curve $\widehat{f}_S^{(i)}$ representing models’ output was plotted against $x_S^{(i)}$, while $x_C^{(i)}$ remained fixed, where N was size of the dataset, i was index of the i th datapoint, x_S was the interpreted feature, x_C was other input features. PDP was defined as average of ICE

$$\widehat{f}_{x_S}(x_S) = \mathbb{E}_{x_C}[\widehat{f}(x_S, x_C)] = \int \widehat{f}(x_S, x_C) d\mathbb{P}(x_C)$$

In practice, the partial function $\widehat{f}_{x_S}$ was estimated by calculating averages on the dataset

$$\widehat{f}_{x_S}(x_S) = \frac{1}{N} \sum_{i=1}^N \widehat{f}(x_S, x_C^{(i)})$$

Alternatively, Shapley Additive Explanation (SHAP), one of the most advanced interpretation methods, were utilised to address both the importance and contribution of a feature to the prediction outcome of machine learning models. The SHAP method considers each feature as a player in a coalitional game, cooperating to explain how the model’s prediction is different from an averaged prediction. Compared to other interpretation methods, SHAP allows quantification of one feature’s contribution in interaction with other

features, which provides a detailed explanation on behaviours of machine learning models. SHAP values were estimated using ‘kernelShap’ package on R²²¹. Technical details about SHAP calculation can be found in ²¹⁸.

2.3.6. Distance ratio

To assess segregation of the isolates, I defined an index named distance ratio

$r(X_1, \dots, X_k, \dots, X_n)$ as the following equation:

$$r(X_1, \dots, X_k, \dots, X_n) = \frac{\mathbb{E} \left(\|x^{(i)} - x^{(j)}\|_2 \right)_{x^{(i)} \in X_k, x^{(j)} \in X_k}}{\mathbb{E}(\|x^{(i)} - x^{(j)}\|_2)_{x^{(i)} \in X_k, x^{(j)} \notin X_k}}$$

where X_k was data group, e.g. resistant vs susceptible, x_i was a datapoint, $\mathbb{E}(\cdot)$ was statistical mean, and $\|\cdot\|_2$ was Euclidean norm. Low distance ratio indicates that data points within one group are closely related while well separated from ones in other groups.

2.4. Chapter 3

Twenty-seven *E. coli* isolates from collection 1 were used for the experiments in Chapter 3.

A laboratory reference strain ATCC 25922 was included in the collection as a susceptible control. For all experiments, the antimicrobials of interest were CIP, AZI, CRO (ceftriaxone), TE, GN (gentamicin), and CIP+CRO.

2.4.1. Growth dynamics

For bacterial growth monitoring, I performed two-fold dilutions of antimicrobials in 100 μ L MHB2 (cation-adjusted MH broth from Merk, Germany or Oxoid, UK) on 96-well

microplates. A range of antimicrobial concentrations were generated from 0 to 512 μ g/mL. Blank controls were included on each microplate. The *E. coli* isolates were recovered overnight. For each isolate, one colony was inoculated in 5mL LB broth (Oxoid, UK) at 37°C, 200rpm, overnight. On the next day, bacterial suspensions were diluted 1/100 in fresh MHB2 and incubated at 37°C, 200rpm, for 3 hours to reach an OD₆₀₀ of approximately 0.3. Those suspensions were diluted 1/1,000 in fresh MHB2 and aliquoted 100 μ L/well on the prepared microplates with two-fold dilutions. As a result, the antimicrobial concentrations halved from 0-512 μ g/mL to 0-256 μ g/mL. Bacterial densities in the microplates were measured at the wavelength of ~600nm using the FLUOstar OPTIMA microplate reader (BMG Labtech, Germany). Those densities were measured every 15 minutes for 24 hours.

To model growth dynamics of the isolates, the modified Gompertz's (eq. 1), Baranyi's (eq. 2) and three-phase linear (eq. 3) models were fit to the experimented OD₆₀₀ data. The three models have been used to describe the growth curve of bacteria in previous studies^{222–225} with high accuracy. Of those, the modified Gompertz's and Baranyi's equation describe the growth dynamics of bacteria as logistic functions. The three-phase linear model uses three linear equations to model bacteria growth, corresponding to lag, log, and stationary phase. The three models were described as the following

$$N(t) = N_0 + A \exp \left\{ -\exp \left[\frac{\mu e}{A} (\lambda - t) + 1 \right] \right\} \quad (\text{eq. 1})$$

$$N(t) = N_0 + \mu C(t) - \ln \left\{ 1 + \frac{e^{\mu C(t)} - 1}{e^A} \right\} \quad (\text{eq. 2})$$

$$N(t) = \begin{cases} N_0, & t \leq \lambda \\ N_0 + \mu(t - \lambda), & \lambda < t < t_s \\ A, & t \geq t_s \end{cases} \quad (\text{eq. 3})$$

where $N(t)$ was the cell density, measured in OD₆₀₀, at timepoint t , N_0 was the cell density at $t = 0$, A was the maximum cell density, μ was the maximum growth rate (per hour), λ was the lag time (in hours), t_s was the time to reach stationary phase (in hours), and $C(t) = t + \frac{1}{\mu} \ln[e^{-\mu t} + e^{-\mu \lambda} - e^{-\mu(t+\lambda)}]$. AUC was estimated from the growth curves using the Reimann summation method. The parameters A , μ , λ , and AUC were referred as growth features of a growth curve. Using ‘optim’ function in R, the three models were fit to the OD₆₀₀ data by minimising the mean squared error (MSE) between the actual and modelled OD values.

2.4.2. Bright-field imaging

The isolates were recovered overnight on nutrient agar plates. On the next day, about 10 colonies for each isolate were inoculated in 0.85% saline to each an OD⁶⁰⁰ of approximately 0.6 ($\sim 10^8$ CFU/mL). To reduce bacterial density in each field-of-view, suspensions of resistant isolates were diluted 1/20 in saline. At CLSI breakpoint concentrations¹⁹³, antimicrobials of interest were supplemented to the bacterial suspensions and to thin nutrient agar plates (8mL/plate). Next, 100 μ L of the suspensions were spread onto the thin plates. The plates were then incubated at 37°C for 3 hours. Finally, small pieces of agar were excised from the thin agar plates for bright-field microscopic imaging with the Nikon Ni-e upright microscope (Nikon, Japan). Digital images were acquired with the Element-Br software (Nikon, Japan) for further qualitative analysis.

2.5. Chapter 4

2.5.1. *gyrA* spontaneous mutants

To isolate spontaneous nalidixic acid mutant lines from *S. Typhimurium* isolates SL1344 and D23580, bacterial cultures were grown overnight in LB broth, and 100µl of this was spread onto LB agar containing 100µg/ml nalidixic acid for initial spontaneous mutant generation. After overnight incubation at 37°C, single colonies that had grown were sub-cultured on LB agar also containing 100µg/ml nalidixic acid. Any colonies that were present on these agar plates were then sub-cultured serially onto agar plates harbouring increasing concentrations of nalidixic acid up to 400µg/ml, then these were switched to plates containing CIP, harbouring from 0.1µg/ml up to 1.0µg/ml. Once colonies were able to grow stably on 1.0µg/ml CIP, pure cultures were established and stored as frozen stocks. From the frozen stocks, overnight cultures were grown for genomic DNA purification and were purified using the Wizard DNA Purification Kit (Promega, UK). Following purification, DNA was PCR-amplified to check for mutations in the *gyrA* gene using primers: 5'-GAGATGGCCTGAAGC-3' for nucleotides 108 to 127 and 5'-TACCGTCATAGTTATCCACG-3' for nucleotides 435 to 454, forward and reverse, respectively²²⁶. To evaluate genetic differences between parent and isogenic strains, the isogenic and parent strains were grown for 24h in LB broth prior to genomic DNA isolation for whole genome sequencing.

2.5.2. Time kill curves

Four *S. Typhimurium* isolates were chosen for the initial time kill curve analysis, performed as in (ref. ¹⁵³). These isolates were D23580, SL1344, SL1344*gyrA*, VNS20081^{201,227-229}. Initially, colonies from plates were inoculated into 10mL of MHB2 broth and these were shaken at 200rpm, 37°C, overnight. Ten microlitre of the subsequent culture was then added

to 990 μ L of 1x PBS to make a 1:100 dilution for the inoculum. One hundred microlitre of this preparation was added to 10mL of MHB2 containing different concentrations (0x, 1x, 2x, 4x MIC) of CIP according to the predetermined MIC of each isolate. The starter inoculum was between 1 and 5 x 10⁵ CFU/mL. Cultures were incubated shaking at 37°C and aliquots were taken to determine CFU at 0, 2, 4, 6, 8, and 24 hours. For this analysis, serial dilutions were made using samples of each culture, and a total of 50 μ L of each dilution was plated using 10 μ l spots of inoculum onto LB agar. CFUs were counted and determined as CFU/ml. Means and standard deviations (SD) of three replicates per isolate were calculated.

2.5.3. High-content imaging

Two separate HCI experiments were performed. The first experiment was a 24-hour evaluation of bacterial growth under four CIP concentrations (0x, 1x, 2x, 4x MIC) at two-hour increments (2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24) conducted on four isolates SL1344, SL1344gyrA, D23580, and VNS20081, which has been previously described in Sridhar *et al.*¹⁵³ Briefly, 150 μ L of each isolate was inoculated 1:1,000 (v/v) in 150mL MHB2 with each appropriate concentration of CIP in a 200-mL flask and incubated at 37°C shaking at 200rpm. Following 2 hours of growth, 10mL of each culture was removed from the flask, and the flask was returned to the incubator. The 10mL sample was centrifuged at 3,000rpm (~1230g) for 7 minutes at 4°C, and the supernatant was decanted. The pellet was resuspended in 100 μ L PBS and 50 μ L of the suspension was added to two wells of a vitronectin-coated CellCarrier Ultra-96 microplate (PerkinElmer, Germany). The plate was statically incubated at 37°C for 10 minutes, after which the bacterial cultures were aspirated and fixed in 4% paraformaldehyde (PFA) for 10 minutes and washed with PBS. After fixation, the plate of fixed bacteria was kept at 4°C until the next time point. The same protocol was followed for each time point with this exception: once there was sufficient bacterial growth (as assessed

visually), 10mL of bacteria was still removed from the flask at each time point; however, 50 μ L of bacterial suspension was added neat to the wells instead. Upon completion of the 24-hour period, wells were incubated with 2% bovine serum albumin (BSA) for 30 minutes and then for 1h at ambient temperature in the dark with CSA-Alexa Fluor 647 (Novus Biologicals, USA) at 1:1,000 in BSA (v/v). Wells were aspirated and then incubated with solutions containing 1:100 4',6-diamidino-2-phenylindole (DAPI) (Invitrogen, UK) and 1:200 SYTOX green (Invitrogen, UK) for 20 minutes. Photophysical properties of those fluorophores were described in Table 2-1. Wells were washed with PBS; plates were sealed and imaged. Three biological replicates of this experiment were performed for D23580 and VNS20081; two biological replicates were performed for SL1344 and SL1344gyrA.

The second HCI experiment used 16 of the 17 isolates detailed in Table 4-1, including all except D23580. Here, the only time point assessed was 22 hour and no ciprofloxacin treatment was used. However, to maintain experimental consistency with the previous experiment, cultures were grown like before in 150mL MHB2, and 10mL of culture was removed every two hours to mimic the change in growth condition. At the 22 h time point, 50 μ L of each culture was added (neat) to two wells of a vitronectin-coated Opera CellCarrier Ultra-96 plate, and the same fixation and staining protocol as above was used.

Table 2-1. Photophysical properties of fluorescence dyes/stains used in chapter 4, 5, and 6

Fluorophore	Excitation wavelength (max)	Emission wavelength (max)
CSA-Alexa Fluo 647	650nm	665nm
DAPI	350nm	470nm
Sytox green	504nm	523nm
Syto 41	430nm	454nm
FM 4-64	515nm	640nm

2.5.4. Imaging analysis

Analysis was performed using a Perkin Elmer Harmony software analysis pipeline designed for *S. Typhimurium*, as previously described^{145,153}. Briefly, inputted images were subjected to flatfield correction, and images were calculated using the DAPI and AlexaFluor 647 channels. Image calculations were refined by size and shape characteristics. A linear classifier was applied to the filtered population, and single bacteria were identified, and morphology and intensity characteristics were calculated. The output of the Harmony analysis was tabulated by plate and object, and results were further analysed and visualised in R (version 4.0.5)²⁰⁶ using ggplot2 package²¹⁰. Adobe Illustrator was used to format images and graphs for presentation.

2.6. Chapter 5 and Chapter 6

2.6.1. Intact protein analysis with MALDI-TOF MS

2.6.1.1. Sample preparation

There were two major approaches to prepare samples for protein analysis with MALDI-TOF mass spectrometry. For the first method, bacterial cells were inactivated in either alcohol (section 2.6.1.1.1)²³⁰ or acid (section 2.6.1.1.2)²³¹. Regarding the second method, proteins were extracted using acidic solutions (section 2.6.1.1.3²³² and 2.6.1.1.4²³³). Next, the inactivated cells and the protein extracts were deposited on a target plate for co-crystallisation with matrix solution and further measurement on MALDI-TOF MS.

2.6.1.1.1. Bacteria inactivation with Ethanol

Bacterial suspensions in Eppendorf tubes were centrifuged at 3,000rpm (~845g) for 5 minutes. Supernatant was discarded; containers were tapped onto paper towel to remove excessive liquid. To inactivate bacterial cells, pellets were resuspended in 100 μ L 35% ethanol. The suspension was then centrifuged for 5 minutes at 3,000rpm (~845g); supernatant was discarded. Bacterial cell pellets were washed twice with 100 μ L 35% ethanol. Pellets were resuspended in 100 μ L 35% ethanol and stored at 4°C till being uses.

2.6.1.1.2. Bacteria inactivation with trifluoroacetic acid

Similarly to the above method with ethanol, bacterial suspensions in Eppendorf tubes were centrifuged at maximum speed for two minutes. Supernatant was discarded and excessive liquid was removed by tapping the containers onto paper towel. Bacterial pellets were resuspended in 100 μ L Trifluoroacetic acid (TFA). The suspension was shaken at 2,000rpm for 2 minutes, mixed with 900 μ L water (HPLC grade) and stored at 4°C till being used.

2.6.1.1.3. Protein extraction with Acetonitrile

Another method to prepare samples for experiments with MALDI-TOF is protein extraction with acetonitrile. Deep-well plates or Eppendorf tubes containing bacterial suspensions were centrifuged at maximum speed for two minutes. Supernatant was removed by tapping the containers onto paper towel. Bacterial pellets were resuspended in 300 μ L 0.5% TFA and 700 μ L acetonitrile (ACN) was added to the suspension, which was subsequently shaken at 2,000rpm for 30 minutes, followed by centrifugation at maximum speed for 15 minutes. Supernatant was stored at 4°C till being used.

2.6.1.1.4. Protein extraction with ethanol/formic acid/acetonitrile

The final method for sample preparation is protein extraction using ethanol/formic acid/acetonitrile. Bacterial suspension in deep-well plates or Eppendorf tubes were centrifuged at maximum speed for two minutes. Supernatant was discarded; containers were tapped onto paper towel to remove excessive liquid. Remaining pellets were resuspended in 100 μ L absolute ethanol to inactivate bacterial cells. Subsequently, the bacterial suspension was centrifuged; supernatant was discarded. Bacterial cell pellets were airdried and resuspended in 50 μ L 70% formic acid (FA) and 50 μ L ACN. The final suspension was centrifuged at maximum speed for 2 minutes; supernatant was stored at 4°C till being used.

2.6.1.2. Matrix preparation

To select the optimal matrix for MALDI-TOF experiments, I prepared six different matrix solutions which were combined with different sample preparation and deposition methods. CHCA (α -Cyano-4-hydrocinnamic acid, Sigma-Aldrich) was dissolved in Bruker Standard

Solvent (50% ACN, 47.5% water, 2.5% TFA, Sigma-Aldrich, Germany), with vigorous shaking, at the concentration of 10mg/mL to obtain standard matrix solution (CHCA-Std). Saturated CHCA was made by mixing at least 30mg CHCA into 1mL acetone and shaking for 20 minutes at 2,000rpm. The solution was then centrifuged at maximum speed for two minutes; supernatant was collected and used as saturated CHCA solution. DHB standard solution (DHB-Std) was prepared by dissolving super DHB (2,5-dihydroxybenzoic acid, Sigma-Aldrich, Germany) in Bruker Standard Solvent at 20mg/mL. SA standard solution (SA-Std) was made with 20mg SA (sinapinic acid) dissolved in 1mL Standard Solvent. SA was also dissolved in acetonitrile to make up 20mg/mL SA-ACN solution. CHCA-DHB solution was prepared by mixing 1:1 (v/v) solution A and solution B. For solution A, 20mg CHCA was dissolved in 1mL of 7:3 (v/v) ACN and 5% FA. Solution B was made with 20mg super DHB in 1mL of 7:3 (v/v) ACN and 0.1% TFA.

2.6.1.3. Intact protein analysis with MALDI-TOF MS

I Have evaluated different combinations of sample preparation, matrix preparation, and matrix deposit methods to select the optimal procedure for intact protein analysis, described as the following. Deep multi-well plates containing bacterial suspension were centrifuged at 3,000rpm (~1230g) for five minutes. Supernatant was removed by pouring out and tapping the plates onto paper towel. Pellets were resuspended in 100 μ L of 35% ethanol and shaken for two minutes at 2,000rpm to inactivate bacterial cells. The deep plates were subsequently centrifuged at 3,000rpm (~1230g) for five minutes; supernatant was decanted to collect bacterial pellets. Subsequently, the cell pellets were wash once in 100 μ L ethanol (35%). The Bacterial pellets were resuspended in 100 μ L of 35% ethanol, and transferred to 384-well microplates, which is appropriate for high-throughput deposition of prepared samples onto MALDI target plates. The liquid transfer was facilitated by an Eppendorf Explorer® Plus 24-

channel electronic multi-pipette (Eppendorf, Germany). The bacterial suspension was centrifuged again at 3,000rpm (~1230g) for five minutes and supernatant was removed. Cell pellets were resuspended in 10 μ L of 35% ethanol, followed by addition of 10 μ L CHCA-DHB matrix solution (section 2.6.1.2). The bacterial suspension was stored at 4°C till further steps. Prior to sample deposition, the MALDI target plate was coated with a thin layer of saturated CHCA solution in acetone (section 2.6.1.2). Using the 24-channel multi-pipette at dispenser mode, 2 μ L of the prepared bacterial samples were deposited on the precoated MALDI target plate. Bacterial samples were airdried on the plate and measured with MALDI Biotyper[®] (Bruker Dalton, Germany) for mass spectrum acquisition of intact proteins.

2.6.1.4. Data pre-processing for mass spectra

Raw data output from the MALDI Biotyper[®] was read and converted to csv files by the ‘MALDIQuantForeign’ package²³⁴ on R. The csv files were subsequently sampled with the ‘Mass Spectrometry Data Analysis’ toolbox on Matlab to remove inconsistency in data. Then, the sampled csv files were imported to R and analysed with the ‘MALDIQuant’ package²³⁴. Briefly, the spectra intensity was undergone root-square transformation, smoothing and background correction. Each pre-processed spectrum was normalised to the total ion current (TOC) and aligned to a set of reference spectra. Finally, mass peaks were detected and binned, resulting in a peak matrix for further data analysis.

2.6.2. High-content imaging with Opera Phenix

2.6.2.1. Fluorescent staining and cell immobilisation

2.6.2.1.1. Sridha's method

An imaging plate (CellCarrier Ultra microplate, PerkinElmer, Germany) was coated with vitronectin (Stem Cell Technologies, UK) one night prior to the experiment. Briefly, 50 μ L of 0.01mg/mL vitronectin in dilution buffer (Stem Cell Technologies, UK) was added to each well of the imaging plate. The plate was incubated overnight at 37°C. On the following day, the wells were wash with the dilution buffer (50 μ L/well). Next, 50 μ L of bacterial suspension was added to the experimented wells of the imaging plate. The plate was statically incubated at 37°C for 10 minutes, after which the bacterial cultures were aspirated and fixed in 50 μ L of 4% PFA for 15 minutes and washed with 50 μ L PBS. The fixed cell was stained with 100 μ L of fluorescence dye mixture (1 μ M syto 41 + 1 μ M sytox green + 4 μ g/mL FM4-64 in HBSS) for 20 minutes in the dark, followed by a wash with 50 μ L PBS. Finally, 50 μ L of PBS was added to each imaging well.

2.6.2.1.2. "Sandwich" method

Prior to fluorescent staining, an imaging plate (CellCarrier Ultra microplate, PerkinElmer, Germany) was coated with a thin layer of gelatine. The coating was performed by adding 50 μ L of 2% gelatine (Bio-Rad, UK) to the individual wells of the imaging plate and then aspirating out the gelatine solution; the plate was incubated for at least three hours at room temperature (or incubated overnight at 4°C). Bacterial suspension in deep-well microplate was centrifuged at 3,000rpm (~1230g) for five minutes; supernatant was decanted. To fix bacterial cells, 50 μ L of 4% PFA was added to each well. The plate was shaken for two minutes at 2,000rpm, incubated at room temperature for 3 minutes, and centrifuged for 10 minutes. PFA was removed by tapping the plate onto a paper towel. The bacterial cell pellets

were washed with PBS (1mL/well) to quench the fixing reagent. Next, the cell pellets were resuspended in 100 μ L of fluorescence dye mixture (1 μ M syto 41 + 1 μ M sytox green + 4 μ g/mL FM4-64 in HBSS) and transferred to the precoated imaging plate. Photophysical properties of the fluorescence dyes were described in Table 2-1. The plate was incubated in the dark, at room temperature for 10 minutes, followed by centrifugation at 1,000rpm (~137g) for 10 minutes. Dye mixture was aspirated out; the imaging plate was centrifuged again for two minutes at 1,000rpm (~137g). The bacterial cells were immobilised on the bottom of the imaging plate by addition of 50 μ L melted agarose (1%) to each well. The imaging plate was sealed with aluminium foil (ThermoFisher, UK); image acquisition was performed on Opera Phenix high-content imaging platform (PerkinElmer, Germany).

2.6.2.2. Image processing

The workflow for image analysis goes from the top to the bottom of Figure 5-7. For maximum projection, a fluorescence z-stack (3D image) was transformed into a single (2D) image. In that, each pixel on the projected (2D) image took the highest intensity from corresponding pixels on the stacked images. For cell segmentation, bacterial cells were detected based on the stained nucleoids on syto 41 channel. At this stage, the workflow detected thousands of bacterial cells per experimental condition. Processing all of them is computationally consuming. Therefore, cells with averaged intensity lower than 1,000 were filtered out to remove low-signal cells and allow high-quality images of cells entering the further steps. That process reduced computing resources required for feature extraction at the end of the workflow. The remaining cells were used to segment cell membrane on FM4-64 channel and nucleoids on syto 41 and sytox green channel. Segmented cell components were then subjected into a series of morphological operators, including dilation and erosion, to smooth the boundaries. A small set of texture features was extracted from the segmented cell

components. Feature extraction is a process that transforms an image object into a numerical data entry (a row of data). The extracted features were used to build a linear classifier to remove cells with irregular shape. The purpose of this step is to reduce the computing cost in the next step. Then, a full set of features, including morphology, intensity and texture, were extracted on the remained cells. Finally, a linear classifier was employed again to discard irregular-shape cells based on the full features.

2.6.3. Integrating imaging and omics to infer antimicrobial susceptibility: method development

Frozen *E. coli* stocks were cultured on LB plates and incubated overnight at 37°C. On the next day, three single colonies, corresponding to three biological replicates, for each isolate was inoculated in 1mL of MHB2 (Milipore, Germany) in a deep-well microplate and incubated overnight, at 37°C, 200rpm. The bacterial suspensions were diluted 1/1,000 (v/v) with fresh MHB2 in two new deep-well plates (1,350µL/well); one was for intact protein analysis, and another was for HCl. The two microplates were incubated for 2 hours, at 37°C, 200rpm. Next, 150µL 10x stock of three antimicrobials were added to the bacterial suspension, followed by incubation for 3 hours, at 37°C, 200rpm. The three antimicrobials included CIP, AZI, and TE; final concentrations were at CLSI breakpoints. The experiment layout was described as in Figure 2-1. Finally, one deep-well microplate was used for analysis of intact proteins (section 2.6.1.3) and the remaining plate went through HCl for analysis of bacterial population (section 2.6.2).

		Columns											
Treatment	Rows	1	2	3	4	5	6	7	8	9	10	11	12
CIP	A	Blank	ATCC 25922	CM3585	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
AZI	B				Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
TE	C				Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
No drug	D				Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
CIP	E	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
AZI	F	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
TE	G	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3
No drug	H	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3	Replicate 1	Replicate 2	Replicate 3

Figure 2-1. Plate layout for HCI and intact protein analysis of pathogenic *E. coli* in one batch of experiment

White wells, on row A, B, C, and D, were reserved for standardised controls, including the blank on column 1, the susceptible control (ATCC 25922) on column 2, and the resistant control (CM3585) on column 3. Blue, orange, green, red, violet, brown, and pink wells contained bacterial suspension of seven different experimented isolates. For each isolate, experiments on three biological replicates were performed. Row A and E were treated with CIP; row B and F were treated with AZI; row C and G were treated with TE.

2.6.4. Molecular modelling and gene network analysis

The X-ray crystallography structure of CydABXH (Uniprot ID: 6RKO) and KdpFABC (Uniprot ID: 5MRW) complexes were retrieved from the Uniprot database. AlphaFold²³⁵ was used to predict 3D structure of hypothetical proteins whose X-ray crystallography has yet been solved. The predicted 3D structures were input to deepFRI²³⁶, a deep learning web-based application, to predict the molecular and cellular functions of those proteins. AutoDock Vina²³⁷ was used with default parameters to predict the interaction energy between a CIP molecule and OmpG porin protein (Uniprot ID: 2F1C). Molecular visualisation of all proteins has been done with PyMOL. On STRING database²³⁸, gene interaction network was constructed by querying all the genes related to the two proteins YnhF and KdpF. Results from text mining were excluded to ensure high confidence of the network inference.

Chapter 3. Genomic, physiological and morphological characterisation of pathogenic *Escherichia coli* to study antimicrobial resistance

Study conceptualisation and experiment design were performed by me. Recovery of isolates and susceptibility test were done by me with training and support from Ha Thanh Tuyen, Le Minh Triet, and Tran Hanh Triet. Microscopic observation of bacteria was carried out by me and Ha Thanh Tuyen. Growth dynamics was done by me, Ha Thanh Tuyen and Tran Hanh Triet. WGS was performed by Ha Thanh Tuyen and Ho Ngoc Dan Thanh. I analysed data and wrote the manuscript, which was published in Antibiotics as the following.

Tuan-Anh, T. et al. Pathogenic *Escherichia coli* Possess Elevated Growth Rates under Exposure to Sub-Inhibitory Concentrations of Azithromycin. *Antibiotics* 9, 735 (2020).

3.1. Introduction

Although culture-based AST is the gold standard approaches to identify and describe resistant bacteria, it provides no insight into the molecular and cellular response triggered by antimicrobial exposure. There have been previous investigations into the molecular mechanisms of AMR in specific bacteria ¹⁰⁶ and their morphological ^{239–241} and physiological characteristics ²²² under antimicrobial treatments. However, there are few studies directly linking molecular mechanisms of AMR with the morphology and physiology of antimicrobial resistant organisms. Given the high burden of resistant *E. coli* in LMICs and the well-established studies on physiology and genomics of the bacterium, I chose *E. coli* as a model organism for this study. I designed a holistic investigation to provide an in-depth

understanding of AMR in bloodborne *E. coli* from both a cellular and molecular perspective using microscopic techniques and whole genome sequencing. Additionally, I studied the growth dynamics of pathogenic *E. coli* to measure the physiological changes of *E. coli* with a diverse compendium of AMR genes upon short-term exposure to different antimicrobial classes.

3.2. Resistome and morphology associated with AMR

Using AST, I found that highly resistant pathogenic *E. coli* was prevalent in blood samples from children with sepsis. Particularly, all the clinical isolates were resistant to at least one antimicrobial (Table 3-1). Among those, 70.4%, 70.4%, 88.9%, 74.1%, and 64.0% of the *E. coli* isolates were resistant to CIP, AZI, CRO, TE, and CN, respectively (Table 3-1). Additionally, 22/27 (81.5%) isolates were resistant to at least three different classes of antimicrobials and defined as MDR²⁴².

To investigate the AMR gene profiles of the collection, WGS was performed on the 27 clinical *E. coli* isolates. Genomic analysis was conducted on the WGS data and the complete genome of *E. coli* ATCC 2522 obtained from NCBI database. The collection showed an appreciable diversity in genomic background of pathogenic *E. coli*. In total, the 27 clinical isolates fell into 12 different sequence types (STs) (Figure 3-1b). The most ST was ST1193, which accounted for 8/27 isolates (29.6%).

Table 3-1. MICs and antimicrobial resistance of ATCC 25922 and 27 clinical isolates used in this study

Isolate ID	Minimum inhibitory concentrations (mg/L)				
	CIP	AZI	CRO	TE	CN
E3585	32 (R)	48 (R)	256 (R)	256 (R)	256 (R)
E3823	32 (R)	48 (R)	256 (R)	256 (R)	256 (R)
E4569	0.064 (S)	12 (S)	32 (R)	4 (S)	1 (S)
E6348	0.38 (S)	8 (S)	256 (R)	4 (S)	256 (R)
E8964	0.38 (S)	64 (R)	256 (R)	256 (R)	64 (R)
E9833	0.38 (S)	16 (S)	256 (R)	256 (R)	1 (S)
E10085	32 (R)	48 (R)	256 (R)	16 (R)	256 (R)
E10487	32 (R)	256 (R)	256 (R)	256 (R)	96 (R)
E10996	0.094 (S)	16 (S)	256 (R)	256 (R)	1 (S)
E11030	32 (R)	4 (S)	0.094 (S)	256 (R)	64 (R)
E12236	0.25 (R)	256 (R)	1.5 (I)	256 (R)	1 (S)
E12241	0.094 (S)	8 (S)	0.064 (S)	128 (R)	1 (S)
E12674	32 (R)	6 (S)	64 (R)	128 (R)	1.5 (S)
E14252	12 (R)	16 (S)	256 (R)	256 (R)	32 (R)
E14488	32 (R)	256 (R)	256 (R)	4 (S)	16 (R)
E15475	32 (R)	256 (R)	256 (R)	256 (R)	1 (S)
E15476	32 (R)	256 (R)	256 (R)	256 (R)	1.5 (S)
E15583	32 (R)	256 (R)	256 (R)	256 (R)	1 (S)
E1456	0.25 (S)	48 (R)	256 (R)	64 (R)	256 (R)
E1965	32 (R)	256 (R)	256 (R)	4 (S)	256 (R)
E2408	32 (R)	256 (R)	256 (R)	4 (S)	64 (R)
E2542	32 (R)	256 (R)	256 (R)	4 (S)	256 (R)
E4751	32 (R)	256 (R)	256 (R)	256 (R)	256 (R)
E5306	0.38 (S)	48 (R)	256 (R)	256 (R)	256 (R)

E5610	32 (R)	256 (R)	256 (R)	4 (S)	256 (R)
E5896	0.38 (S)	256 (R)	128 (R)	256 (R)	256 (R)
E6227	32 (R)	256 (R)	256 (R)	256 (R)	1.5 (S)
ATCC25922	0.094 (S)	4 (S)	0.125 (S)	4 (S)	1.5 (S)
Susceptible	8/27 (29.6%)	8/27 (29.6%)	2/27(7.4%)	7/27 (25.9%)	10/27 (37.0%)
Intermediate	0/27 (0.0%)	0/27 (0.0%)	1/27 (3.7%)	0/27 (0.0%)	0/27 (0.0%)
Resistant	19/27 (70.4%)	19/27 (70.4%)	24/27 (88.9%)	20/27 (74.1%)	17/27 (63.0%)

S: susceptible, I: intermediate, R: resistant

A resistome analysis demonstrated that resistance to AZI was associated with the presence of *mphA* (Figure 3-1b). Resistance to TE was associated with the *tet* gene variants (*tetA/tetR* or *tetB*). All isolates harbouring *aac(3)-IIa* were resistant to gentamicin. In contrast, the resistome data was insufficient for explaining CIP resistance, which is commonly attributed to mutations in the QRDR⁶⁵. Eleven and eight of 27 clinical isolates respectively harboured *bla*_{CTX-M-1} and *bla*_{CTX-M-9}, which, as extended spectrum beta lactamases, confer resistance to third generation cephalosporins. Ceftriaxone MICs were positively correlated with the number of acquired resistance genes to beta-lactams in these isolates (i.e., the more *bla* resistance genes the higher the MIC; $\rho = 0.66$, $p < 0.001$). Moreover, three isolates (ST69 and ST410) carried *bla*_{NDM-1}, which is responsible for carbapenem resistance.

Aiming to explore the impacts of AMR and AMR genes on morphological changes, the 28 *E. coli* (27 clinical isolates and the reference strain ATCC 25922) were subjected to antimicrobial exposure for three hours at CLSI breakpoint concentrations¹⁹³ and examined by light microscopy. The microscopy images were qualitatively classified into four principal categories: lysis, ellipsoid, rod-like shape, and elongation (Figure 3-1a). Figure 3-1b demonstrates that *E. coli* cells responded to antimicrobial exposure with differential modifications in their morphology, with some showing an association with a particular AMR phenotype. Specifically, CRO-susceptible and -intermediate resistant isolates were elongated while resistant isolates retained rod-like shape when exposed to ceftriaxone (Figure 3-1b). Similarly, the majority of *E. coli* that were resistant to AZI, TE, and CN retained a rod morphology, while most susceptible organisms were ellipsoid when exposed to these antimicrobials (Figure 3-1b).

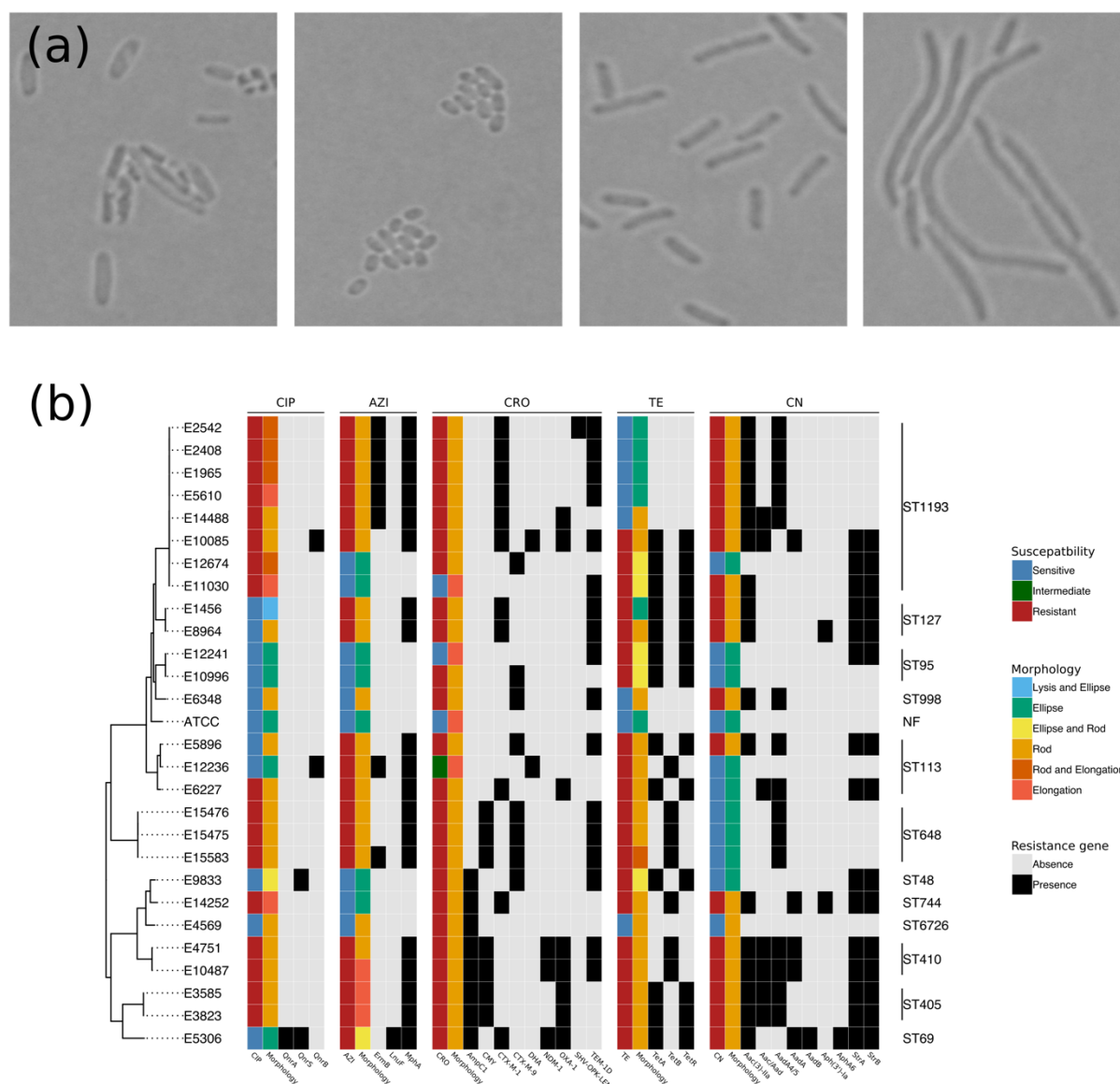


Figure 3-1. Bacterial morphology associated with AMR and resistome.

(a) The morphological features of *E. coli* after 3h being treated with ciprofloxacin.

Brightfield images of *E. coli* in different morphologies, including, from left to right, cell lysis, ellipsoid, rod-like shape, and elongation, were acquired with a Nikon Ni-E upright microscope and a 100X objective lens. Morphology of the reference strain is in rod-like shape in drug-free condition. (b) The phylogenetic structure of *E. coli* with corresponding

susceptibility, AMR genes, and morphology. (Left) The unrooted phylogenetic tree constructed from core-genomes of the 28 *E. coli* isolates. (Right) Sequence types of those isolates obtained by mapping the whole genome reads to a database of seven housekeeping genes. (Middle) Five heatmap blocks, each of which demonstrates (first column) antimicrobial susceptibility, (second column) morphology under exposure to antimicrobials, and (remaining columns) the presence (black) and absence (grey) of resistance genes against quinolones, macrolides, β -lactams, tetracyclines, and aminoglycosides. NF: non-identified ST.

Upon exposure to AZI and TE, morphological changes were more heterogeneous and not consistent with the respective susceptibility profiles of the organisms. For example, of the eight TE-resistant *E. coli* isolates, five were ellipsoid and three were rod-like shape. Ciprofloxacin exposure induced the highest level of morphological heterogeneity, in which resistant organisms were rod-like shape, elongated or rod-elongate mixture; CIP susceptible organisms were ellipsoid, rod shape or lysed when exposed to CIP. I did not observe any obvious association between AMR gene content and bacterial morphological changes under exposure to antimicrobials (Figure 3-1b), with the exception of the aforementioned rod morphology with *mphA*, *aac(3)-IIa* gene and *tet* gene variants conferring resistance to AZI, CN and TE, respectively.

3.3. Assessing the growth dynamics of *E. coli* under exposure to antimicrobials

The 28 *E. coli* isolates were exposed to different concentrations (0 to 256mg/L) of CIP, AZI, CRO, TE, and CN; growth curves were performed over 24 hours. One isolate was chosen to perform three technical replicates (Figure 3-2) and the pair-wise correlation coefficients among three replicates were calculated. An average correlation coefficient of 0.89 was determined. Additionally, each isolate was exposed to AZI and TE with additional 1-2 biological replicates to confirm the observed effects of resistance on their growth dynamics.

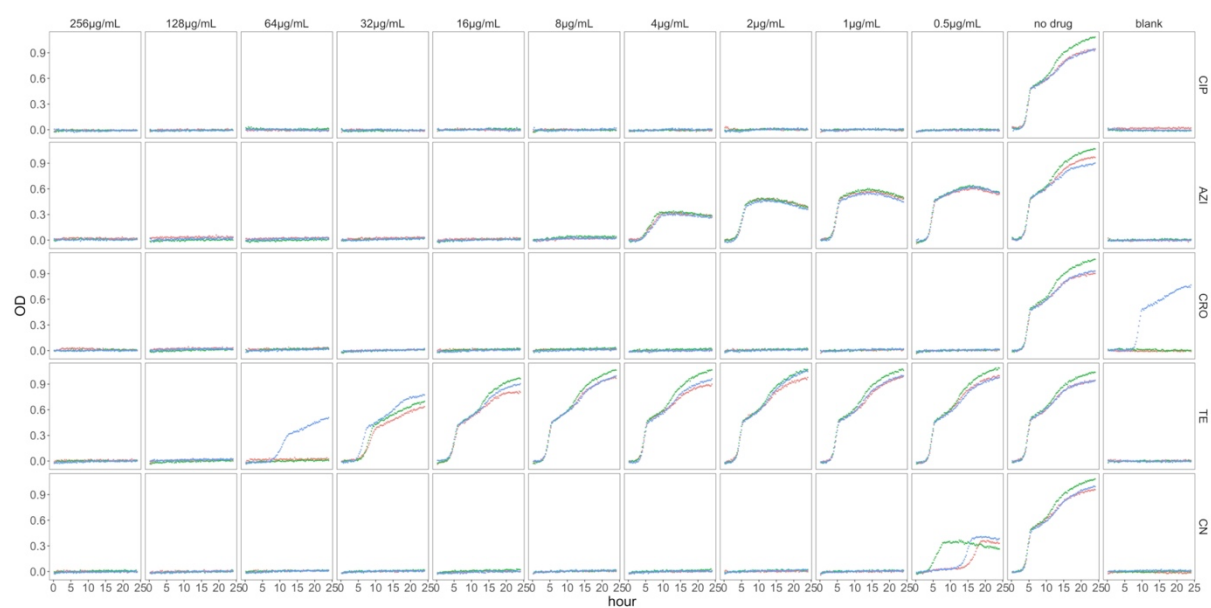


Figure 3-2. Three replicates of growth dynamics monitoring

A clinical *E. coli* isolate was treated, three times, with CIP, AZI, CRO, TE, and CN at different concentrations. Red dots: replicate #1, green dots: replicate #2, and blue dots: replicate #3.

More than 2,500 growth curves from the 28 *E. coli* in different antimicrobial exposure conditions were generated, and Gompertz's, Baranyi's and three-phase linear model were fit to the growth curves. The Gompertz's and Baranyi's model fit with the lowest mean MSE and MAE; the highest ρ and R^2 (Figure 3-3). However, performance of the Gompertz's model was superior to that of the Baranyi's model, which indicates that the Gompertz's (with three parameters) was the most suitable model for these data. Consequently, I employed the Gompertz's model to estimate the growth parameters of each curve for further analysis.

The growth features were plotted separately for five experimental conditions, and against different antimicrobial concentrations as relative to the isolate's respective MIC (Figure 3-4). Generally, antimicrobials suppressed the growth of *E. coli*, which could be observed as the inferred A reduced gradually and AUC reduced rapidly with increasing drug concentrations. Alternatively, increased antimicrobial concentrations induced a prolonged λ in all cases. The μ against relative drug concentration was more heterogeneous and largely dependent on antimicrobial class. The μ of *E. coli* treated with CIP and TE declined inversely with respect to an increase in the respective antimicrobial concentrations. However, when organisms were exposed to ceftriaxone, the μ remained constant, with some organisms showing increased μ at higher concentrations. Notably, under AZI exposure, the general trend of the 28 *E. coli* was that μ increased in lower concentrations of the macrolide and then decreased gradually at higher drug concentrations. These data suggest that sub-inhibitory concentrations of AZI induce an increased rate of *E. coli* replication.

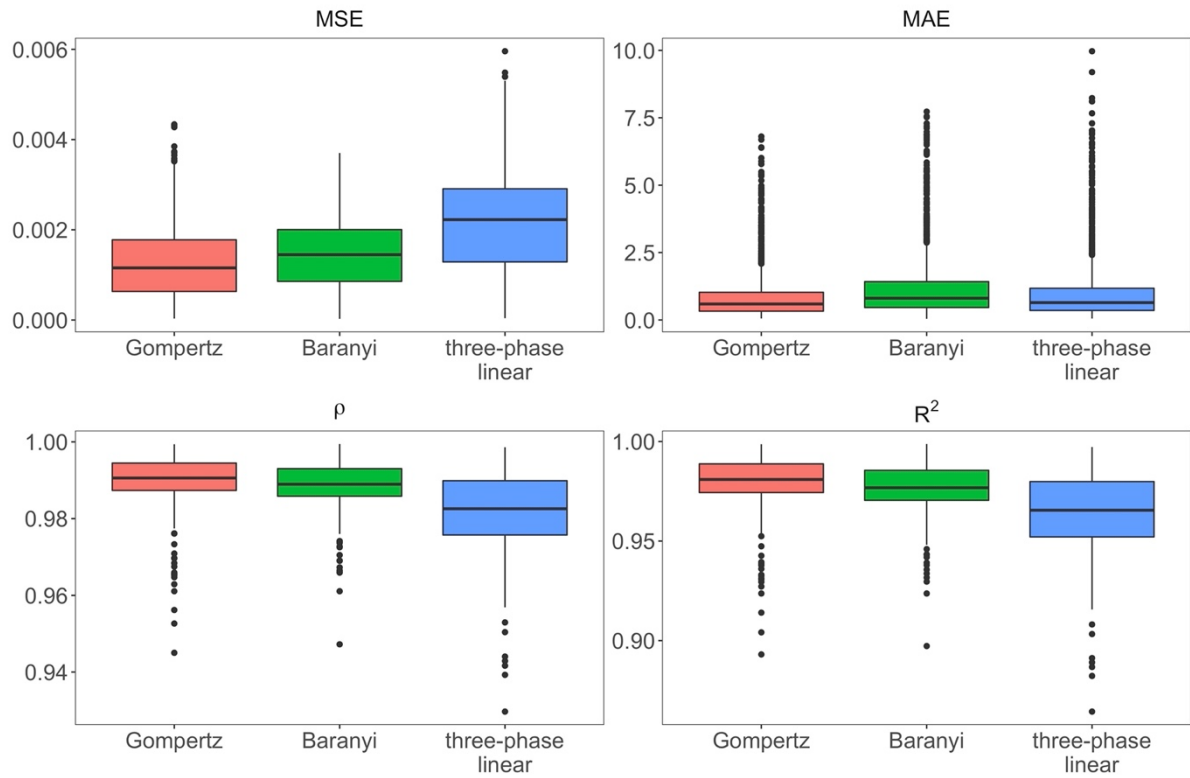


Figure 3-3. Comparison between performance of Gompertz's, Baranyi's and three-phase linear model.

For each growth curve, the three models were utilised to predict OD values at every timepoints from 0 to 24h. The actual and predicted values were used to calculate the MSE, MAE, ρ , and R^2 . The lower MSE and MAE; higher ρ and R^2 , the better a model is.

On further investigation I also found that TE-resistant isolates possess a fitness advantage over susceptible isolates in both antimicrobial-free conditions and at 0.5×MIC (Figure 3-5). Specifically, at 0.5×MICs of TE, μ ($p < 0.0001$), A ($p < 0.001$), and AUC ($p < 0.0001$) of the TE resistant *E. coli* were all significantly higher than those of the susceptible organism (Figure 3-5). Without antimicrobial exposure, the AUC of TE-resistant isolates were significantly higher than that of susceptible isolates ($p < 0.05$). Conversely, AZI susceptible *E. coli* in drug-free conditions had a fitness advantage over AZI resistant *E. coli* with significantly higher AUCs ($p < 0.05$) (Figure 3-5).

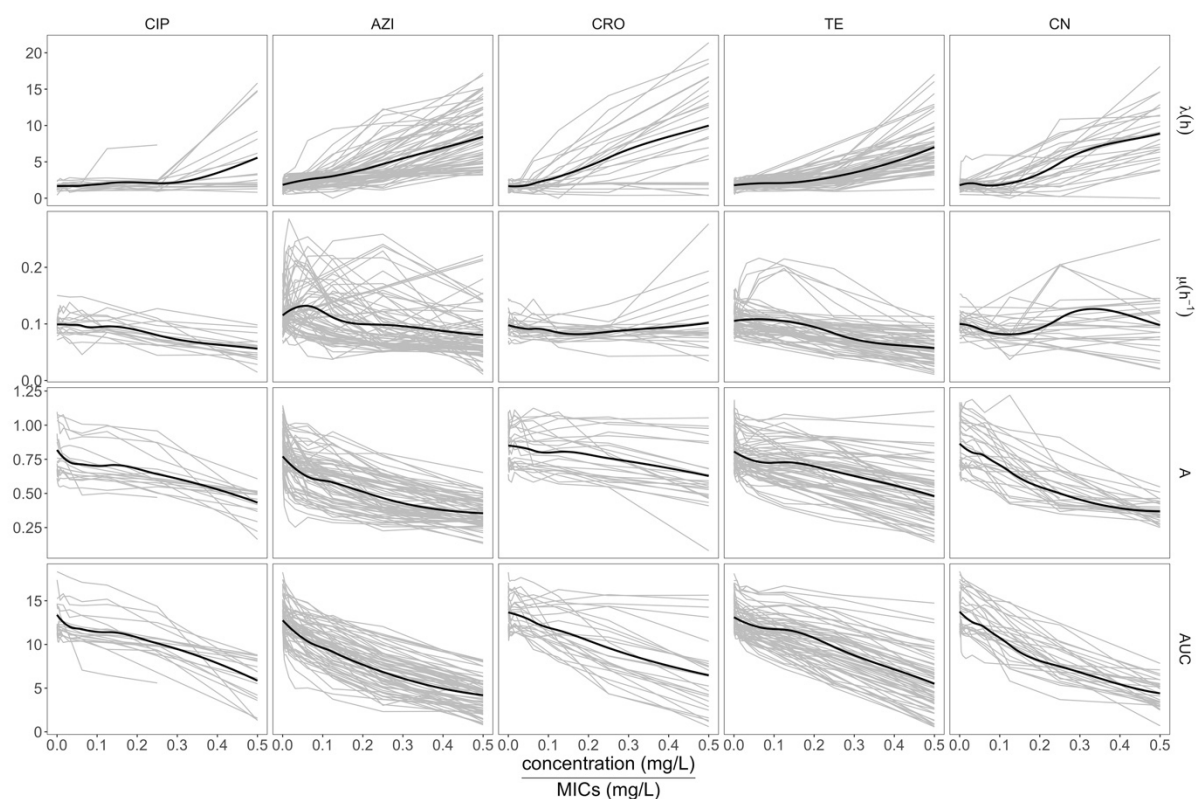


Figure 3-4. The dynamics of the growth features

From top to bottom, lag-phase period (λ), maximum growth rate (μ), maximum cell density (A), and area under the curve (AUC) corresponding to antimicrobial concentration relative to MICs of 28 *E. coli* isolates from clinical samples. From left to right, CIP, AZI, CRO, TE, and CN. The horizontal axis is antimicrobial concentrations relative to MICs of each isolate. The vertical axis represents the values of growth features. Bold lines are LOESS regression and thin lines are growth features of each isolate.

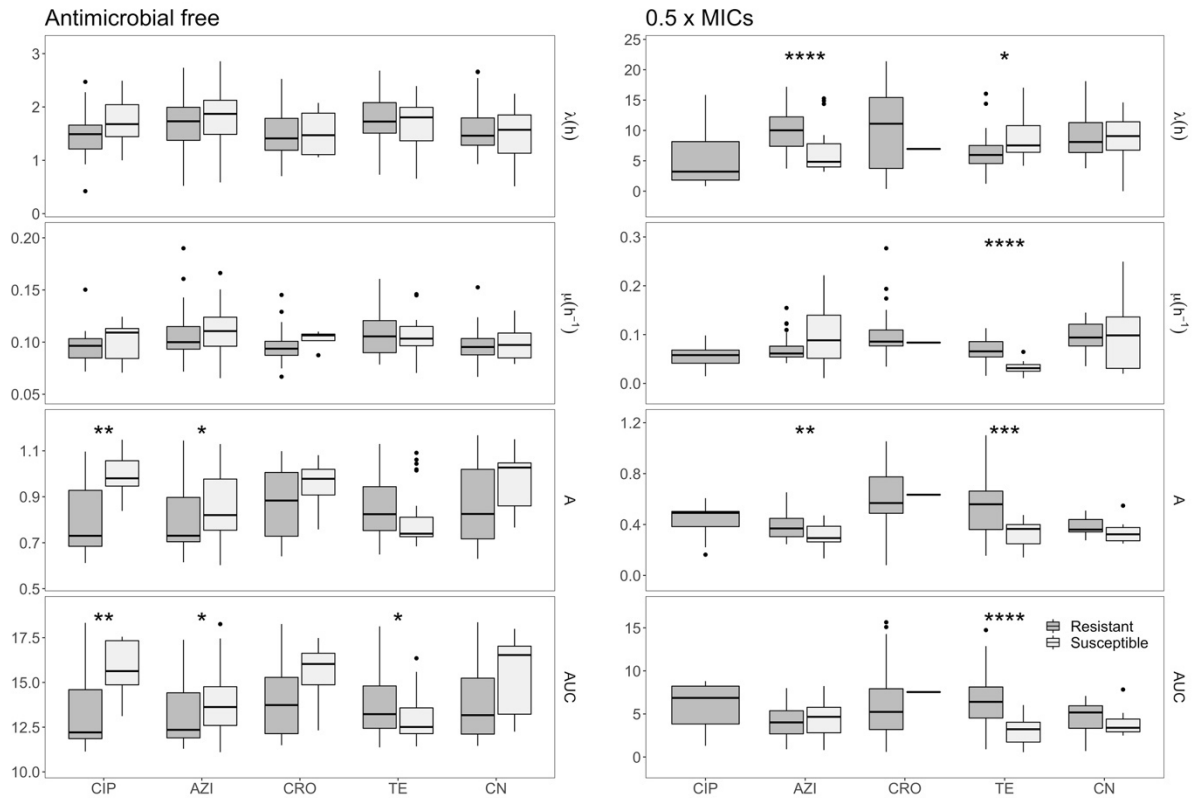


Figure 3-5. A comparison between the growth features for susceptible isolates and those for resistant isolates.

From top to bottom, maximum cell density (A), maximum growth rate (μ), lag-phase period (λ), and area under the curve (AUC) of resistant isolates (dark grey) and those of susceptible isolates (light grey) without antimicrobial (left) and at half MICs (right). *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001, ****: p-value < 0.0001.

3.3.1. Effects of azithromycin and tetracycline resistance on growth dynamics of *E. coli*

To further investigate the impact of AZI and TE resistance on the growth dynamics of *E. coli*, I plotted the growth features of resistant and susceptible isolates at different relative concentrations of these antimicrobials. Figure 3-6 shows a comparison between trends of growth features (as inferred by Gompertz's model) on data from three replicates of resistant and susceptible organisms grown in increasing concentrations of AZI and TE. I found that the λ of TE-resistant organism was lower than that of the susceptible organisms, while the μ and AUC of organisms resistant to TE were higher than those of susceptible organisms. Conversely, the AZI-resistant organisms exhibited a longer λ , a higher A , and a lower μ than the AZI susceptible organisms. Additionally, I found that μ of both AZI resistant and susceptible *E. coli* peaked at low antimicrobial concentrations relative to MIC, which was supported by both Gompertz's, Baranyi's and three-phase linear model (Figure 3-7). These data suggest that sub-inhibitory concentrations of AZI accelerate the growth of both susceptible and resistance *E. coli*.

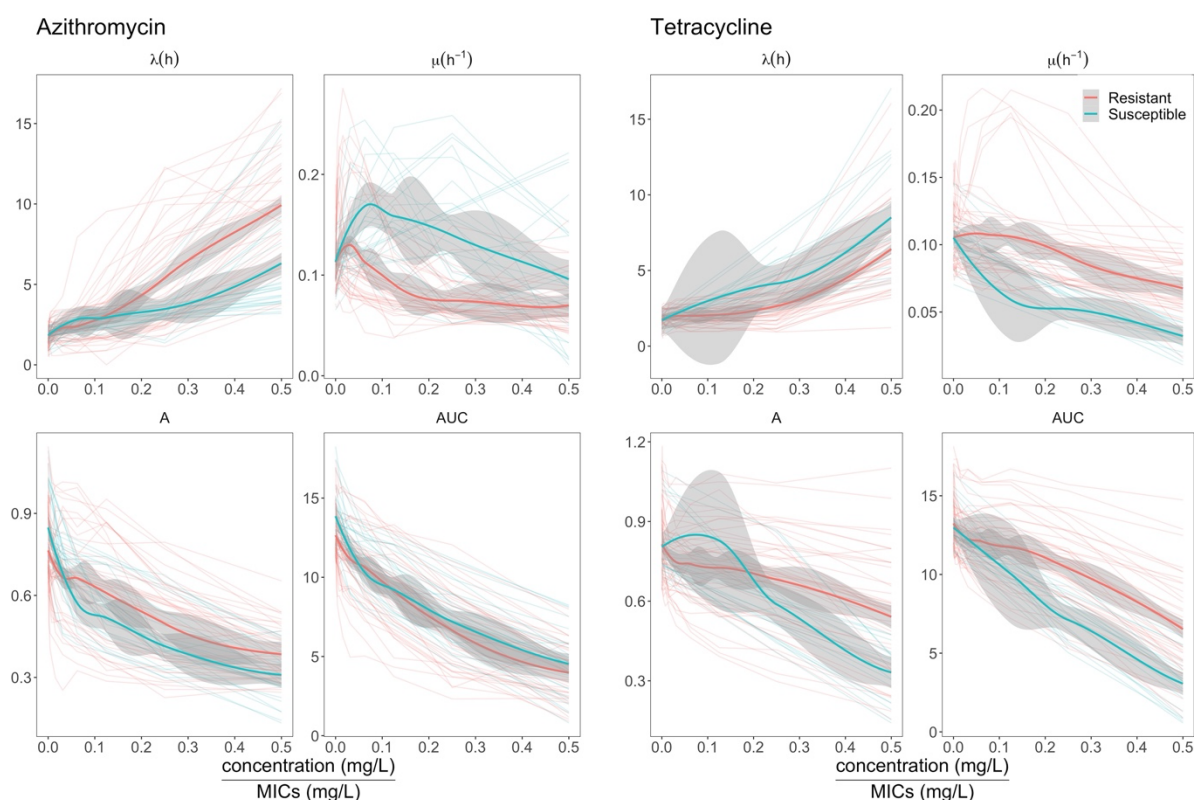


Figure 3-6. A comparison between growth features of resistant and susceptible isolates under exposure to azithromycin and tetracycline

The horizontal axis is antimicrobial concentrations relative to MICs. The vertical axis is values of the four growth features, including lag time (λ), maximum growth rate (μ), maximum cell density (A), and area under the curve (AUC). For each isolate, growth features were plotted against relative concentrations of antimicrobials as thin lines. Bold red (resistant isolates) and blue (susceptible isolates) lines indicate LOESS regression. Grey shading represents 95% confident intervals.

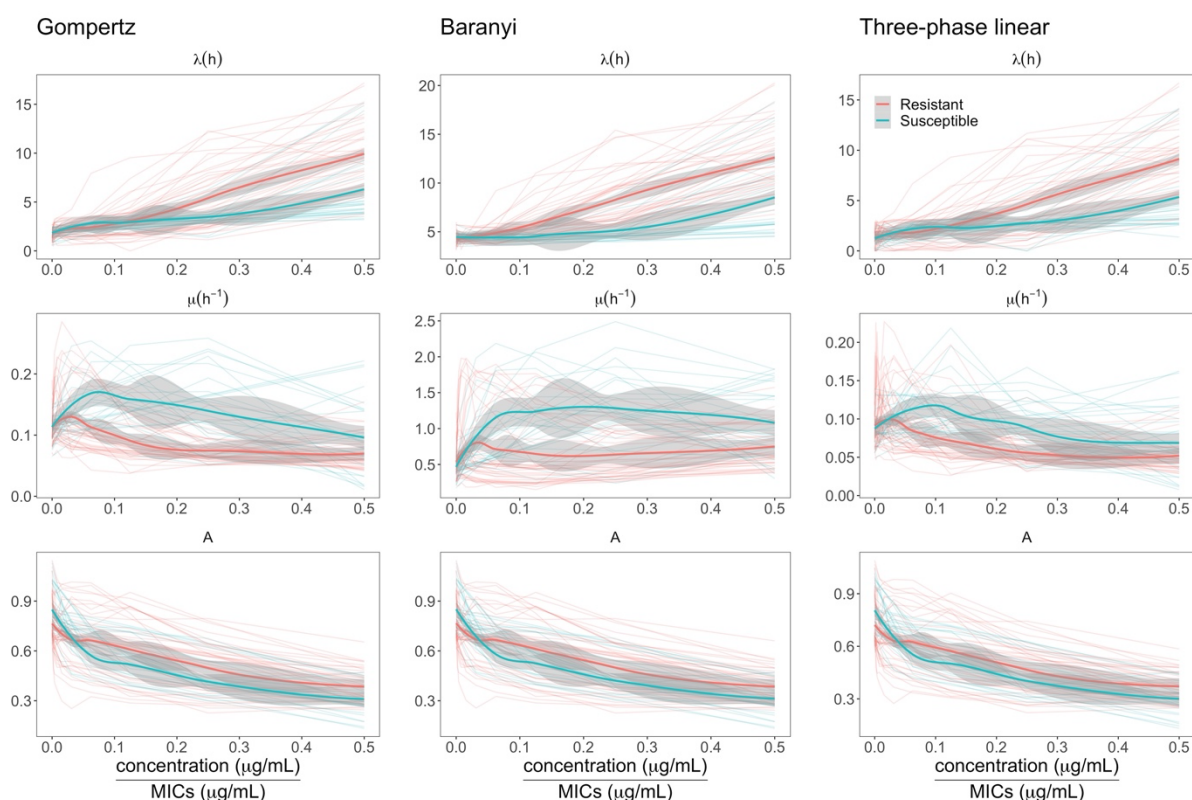


Figure 3-7. Comparison between growth parameters estimated from Gompertz's, Baranyi's and three- phase linear model.

Parameters include lag-phase period (λ), maximum growth rate (μ), and maximum cell density (A). The horizontal and vertical axes are relative concentration of antimicrobials and values of the growth parameters, respectively. Bold red (resistant isolates) and blue (susceptible isolates) lines are LOESS regression.

3.3.2. Elevation of growth rate in *ermB*⁻ *E. coli* at sub-MIC concentrations of azithromycin

All AZI-resistant *E. coli* isolate (n=19) carried *mphA*; additionally, seven harboured *ermB* and one isolate harbored *lnuF*. In order to investigate the influence of *ermB*, which was prevalent among *E. coli* resistant to AZI, on the growth features of *E. coli*, I compared growth features of the *E. coli* with the presence (*ermB*⁺) and absence (*ermB*⁻) of this gene. The μ , A , and AUC decreased in both *ermB*⁺ and *ermB*⁻ isolates, while λ was elevated with increased relative concentrations of AZI (Figure 3-8a). I found that *ermB*⁺ *E. coli* exhibited a higher A and AUC, shorter λ , and lower μ than those of *ermB*⁻ *E. coli*. At low concentrations of AZI relative to MICs, μ in *ermB*⁻ isolates increased rapidly (peaked at $\sim 0.03 \times \text{MIC}$) and then gradually decreased. The highest μ value from the *ermB*⁺ organisms was lower than that from *ermB*⁻ organisms, which peaked at a greater relative concentration of AZI. I further found that at lower concentrations of AZI (absolute concentrations of 1mg/L and 2mg/L), the *ermB*⁻ organisms exhibited significantly higher μ than those of *ermB*⁺ organisms (Figure 3-8b), which was in concordance with data generated with relative AZI concentrations. Lastly, I found that MICs of *ermB*⁺ isolates were significantly higher than those of *ermB*⁻ isolates ($p < 0.05$) (Figure 3-8c). My data show that MICs against AZI are associated with the accumulation of macrolide resistance gene.

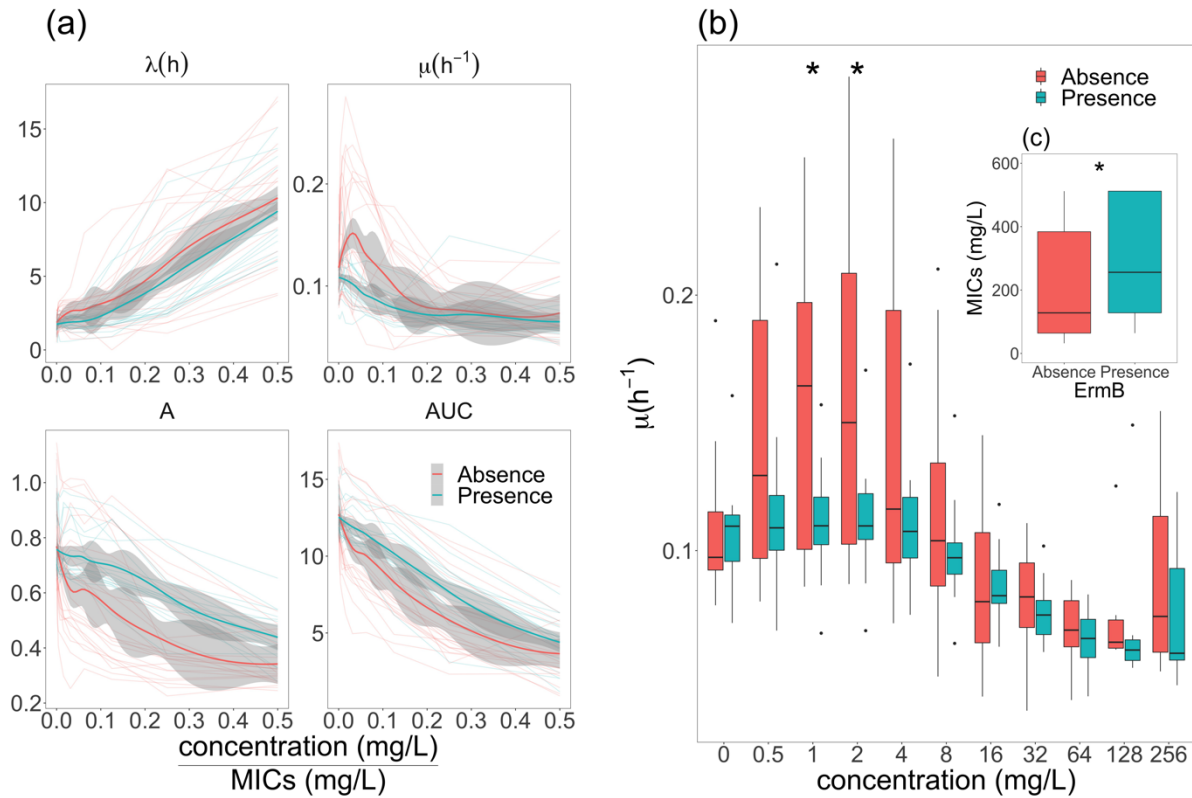


Figure 3-8. The impact of *ermB* gene on growth dynamics of azithromycin-resistant *E. coli*.

(a) Growth features of *E. coli* in the presence (blue thin lines) and absence (red thin lines) of *ermB*. Vertical axis is values of λ : lag time, μ : maximum growth rate, A: maximum cell density, and AUC: area under the curve. Horizontal axis is antimicrobial concentration relative to MICs. Bold red and blue lines are LOESS regression. Grey shading is 95% confident intervals. (b) Maximum growth rate (μ) of *ermB*⁻ *E. coli* (red boxes) compared to that of *ermB*⁺ *E. coli* (blue boxes). Horizontal axis is the absolute concentration of AZI (mg/L). (c) Comparison MICs of *ermB*⁺ and *ermB*⁻ *E. coli*. *: p-value < 0.05.

3.4. Discussion

This study systematically assessed the physiological and morphological responses of AMR *E. coli* to clinically relevant antimicrobials. Regarding morphology, I found a relationship between resistance to ceftriaxone and cell elongation, which has been previously reported^{243–248} and also utilized to rapidly detect resistant isolates^{239,240}. However, I found that the molecular basis of AMR was not associated with antimicrobial-induced morphological responses. This observation suggests that morphology-based rapid antimicrobial susceptibility testing may be performed without prior knowledge of resistance mechanism, which is a drawback of nucleic acid-based and protein-based methods. A limitation of morphology-based testing is the highly diverse dependence of that morphological changes on types of antimicrobials, even for antimicrobials belonging to the same family. Considering the β -lactams as an example, meropenem induces a spindle morphology in *E. coli*, while moxalactam induces filamentous cells¹⁶⁴. Thus, for each antimicrobial, a prior imaging dataset of both resistant and susceptible organisms is required to infer the susceptibility of newly isolated organisms. For other antimicrobials used in this study, the morphology-resistance correlation was less consistent, which may be addressed by additional time data provided by live-cell imaging, which may be required to build highly accurate antimicrobial susceptibility tests^{240,249}. The heterogeneous responses to CIP, AZI, TE, and CN might be explained by the highly diverse genetic background of the *E. coli* collection comprising of 11 different STs. My results showed antimicrobial-induced morphological changes in both resistant and susceptible isolates, suggesting that bacterial cells response to drug pressure despite the resistance to antimicrobials. This is probably due to the residue, in cytoplasm, of active antimicrobials which is not yet extruded by efflux pumps (e.g. TetA and TetB) or inactivated by enzymes (e.g. β -lactamases). This could also be explained by the weak

interaction between antimicrobials (e.g. CIP) and modified target proteins (e.g. mutated GyrA).

Due to improved manufacturing technology, the price of tetracyclines has decreased dramatically, to the extent where they are the cheapest antimicrobials available. This has made TE the second most widely used antimicrobial worldwide²⁵⁰. Although the usage of tetracyclines in human health is becoming less common, they are still important antimicrobials disease prevention and growth promotion in animals¹⁹. My data demonstrated a high prevalence of TE resistant *E. coli* via the *tet* gene, which is in accordance with previous studies²⁵¹. By comparing growth features of *E. coli* treated with TE, I found that TE-resistant isolates possessed a fitness advantage over susceptible isolates. This fitness advantage in absence of the antimicrobial may explain the ubiquity of *tet* gene carriage in the gut microbiota of children and the sustained persistence of TE resistance²⁵¹. My observations concur with a pharmacodynamic model²⁵², in which the fitness of resistant *E. coli* did not differ from susceptible *E. coli* in antimicrobial-free conditions. However, a recent study using growth curve analysis found that TE resistance may exert a fitness cost of up to 10% on *E. coli* compared with a wild-type strain²⁵³.

Azithromycin has been listed by WHO as one of the critically important antimicrobial reagents for human medicine²⁵⁴. Moreover, the WHO have been developing guidelines to prophylactically prescribe AZI to infants and children to reduce child mortality²⁵⁵, which might increase the resistance against macrolide antimicrobials globally²⁵⁶. Therefore, the molecular mechanisms and prevalence of AZI resistance have been studied extensively^{255,257–260}. However, the influence of AZI exposure on the physiology of bacteria is not well understood. Here, I observed an accelerated μ of both resistant and susceptible isolates under

exposure to low concentrations of AZI, raising a concern about potential impacts of this antimicrobial on developing new resistant mutants. These results suggest further investigations are required to evaluate the potential benefits of presumptive antimicrobial treatments compared to the risk of AMR development. I speculate that at sub-MIC concentration, AZI molecules inhibit protein synthesis by binding to P site on the 50S subunit of the bacterial ribosome⁷⁶, inducing up-regulation of ribosome synthesis which is associated with higher μ ^{261,262}. With increased concentration of the macrolide, more ribosomal molecules are bound by AZI, which leads to insufficient protein synthesis and reduction in μ .

Notably, the trend of μ in AZI resistant *E. coli* was lower than the trend in AZI susceptible isolates. This observation may be experimental evidence supporting a mathematical model for ribosome-targeting antimicrobials, which showed that faster growth induces greater drug susceptibility²⁶³. Moreover, amongst the AZI-resistant *E. coli* harbouring *mphA* gene, the *ermB*⁺ isolates exhibited lower μ than those of *ermB*⁻. I hypothesize two feasible causal pathways of how *ermB* influences μ : (1) the presence/absence of *ermB* directly impacts on the μ trend or (2) the presence of *ermB* increases the MICs of *E. coli* and consequently changes the trend of μ against relative concentration of AZI. The lower μ in *ermB*⁺ may be due to reduced ribosome synthesis as *ermB* allows constitutive methylations of 23S rRNA²⁶⁴, which protects the complexes from AZI binding and potentially interferes with ribosome assembly. This suggests that the presence/absence of *ermB* is likely to have a direct impact on μ .

This study has limitations. First, using these approaches I were unable to develop an algorithmic tool to quantitatively measure morphological changes and to categorize cell

shapes. Therefore, the cell-shape classification may be subjective. Second, the relationship between morphology, growth dynamics and cell viability were not investigated. Third, the correlation between the live cell counts, measured CFUs and OD values, representing the density of both live and dead cells, was not verified. Although bacterial enumeration provides information on the actual growth rates of the cultured bacteria, I opted to analyse the growth of *E. coli* by monitoring OD as it enables high-throughput experiments with biological replications. Lastly, the 27 *E. coli* isolates possessed potential cross-interaction among AMR genes and high diversity in genetic background, which I was not able to address.

In conclusion, I exploited growth dynamics and bacterial morphology to study the impact of AMR in clinical *E. coli* isolates. The morphological responses to ceftriaxone were consistently associated with the resistance to the antimicrobial, while the responses to other antimicrobials was more heterogeneous. In the absence of the drug, TE resistance confers a fitness advantage on *E. coli* while the other antimicrobials exert a measurable fitness cost (lower AUC and μ) on these bacteria. At low concentrations of AZI, μ was accelerated and AZI susceptible isolates possessed higher μ than resistant isolates. I additionally observed an impact of AMR genes on growth dynamics of bacteria, in which *E. coli* harbouring *ermB* gene exhibited lower μ than *ermB*⁻ *E. coli*. My study raises a concern regarding the persistence of AMR and the development of new resistance phenotypes under antimicrobial exposure, showing the need for a better understanding of how morphological and physiological response contribute to the AMR and tolerance in bacterial pathogens.

Chapter 4. Application of machine learning with high-content imaging to infer antimicrobial resistance phenotypes in *Salmonella* Typhimurium

The study was conceptualised by Prof. Stephen Baker. Experiments were designed and conducted by Dr. Sushmita Shridha. Data analysis, construction of machine learning models and figure preparation were carried performed by me. I and Dr. Sushmita Shridha wrote the manuscript as the following.

Tuan-Anh, T. et al., Application of machine learning to high content imaging to infer antimicrobial resistance phenotypes in *Salmonella* Typhimurium. (in progress)

4.1. Introduction

Ciprofloxacin has been listed by WHO as a critically important antimicrobial reagents for human medicine ²⁵⁴. However, the rapid spread of (fluoro)quinolone resistance bacteria threatens the efficacy of the drug. The rapid detection of bacteria that are resistant to the drug can play a crucial role to preserve the availability of CIP. In Chapter 3, using a conventional brightfield microscope, I found an association between resistance to CIP and the morphological changes in *E. coli*. However, this association was not consistent across all the isolates and hence could not be used to infer susceptibility. Therefore, I attempted to use HCI, which is a more advanced microscopic technique, to observe the morphological responses of bacteria treated with CIP.

Aiming to better understand the relationship between cellular morphology and AMR, I sought to link HCI with machine learning algorithms to identify the key characteristics that

may predict how a bacterial population responds to an antimicrobial. To address this problem, I selected *S. Typhimurium*, a classical enteric pathogen, which can induce gastroenteritis, diarrhoea, and in some cases, systemic disease, as the model organism^{115–118}. Pertinently, *S. Typhimurium* is becoming increasingly resistant to key antimicrobials, such as the broad-spectrum fluoroquinolone, CIP. Selecting two laboratory typed strains and two clinically relevant *S. Typhimurium* isolates I exposed them to increasing concentrations of CIP over a 24-hour timeframe and periodically subjected them to HCl. After subjecting these data to detailed machine learning algorithms, I found that *S. Typhimurium* have distinct morphological characteristics that can be exploited to predict CIP susceptibility, without prior knowledge of phenotype, or exposure to the antimicrobial.

4.2. *S. Typhimurium* isolates display variable growth and morphological characteristics following ciprofloxacin exposure

I selected two clinical isolates and two isogenic laboratory strains of *S. Typhimurium* with a range of CIP susceptibilities (D23580, VNS20081, SL1344, and SL1344gyrA) to subject to sustained CIP exposure (Table 4-1). I performed time kill curves with 1X MIC CIP (and without antimicrobial), enumerating colony forming units (CFU) every two hours for 8 hours and after 24 hours (Figure 4-1a).

All four organisms displayed comparable growth dynamics in the absence of CIP, but predictably I observed greater variation in the growth trajectories between isolates at 1X MIC of CIP. Overall, growth was slower in the presence of CIP, with D23580 showing the greatest reduction in growth rate compared to other isolates and to D23580 without the drug (Figure 4-1a, right panel). Despite a reduced growth rate over the initial 8 hours of CIP exposure, all

isolates treated with 1X MIC CIP rebounded to a CFU comparable to that of no CIP treatment by the 24-hour timepoint. These dynamics show that bacterial isolates with a different genetic composition exhibit distinct growth trajectories in response to antimicrobial exposure. I hypothesized that the observed differences may be measured more precisely by assessing changes in cellular morphology; therefore, I subjected the bacteria to HCI to identify morphological differences associated with CIP exposure.

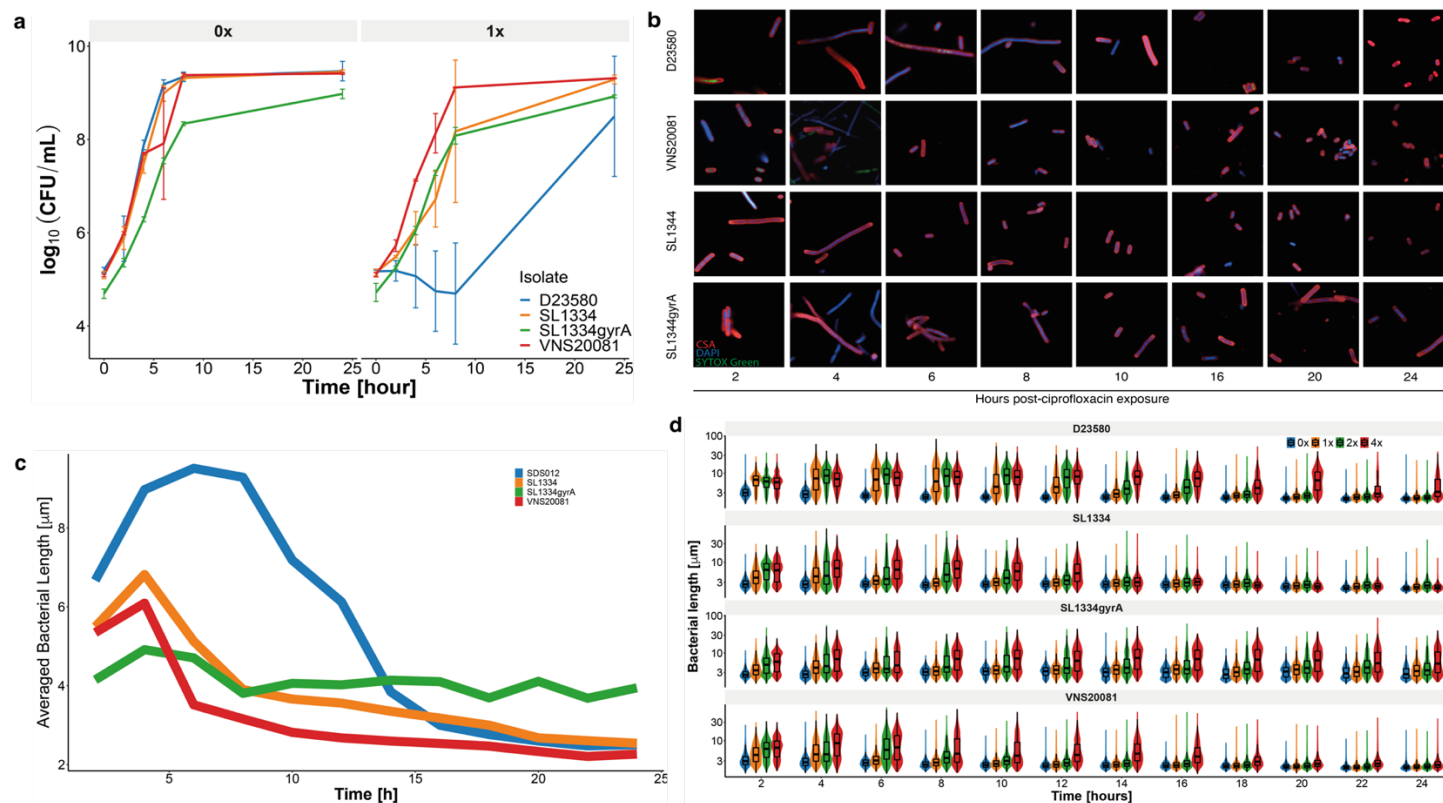


Figure 4-1. *S. Typhimurium* responses to ciprofloxacin treatment

(a) Time-killing curves of the four isolates at 0xMIC (left subplot) and 1xMIC (right subplot) against exposure time (horizontal axis). Vertical axis is the average of colony forming units in logarithm scale. Vertical bars represent standard deviation. (b) Morphological responses of *S. Typhimurium* isolates at 1xMIC against increasing exposure time (from left to right). CSA, SYTOX Green, and DAPI stains give red, green, and blue fluorescence, respectively. (c) Averaged length over all bacterial individuals at 1xMIC. Horizontal axis is exposure time in hours. Vertical

axis is averaged length in micrometres. Blue line is D23580, orange line is VNS20081, green line is SL1344, and red line is SL1344gyrA. (d)

Single bacterial length over time.

Table 4-1. Resistome analysis of the isolates in relation to fluoroquinolone susceptibility

Isolate	Ciprofloxacin Susceptibility	Ciprofloxacin MIC (µg/ml)	<i>gyrA</i> mutation	<i>gyrB</i>	<i>qnr</i>	<i>acrA/acrB</i>	<i>oqxA/oqxB</i>
SL1344	S	0.015	no	no	no	no	no
SL1344gyrA	R	1.5	D87Y	no	no	no	no
D23580	S	0.03	no	no	no	no	no
VNS20081	R	1.0	D87N	no	no	no	yes
VNB1779	I	0.75	D87N	no	no	no	yes
VNB2315	S	0.03	no	no	no	<i>acrB</i>	no
gha113289	I	0.75	no	no	no	<i>acrB</i>	no
gha200597	S	0.016	no	no	no	no	no
2101	S	0.032	no	no	no	<i>acrB</i>	no
16755_3	I	0.375	S83Y	no	no	no	no
5390_4	I	0.5	no	no	<i>qnrS</i>	no	no
8314_12	I	0.19	no	S464F	no	no	no
8599_13	S	0.023	no	no	no	no	no
319_8	I	0.25	D87Y	no	no	no	no
1304	S	0.023	no	no	no	no	no
D23580gyrA	I	0.5	D87G	no	no	<i>acrA</i>	no
10433_3	S	0.03	no	no	no	<i>acrB</i>	no

S: susceptible, I: intermediate, R: resistant. MIC: minimal inhibitory concentration.

The four organisms were again grown aerobically in liquid media at 0X, 1X, 2X, and 4X MIC CIP for 24 hours; samples were taken every two hours for imaging. Over the course of the experiment, I observed a temporal change in bacterial length under 1X MIC CIP treatment up to 24 hours (Figure 4-1b). This change in bacterial cell length was more pronounced between 2 and 8 hours, although SL1344*gyrA* could be distinguished by sustained elongation at 24 hours (Figure 4-1b, right-most panel).

My HCI platform captured 65 morphological, intensity, and texture features for each individual bacterium within the selected field-of-view¹⁵⁴, generating the power to quantify mean bacterial length for an entire field at each timepoint. At 1X MIC CIP, the mean cellular length peaked at $9.51 \pm 0.20 \mu\text{m}$ (mean $\pm$ standard error) within six hours, which was a predicted consequence of fluoroquinolone exposure²⁶⁵, and subsequently decreased below the initial length at time point 0 hours (Figure 4-1c). I observed several isolate-specific disparities: D23580 exhibited greater elongation within six hours, and SL1344*gyrA* exhibited only a modest decrease in mean cellular length after eight hours of exposure to 1X MIC of CIP. Notably, D23580 and VNS20081 grown with CIP supplementation displayed considerable heterogeneity in cell length over the course of the experiment, with the degree of cell length heterogeneity being dependent on the isolate, the CIP concentration, and also the exposure time (Figure 4-1d). Overall, the general trend across all isolates at 1X MIC of CIP was growth (as measured by CFU), a significant increase in bacterial length with time followed by a decrease, and a measurable change in the morphological characteristics of the exposed bacteria.

4.3. Development of a random forest classifier to discriminate between isolates and between growth conditions

I next sought to disaggregate the role of bacterial isolate and treatment condition. For every treatment condition (of exposure time and drug concentration), morphological, intensity, and texture features were averaged across individual bacteria on a single focal plane within a microwell. The averaged data was then collated as one datapoint, hereafter referred to as imaging data/features. That process resulted in 65 imaging features for each datapoint. Subsequently, I normalised the imaging data to a z-score, which was visualized on a heatmap plot with hierarchical clustering (Figure 4-2a). The resulting data was segregated into two main groups, with the majority of the 0X MIC and 1X MIC imaging datapoints in one major cluster and the 2X MIC and 4X MIC datapoints in another (Figure 4-2a). This clustering suggested that, of the factors measured, drug concentration had the greatest effect on the cellular response to CIP exposure. To compare the contribution of time and CIP concentration to patterns of the data, the data was projected into two-dimensional space using principal coordinate analysis (PCoA) (Figure 4-2b). The two principal axes explained 98.6% variation of the data, and I observed a time-associated left-to-right (purple to yellow) pattern for the four isolates with increasing exposure time to CIP. However, there was no clear pattern associated with CIP concentration across the four isolates, suggesting that the bacterial response to CIP was specific to the isolate.

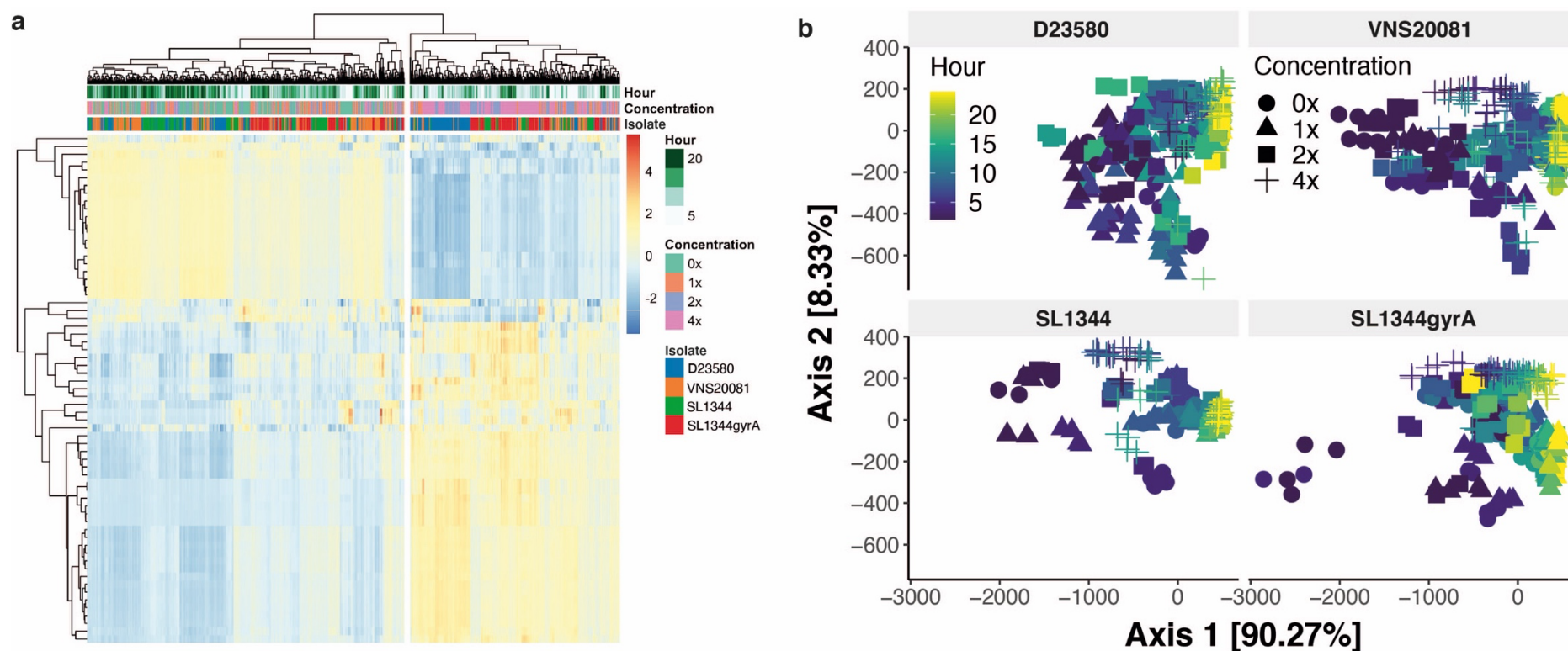


Figure 4-2. Morphological features associated with ciprofloxacin treatments

(a) Feature heatmap with hierarchical clustering. Horizontal axis is datapoints, vertical axis is features, and each cell in the heatmap is the z-score value of a feature of a datapoint. (b) Two-dimensional scatter PCoA plot of *S. Typhimurium* treated with ciprofloxacin at 0xMIC (circle), 1xMIC (triangle), 2xMIC (square) and 4xMIC (cross). Colour scale represents time of exposure

I next trained a random forest classifier on the imaging data from each isolate and on merged data from all isolates to classify the treatment conditions, which were a combination of exposure time and CIP concentrations (i.e. 0X-2h, 0X-4h,..., 4X-24h etc). The classifier was evaluated by the out-of-bag (OOB) error rate with lower error rate indicating higher predictive accuracy. The random forest yielded an acceptable predictive performance with an OOB error rate of 0.25 when using the combined dataset of all isolates, signifying that this model would allow us to effectively predict important imaging features in the data (Figure 4-3a). The ten most important features for each isolate and the combined isolate dataset, corresponding with the highest calculated prominence, were determined by each random forest classifier. Notably, the importance score calculated for each feature displays how much the feature contributes to the predictive performance of a machine learning model, most notably the random forest model in this study. To compare the isolates against each other, I aimed to identify the most important features common across the four isolates and within the combined data. Intersecting the most important features from the five random forests models, trained on data from each of the four isolates and the combined data, resulted in the generation of four common features (Figure 4-3b). Three of these common features were associated with fluorescence (FITC relative radial deviation, FITC intensity mean, and CSA radial mean), the fourth was cell length.

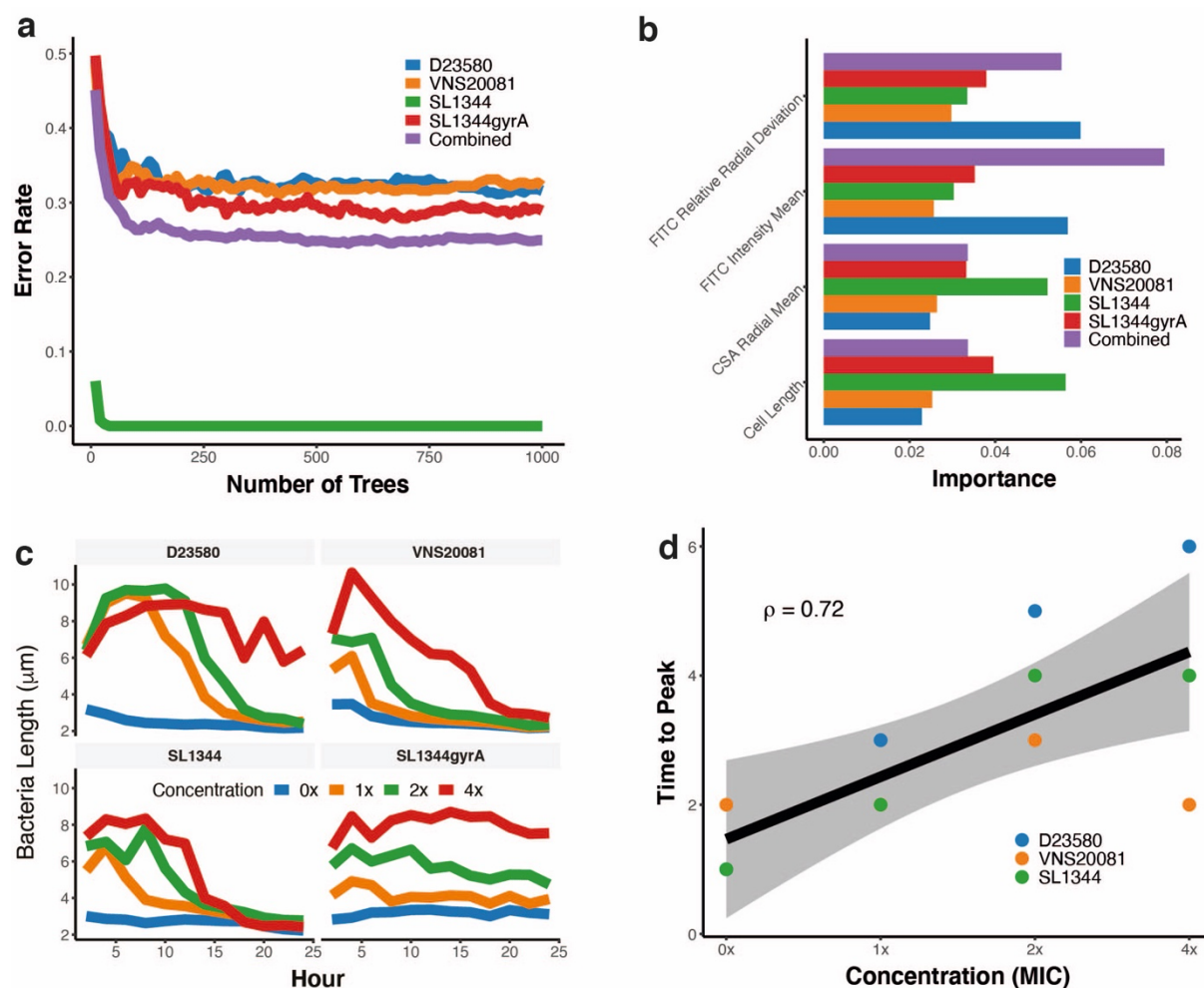


Figure 4-3. Random forest model to interpret the association between morphology and antimicrobial exposure

(a) Random forest OOB error rate against number of decision tree used in ensembles. (b) Feature importance extracted from random forest model with 1,000 decision trees to distinguish isolates at different treatment conditions. The horizontal axis is importance, and the vertical axis is the five most important features. (c) Averaged bacterial length over 24h of exposure to CIP at varied concentrations. Horizontal and vertical axes are exposure time (hours) and averaged length of bacteria (μm) respectively. (d) Correlation between CIP concentration relative to MICs (horizontal axis) and required time for each isolate to elongate to maximum length (vertical axis).

I next plotted the mean bacterial length at different CIP concentrations exposure time in order to build upon our earlier analysis of mean bacterial length at 1X MIC of CIP (Figure 4-3c). These results exhibited a time-dependent trend for bacterial length; specifically, D23580, VNS20081, and SL1344 responded to CIP exposure in a biphasic manner: (1) bacterial cells initially elongated, and then (2) reduced with time. In contrast, the bacterial length trend for the SL1344gyrA mutant was uniform. These trends in bacterial cell length were comparable for the conditions of 1X, 2X, and 4X CIP MIC. I observed a linear correlation ($\rho = 0.72$) between drug concentration and the time required for each isolate required to reach its peak average length (Figure 4-3d). This finding shows how interrelatedness of the exposure time and concentration effects bacterial length. I additionally observed that bacterial length was a convergent trend over time in D23580, VNS20081, and SL1344, suggesting that there was a decline in the morphological distinction between *S. Typhimurium* isolates treated with different CIP concentrations with increased exposure time (Figure 4-3c). Images of D23580 and VNS20081 from discrete time points and across all CIP concentrations indicated a similar convergence of bacterial length, although this was the most pronounced in VNS20081 (Figure 4-4). Overall, I could train a random forest classifier to identify important features common across bacterial isolates, and, using cell length as an exemplar of an important feature, I was able to examine how cell length changed over time and differed between concentrations of CIP.

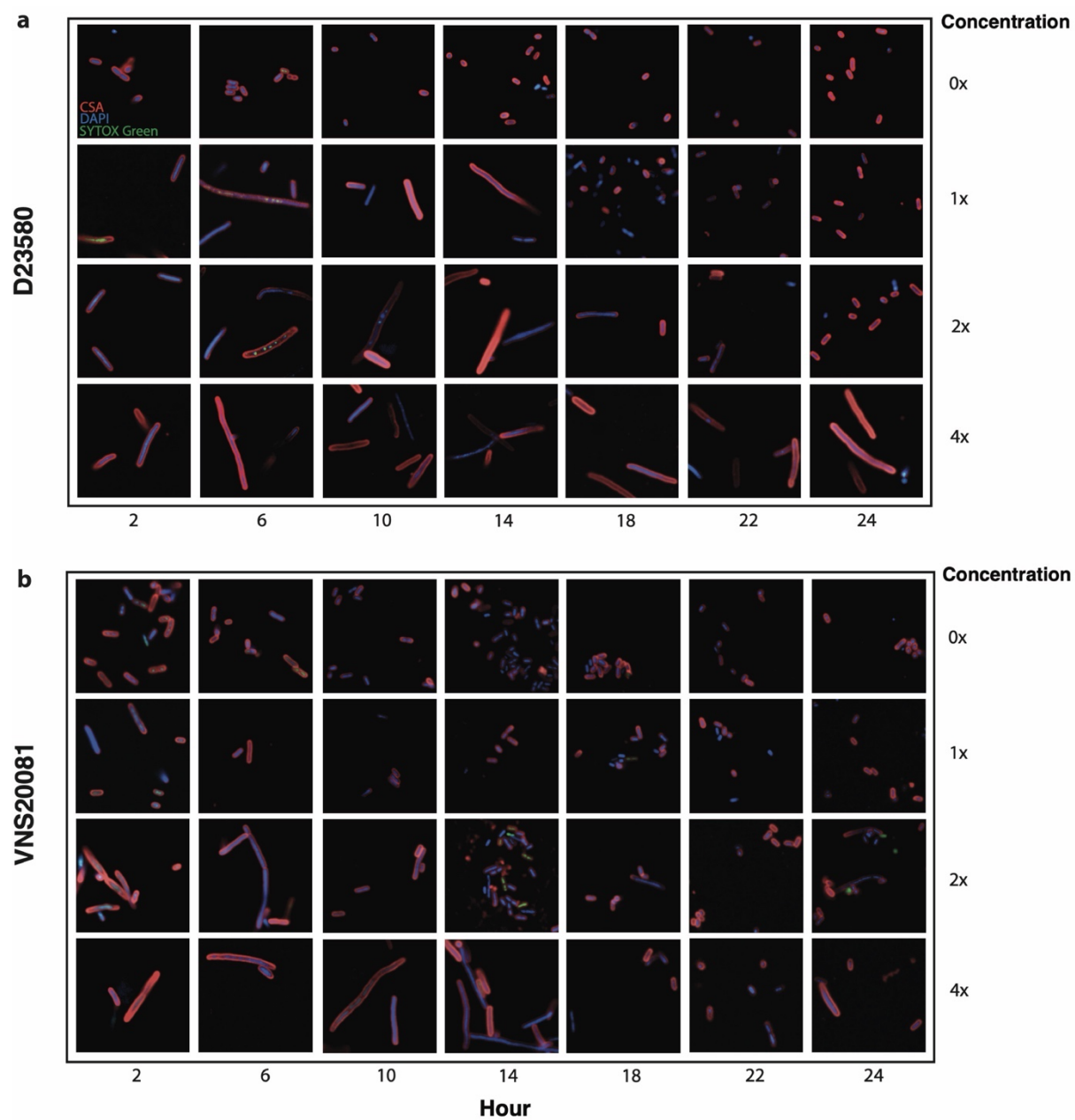


Figure 4-4. Morphologies of (a) D23580 and (b) VNS20081 over the 24-h exposure to CIP at different concentrations

4.4. *S. Typhimurium* SL1344 and its *gyrA* mutant display distinct differences in morphological features

While it was important to recognize that four isolates of *S. Typhimurium* had distinct morphological features, I additionally wanted to investigate if SL1344 and its CIP-resistant *gyrA* mutant derivative exhibited observable differences in cellular morphology. To identify the treatment condition that segregated SL1344 and SL1344gyrA with the highest degree of distinction, the distance ratio at the four CIP concentrations were plotted against exposure time. The distance ratio minimized at 4X MIC and 20 hours (4XMIC-20h), corresponding to the most separation between the two isolates (Figure 4-5a). Therefore, subsequent analyses were conducted at 4X MIC at the 20-hour time point. I found that SL1344 and SL1344gyrA segregated well at 4X MIC-20h, with two independent 95%-confidence ellipses (Figure 4-5b). A random forest classifier was trained to classify the two organisms, yielding an OOB error rate of 0, equivalent to 100% predictive performance. I then derived the ten most important features from the trained random forest classifier and plotted them against the corresponding highest calculated importance, shown for SL1344 and SL1344gyrA combined, and within each organism (Figure 4-5c). I plotted the ten most important features of SL1344 and SL1344gyrA individually, finding that all comparisons were highly significant ($p < 0.001$) (Figure 4-5d). A comparison of representative images from 4XMIC-20h showed that SL1344 and SL1344gyrA were distinct (Figure 4-5e). These ten important features were combined into a radar chart and exhibited in observable differences between organisms (Figure 4-5f). These data indicated that while SL1344 and SL1344gyrA are isogenic, the *gyrA* mutation results in substantial morphological differences when the bacteria were exposed to CIP. More broadly, these data suggest that the mutation in *gyrA* may have further global effects in addition to reduced CIP susceptibility, and the ability to identify important morphological features for a given isolate, may provide novel insights for inferring CIP susceptibility.

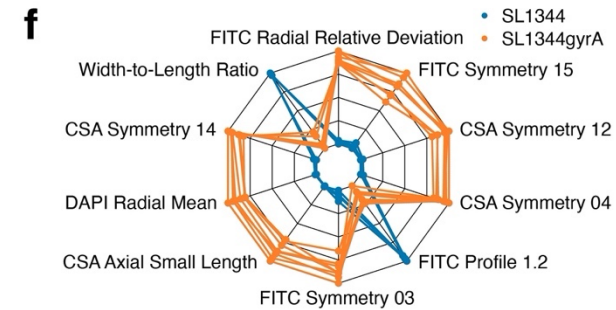
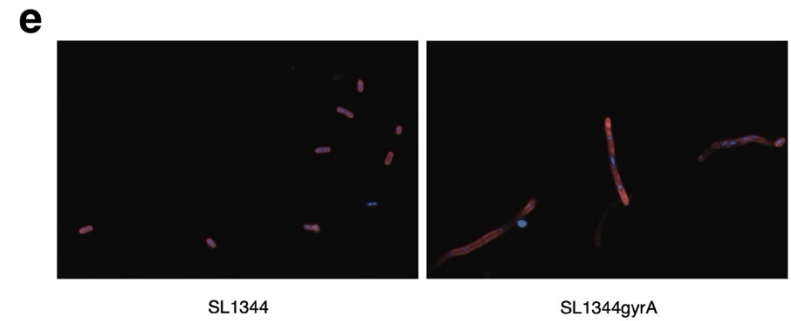
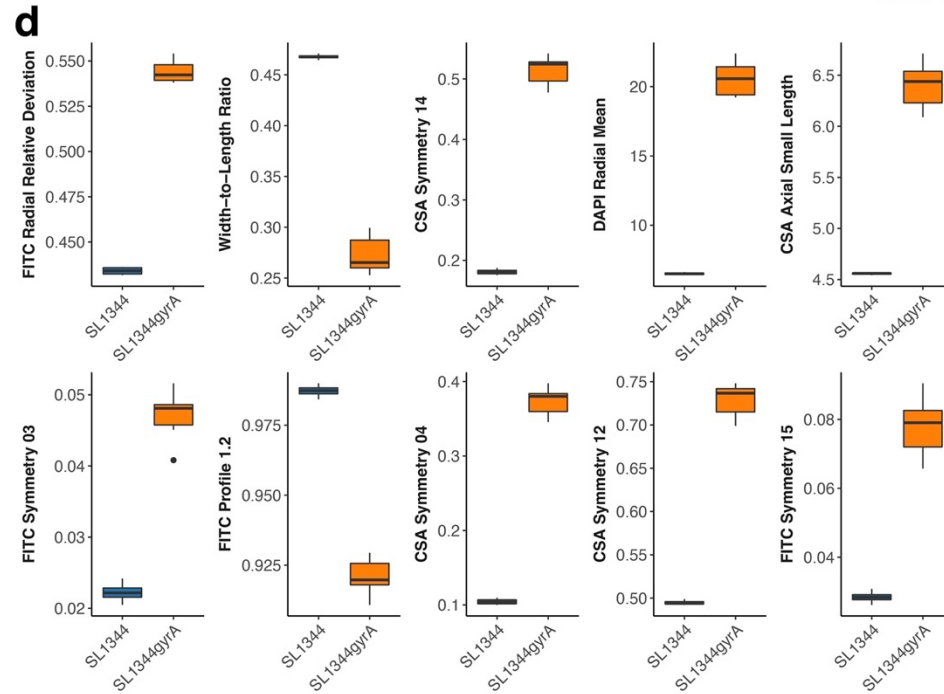
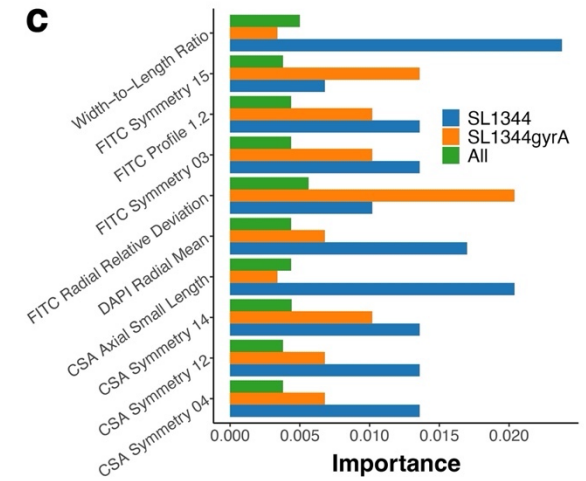
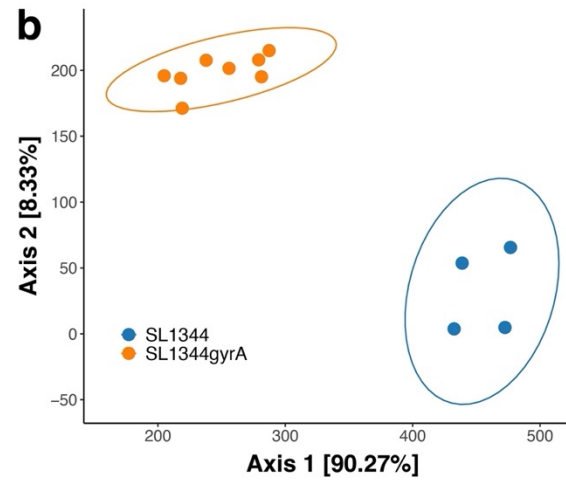
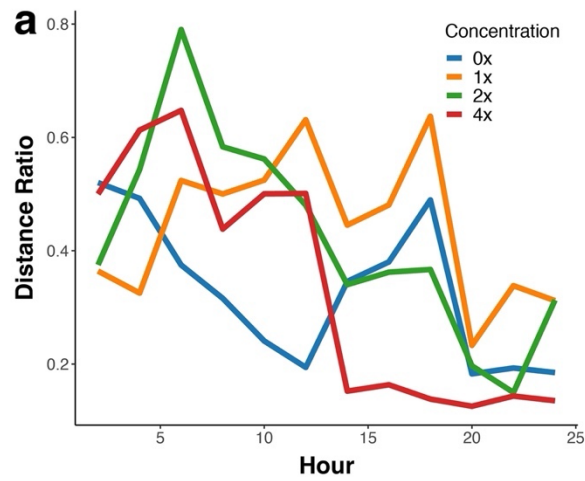


Figure 4-5. Morphological distinction between SL1344 and SL1344gyrA

Distance ratio, between datapoints grouped by the isolate SL1344 and its mutant derivative, against CIP exposure time. (b) PCoA plot, derived from a subset of the data shown in Figure 4-2b, of the two isolates at 4xMIC-20h with 95%-confident ellipses. (c) The ten most important features (vertical axis) for random forest model, built from 1,000 decision trees, to distinguish SL1344 from SL1344gyrA with corresponding importance index (horizontal axis). (d) Boxplots for the ten most important features derived from the random forest. All the boxplot pairs were significant different with $p\text{-value} < 0.001$ (Kruskal-Wallis test). (e) Morphologies of *S. Typhimurium* SL1344 and SL1344gyrA at 4xMIC-20h. Red, green, and blue fluorescence corresponds to CSA, SYTOX Green, and DAPI stains, respectively. (f) Radar chart for z-score transferred value of the most important features. Each line represents a datapoint of either SL1334 (blue) or SL1334gyrA (orange).

4.5. Ciprofloxacin susceptible and resistant *S. Typhimurium* isolates can be distinguished by morphological features

Having quantitatively distinguished *S. Typhimurium* SL1344 from a CIP resistant SL1344 *gyrA* mutant, I sought to discriminate between CIP susceptible and resistant isolates with diverse genetic backgrounds. Notably, CIP resistant *S. Typhimurium* exhibited the most distinction from susceptible isolates after 22 hours of culture without exposure to CIP, which was quantified by the lowest distance ratio at the condition of 0X MIC-22h (Figure 4-6a). This observation suggests that *gyrA* mutations may interact with other cellular pathways, which ultimately alter morphology. I subsequently investigated how susceptibility to CIP affects the distribution of imaging data at 0X MIC-22h. Projecting all 65 imaging features at 0X MIC-22h into a PCoA plot (explaining 98.6% variation), I found resistant organisms clearly segregated from susceptible organisms (Figure 4-6b). I next trained a random forest model to select the ten most important features associated with CIP resistance at 0X MIC-22h for further analysis. The random forest model possessed a high predictive-performance at 0X MIC-22h in the best segregation of CIP susceptible and resistant isolates. Therefore, the feature-selection interpretation from the model was associated with high confidence; the ten most relevant important features were measurements of fluorescence (Figure 4-6c). These important imaging features were plotted to a heatmap using hierarchical clustering (Figure 4-6d). I again found that CIP resistant *S. Typhimurium* intrinsically segregated from susceptible organisms (Figure 4-6d), and had distinctive imaging data patterns (Figure 4-6e), suggesting that the clustered structure of the data was driven principally by these important features. These data indicate that CIP-susceptible and -resistant isolates can be distinguished based on a small number of morphological features without CIP exposure.

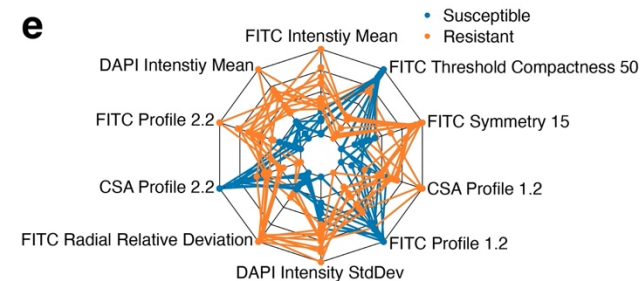
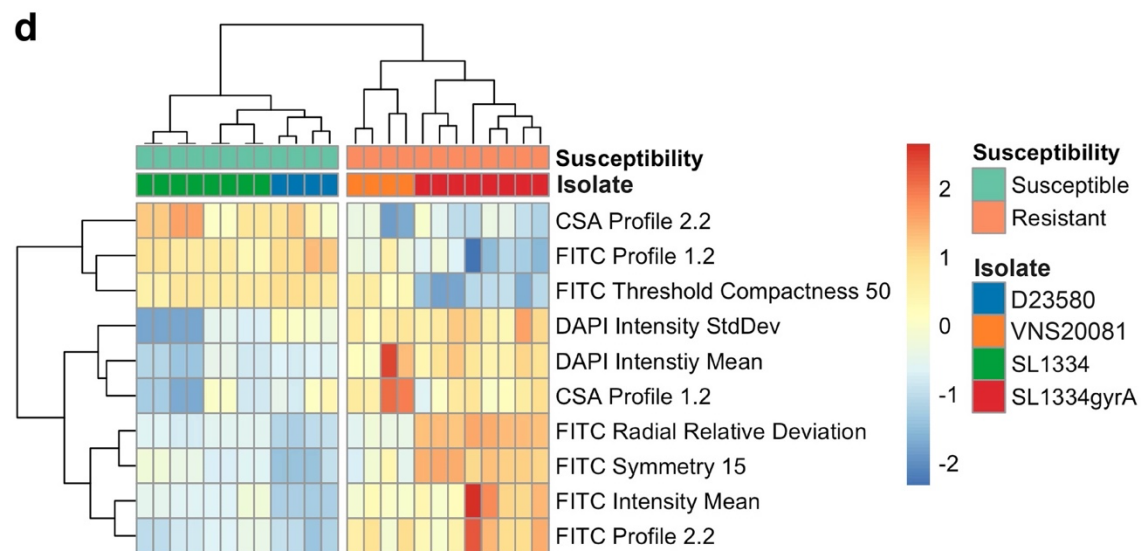
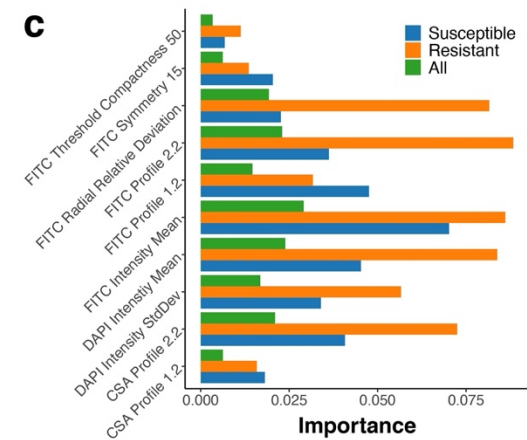
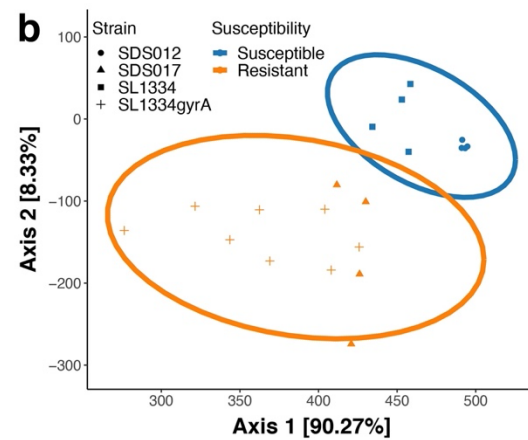
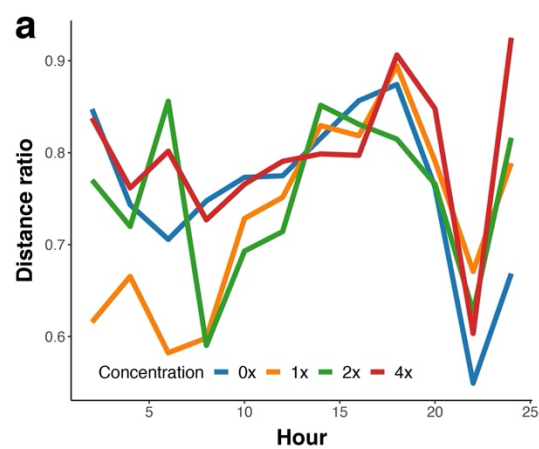


Figure 4-6. Distinguish resistant and susceptible isolates at 0xMIC-22h

(a) Distance ratio (vertical axis), of data grouped by antimicrobial susceptibility, at different drug concentration against exposure time (horizontal axis). (b) PCoA plot for all datapoints at 0xMIC-22h. A subset from PCoA data (Figure 2b) at 0xMIC-22h was reused for this plot. Blue (susceptible) and orange (resistant) ellipses are 95%-confident ellipses. (c) The ten most important features derived from random forest, with 1,000 decision trees, identifying resistant isolate at 0xMIC-22h. Vertical axis is the important features against the importance on horizontal axis. (d) Heatmap of the ten most important features with hierarchical clustering on horizontal axis of datapoints and vertical axis of features. Each cell in the heatmap is the z-score value of a feature of a datapoint. A subset of z-score data from previous heatmap (Figure 4-2a) was reused for the plot. (e) Spider chart for the ten most important features to differentiate susceptible (blue) and resistant (orange) isolates. Data was transferred to z-score and each line is a datapoint.

4.6. Random forest classifier can distinguish between ciprofloxacin susceptible and non-susceptible isolates without ciprofloxacin exposure

To generalize the distinction between CIP-susceptible and resistant isolates using an extended dataset, I generated new imaging data of 13 additional and the four prior *S. Typhimurium* isolates at 0X MIC-22h (Table 4-1). Data from these 17 isolates were combined and analysed (Figure 4-7a). All important features of these isolates were ranked in descending order, based on their relative contribution from the random forest model. The ranked features were then input stepwise into machine learning classifiers in an increment of five features to identify the best classifiers for these data. The classifiers included Naïve Bayes classifier, K-nearest neighbour classifier, support vector machine (SVM), random forest, gradient boosting of decision trees (CatBoost), and artificial neural network. Incorporating a large number of features in the machine learning classifiers diminished the predictive performance of those models, and the accuracy of most of the classifiers declined as features were added. Ultimately, I found that only five morphological features were required to reliably distinguish antimicrobial susceptibility (Figure 4-7b).

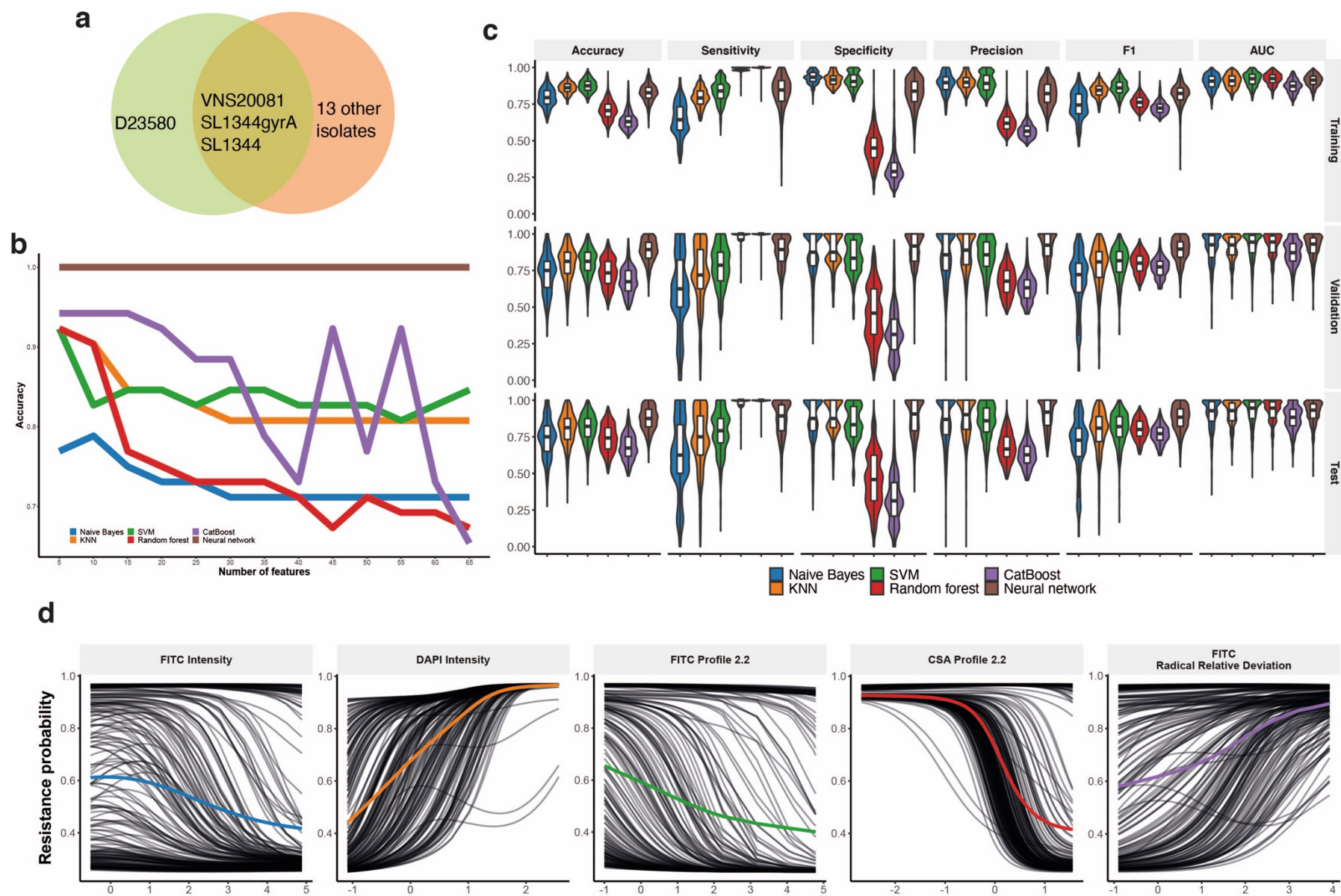


Figure 4-7. Machine learning classifiers to identify resistant isolates at 0xMIC-22h

(a) Venn diagram of isolates used to train machine learning classifiers. Green circle represents the four main isolates and orange circle represents 16 isolates to test the generalization of the classifiers. (b) Predictive performance of machine learning models set against number of most important features. Horizontal axis is number of features ranked by importance and vertical axis is accuracy on validation set. (c) Boxplots and violin plots for predictive performance of machine learning models on 1,000 random training/validation/test partitions. Vertical axis is the values of performance metrics. (d) Partial Dependence Plot of the five most important features (colour lines) against resistance probability calculated by the neural network. Gray lines are Individual Conditional Expectation plot of each datapoint, which shows how predicted resistance probability changes when one feature varies. Horizontal axis is z-score of the features and vertical axis is corresponding resistance probability.

To compare predictive performance among the classifiers, the machine learning models were trained with the five most important features, and performance metrics were recorded accordingly on 1,000 randomly split training, validation, and testing sets (Figure 4-7c). The metrics consisted of accuracy, sensitivity, specificity, precision, F1 score, and receiver operating characteristic (ROC) area under the curve (AUC). Among all experimental classifiers, the neural network exhibited the highest performance metrics compared to other classifiers. Specifically, on the test sets, the neural network exhibited an average accuracy of 0.87 ± 0.08 (mean $\pm$ standard deviation), sensitivity of 0.87 ± 0.11 , specificity of 0.89 ± 0.12 , precision of 0.90 ± 0.1 , F1 score of 0.87 ± 0.08 , and ROC AUC of 0.91 ± 0.07 (Figure 4-7c, brown violin plots).

I then sought to interpret the causal relationship between input features and the output of resistance probability from the neural network classifier (Figure 4-7d). Given imaging features of an isolate, neural network predicts the corresponding resistance probability, such that higher probability indicates greater likelihood of the isolate resistant to CIP. Overall, resistance probability responded to the important features in a nonlinear fashion as a result of the complex neural network. Notably, the signal from SYTOX green stain captured by FITC channel comprised 60% of the important features. As SYTOX green, a DNA stain, cannot penetrate intact cells, FITC intensity is associated with cell permeability of *S. Typhimurium*. Specifically, low FITC intensity, indicating membrane integrity, was associated with high resistance probability (Figure 4-7d). This finding suggests that under exposure to CIP, resistant isolates were more likely to survive and to remain intact membrane, while susceptible isolates died, leading to higher membrane permeability.

In contrast to FITC intensity, low DAPI intensity, indicating bacterial DNA content, correlated with low resistance probability, which may be explained by the membrane integrity and hypothetically by competitive DNA binding with SYTOX green (Figure 4-7d). The neural network showed that high FITC profile and CSA profile were associated with low probability for resistance and high FITC radial relative deviation correlated with high probability of resistance. Overall, classification using the neural network on 13 additional *S. Typhimurium* isolates of differing CIP susceptibilities confirmed that CIP resistant *S. Typhimurium* isolates are morphologically distinct from the susceptible CIP isolates, and this distinction can be identified by assessing specific morphological features without any prior antimicrobial exposure or knowledge of their susceptibility.

4.7. Discussion

Here, I combined HCI and machine learning models to determine phenotypic differences between isolates of *S. Typhimurium* exposed to CIP. I found that CIP-resistant and -susceptible isolates have distinct morphological characteristics, independent of CIP exposure, that can be distinguished using machine learning classifiers. To my knowledge, this is the first instance of machine learning techniques being utilised to predict antimicrobial-resistant and -susceptible bacterial isolates using imaging data. The proposed machine learning classifiers can infer the inter-relation between morphological data and the complex combinatory treatments of different CIP concentrations and CIP exposure time.

Compared to many other methods to study AMR^{266–268}, HCI analysis is high-throughput and can be performed using small culture volumes. HCI also provides a high-resolution overview of bacterial populations at the scale of an individual organism. My study has shown that HCI

does not require prior knowledge of resistance mechanisms or phenotypic susceptibility by traditional antimicrobial susceptibility testing to determine resistant isolates. This makes HCI a potentially valuable technology for future diagnostic laboratories. Given that other Gram-negative rod-shaped bacteria undergo similar phenotypic changes in response to CIP, it may be possible to extrapolate our findings to infer the responses of other Gram-negative bacteria to CIP and other antimicrobials. For that reason, an imaging database at family level may be sufficient to train machine learning models to diagnose an infection caused by AMR organisms, without labour-intensive efforts requiring the development of databases for individual species.

The majority of contemporary studies have coupled HCI with dimensional analyses to distinguish antimicrobials with different MoA^{2,3,4,5,7} and pathogenic isolates with different AMR profiles¹⁵⁶. The disadvantage of such methods is that they do not allow quantitative prediction of new queried data points, especially ones that lie in the middle of pre-defined clusters. For example, dimensional analysis is unable to predict the resistance profile of an isolate with corresponding imaging datapoint falling between resistant and susceptible clusters. Alternatively, very few studies have employed machine learning classifiers to quantitatively predict antimicrobial MoA^{155,159} and target proteins¹⁶⁴. More importantly, HCI-based AMR identification using machine learning classifiers has not previously been published. Due to their high complexity, machine learning, especially nonlinear, models are considered to be ‘black-box’ tools lacking interpretability, which might prevent scientific understanding of the results²⁶⁹. In this study, I designed and implemented various ‘white-box’ machine learning classifiers to identify AMR isolates of *S. Typhimurium* and interpreted nonlinear relations between morphological features and AMR probability. The interpretable machine learning analysis proposed in this study could be a model for other HCI analyses,

promoting more popular use of machine learning in the field of biology, particularly clinical microbiology.

This study has limitations. While I created an HCI analysis pipeline suitable for bacteria, it likely needs further refinement to optimally segment bacterial cells. An improved image analysis algorithm specialised for bacterial imaging data may provide better differentiation between resistant and susceptible organisms at the single-cell level. My study is also limited by experimental sampling methods in which averaged morphological data was extracted from images acquired on cross-sectional samples of *S. Typhimurium* populations at discrete timepoints. As a result, it is not possible to measure the longitudinal transformation of specific bacterial morphologies and reduces the single-cell resolution of the assay. Furthermore, bacteria were cultured in liquid medium in a laboratory setting, which makes the study less clinically relevant. However, this approach facilitates further studies on *S. Typhimurium* at the molecular and cellular levels, which could be leveraged for rapid antimicrobial susceptibility testing.

Using various machine learning algorithms, I have conceived an analysis pipeline for interpreting and predicting CIP resistance phenotypes of *S. Typhimurium* based on high content confocal imaging data, which can serve as a framework for further similar studies. This framework can be generalized to bridge the gap between molecular information and bacterial cellular responses, for example using multi-omics data and morphological changes. In the field of diagnostics, this could be exploited to develop high-accuracy machine learning classifiers for rapid antimicrobial susceptibility testing. Moreover, this framework, coupled with increasingly accessible microfluidic technology, could facilitate high-throughput diagnoses, thus reducing time, labour, and reagents. While the clinical application may still

be distant, this imaging and machine learning-based approach may be an important tool in the future for rapidly and accurately assessing drug resistance in the hospital context.

Chapter 5. Integrating Imaging and Omics with Machine Learning Approaches to Infer Antimicrobial Susceptibility in Gram-negative Bacteria: Method Development

I declared that all the work in this chapter was performed by myself.

5.1. Introduction

In the last two decades many studies have attempted to exploit the potential of intact protein analysis using MALDI-TOF MS to identify antimicrobial resistant bacteria²⁷⁰; however, studies applying HCI for AMR detection are scarce²⁷¹. Moreover, research that incorporates MALDI-TOF and HCI to study AMR has not yet been conducted. Hence, I aimed to develop a novel method integrating these two techniques (MALDI-TOF and HCI) to infer antimicrobial susceptibility in bacteria and to study AMR (Figure 5-1a and section 2.6.3).

Previously, in Chapter 3, I used 27 isolates of *E. coli* from blood samples (collection 1) to study the association between AMR profile and bacterial morphology. Here, I combined an additional 43 *E. coli* isolated from stool samples (collection 2) with the 27 blood isolates, to improve the analysis power of the study. As the bacteria were isolated from two different sources, I considered it would provide more diverse genetic information to add to the analysis.

As shown in Chapter 3, AZI was found to accelerate the growth rate of *E. coli*, while TE resistance was found to confer fitness advantage on the population. For their defined effects on bacterial physiology, AZI and TE were included in this experiment as antimicrobials of interest for further analysis. In Chapter 4, HCI data was found to associate with CIP exposure

and CIP resistance on a small collection of *S. Typhimurium*. Here, I sought to re-examine this association on a larger collection of *E. coli* and hence, CIP was included to the experiment. Testing the association on *E. coli* was also a proof-of-concept providing data that supports the use of HCI to study AMR across different bacterial species.

Here, in this chapter, I developed and evaluated optimum methods for intact protein analysis with MALDI-TOF MS and Opera-Phenix HCI analysis of AMR pathogenic *E. coli*. The optimised method for MALDI-TOF were subsequently integrated with the superlative method for HCI with Opera Phenix. This approach was expected to produce an enormous multi-perspective dataset to study and identify AMR. Additionally, I evaluated the reproducibility, robustness, and bias of the methods.

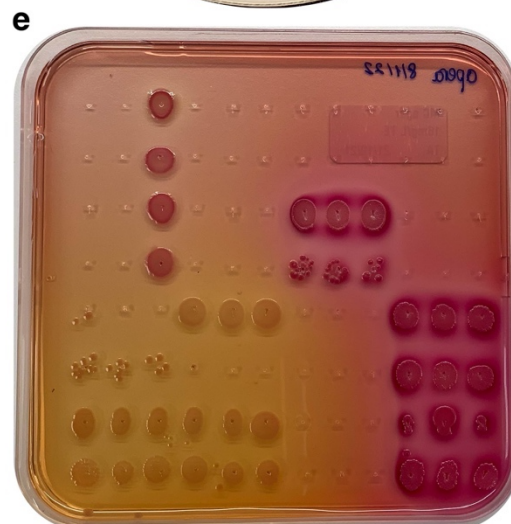
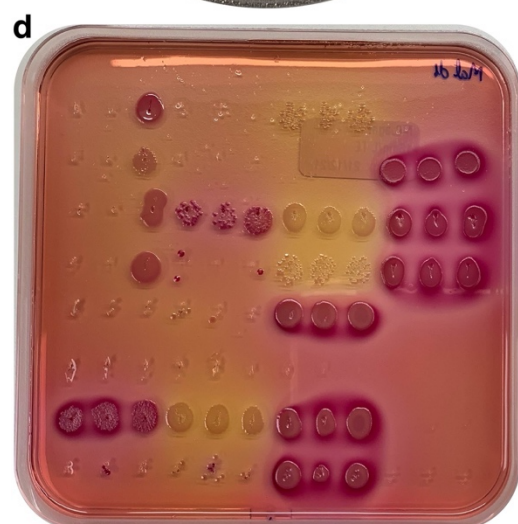
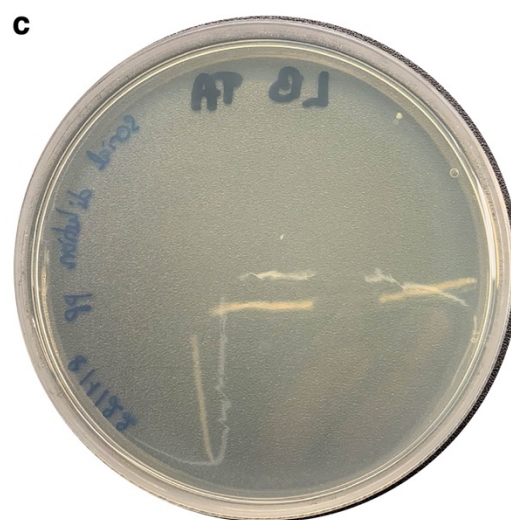
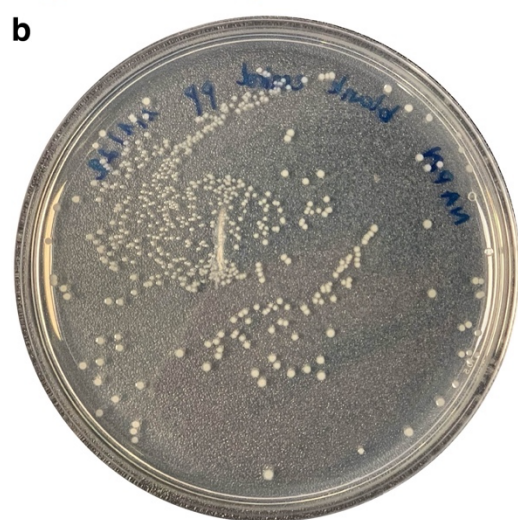
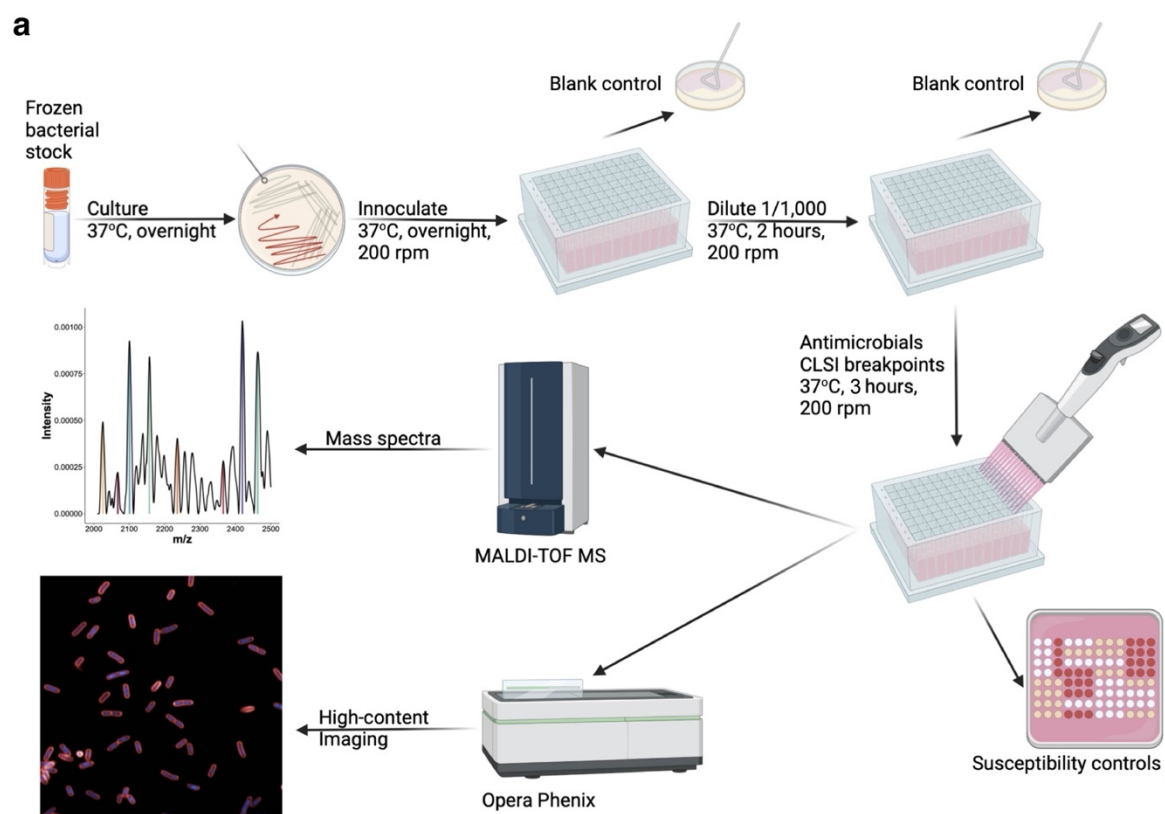


Figure 5-1. Integrating intact protein analysis by MALDI-TOF with HCI by Opera Phenix

(a) Experiment overall workflow. Details were described in section 2.6.3. Briefly, isolates were recovered overnight at 37°C and inoculated in a deep-well plate, followed by overnight incubation at 37°C. Blank controls were carried out that this step, e.g. sub-figure (b).

Bacterial suspension was diluted in fresh medium and incubated for 2h and blank controls were executed, e.g. sub-figure (c). Subsequently, the bacteria were exposed to CIP, AZI, or TE for 3h. Finally, the bacterial suspension was input to both intact-protein analysis with MALDI-TOF MS and HCI with Opera Phenix. (b) Contaminated pure plate for overnight culture, batch 3. (c) Contaminated pure plate for bacterial dilution, batch 4. (d) TE susceptibility control, batch 3. (e) TE susceptibility control, batch 4. (d) and (e) Each position on the susceptibility control plates corresponded to one well on the experimented deep-well plates with the layout described in Figure 2-1.

5.2. Intact protein analysis with MALDI-TOF MS

5.2.1. Assessing sample preparation methods

Aiming to evaluate different methods for sample preparation, I cultured the laboratory reference strain *E. coli* ATCC 25922 in liquid media. Bacterial suspension was used as an input for four preparation protocols, including (1) ethanol inactivation, (2) TFA inactivation, (3) ACN extraction, and (4) ethanol/FA/ACN extraction prior to MALDI-TOF MS. Prepared samples were measured with MALDI-TOF MS to acquire protein spectra. To compare the methods, four evaluation metrics were calculated for each spectrum. The metrics were number of mass peaks, averaged signal-to-noise ratio (SNR), averaged normalised-intensity, and total normalised-intensity.

Samples, prepared with the ACN method, yielded significantly low-quality signals, such that the MALDI-TOF machine was not able to acquire spectra. For the three remaining methods, spectra were acquired and subsequently subjected to signal pre-processing, peak detection, and calculation for evaluation metrics, Figure 5-2. Among the preparation methods, ethanol inactivation displayed the highest SNR, total intensity, and averaged intensity. Although the TFA method possessed the highest number of peaks, it also displayed the lowest SNR. This result was expected since low SNR might result in detecting more false peaks. Overall, ethanol inactivation (section 2.6.1.1.1) was selected as the optimum method for sample preparation and was employed for further experiment set-up and analysis.

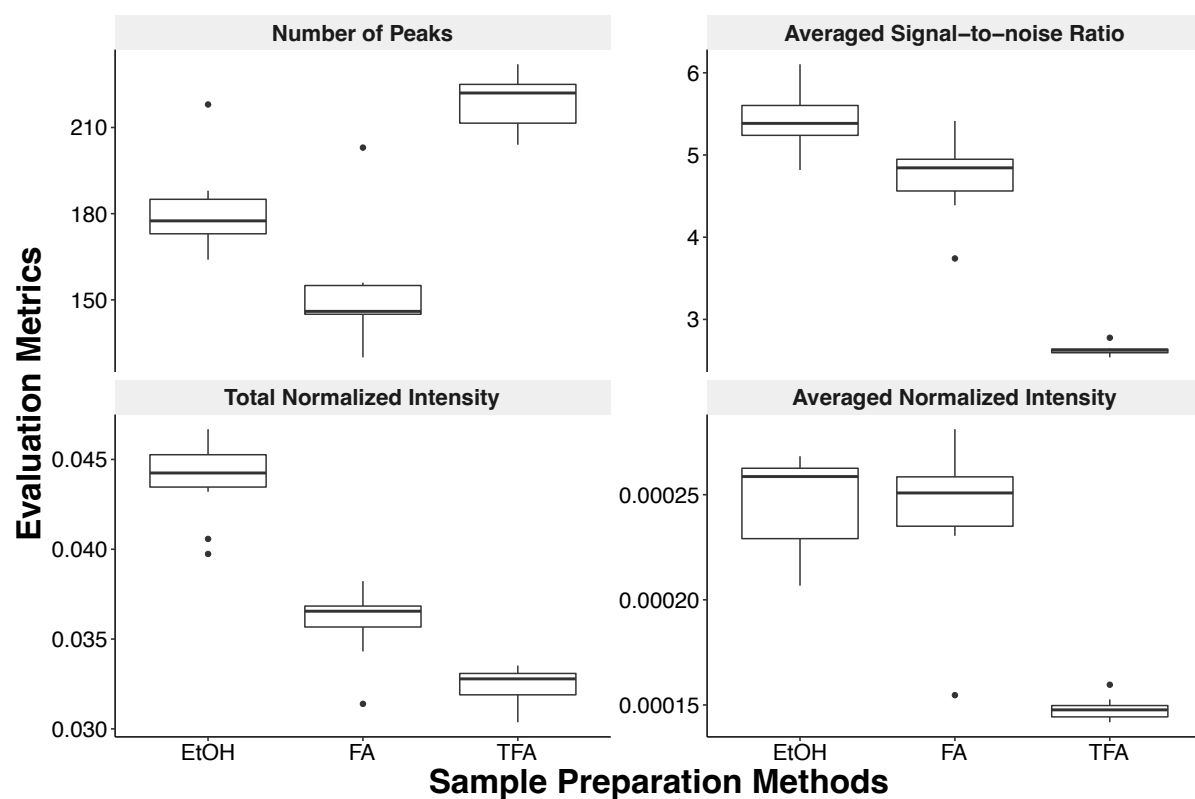


Figure 5-2. Comparing sample preparation methods.

Boxplots shows spectrum's evaluation metrics (vertical axes) regarding the preparation methods (horizontal axis). ACN: protein extraction with acetonitrile. EtOH: bacterial cell inactivation with ethanol. FA: protein extraction with ethanol/formic acid/acetonitrile. TFA: bacterial cell inactivation with trifluoroacetic acid.

5.2.2. Assessing matrix solution and deposit methods

Samples from *E. coli* ATCC 25922 were prepared with ethanol inactivation method (section 2.6.1.1.1). The samples were mixed with different matrix solutions (section 2.6.1.2) in a 1:1 ratio (v/v) before depositing onto the MALDI target plate. A volume of 1 μ L to 2 μ L of the mixed samples was deposited on the target plate for each experimental matrix solution. Prior to the experiment, the plate was either coated with a thin layer of saturated CHCA or CHCA-Std or left without coating. Evaluation metrics, for varied combinations of matrix solutions and target plate coatings, were calculated and summarised in Figure 5-3.

Assessing the four metrics, although DHB-Std with no coating displayed the highest SNR and spectra intensity, it generated the lowest number of peaks. Alternatively, CHCA-DHB with saturated CHCA coating was a more optimum method generating the second highest intensity, the third highest number of peaks and a reasonable SNR. For the protein spectra analysis, I aimed to calculate not only high-quality spectra with great SNR and intensities, but also those with the greatest number of peaks, which provides deeper insight on differential protein expression. For this reason, CHCA-DHB with saturated CHCA was selected as standard method for intact protein analysis. Altogether, the final optimum protocol for intact protein analysis with MALDI-TOF MS was summarised in section 2.6.1.3.

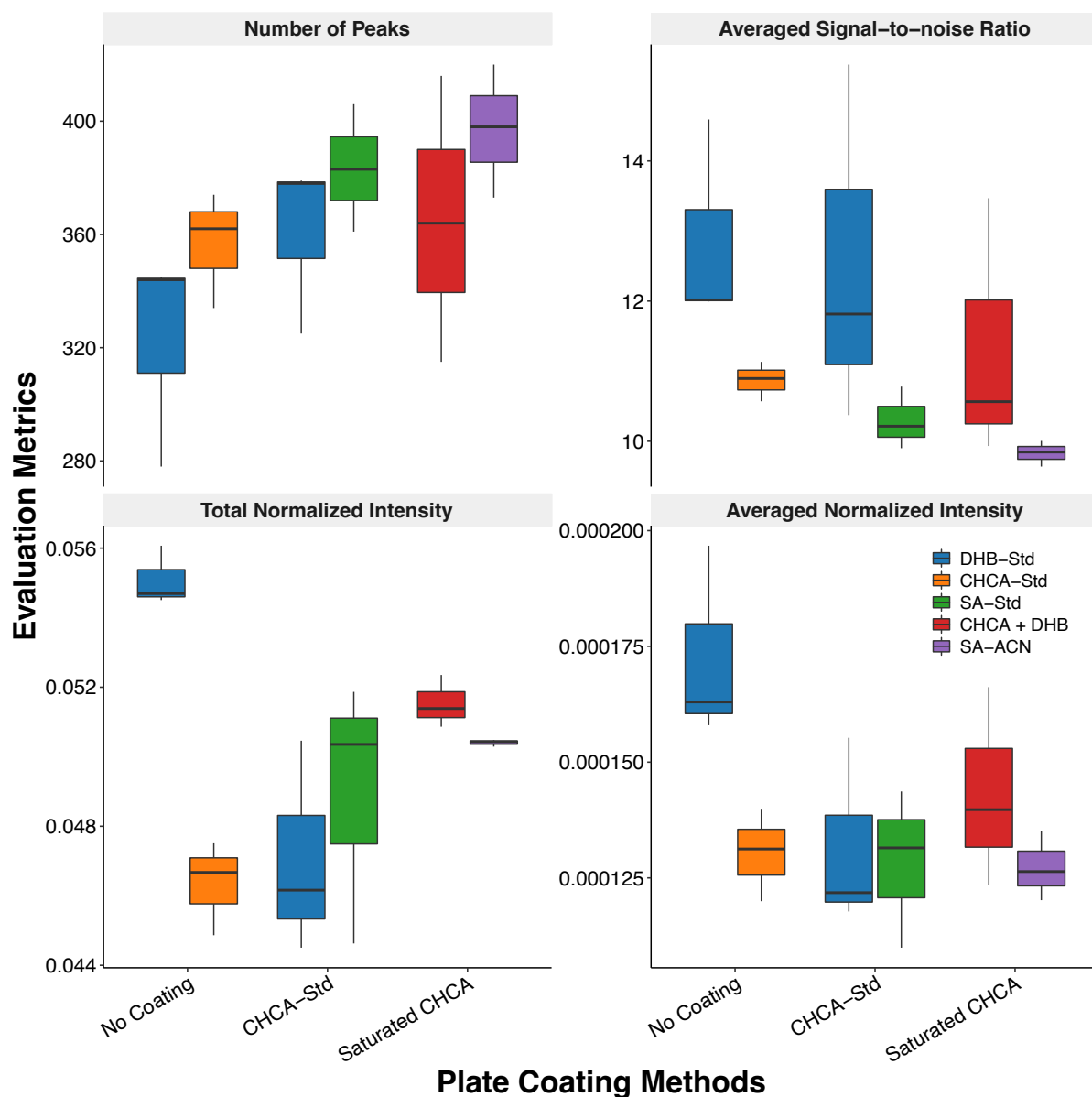


Figure 5-3. Evaluating different combinations of matrix solutions and target plate coatings.

Each boxplot was a combination of a matrix solution and a coating method for target plate. Horizontal axis showed plate coating method and vertical axis displayed evaluation metrics of the acquired mass spectra. For matrix-solution preparation, blue: DHB-Std, orange: CHCA-Std, green: SA-Std, red: CHCA+DHB, and violet: SA-ACN.

5.3. High-content imaging

5.3.1. Fluorescent staining

5.3.1.1. Method development

Fluorescent staining is the most critical step for HCI as it determines the quality of outcome images. Poor quality images can lead to inaccurate cell segmentation and thus, leads false discovery. In this work, I sought to develop a robust method to fluorescently stain bacterial cells and immobilise them onto the bottom surface of multi-well imaging plates. I also attempted to reproduce the method, proposed by Sridha and colleagues in 2021^{145,153}, as a baseline for comparison.

To enhance the adhesion of bacterial cells on imaging plates, I proposed a ‘sandwich’ method to immobilise the stained cells onto the bottom surface of imaging plates (described in Figure 5-4). Compared to Sridha’s method, the ‘sandwich’ was able to immobilise more bacterial cell on the horizontal plane of the imaging plate’s bottom (Figure 5-5a). The ‘sandwich’ structure also stabilised the cells during z-stack image acquisition, while cells moved more freely with Sridha’s method (Figure 5-5b). As a result, the ‘sandwich’ method produced much higher contrast images. Moreover, maximum projection of images, acquired with ‘sandwich’ method, showed that most cells were in-focus with crisp and sharp edges. Additionally, since cells were fixed and stained prior to immobilisation onto the imaging plate, there were less cell aggregations in a field of view. To all, the ‘sandwich’ method showed ability to produce high quality images facilitating more accurate cell segmentation at the image analysis step.

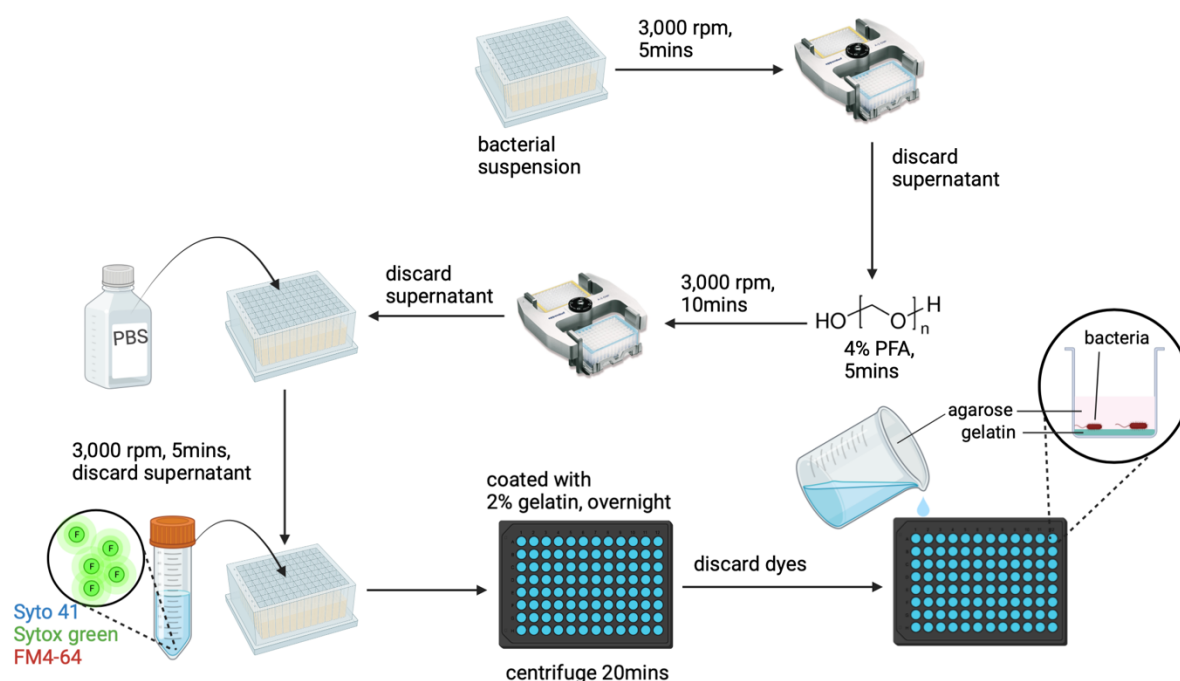


Figure 5-4. Fluorescent staining and cell immobilisation method

Bacterial cells were fluorescence-stained as in section 2.6.2.1.2. Briefly, antimicrobial-treated bacterial suspension was centrifuged, followed by cell fixation with 4% PFA for 15 minutes. Bacterial cells were washed with PBS, stained with fluorescent dyes for 20 minutes, and deposited on an imaging plate precoated with 2% gelatin. Finally, melted agarose solution was aliquoted into the imaging plate.

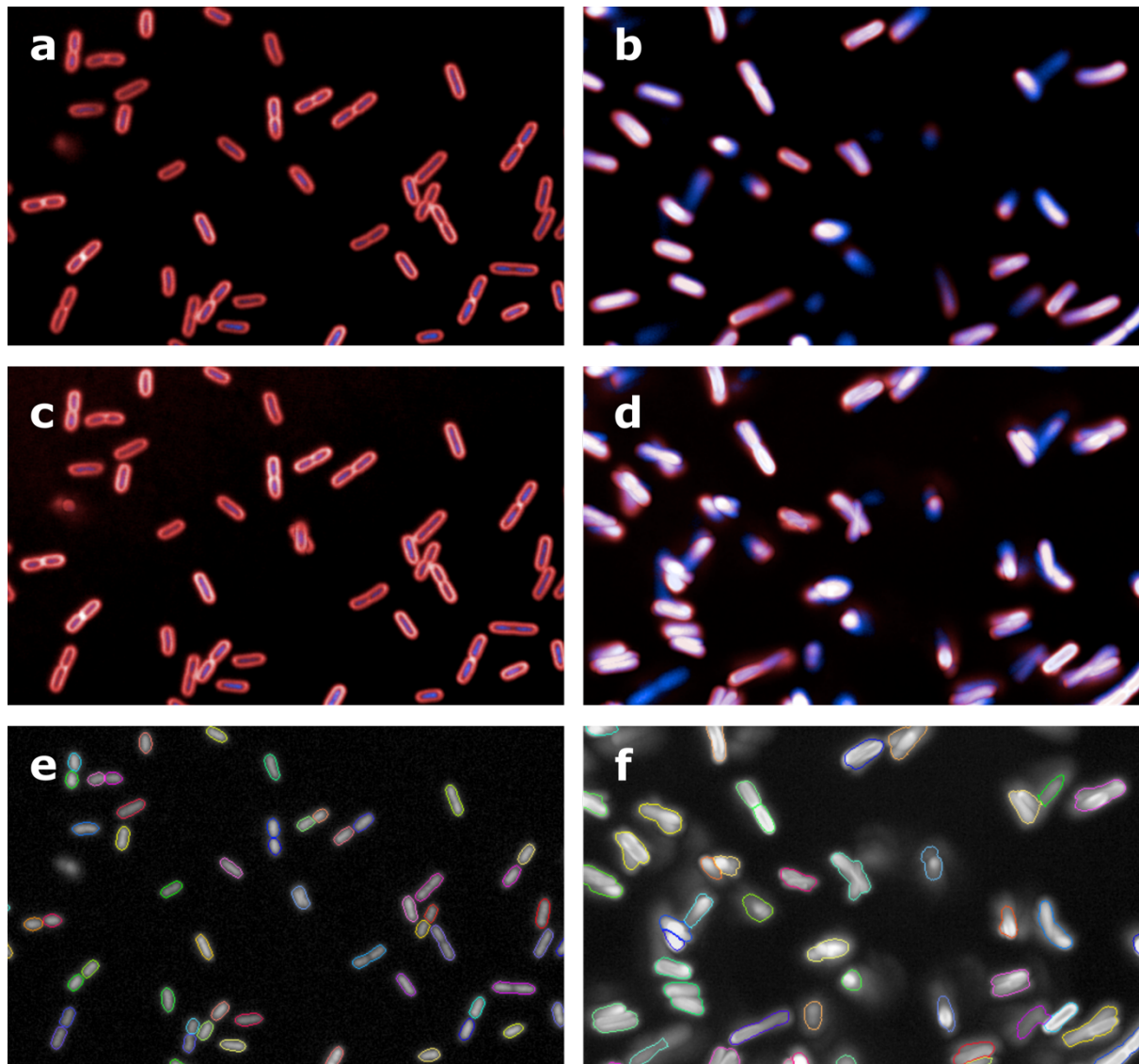


Figure 5-5. Comparing the proposed ‘sandwich’ method and Sridha’s method for fluorescent staining and cell immobilisation

Fluorescent image of bacterial cells and cell segmentation produced with (a, c, e) the ‘sandwich’ method and (b, d, f) Sridha’s method. (a, b) Fluorescent images on a single plane of z-stacks. (c, d) Maximum projection of images on three z-stacks. (e, f) Cell segmentation from maximum-projected images. Red channel: FM4-64. Green channel: sytox green. Blue channel: syto 41.

5.3.1.2. Method evaluation

To assess the reproducibility of the ‘sandwich’ method, *E. coli* ATCC 25922 was cultured in liquid media and stained on a 96-well imaging plate. The experiment was conducted with three biological replicates. Acquired images were subjected to image analysis workflow (section 2.6.2.2) for cell segmentation and feature extraction. Imaging features of the three biological replicates were demonstrated in a PCoA plot (Figure 5-6a). I observed that the three replicates had similar distribution and datapoints from the replicates clustered within the same area of the PCoA plot. These observations suggest that the ‘sandwich’ method is highly reproducible.

High-throughput assays using multi-well plates are susceptible to plate-layout effect in which samples under the same conditions behave differently depending on the positions within the plates²⁷². To detect the plate effect of the ‘sandwich’ method, the data from the above paragraph was examined and plotted on Figure 5-6b. I found that the number of detected cells distributed randomly on the imaging plates. Fluorescent intensities on three channels were also found randomly distributed across the plates (Figure 5-6c). Those suggests that the method yielded spatially unbiased measurements and there was no sign of plate effect.

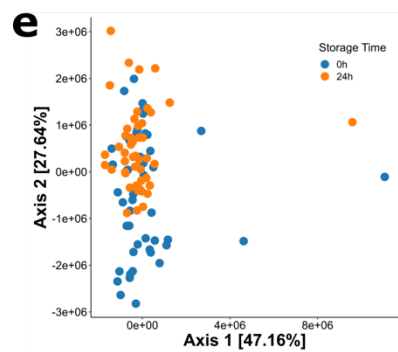
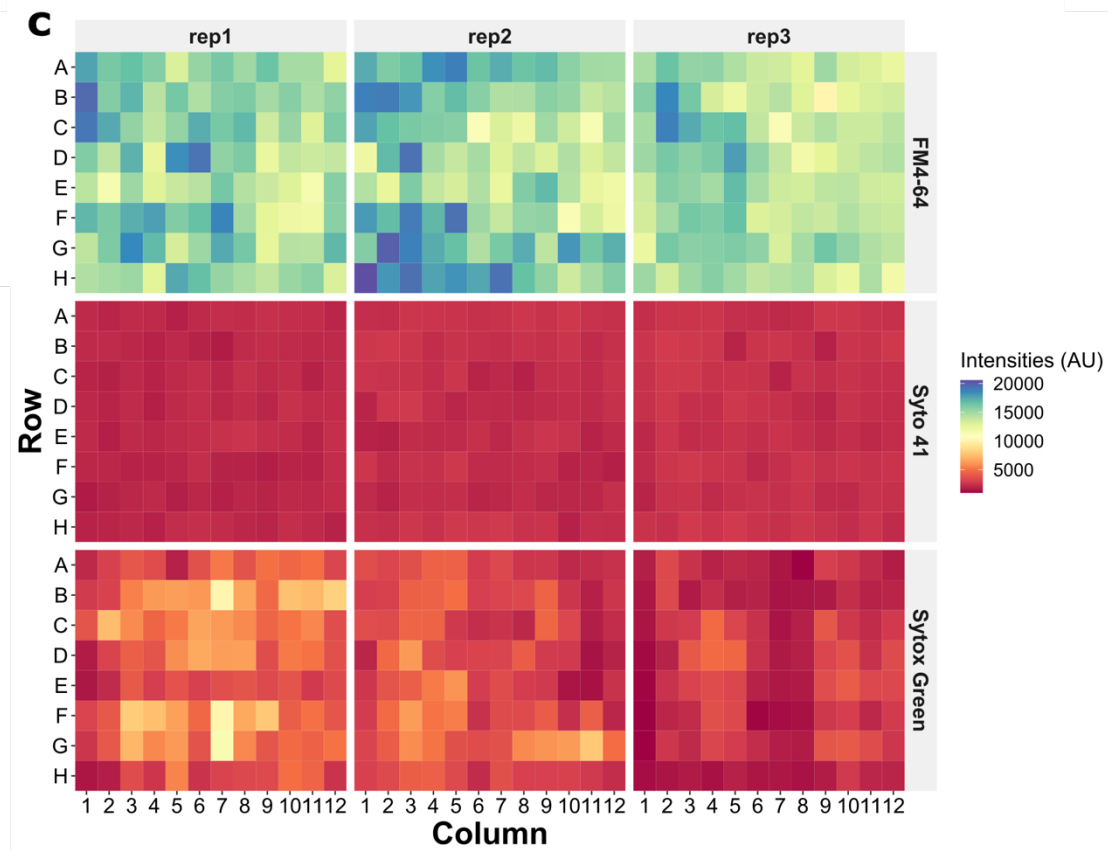
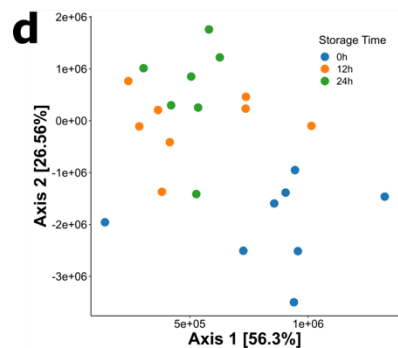
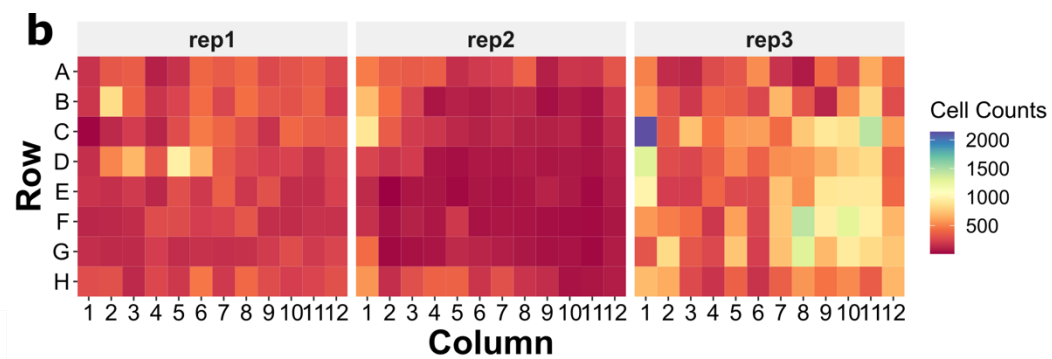
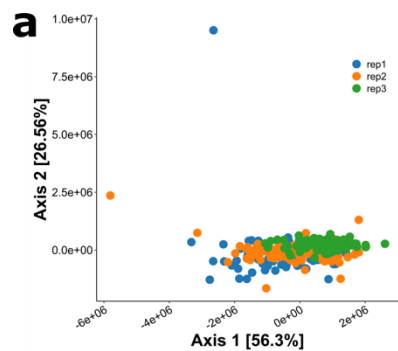


Figure 5-6. Evaluating the ‘sandwich’ method

(a) PCoA plot for three biological replicates showed appreciable reproducibility of the ‘sandwich’ method. Each datapoint represented averaged imaging features of one well on a multi-well imaging plate. Each colour of blue, orange, and blue corresponded to one replicate. (b) Heatmaps displayed numbers of segmented bacterial cells for three replicates. Every individual cell on the heatmap represented the numbers of bacterial cells on respective well. Horizontal and vertical axes were columns and rows of a multi-well plate. (c) Heatmap demonstrated the averaged fluorescent intensities on three channels, i.e. FM4-64, Syto 41, and Sytox green. (d) Distribution of datapoints with respect to storage time. Bacterial cells were fluorescent-stained and store at 4°C for 12h and 24h prior to HCI measurement. Horizontal and vertical axes were the first and second axis of a PCoA. (e) Similar distribution and positions of datapoints corresponding to storage time in PBS. Bacterial cells were fixed and stored in PBS at 4°C prior to staining and image acquisition with the HCI platform.

HCI on multi-well plates is time consuming. Measurement on one single plate can acquire thousands of fluorescent images in six dimensions (X, Y, Z, channels, time, fields of view). One typical experiment could occupy from 100 to 300GB of storage space and could take up to six hours depending on the details of the measurements. It is likely that while one plate is being measured, fluorescent samples on other plates would degrade. Hence, examining the stability of stained cells over time is of important. To do that, stained bacterial cells were stored at 4°C and imaged with Opera Phenix over time. Features were extracted from the Opera images (section 2.6.2.2) and were summarised on Figure 5-5d. I found that longer storage time resulted in more alternation in fluorescent signal. For example, datapoints of 12-hour storage more resembled to data of freshly stained cells, while after 24h of storage, the data showed much distinction. This observation indicates that fluorescent signal of stained cells is unstable with time and HCI measurements should be acquired soon after the staining procedure.

To overcome the instability of fluorescent stained cells, the staining protocol should be paused after the fixation step (Figure 5-4). Fixed cells can be stored in PBS at 4°C and fluorescent-stained prior to HCI measurement. Datapoints from stored cells were found to cluster along with the freshly stained cells (Figure 5-5e). Distribution of the stored-cell data were also found to be comparable to that of the fresh cell. Therefore, storing fixed cells at 4°C does not alternate the fluorescent signals compared to freshly stained cells. This would facilitate high-throughput experiment with the ‘sandwich’ method on multiple plates on one go, without compromising the high quality and consistency of data.

5.3.2. Image processing

After acquiring high-quality images, I next analysed the images with the workflow described in section 2.6.2.2 and Figure 5-7. The purpose of this step was to represent each bacterial cells in the images with numerical data called features. Those features would subsequently be used for AMR inference with machine learning approaches and standard statistical methods. Compared to the most recent image processing workflow published by Sridha *et al.*, I introduced additional morphological operators to smooth the boundaries of the segmented cells. This approach was hypothesised to reduce bias caused by irregular boundaries when calculating texture features. I also introduced cell filter and classifiers to the workflow, removing cells with low signal and irregular shape from the final population. This approach resulted in less cells detected in the population, but higher quality of cell segmentation (Figure 5-8). Since thousands of cells were imaged in each experimental conditions, the image analysis workflow provided good representation of the bacterial populations although a high number of irregular cells were discarded from the final analysis.

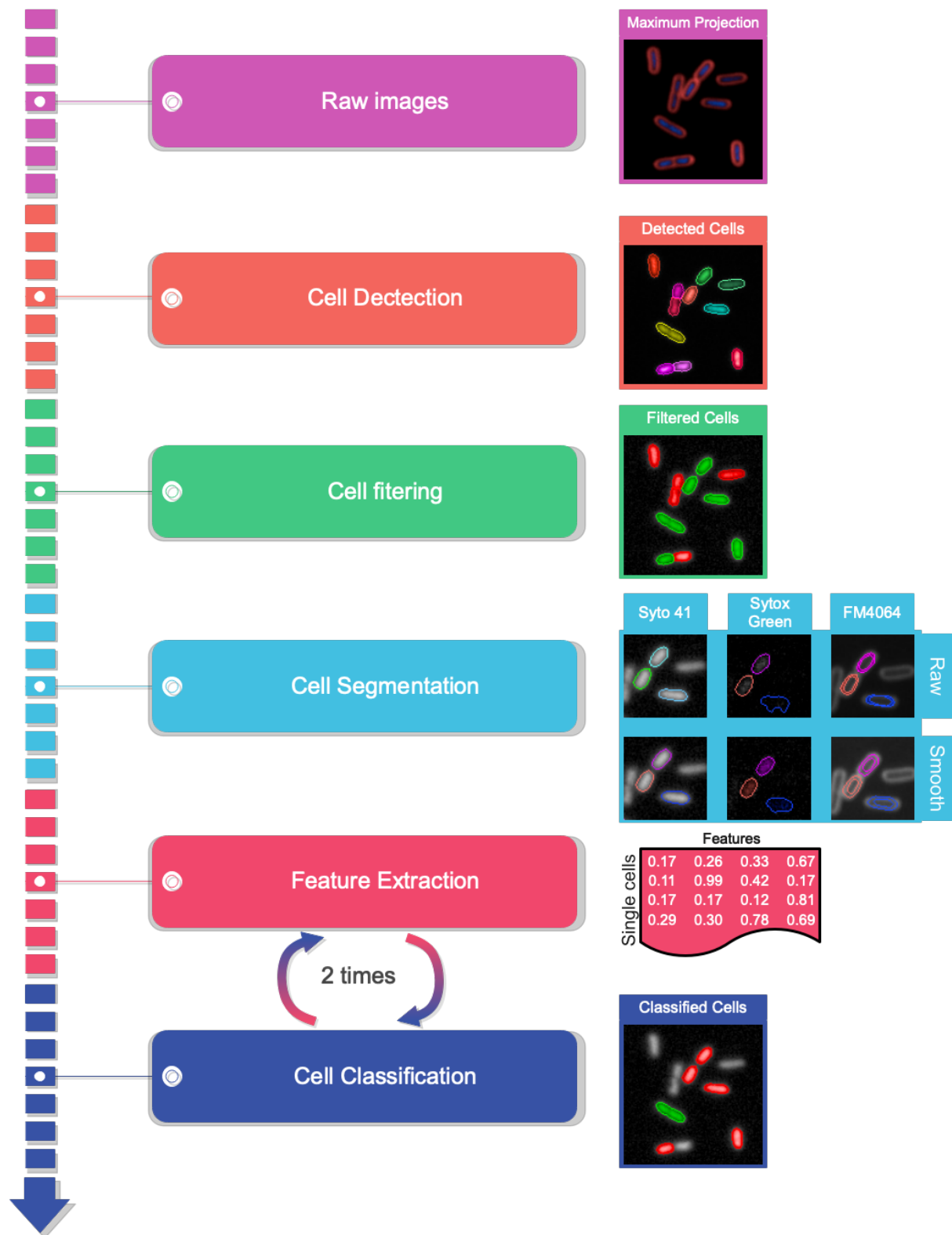


Figure 5-7. Image analysis workflow

The workflow for image analysis goes from the top to the bottom. (Cell detection) Bacterial cells on syto 41 channel. Detected cells were displayed in solid-coloured objects. (Cell

filtering and Cell classification) Remove cells with irregular shape. Red: discarded cells.

Green: remained cells. (Cell segmentation) Segmented cell components. From left to right, nucleoids on syto 41 and sytox green channel, and bacterial membrane on FM4-64 channel.

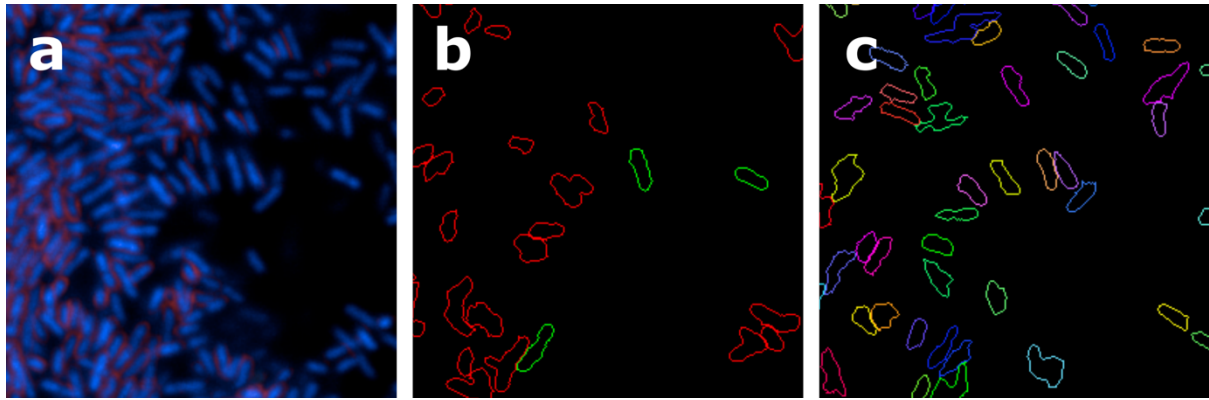


Figure 5-8. Compare the proposed image analysis workflow with Sridha's workflow

(a) Maximum-projected image. (b) Cell segmentation with the proposed workflow. Red: irregular cells not included in the final population. Green: selected cells. (c) Irregular cells segmented with Sridha's method. Cells were marked with coloured objects.

5.4. Integrating imaging with omics

Having standardised protocols for intact protein analysis and HCI with a laboratory standard *E. coli* strain, I aimed to combine the two methods to (1) rapidly identify resistant bacteria with machine learning inference and (2) to discover the complex relation between AMR and protein expression and bacterial morphology. The experiment workflow was described on Figure 5-1. Briefly, seventy *E. coli* isolates (from collection 1 and collection 2) were cultured in liquid media and exposed to three different antimicrobials: CIP, AZI and TE. Non-exposure condition was also investigated. Subsequently, the bacterial cultures were used for intact protein analysis and HCI.

Aiming to determine antimicrobial susceptibility in *E. coli*, I optimised exposure time of the bacteria to different antimicrobials (Figure 5-1). The optimisation was predicted to reduce the experimental time without compromising the predictive performance of the assay. Briefly, a susceptible isolate (ATCC 25922) and a resistant isolate (CM3585) were exposed to antimicrobials for different time lengths and subjected to HCI analysis. The CM3585 was resistant to all the experimented antimicrobials. The time lengths were 0.5, 1, 2, and 3 hours and the experiments were conducted with nine technical replicates. Subsequently, the curated HCI data was input to a dimensional analysis with PCoA and summarised in Figure 5-9. Overall, increased exposure time to the three experimented antimicrobials resulted in more segregation between the susceptible and the resistant isolates. Indeed, I found the most noticeable segregation after three hours of exposure to all the antimicrobials. Hence, three hours was selected as the optimal exposure time for further experiments.

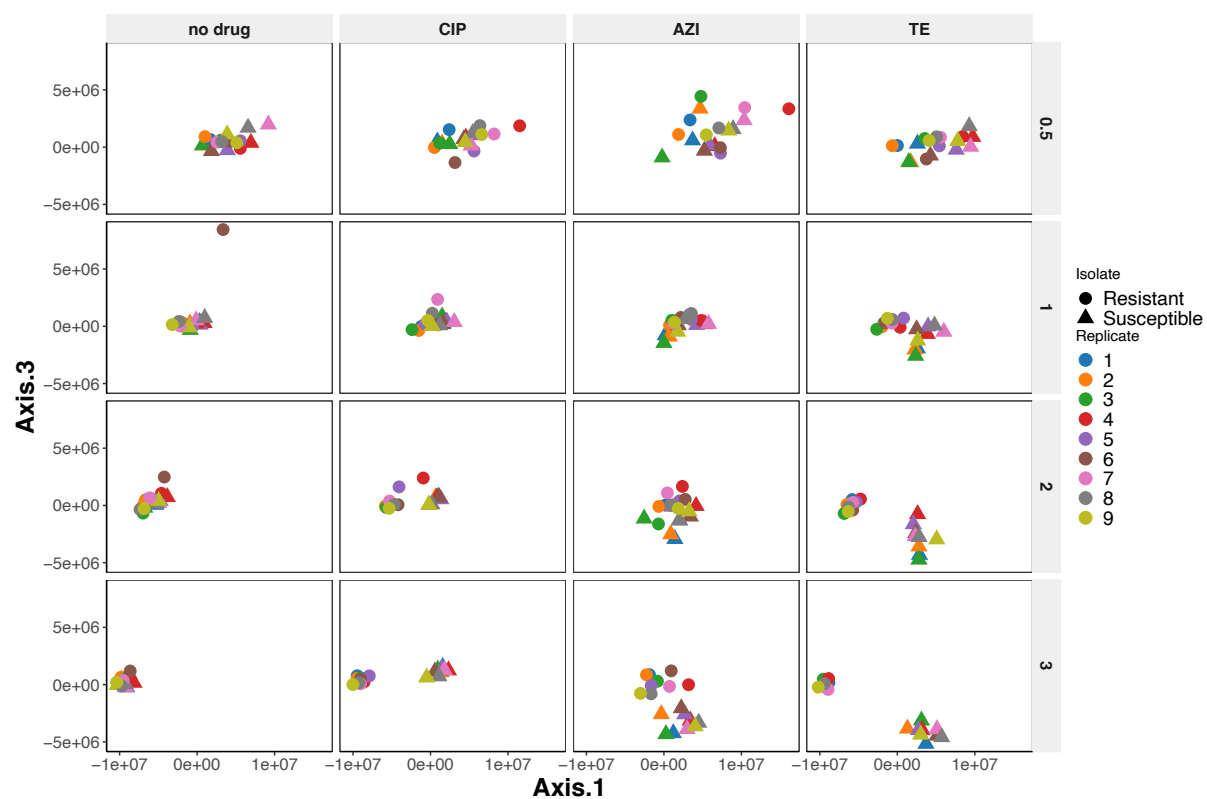


Figure 5-9. Resistant *E. coli* segregated from susceptible ones with increased exposure time to different antimicrobials.

PCoA plots with the first and the third principal axes, explaining 85.68% variation of the HCl data. From left to right, no drug exposure, CIP, AZI, and TE exposure. From top to bottom, 0.5, 1, 2, and 3 hours of exposures.

To establish a standardised control for the experiment, the laboratory reference strain ATCC 25922 was selected as a negative control. The clinical isolate CM3585, which was resistant to all the experimental drugs, was selected as a positive control. The experiment was conducted on the controls with three biological replicates and 12 technical replicates for all the exposure conditions (hereafter called standard controls). On the PCoA plots, all replicates of the controls aggregated within constrained areas based on the bacterial susceptibility, Figure 5-10a and b. These data again indicated that the developed methods for protein analysis and HCI were appreciably robust and reproducible.

Thereafter, the 70 selected clinical isolates were split into ten batches and subjected to the experiment with three biological replicates. For each batch, one replicate of negative, positive, and blank control was included in the experiments. To evaluate the validity of the batches, the controls on each batch were compared to the standardised controls. I found that all the experimental controls were in accordance with the standard, Figure 5-10c and d. These data shows that the batch experiments were valid for further analysis.

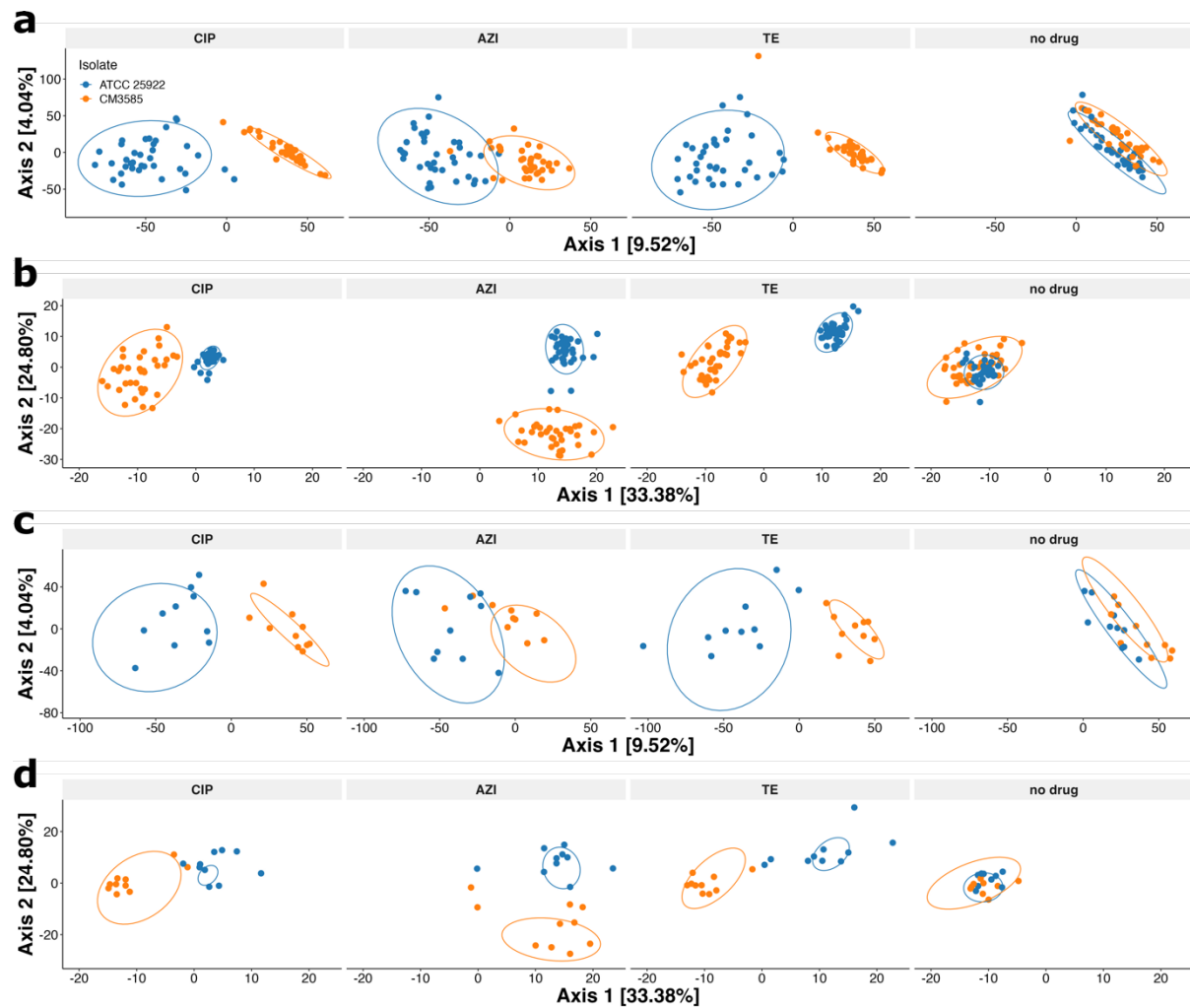


Figure 5-10. Standardisation of controls

Standardised controls for (a) MALDI-TOF MS data and (b) HCI data with 95%-confident ellipses. Each dot was a datapoint within three biological and 12 technical replicates. (c and d) Comparing controls on batch experiments to the standardised controls. (c) MALDI-TOF MS data. (d) HCI data. Confident ellipses were calculated from the standardised controls for the purpose of comparison. Dots represented controls on each batch of experiments. Blue: ATCC 25922. Orange: CM3585.

As with other high-throughput assays, this experiment was prone to batch effect, which introduces misinterpretation and false conclusion to the study. To correct for the batch effect, data was transferred to z-score across all the features with mean and standard deviation calculated within an experimented plate. The corrected MALDI-TOF and HCI data were integrated and subjected to dimensional reduction with PCA. Overall, the distribution of the batch-corrected data strikingly differed from the distribution of the uncorrected data. Principally, I found that with the uncorrected data, the *E. coli* isolates were segregated by their AMR profiles and their source, e.g. blood vs stool (Figure 5-11 left). However, there was no clear evidence of the blood-stool-isolate segregation on the batch-corrected data, although the segregation based on resistance profile was still noticeable (Figure 5-11 right). This result means the blood-stool segregation was a consequence of the batch effect, probably arising from the device calibration and was neglected by the correction. For this reason, batch correction was used to normalise data for further analysis.

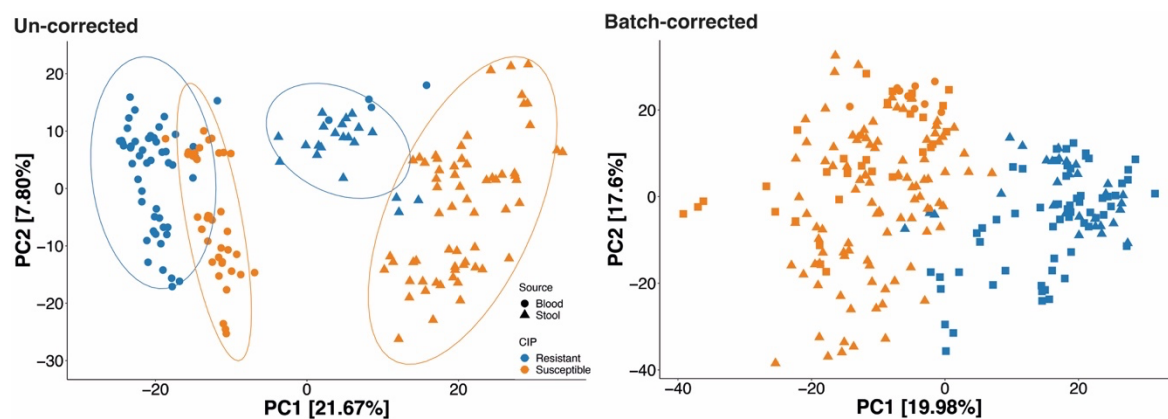


Figure 5-11. Effects of batch-correction

PCoA plots of MALDI-HCI integrated data. From left to right, original data and batch-corrected data. Circle dots: bloodborne isolates. triangular dots: stool isolates, blue: resistant, orange: susceptible.

Additionally, since the experiment flow involved a long series of handling viable microorganisms with multi-channel pipettes (Figure 5-1), the experiment was prone to contamination. To control for contamination, pure plates of blank controls were executed for the overnight culture and bacterial dilution. The resistance profile of each isolate was also assessed on solid media (MC agar) supplemented with the experimental antimicrobials. I found that pure plates on batch 3 and batch 4 showed signs of contamination (Figure 5-1b and c), but there was no contamination on respective resistance assessment plates (Figure 5-1d and e). This observation suggested that contamination could originate from experiment implementation, but not from the liquid media. Additionally, all resistance assessment on solid media was in accordance with resistance profiles previously curated using E-test strips. That suggested no evidence of cross contamination amongst the wells within one experimented plate. I also investigated the impacts of the suspected contamination on the batch effects. The PCA plot on uncorrected data showed a strong cluster of bloodborne isolates from batch 1, 2, 3, and 4. While pure plates in batch 3 and 4 were suspected to be contaminated, those in batch 1 and 2 showed no sign of contamination. Therefore, no strong evidence was found supporting that the suspected contamination contributed to the batch effects. Overall, the contamination was less likely and should not alter any downstream results. Data from the two batches were valid unless abnormality would be found in further analysis.

In conclusion, it was the first time MALDI-TOF has been integrated with HCI to study AMR on a bacterial collection in clinical setting. I have developed and evaluated the optimum protocols for both the two techniques. The data indicated that the proposed methods were unbiased, robust, and highly reproducible. Moreover, experimental controls were

standardised, showing that the method produced valid data for the next analysis in Chapter 6.

Major results and general discussions for chapter 5 and chapter 6 were described in chapter 6.

Chapter 6. Integrating Imaging and Omics with Machine Learning Approaches to Infer Antimicrobial Susceptibility in Gram-negative Bacteria: Results

I declared that all the work in this chapter was performed by myself.

6.1. Introduction

In the previous chapter, seventy *E. coli* isolates were exposed to three different antimicrobials and subjected to intact-protein analysis and HCI for proteomic and cellomic profiling of bacteria respectively. The antimicrobials were CIP, AZI, and TE. The susceptibility profiles of the isolates against these antimicrobials were generated using E-test. Whole genomes of the isolates were also sequenced. Here, I set to employed different machine learning models to infer AMR from the multi-omic profiling. I also aimed to interpret the association between the multi-omic data and AMR, using explainable machine learning techniques.

6.2. Phylogeny and resistome analysis

Overall, the 70 *E. coli* isolates belonged to 37 distinct STs, of which ST1193 accounted for the majority (12.86%) of the collection; ST131 and ST648 equally contributed 10%. To explore the genetic relations among the ST, WGS data of those isolates was exploited to construct a phylogenetic tree (Figure 6-1). According to the phylogenetic tree, the *E. coli* isolates spectated with the phylogenetic tree based on their ST. The isolates were clustered into two main clades, of which clade (II) comprised most of the bloodborne isolates in the collections.

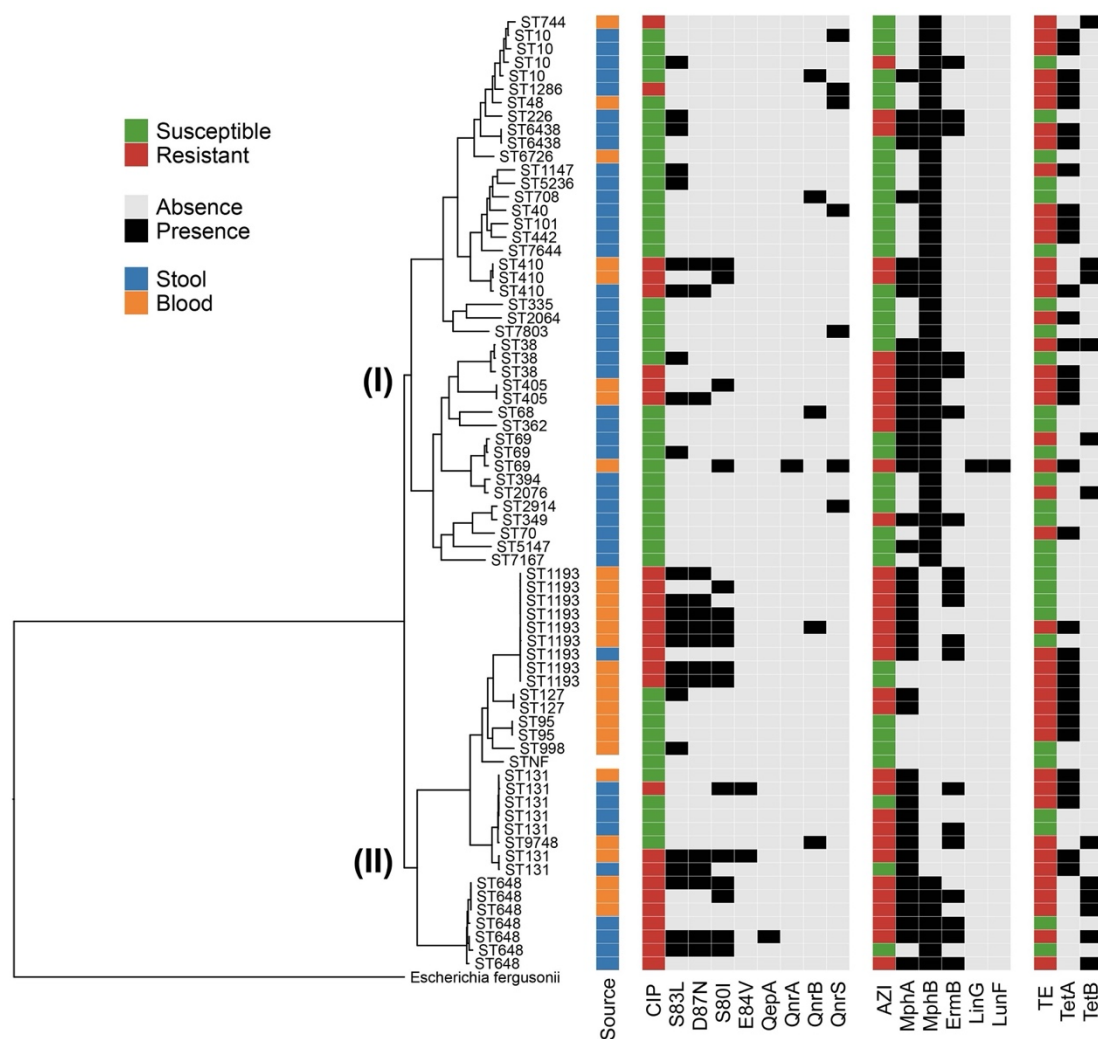


Figure 6-1. Phylogenetic structure of the 70 *E. coli* isolates with respective sequence type, source of isolation, susceptibility, and resistome profiles.

The tips of the phylogenetic tree represented sequence types; *Escherichia fergusonii* was used as an outgroup for tree rooting. The heatmap was divided into four major blocks. From left to right, the first block showed the source of isolates with orange cells were blood isolates and blue cells were stool isolates. The remaining blocks represented the susceptibility and resistome profiles against CIP, AZI, and TE. For each block, the first column displayed the susceptibility profiles with red cells were resistant and green cells were susceptible. The remaining columns of each these blocks showed the presence (black) and absence (grey) of AMR genes.

To explore the distinct behaviours the bloodborne and stool isolates upon exposure to the experimented antimicrobials, various machine learning models were trained on the proteomic and cellomic profiles separately and collectively to distinguish the source of isolation. Of those models, SVM trained on proteomic profiles, under exposure to CIP, gave the highest performance, with an accuracy of 0.79, 0.85 AUC, 0.84 sensitivity, 0.72 specificity, and 0.82 F1 score (Figure 6-2a). The ten most important features were extracted from the SVM model with the mass peaks ranging from 2,283.95 (m/z) to 4,084.65 (m/z) (Figure 6-2b). To understand how bloodborne isolates behaved differently from the stool isolates, a minimum spanning tree (MST) was constructed on the ten most important features (Figure 6-2c). I found that the tree was split into two major branches; the majority of the stool isolates laid on the left branch and the majority of the bloodborne isolates were distributed along the right branch (Figure 6-2c). This finding suggests that the bloodborne isolates display distinct behaviour under exposure to CIP, compared the those of the stool isolates. In a in-depth analysis, the protein profiles for ten most important features are illustrated on a radar chart (Figure 6-2d), showing significantly different patterns of proteomic profiles between the blood and stool isolates (MANOVA, p-value <0.001). Of these features, the stool isolates displayed significantly lower peaks at 2,283.95, 2,706.24, and 3,652.01 m/z, than those of the blood isolates (Figure 6-2d). Overall, based on the source of isolation, pathogenic *E. coli* behaved differently under exposure to CIP, with significantly distinct proteomic profiles. However, I was unable to identify the proteins associated with these peaks.

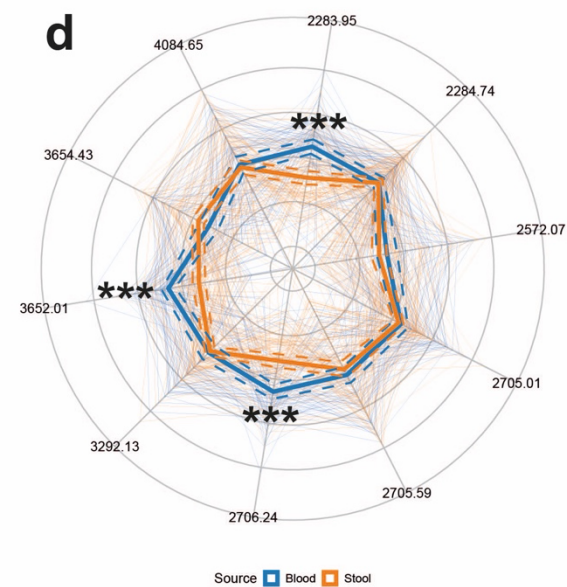
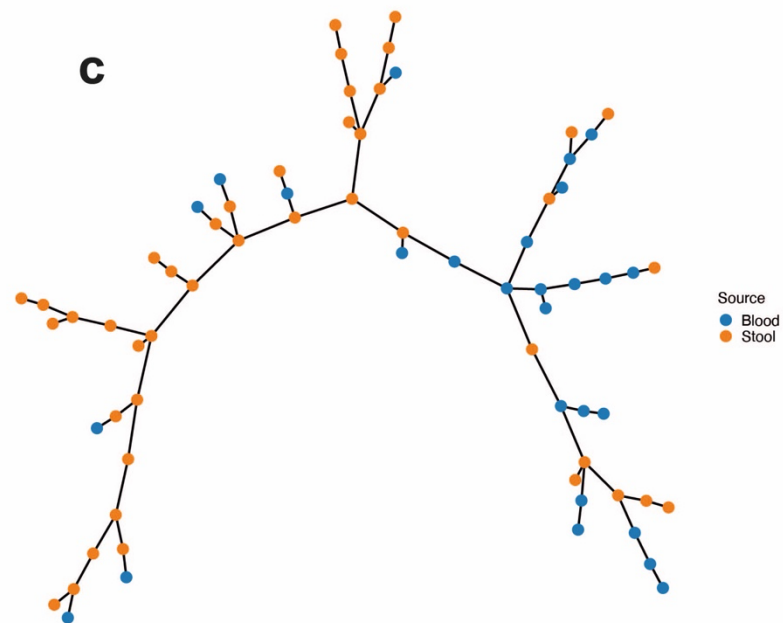
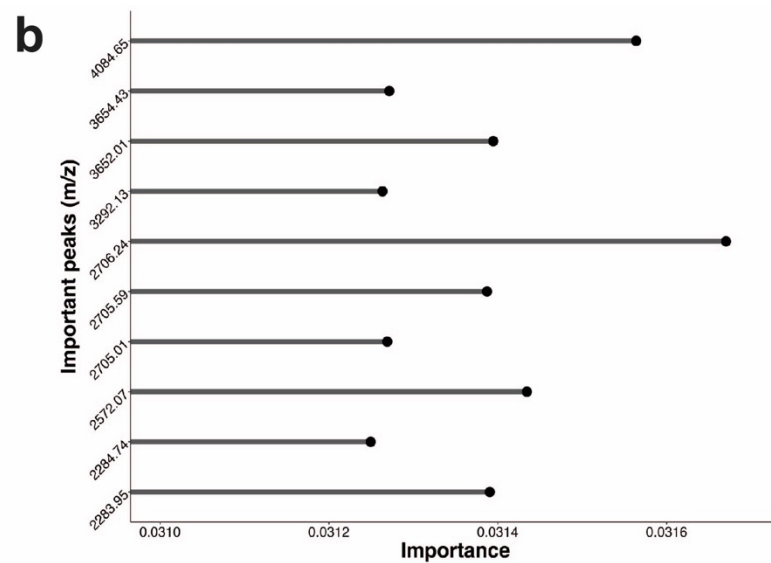
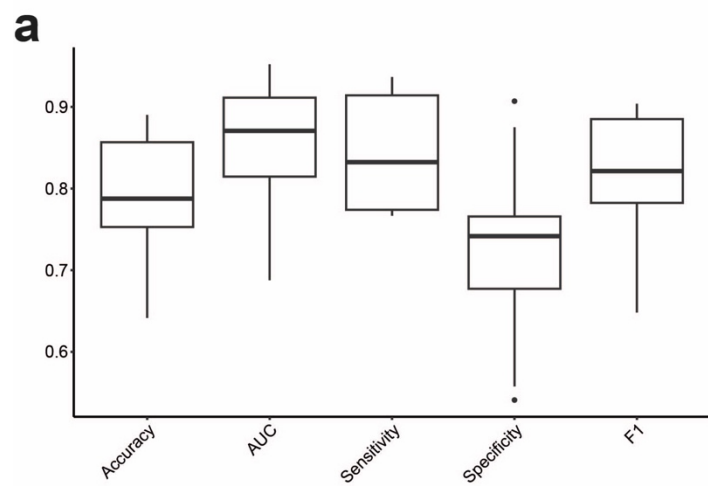


Figure 6-2. Bloodborne and stool isolates behaved differently under exposure to CIP

(a) Performance of the SVM model to predict the source of isolates. Horizontal axis was performance metrics, including accuracy, AUC, sensitivity, specificity and F1 score. Vertical axis represented values of the metrics. (b) Ten important features of mass peaks (y-axis, m/z) with respective calculated importance score (x-axis). (c) Minimum spanning tree built on the ten important peaks of the 70 *E. coli*. Each circle represented one isolate with bloodborne isolates in blue and stool isolates in orange. (d) Radar (spider) chart for the ten important proteomic features of the isolates with each thin lines represents one isolate. Bold lines show average values across bloodborne isolates (orange) or stool isolates (blue). Dashed lines were 95%-confident intervals. Wilcoxon rank-sum test, ***: p-value < 0.001

Using AST, I found a high prevalence of resistance in pathogenic *E. coli* in Vietnam. Overall, 82.86% of the 70 clinical isolates were resistant to at least one antimicrobial, while 45.71% were resistant to two antimicrobials and 20% were resistant to three different antimicrobials (MDR), Figure 6-1. Note that the prevalence of MDR may be much higher since I only tested for three antimicrobials. Blood isolates generally had a higher level of resistance, compared to stool isolates.; 92.59% of blood isolates were resistant to at least one antimicrobial, whereas 76.74% of stool isolates were resistant to at least one antimicrobial. Of the 70 isolates, 38.57%, 48.57%, and 61.43% were resistant to CIP, AZI, and TE, respectively.

I next conducted a resistome analysis, identifying all of the ARGs associated with resistance to CIP, AZI and TE. Among the three experimented antimicrobials, TE resistance displayed the highest correlation with the resistome data, in that all clinical isolates harbouring *tetA* or *tetB* were resistant to TE (Figure 6-1). However, AZI resistance demonstrated a less consistent association with the presence of an associated ARG. Briefly, all blood isolates harbouring *mphA* gene were resistant to AZI, while only 14/24 stool isolates with the *mphA* exhibited resistance to AZI (Figure 6-1). Similarly, isolates having mutations at position D87N in *gyrA* and at S80I in *parC* displayed resistance to CIP. However, there was no evidence in this analysis showing that the presence of well-known fluoroquinolone resistance genes was associated with resistance to CIP (Figure 6-1). Overall, these findings confirm the relation between resistome profile and antimicrobial susceptibility profile in clinical setting, which has been extensively studied in previous publications^{58,106,273}.

6.3. Machine learning to explore the association of proteomics and cellomics with AMR

Although the AMR profiles were generally concordant with the genetic information, genetic-based susceptibility testing is expensive and time-consuming. This method also requires prior

knowledge about resistance mechanisms to identify AMR, which may fail to detect novel resistance of an unknown mechanism. Therefore, I set to employ proteomic and cellomic data, as rapid indicators, using machine learning models to infer susceptibility of the *E. coli* isolates against the selected experimented antimicrobials. Note that if a machine learning model, trained on a dataset, successfully predicts the resistance against an antimicrobial with high accuracy, the resistance strongly associates with the training dataset. In this study, the training data were 1) proteomic data, 2) cellomic data, and 3) the integration of both.

The proteomic (874 features) and cellomic (379 features) datasets were separately or collectively (1253 features) input to different machine learning models, whose evaluation of performance was summarised in Figure 6-3. Overall, the machine learning models successfully captured the association of AMR profiles to the training data with high accuracy. I found that resistance to TE was most strongly associated with the proteomic data, while AZI resistance was most strongly associated with the cellomic data, and resistance to CIP was most strongly associated with the integration of both datasets. More specifically, the random forest model trained on the proteomic data gave the highest accuracy of to predict resistance to TE (0.88), while for AZI resistance, the random forest trained on cellomic data gave the highest performance of 0.81 accuracy. For CIP, the random forest trained on the individual omic datasets possessed considerably high accuracies of ≥ 0.94 . However, random forest trained on the integrated omic data yielded an accuracy of 0.98 to predict CIP resistance, which outperformed the models on the individual datasets. In short, the machine learning models determined a strong association of AMR with bacterial proteomic and cellomic profiles with high predictive performances. These performances are task-specific, which means multiple types of data, i.e., proteomic, cellomic or integrated data, are required

to explain different resistance to different antimicrobials. Therefore, training machine learning on diverse datasets is crucial to improve predictive performance of the models.

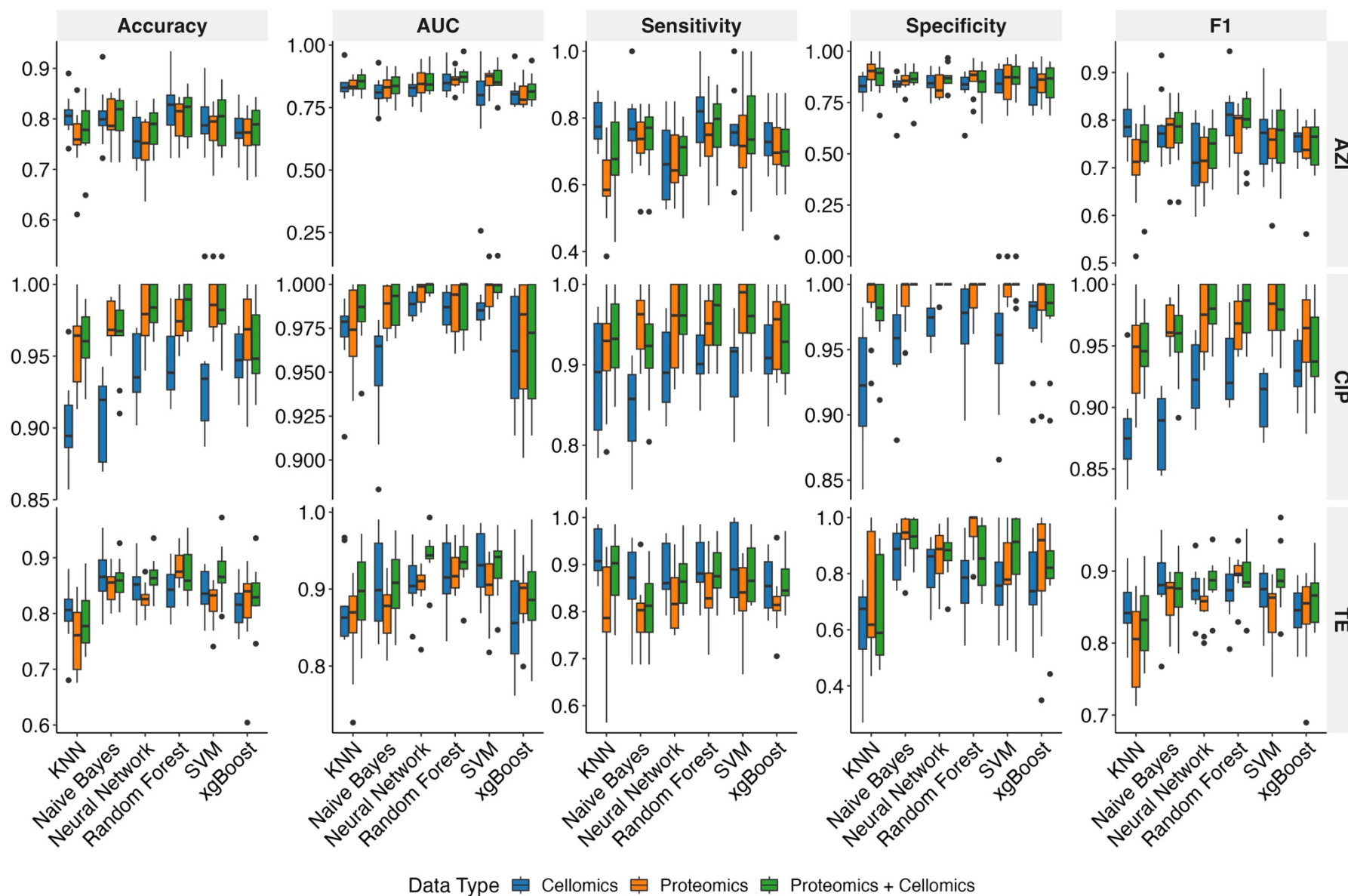


Figure 6-3. Performance of various machine learning models to predict AMR

The models were trained on proteomic (orange), cellomic (blue), or integrated data (green) to predict resistance to AZI (top row), CIP (middle row), and TE (bottom row). Horizontal axis represented different machine learning models, while vertical axis represented the performance metrics, including (from left to right columns) accuracy, AUC, sensitivity, specificity and F1 score.

Given the high prevalence of MDR pathogens found in this study and previous publications^{58,106,274–278}, it is important not only to rapidly provide resistance profiles of the pathogens, but also to quantify the MICs against different antimicrobials. This information is helpful to optimise the use of antimicrobials, reduce the risk of emerging resistance, and improve the treatment outcomes. Here, the proteomic, cellomic and integrated data were input to different machine learning models to predict the MICs of the 70 *E. coli* isolates. The models were evaluated with different metrics, including MSE, MAE, relative absolute errors (RAE), and $\rho_{Spearman}$. Overall, the random forest models possessed the highest predictive performance, with the MSE as low as 0.017, 0.036, and 0.064 to predict MICs against CIP, AZI, and TE respectively (Figure 6-4). In accordance with the AMR prediction analysis, this result indicates that the proteomic and cellomic data were both strongly associated with the bacterial MICs. The result also suggests that under exposure to antimicrobials, genetic resistance factors might manifest as the measurable perturbation in protein expression and cellular morphology, which is informative to predict resistance against various antimicrobials.

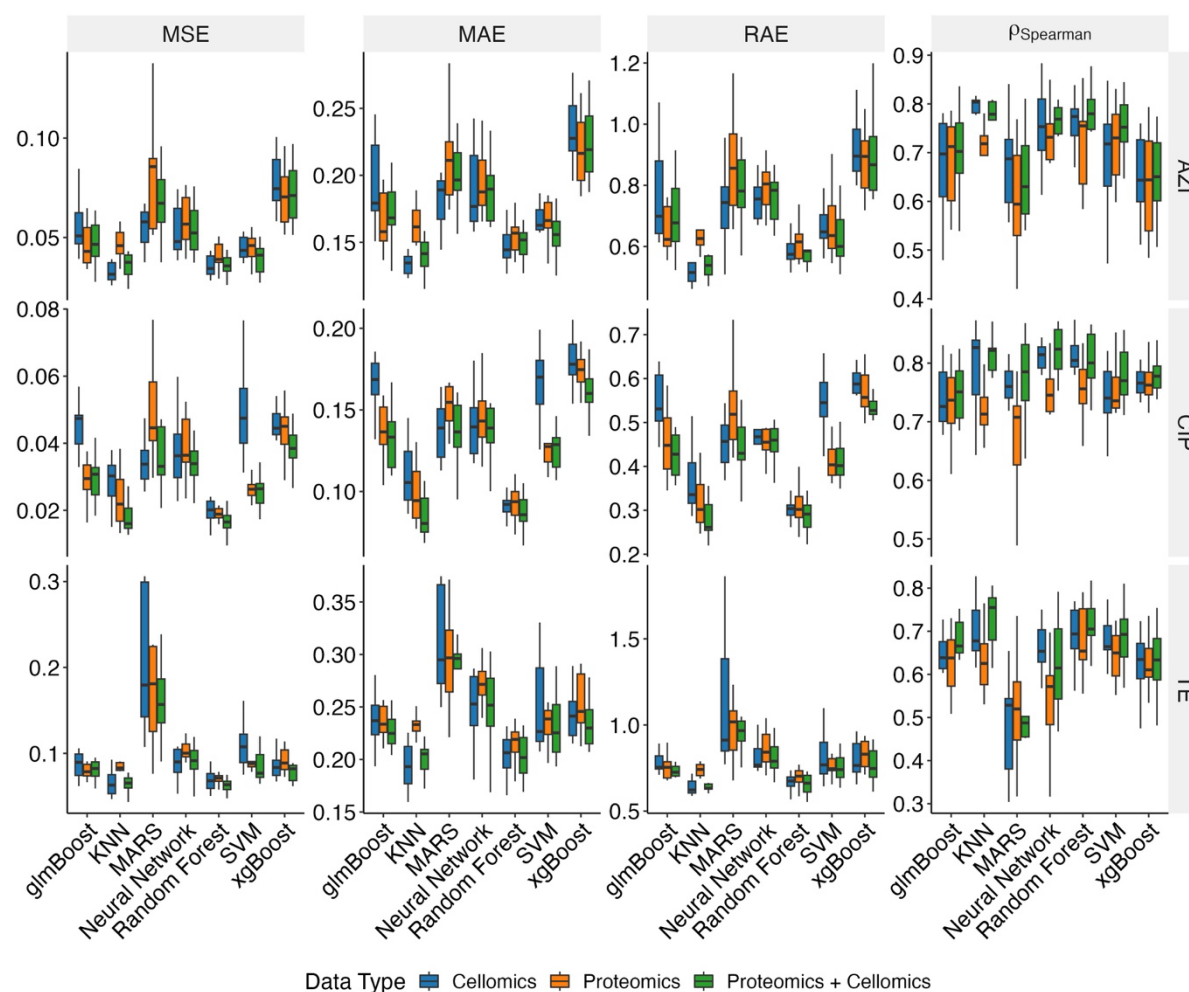


Figure 6-4. Performance of various machine learning models to predict MICs

The models were trained on proteomic (orange), cellomic (blue), or integrated data (green) to predict MICs against AZI (top row), CIP (middle row), and TE (bottom row). Horizontal axis represented different machine learning models, while vertical axis represented the performance metrics, including (from left to right columns) MSE, MAE, RAE, and ρ .

6.4. Proteomic and cellomic analysis with machine learning reveals modifications in protein expression and plasma membrane

As the random forest model trained on the integrated data yielded the highest predictive performance, I used the SHAP values to interpret this model. Considering the random forest model that predicted the likelihood of resistance to CIP, the SHAP values describe how each feature contributes to the predicted outcomes of the model. Typically, more positive SHAP values indicate higher likelihood to be predicted as resistance, whereas more negative SHAP values means higher likelihood to be susceptible. Moreover, the absolute values of SHAP demonstrates how important each feature contributes to the model outcome.

Of 1,253 features from the integrated data, I selected 10 most important features with the highest respective absolute-SHAP-values. Of those ten important features, seven features were from the proteomic profiles, while three features were from the cellomic profiles (Figure 6-5a). This finding means, the proteomic and cellomic profiles both made contributions to the model outcomes. The first important feature from the cellomic profiles was the standard deviation of the dye intensity on the bacterial membrane (Std FM4-64 membrane), which indicates the granularity of the membrane structures. I found that lower Std FM4-64 membrane values were associated with higher likelihood of CIP resistance, Figure 6-5b. This finding suggests that the resistant isolates was able to maintain more intact, or in another word “smoother”, membrane compared to the susceptible isolates, which might be a result of “better” cellular adaptation against CIP exposure. The second and the third important cellomic feature were standard deviation and maximum of the membrane dye intensity on the whole cell, respectively. In accordance with section 4.6, these results suggest that the membrane of susceptible isolates treated with CIP was more permeable compared to the resistant isolates, which allowed the membrane dye to enter the bacterial cytoplasm.

Overall, exposure to CIP inflicted observable perturbations on bacterial membrane, which could be utilised as indicators to identify resistant bacteria.

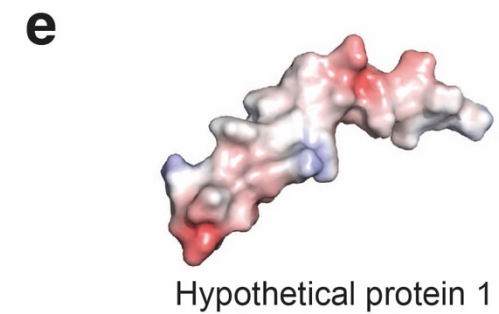
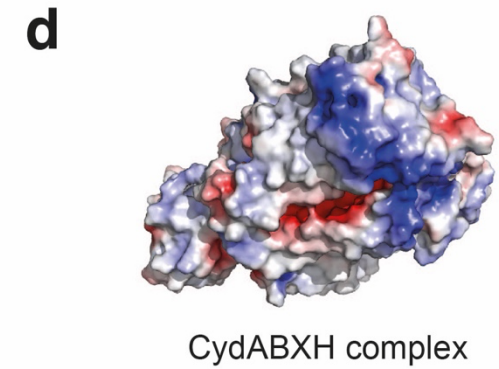
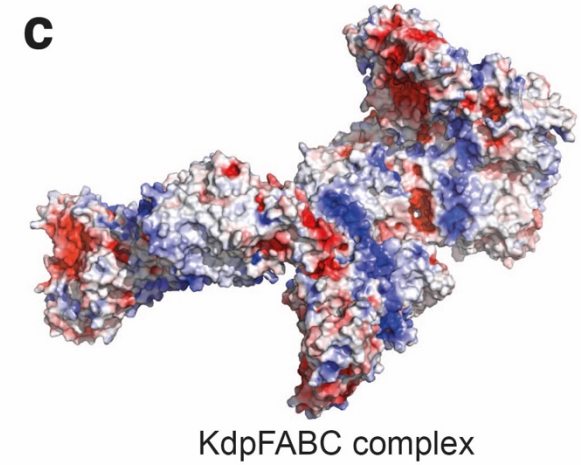
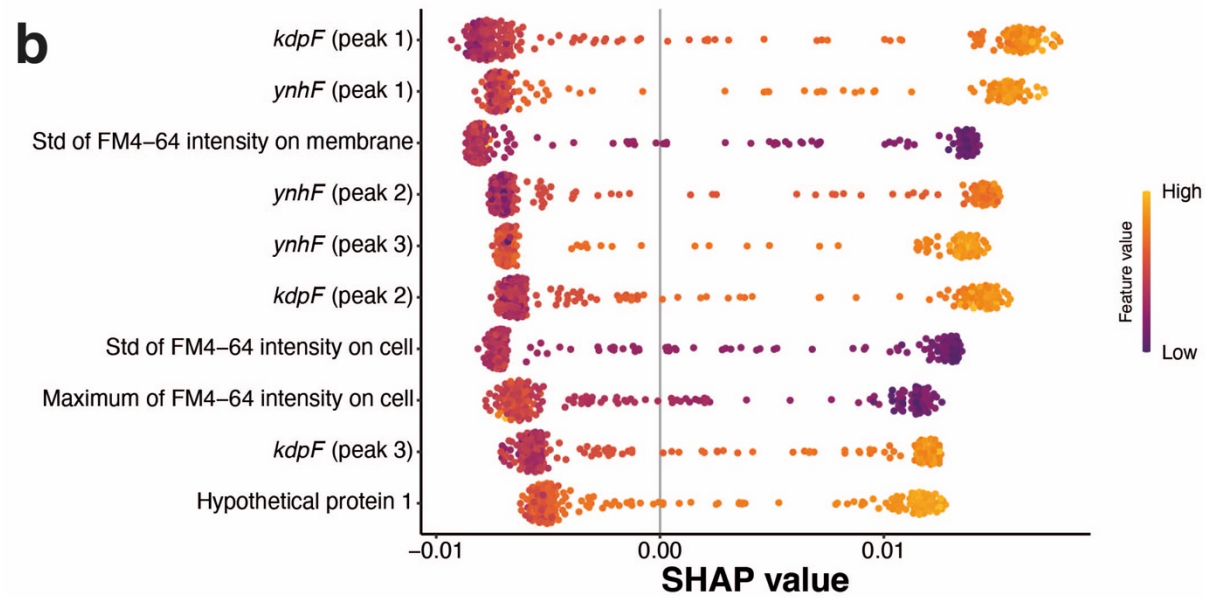
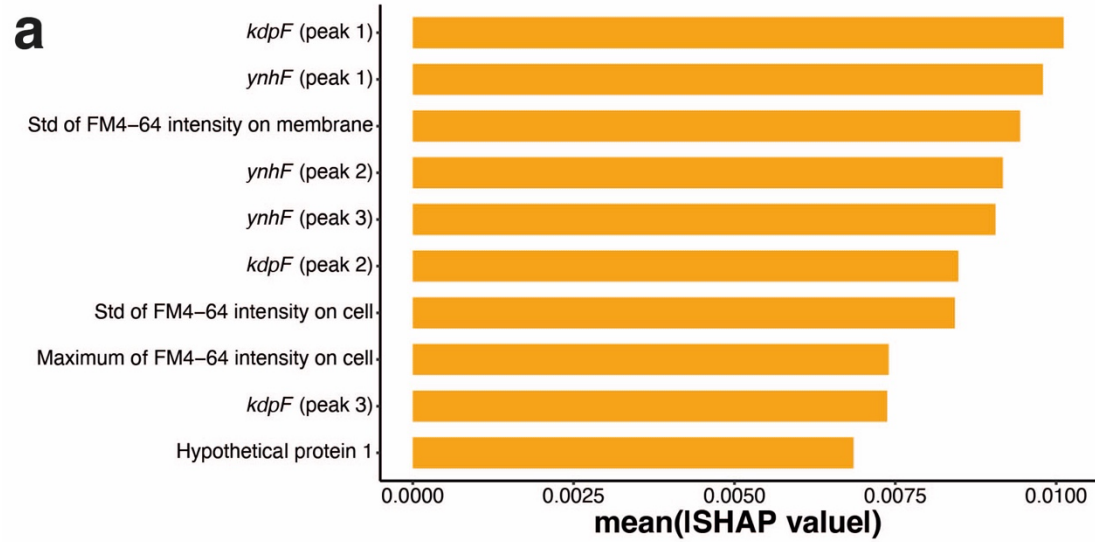


Figure 6-5. Interpretation of the random forest model to predict resistance to CIP

(a) Average of SHAP values for the important features across the collection. Horizontal axis and vertical axis represented mean of absolute SHAP values and the important features, respectively (b) Bee swarm plot for the ten important features. Each dot was the SHAP value for one important feature of one isolate. Magnitude of the feature was showed on the pseudo-colour legend with violet for lower values and yellow for higher values. Horizontal axis represented SHAP values, while vertical axis represented the important features. (c-e) Protein structures of the KdpFABC, CydABXH complex, and HP1. Red area on the protein structures showed negative charges, blue was positive charges, and white was neutral.

Of the ten most important features, the seven remaining features were proteomic profiles, which were measured as mass peaks (in m/z). Using molecular weight matching, I assigned these seven mass peaks to three distinct proteins, these were the hypothetical protein 1 (HP1), YnhF, and KdpF, Figure 6-5a. I found that higher concentrations of these three proteins were associated with higher likelihood of resistance to CIP, Figure 6-5b. Since the HP1 has not been previously studied, I used AlphaFold²³⁵ to predict its 3D structure, which was illustrated in Figure 6-5e. From the predicted structure, I used deepFRI²³⁶, a deep learning algorithm, to predict the molecular and cellular functions of the HP1. The protein was predicted to act as a transferase (GO:0016740), which involves in nitrogen compound metabolic process (GO:0006807).

YnhF, or synonymously CydH, is a transmembrane protein of 29 amino acids. YnhF has no reported biological functions, but is a subunit of the cytochrome bd-I ubiquinol oxidase complex CydABXH²⁷⁹ (Figure 6-5d, Uniprot ID: 6RKO), which is responsible for the stress response in *E. coli*. Although having no known specific molecular and biological functions, YnhF was reported to be overexpressed under oxidative stress, which again emphasises the role of YnhF in bacterial stress response. A study in *S. enterica* showed that deletion of the subunit CydX led to resistance against certain antimicrobials²⁸⁰, but the role of YnhF (or CydH) is yet to be fully determined. Moreover, a gene network analysis on the STRING database v11.5²³⁸ showed that the YnhF was related to the inner membrane transport protein YdhP^{281,282}, which is an efflux pump belonging to the MFS family. However, the contributions of the YdhP to AMR is also unknown. In short, although the YnhF and its related proteins have not been previously reported to confer AMR, they have the potential to be novel mechanisms of resistance, which requires further experimentation to confirm.

KdpF is a subunit of the high-affinity ATP-driven potassium transport system KdpFABC, which transports potassium into the bacterial cytoplasm (Figure 6-5c, Uniprot ID: 5MRW)²⁸³. Using the gene network analysis on the STRING database, I found that the protein KdpF was connected to at least 50 transporter proteins (Figure 6-6a). Of these connections, the KdpF was related to the ABC transporter ATP-binding/permease protein YojI conferring resistance to the antimicrobial peptide microcin J25²⁸⁴. KdpF also connected to the L-methionine/D-methionine ABC transporter complex MetNIQ, Figure 6-6a. The complex is responsible for D-methionine transportation into bacterial cytoplasm, which plays an important role in cell wall synthesis²⁸⁵. Additionally, KdpF was reported to have connection with the iron(III) hydroxamate ABC transporter complex FhuADB (Figure 6-6a), which is known to be a virulence factor in pathogenic *E. coli*²⁸⁶. Finally, KdpF was also related to the outer membrane protein OmpG (Figure 6-6a). Unlike other membrane porins, e.g., OmpA, OmpC, and OmpF⁷², OmpG is not well studied. Although the contribution of the OmpG in AMR remain unclear, the porin has been reported to be vital for survival of *E. coli* lacking both OmpC and OmpF²⁸⁷. To investigate the potentials of OmpG in transportation of antimicrobials, I conducted a molecular docking analysis between the porin and a CIP molecule. I found that the molecule was able to enter the aqueous channel of the OmpG, Figure 6-6b. Indeed, the interaction between the CIP molecule and the aqueous channel was energy-preferable with the binding energy of -7.4 kcal/mol. This suggests that the OmpG might be able to transport antimicrobials into bacterial cytoplasm and deletion of the porin might confer resistance to certain antimicrobials.

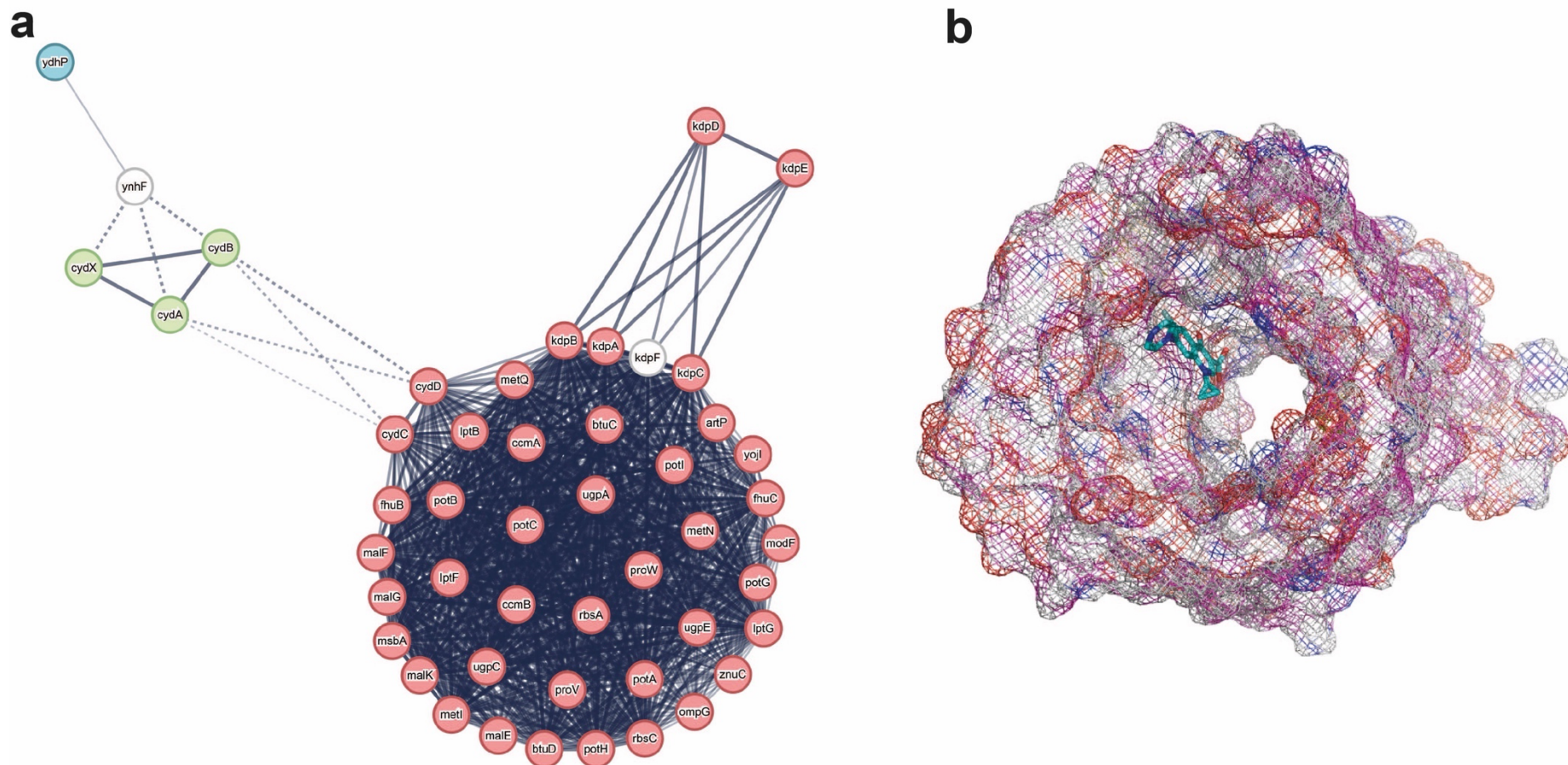


Figure 6-6. Gene network analysis

(a) A gene network analysis on the STRING database with the queried genes on white circles. The network was aggregated in three distinct clusters coloured in red, green and blue. Boundaries between clusters were illustrated as dashed lines. The solid lines between pairs of genes showed the connection of the pair with the thickness of the lines representing the strength of data support. (b) Molecular docking between the

OmpG and a CIP molecule. The OmpG was displayed in a mesh of electrostatic surface, with positively charged area in blue and negatively charged area in red. OmpG structure showed an aqueous channel in the centre of the protein going through one side to another. In the middle of the aqueous channel was the CIP molecule in solid-colour representation.

6.5. Discussion

Machine learning approaches have been widely used in the fields of biology and biomedical science^{173–175}. However, due to their complexity, the interpretation of machine learning models faces certain difficulties, which prevents the technology from being exploited at its full potential in different fields of science²⁶⁹. In recent years, the development of explainable machine learning techniques (a.k.a. “white-box” machine learning) has popularised the use machine learning in the fields of biomedicine^{136,288–293}. In this study, explainable machine learning was used to analyse a complex dataset of bacterial proteomics and cellomics for the first time. Explainable techniques were employed to identify the association between the multi-omic data and the predicted resistance profiles, which provides greater understanding of bacterial adaptation under exposure to different antimicrobials. The use of explainable machine learning in biomedicine and other fields of sciences also promotes a new approach of doing science, so called “data-driven” research²⁹⁴. This study is a typical instance of “data-driven” research, in which the complex dataset of bacterial cellomics and proteomics was analysed to generate novel data-guided hypotheses for future laboratory experiments to work on.

Using explainable machine learning, I generated three hypotheses. First, compared to resistant isolates, susceptible *E. coli* showed considerable difference in membrane granularity under exposure to CIP. Specifically, the membrane dye FM4-64 indicated that susceptible isolates possessed more granular membrane, compared the that of the resistant isolates. Although the effects of CIP exposure on cell wall synthesis and bacterial morphology have been well established^{265,295–297}, the effects on membrane synthesis and structure have been scarcely studied^{298,299}. Indeed, this study is one of the few studies that has reported the modification of bacterial membrane under CIP treatment^{298,299}. I hypothesised that membrane

structure of the susceptible isolates might suffer from a greater perturbation upon exposure to CIP; this hypothesis could be tested with high-resolution electron microscopy.

Additionally, susceptible isolates displayed lower intensity profiles of the transmembrane proteins KdpF, YnhF, and HP1, in comparison to the resistant isolates. I hypothesised that since CIP can interact with DNA topoisomerase and fragment chromosome of bacteria³⁰⁰, the drug induces a reduction on biosynthesis of whole bacterial proteins, including these three important transmembrane proteins. An alternative hypothesis is that exposure to CIP downregulates specific proteins, including these three transmembrane proteins. Further RNA sequencing and protein quantification are required to confirm both hypotheses.

OmpG is an outer membrane protein, which non-selectively transport extracellular solutes into bacterial cytoplasm²⁸⁷. While deletion of other outer membrane protein contribute to resistance against certain antimicrobials⁷², the impacts of OmpG on AMR have not been well studied. Here, I conducted a molecular docking analysis, which suggests that OmpG may transport antimicrobials into bacterial cytoplasm. Hence, I hypothesised that deletion of OmpG might reduce the susceptibility of bacteria against certain antimicrobials. A genetic-engineering experiment could be performed to test this hypothesis.

This study has certain limitation. First, MALDI-TOF MS cannot intrinsically distinguish proteins of (mostly) identical molecular weights. Therefore, proteomic profiling using MALDI-TOF faces difficulty to assign mass peaks to single proteins. In section 6.4, a molecular-weight matching algorithm was implemented on the pangenomes of *E. coli* collection and the three proteins KdpF, YnhF and HP1 were found to have unique molecular

weight; consequently, I was able to assign the seven important mass peaks to these proteins. Other mass peaks might be assigned to multiple proteins or alternatively, may not be assigned to a protein. Second, whilst acquiring snapshots of bacterial population, HCI with Opera Phenix was not able to capture the dynamics of the population, which might provide in-depth understanding on cellular behaviour to the cellomic data, which was previously shown to be informative for rapid identification of resistance^{240,249,301–303}. Third, due to the diffraction limits, the Opera Phenix produces fluorescence images at a resolution no less than 180-250nm¹⁶⁶. Given that the size of a phospholipid molecule is 2nm³⁰⁴, the Opera Phenix failed to describe in detailed the alternation of bacterial membrane under exposure to antimicrobials. However, experiments with the MALDI-TOF MS and Opera Phenix HCI platform are relatively cost effective, which makes these instruments suitable for screening a large collection of bacterial isolates. These techniques are also suitable for high-throughput manipulation, which produces data rapidly and thus, are potential for rapid antimicrobial susceptibility testing.

In conclusion, within a collection of 70 pathogenic *E. coli*, I identified a high prevalence of resistant isolates, especially MDR isolates, which is in concordance with previous studies^{106,273}. The proteomic and cellomic data of those *E. coli* was analysed to predict resistance to three antimicrobials, including CIP, AZI, and TE, using multiple machine learning models with high performance. I found a strong association between the AMR and multi-omic profiles, which generated novel hypotheses about the mechanism of the identified protein. According to the proteomic data, I also found that the blood-borne *E. coli* behaved against antimicrobials differently from the stool isolates. Moreover, three hypotheses were generated based on the in-depth data analysis in this study. For future work, RNA and protein sequencing should be conducted with electron microscopy to examine the data-driven

generated hypotheses. Since distinct bacterial species respond to antimicrobial exposures differently²⁴⁰, the proposed method should be repeated on a more diverse collection of bacteria from multiple species. This further experiment is expected to improve the accuracy, robustness, and availability of the method, which makes it a potential assay for rapid detection of AMR.

Chapter 7. Outlook

AMR is a major global health challenge, which was associated with a predicted 4.95 million deaths in 2019⁵⁸. MALDI-TOF MS has been employed to perform routine bacterial identification at hospitals³⁰⁵, which can provide results within 30 minutes¹³³, compared to 24 hours of classical biochemical tests. HCI platforms, e.g. Opera Phenix, have been exploited in pharmaceutical industry for molecular and genetic screening of eukaryotes^{147,149–152}. In this DPhil, I successfully repurposed the two technologies that could be better exploited to contribute to the fight against AMR. For the first time, these technologies have been employed to produce an enormously complex dataset of bacterial proteomic and cellomic profiles, which were subsequently used for identification of resistant pathogens. The proposed method yielded a high predictive accuracy of 0.98 to detect CIP resistance with the exposure time of bacteria as little as 3 hours. Looking beyond the development of novel AST, these technologies can be use in harmony to screen novel drugs and antibodies for AMR treatments^{154,155}, as well as to identify the molecular mechanisms of these novel therapeutic approaches^{157,159,306}.

Machine learning technology has been commonly used in the field of biomedical science^{173–175}. However, this DPhil is the first instance of applied explainable machine learning on the integration of proteomic and cellomic profiles to rapidly identify AMR (chapter 6). These machine learning models provided high predictive performance to further development of novel rapid AST. Moreover, these explainable models identified the associations of AMR to the proteomic and cellomic profiles, as well as generating novel biological hypotheses for future experiments. However, machine learning is not fully capable of modelling the dynamics of biological processes. Alternatively, mathematical modelling with differential equations has been widely used to describe the dynamics of living systems^{307–310}. Recently,

differential neural network and symbolic machine learning have been developed as a combination of mathematical modelling and machine learning to solve problems in physics^{311–314}. Although applications of these methods in biology and biomedicine are still limited, these methods are expected to contribute more to the better understand AMR and improve human health.

I found that exposure to sub-inhibitory concentrations of AZI accelerated the growth of bacteria (chapter 3), which potentially leads to rapid accumulation of mutations and hence confers resistance to the drug. Since WHO has been prescribing the drug as prophylactic treatments for children²⁵⁵, further research is recommended to assess the risk of developing new resistance mutants compared to benefits of the prophylactics. In this DPhil, MALDI-TOF MS and HCI technology were used to identify resistant in bacteria after as little as 3 hours of exposure to antimicrobials (chapter 6). Given the low cost of running MALDI-TOF and HCI, these techniques could be used to real-time monitor the emergence of novel AZI-resistant isolates during the prophylactic treatment.

Although *E. coli* (chapter 3) and *S. Typhimurium* (chapter 4) are two different species, they both belong to the family Enterobacteriaceae and elaborated similar elongated morphology under exposure to CIP. This observation suggests that species of the one family might share common cytological adaptation under the drug pressure. To construct an image-based rapid AST, experiments on a wide arrange of species are required to establish a database of all possible cytological changes of bacteria upon exposure to antimicrobials. This task is time-consuming and labour-intensive. However, since species of one family expressed common cytological features, database at family or genus level could be sufficient to construct the

image-based AST without compromising the prediction performance. This solution may be expected to reduce the cost of developing the rapid AST in the future.

Overall, machine learning models trained on proteomic and cellomic profiles yielded high accuracy to identify resistance to CIP (0.98), AZI (0.81), and TE (0.88). Since the training data was limited to only 70 isolates of *E. coli*, it is reasonable to expect that the performance of these machine learning models would significantly improve on an extended collection of different bacteria. In future, this proposed method could be optimised to reduce the time of exposure to antimicrobials. Additionally, the cellomic profiling should be further performed on live-cell HCI of bacterial community in clinical samples, which might result in a direct susceptibility test. Machine learning is a versatile technology which is able to comprehend the hidden patterns on large datasets and provide insights into biological problems without experts' knowledge. With the development of robotic and microfluidic technologies, a large amount data is expected to be produced in a timely manner. While it takes time to fully understand the molecular mechanism of living systems, machine learning can benefit from these technology movements to train highly accurate models solving issues in AMR and biomedicine without the requirement of prior knowledge on biological mechanisms. With the emergence of the novel machine learning techniques, which can reveal biological insights, generate new hypotheses, and guide future experiments, machine learning would play a more important role to the advance of biomedical science, contributing to the fight against AMR and improve global health system.

References

1. Serio, A. W., Keepers, T., Andrews, L. & Krause, K. M. Aminoglycoside Revival: Review of a Historically Important Class of Antimicrobials Undergoing Rejuvenation. *EcoSal Plus* **8**, (2018).
2. Black, J. G. *Microbiology: Principles and Explorations*. *Microbiology: Principles and Explorations* (Wiley, 2004).
3. Tuan-Anh, T. *et al.* Pathogenic Escherichia coli Possess Elevated Growth Rates under Exposure to Sub-Inhibitory Concentrations of Azithromycin. *Antibiotics* **9**, 735 (2020).
4. Vuillemin, P. Antibiose et symbiose. *Assoc. Française Pour l'Avancement Sci.* **2**, 525–543 (1890).
5. Bartolemeo Gosio. Ricerche chimiche e batteriologiche sulle alterazioni del mais. *G. Accad. Med. Torino* **63**, 464–487 (1893).
6. Nicolaou, K. C. & Rigol, S. A brief history of antibiotics and select advances in their synthesis. *J. Antibiot. (Tokyo)*. **71**, 153–184 (2018).
7. Flemming A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. *Br. J. Exp. Pathol.* (1929).
8. HODGKIN, D. C. The X-ray analysis of the structure of penicillin. *Adv. Sci.* **6**, (1949).
9. Sheehan, J. C. & Henbry-Logan, K. R. The total synthesis of penicillin V. *J. Am. Chem. Soc.* **79**, 1262–1263 (1957).
10. Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria. *Proc. Soc. Exp. Biol. Med.* **55**, (1944).
11. Jones, D., Metzger, H. J., Schatz, A. & Waksman, S. A. Control of gram-negative bacteria in experimental animals by streptomycin. *Science (80-.)*. **100**, (1944).
12. Wainwright, M. Streptomycin: discovery and resultant controversy. *Hist. Philos. Life Sci.* **13**, (1991).
13. Bo, G. Giuseppe Brotzu and the discovery of cephalosporins. *Clin. Microbiol. Infect.* **6**, (2000).
14. Newton, G. G. F. & Abraham, E. P. Cephalosporin C, a new antibiotic containing sulphur and D- α -aminoadipic acid [1]. *Nature* vol. 175 (1955).
15. Goldin, B. R. Health benefits of probiotics. *Br. J. Nutr.* (1998).
16. Duggar, B. M. Aureomycin: A product of the continuing search for new antibiotics. *Ann. N. Y. Acad. Sci.* **51**, (1948).
17. Sczesny, S., Nau, H. & Hamscher, G. Residue analysis of tetracyclines and their metabolites in eggs and in the environment by HPLC coupled with a microbiological assay and tandem mass spectrometry. *J. Agric. Food Chem.* **51**, 697–703 (2003).
18. Ma, F., Xu, S., Tang, Z., Li, Z. & Zhang, L. Use of antimicrobials in food animals and impact of transmission of antimicrobial resistance on humans. *Biosaf. Heal.* **3**, 32–38 (2021).
19. Daghrir, R. & Drogui, P. Tetracycline antibiotics in the environment: A review. *Environ. Chem. Lett.* **11**, 209–227 (2013).
20. Brockmann, H. & Henkel, W. Pikromycin, ein bitter schmeckendes Antibioticum aus Actinomyceten (Antibiotica aus Actinomyceten, VI. Mitteil. *Chem. Ber.* **84**, (1951).
21. Oliver, M. E. & Hinks, T. S. C. Azithromycin in viral infections. *Rev. Med. Virol.* **31**, 1–13 (2021).
22. Christ, W. Central nervous system toxicity of quinolones: Human and animal findings. *J. Antimicrob. Chemother.* **26**, (1990).
23. Appelbaum, P. C. & Hunter, P. A. The fluoroquinolone antibacterials: Past, present and future perspectives. *International Journal of Antimicrobial Agents* vol. 16 (2000).

24. Hutchings, M., Truman, A. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
25. Fernandes, P. & Martens, E. Antibiotics in late clinical development. *Biochem. Pharmacol.* **133**, 152–163 (2017).
26. Fernandes, P. Antibacterial discovery and development - The failure of success? *Nat. Biotechnol.* **24**, 1497–1503 (2006).
27. Corey, R. *et al.* Safety, tolerability, and efficacy of GSK1322322 in the treatment of acute bacterial skin and skin structure infections. *Antimicrob. Agents Chemother.* **58**, 6518–6527 (2014).
28. Ross, D. B. The FDA and the Case of Ketek. <https://doi.org/10.1056/NEJMp078032> **356**, 1601–1604 (2007).
29. Blum, M. D., Graham, D. J. & Mc Closkey, C. A. Temafloxacin Syndrome: Review of 95 Cases. *Clin. Infect. Dis.* **18**, 946–950 (1994).
30. Shaw, P. J., Ganey, P. E. & Roth, R. A. Idiosyncratic Drug-Induced Liver Injury and the Role of Inflammatory Stress with an Emphasis on an Animal Model of Trovafloxacin Hepatotoxicity. *Toxicol. Sci.* **118**, 7–18 (2010).
31. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).
32. Spagnolo, F., Trujillo, M. & Dennehy, J. J. Why Do Antibiotics Exist? *MBio* **12**, (2021).
33. Clardy, J., Fischbach, M. A. & Currie, C. R. The natural history of antibiotics. *Current Biology* vol. 19 (2009).
34. Chait, R., Vetsigian, K. & Kishony, R. What counters antibiotic resistance in nature? *Nature Chemical Biology* vol. 8 (2012).
35. Traxler, M. F. & Kolter, R. Natural products in soil microbe interactions and evolution. *Nat. Prod. Rep.* **32**, (2015).
36. Klassen, J. L. Microbial secondary metabolites and their impacts on insect symbioses. *Current Opinion in Insect Science* vol. 4 (2014).
37. Seipke, R. F., Kaltenpoth, M. & Hutchings, M. I. Streptomyces as symbionts: An emerging and widespread theme? *FEMS Microbiology Reviews* vol. 36 (2012).
38. Sengupta, S., Chattopadhyay, M. K. & Grossart, H. P. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in Microbiology* vol. 4 (2013).
39. Dcosta, V. M. *et al.* Antibiotic resistance is ancient. *Nature* vol. 477 (2011).
40. Wand, M. E. *et al.* Characterization of pre-antibiotic era *Klebsiella pneumoniae* isolates with respect to antibiotic/disinfectant susceptibility and virulence in *Galleria mellonella*. *Antimicrob. Agents Chemother.* **59**, (2015).
41. Datta, N. & Hughes, V. M. Plasmids of the same Inc groups in enterobacteria before and after the medical use of antibiotics. *Nature* **306**, (1983).
42. Hughes, V. M. & Datta, N. Conjugative plasmids in bacteria of the ‘pre-antibiotic’ era. *Nature* vol. 302 (1983).
43. Abraham, E. P. & Chain, E. An enzyme from bacteria able to destroy penicillin. *Nature* vol. 146 (1940).
44. Colquhoun, J. & Weetch, R. S. RESISTANCE TO CHLORAMPHENICOL DEVELOPING DURING TREATMENT OF TYPHOID FEVER. *Lancet* **256**, (1950).
45. Fleming, A. Penicillin’s finder assays its future. *New York Times* (1945).
46. Olesen, S. W. *et al.* The distribution of antibiotic use and its association with antibiotic resistance. *Elife* **7**, 1–15 (2018).
47. Sun, D. S. *et al.* Analysis of multiple bacterial species and antibiotic classes reveals large variation in the association between seasonal antibiotic use and resistance. *PLoS Biol.* **20**, 1–18 (2022).

48. López-Lozano, J. M. *et al.* A nonlinear time-series analysis approach to identify thresholds in associations between population antibiotic use and rates of resistance. *Nat. Microbiol.* **4**, 1160–1172 (2019).
49. Yang, L. *et al.* Distinct increase in antimicrobial resistance genes among *Escherichia coli* during 50 years of antimicrobial use in livestock production in China. *Nat. Food* **3**, 197–205 (2022).
50. Musicha, P. *et al.* Trends in antimicrobial resistance in bloodstream infection isolates at a large urban hospital in Malawi (1998–2016): a surveillance study. *Lancet Infect. Dis.* **17**, 1042–1052 (2017).
51. WHO. *Antimicrobial resistance surveillance in Europe.* (2022).
52. Alvseike, O., Leegaard, T., Aavitsland, P. & Lassen, J. Trend of multiple drug resistant *Salmonella* Typhimurium in Norway. *Euro Surveill.* **7**, 5–7 (2002).
53. Athleen, M. K. *et al.* Emergence of Multidrug-Resistant *Salmonella enterica* Serotype Typhimurium DT104 Infections in the United States. <https://doi.org/10.1056/NEJM199805073381901> **338**, 1333–1339 (1998).
54. R. COX, T. SU, H. CLOUGH, M. J. WOODWARD, C. S. Spatial and temporal patterns in antimicrobial resistance of *Salmonella* Typhimurium in cattle in England and Wales on JSTOR. *Epidemiol. Infect.* **140**, 2062–2073 (2012).
55. Wang, Y. *et al.* The temporal dynamics of antimicrobial-resistant-*Salmonella enterica* and predominant serovars in China. *Natl. Sci. Rev.* (2022) doi:10.1093/NSR/NWAC269.
56. Lagacé-Wiens, P. R. S. *et al.* Trends in antimicrobial resistance over 10 years among key bacterial pathogens from Canadian hospitals: results of the CANWARD study 2007–16. *J. Antimicrob. Chemother.* **74**, IV22–IV31 (2019).
57. WHO/Samuel Aranda. WHO Ten threats to global health in 2019. *Who* (2019).
58. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
59. Cassini, A. *et al.* Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.* **19**, (2019).
60. Hampton, T. Report reveals scope of US antibiotic resistance threat. *JAMA - Journal of the American Medical Association* vol. 310 (2013).
61. WHO. *Antimicrobial resistance. Global report on surveillance.* (2014) doi:10.1007/s13312-014-0374-3.
62. O'Neill, J. *Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance* (2014).
63. Fernandes, R., Amador, P. & Prudêncio, C. β -Lactams: Chemical structure, mode of action and mechanisms of resistance. *Rev. Res. Med. Microbiol.* **24**, 7–17 (2013).
64. Correia, S., Poeta, P., Hébraud, M., Capelo, J. L. & Igrejas, G. Mechanisms of quinolone action and resistance: where do we stand? *J. Med. Microbiol.* **66**, 551–559 (2017).
65. Aldred, K. J., Kerns, R. J. & Osheroff, N. Mechanism of quinolone action and resistance. *Biochemistry* **53**, 1565–1574 (2014).
66. Redgrave, L. S., Sutton, S. B., Webber, M. A. & Piddock, L. J. V. Fluoroquinolone resistance: Mechanisms, impact on bacteria, and role in evolutionary success. *Trends in Microbiology* vol. 22 (2014).
67. Hooper, D. C. & Jacoby, G. A. Topoisomerase Inhibitors: Fluoroquinolone Mechanisms of Action and Resistance. *Cold Spring Harb. Perspect. Med.* **6**, a025320 (2016).
68. Fàbrega, A., Madurga, S., Giralt, E. & Vila, J. Mechanism of action of and resistance

- to quinolones. *Microb. Biotechnol.* **2**, 40–61 (2009).
69. Fournier, B., Zhao, X., Lu, T., Drlica, K. & Hooper, D. C. Selective targeting of topoisomerase IV and DNA gyrase in *Staphylococcus aureus*: Different patterns of quinolone-induced inhibition of DNA synthesis. *Antimicrob. Agents Chemother.* **44**, 2160–2165 (2000).
 70. Price, L. B. *et al.* In vitro selection and characterization of bacillus anthracis mutants with high-level resistance to ciprofloxacin. *Antimicrob. Agents Chemother.* **47**, 2362–2365 (2003).
 71. Correia, S. *et al.* Impacts of experimentally induced and clinically acquired quinolone resistance on the membrane and intracellular subproteomes of *Salmonella* Typhimurium DT104B. *J. Proteomics* **145**, 46–59 (2016).
 72. Fernández, L. & Hancock, R. E. W. Adaptive and mutational resistance: Role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* **25**, 661–681 (2012).
 73. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta - Proteins Proteomics* **1794**, 808–816 (2009).
 74. Aldred, K. J., Kerns, R. J. & Osheroff, N. Mechanism of quinolone action and resistance. *Biochemistry* **53**, 1565–1574 (2014).
 75. Kurland, C. G. Molecular characterization of ribonucleic acid from *Escherichia coli* ribosomes: I. Isolation and molecular weights. *J. Mol. Biol.* **2**, 83–91 (1960).
 76. Gomes, C. *et al.* Macrolide resistance mechanisms in Enterobacteriaceae: Focus on azithromycin. *Crit. Rev. Microbiol.* **43**, 1–30 (2017).
 77. Ettayebi, M., Prasad, S. M. & Morgan, E. A. Chloramphenicol-erythromycin resistance mutations in a 23S rRNA gene of *Escherichia coli*. *J. Bacteriol.* **162**, 551–557 (1985).
 78. Condon, C., Liveris, D., Squires, C., Schwartz, I. & Squires, C. L. rRNA operon multiplicity in *Escherichia coli* and the physiological implications of *rrn* inactivation. *J. Bacteriol.* **177**, 4152–4156 (1995).
 79. Anderson, P. & Roth, J. Spontaneous tandem genetic duplications in *Salmonella* typhimurium arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc. Natl. Acad. Sci.* **78**, 3113–3117 (1981).
 80. Chittum, H. S. & Champney, W. S. Ribosomal protein gene sequence changes in erythromycin-resistant mutants of *Escherichia coli*. *J. Bacteriol.* **176**, 6192–6198 (1994).
 81. Zaman, S., Fitzpatrick, M., Lindahl, L. & Zengel, J. Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in *Escherichia coli*. *Mol. Microbiol.* **66**, 1039–1050 (2007).
 82. Roberts, M. C. *et al.* Nomenclature for Macrolide and Macrolide-Lincosamide-Streptogramin B Resistance Determinants. *Antimicrob. Agents Chemother.* **43**, 2823–2830 (1999).
 83. Roberts, M. C. Resistance to macrolide, lincosamide, streptogramin, ketolide, and oxazolidinone antibiotics. *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.* **28**, 47–62 (2004).
 84. Harmer, C. J., Holt, K. E. & Hall, R. M. A type 2 A/C2 plasmid carrying the *aacC4* apramycin resistance gene and the *erm(42)* erythromycin resistance gene recovered from two *Salmonella enterica* serovars. *J. Antimicrob. Chemother.* **70**, 1021–1025 (2015).
 85. Billard-Pomares, T. *et al.* Complete nucleotide sequence of plasmid pTN48, encoding the CTX-M-14 extended-spectrum β -lactamase from an *Escherichia coli* O102-ST405 strain. *Antimicrob. Agents Chemother.* **55**, 1270–1273 (2011).
 86. Gomes, C., Ruiz, L., Pons, M. J., Ochoa, T. J. & Ruiz, J. Relevant role of efflux pumps

- in high levels of rifaximin resistance in *Escherichia coli* clinical isolates. *Trans. R. Soc. Trop. Med. Hyg.* **107**, 545–549 (2013).
87. Gomes, C. *et al.* In Vitro Development and Analysis of *Escherichia coli* and *Shigella boydii* Azithromycin-Resistant Mutants. <https://home.liebertpub.com/mdr> **19**, 88–93 (2013).
 88. Wehmeier, C., Schuster, S., Fähnrich, E., Kern, W. V. & Bohnert, J. A. Site-directed mutagenesis reveals amino acid residues in the *Escherichia coli* RND efflux pump AcrB that confer macrolide resistance. *Antimicrob. Agents Chemother.* **53**, 329–330 (2009).
 89. Roberts, M. C. Update on macrolide–lincosamide–streptogramin, ketolide, and oxazolidinone resistance genes. *FEMS Microbiol. Lett.* **282**, 147–159 (2008).
 90. Ojo, K. K. *et al.* The *mef(A)* Gene Predominates among Seven Macrolide Resistance Genes Identified in Gram-Negative Strains Representing 13 Genera, Isolated from Healthy Portuguese Children. *Antimicrob. Agents Chemother.* **48**, 3451–3456 (2004).
 91. Liu, J., Keelan, P., Bennett, P. M. & Enne, V. I. Characterization of a novel macrolide efflux gene, *mef(B)*, found linked to *sul3* in porcine *Escherichia coli*. *J. Antimicrob. Chemother.* **63**, 423–426 (2009).
 92. Desmolaize, B., Rose, S., Warrass, R. & Douthwaite, S. A novel Erm monomethyltransferase in antibiotic-resistant isolates of *Mannheimia haemolytica* and *Pasteurella multocida*. *Mol. Microbiol.* **80**, 184–194 (2011).
 93. Grossman, T. H. Tetracycline antibiotics and resistance. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
 94. Villa, L., Feudi, C., Fortini, D., García-Fernández, A. & Carattoli, A. Genomics of KPC-producing *Klebsiella pneumoniae* sequence type 512 clone highlights the role of RamR and ribosomal S10 protein mutations in conferring tigecycline resistance. *Antimicrob. Agents Chemother.* **58**, 1707–1712 (2014).
 95. Beabout, K. *et al.* The ribosomal S10 protein is a general target for decreased tigecycline susceptibility. *Antimicrob. Agents Chemother.* **59**, 5561–5566 (2015).
 96. Connell, S. R. *et al.* Mechanism of Tet(O)-mediated tetracycline resistance. *EMBO J.* **22**, 945–953 (2003).
 97. Connell, S. R., Tracz, D. M., Nierhaus, K. H. & Taylor, D. E. Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance. *Antimicrob. Agents Chemother.* **47**, 3675–3681 (2003).
 98. Roberts, M. C. Mechanisms of Bacterial Antibiotic Resistance and Lessons Learned from Environmental Tetracycline-Resistant Bacteria. *Antimicrob. Resist. Environ.* 93–121 (2011) doi:10.1002/9781118156247.CH7.
 99. O, R. New principles for the classification of bacteria. *Zentralbl Bakteriell Parasitenkd Infekt. Hyg Abt* **96**, 273–286 (1937).
 100. Janda, J. M., Abbott, S. L. & others. *The enterobacteria*. (American Society for Microbiology (ASM), 2006).
 101. Farmer III, J. J., Brenner, D. J. & Ewing, W. H. Opposition to recent proposals which would reject the family name Enterobacteriaceae and *Escherichia* as its type genus. *Int. J. Syst. Evol. Microbiol.* **30**, 660–673 (1980).
 102. Jenkins, C., Rentenaar, R. J., Landraud, L. & Brisse, S. Enterobacteriaceae. in *Infectious Diseases* 1565–1578 (Elsevier, 2017). doi:10.1016/B978-0-7020-6285-8.00180-5.
 103. Brenner, D. J. & Farmer III, J. J. Enterobacteriaceae. *Bergey's Man. Syst. archaea Bact.* 1–24 (2015).
 104. Farmer III, J. J., Farmer, M. K. & Holmes, B. The Enterobacteriaceae: general characters. in *Topley & Wilson's microbiology & microbial infections* (eds. SP, B., PR,

- M. & G, F.) vol. 2 1317–1359 (Hodder Arnold, 2005).
105. Peto, L. *et al.* The bacterial aetiology of adult community-acquired pneumonia in Asia: A systematic review. *Trans. R. Soc. Trop. Med. Hyg.* **108**, 326–337 (2014).
 106. Lan, N. P. H. *et al.* Phenotypic and genotypic characteristics of ESBL and AmpC producing organisms associated with bacteraemia in Ho Chi Minh City, Vietnam. *Antimicrob. Resist. Infect. Control* **6**, 1–9 (2017).
 107. Odoki, M. *et al.* Prevalence of Bacterial Urinary Tract Infections and Associated Factors among Patients Attending Hospitals in Bushenyi District, Uganda. *Int. J. Microbiol.* **2019**, (2019).
 108. Liu, L. J., Yang, Y. J., Kuo, P. H., Wang, S. M. & Liu, C. C. Diagnostic value of bacterial stool cultures and viral antigen tests based on clinical manifestations of acute gastroenteritis in pediatric patients. *Eur. J. Clin. Microbiol. Infect. Dis.* **24**, 559–561 (2005).
 109. Madigan, M. T., Martinko, J. M., Stahl, D. A. & Clark, D. P. *Brock Biology of Microorganisms. Benjamin Cummings* (Pearson, 2009). doi:10.1017/CBO9781107415324.004.
 110. Percival, S. L. & Williams, D. W. *Escherichia coli*. in *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks* 89–117 (Elsevier, 2013). doi:10.1016/B978-0-12-415846-7.00006-8.
 111. Poolman, J. T. *Escherichia coli*. in *International Encyclopedia of Public Health* vol. 2 585–593 (Elsevier, 2016).
 112. Gonzales, L. *et al.* Prevalence, seasonality and severity of disease caused by pathogenic *Escherichia coli* in children with diarrhoea in Bolivia. *J. Med. Microbiol.* **62**, 1697–1706 (2013).
 113. Khalil, I. A. *et al.* Morbidity and mortality due to shigella and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* **18**, 1229–1240 (2018).
 114. Percival, S. L. & Williams, D. W. *Salmonella*. *Microbiology of Waterborne Diseases: Microbiological Aspects and Risks: Second Edition* (Elsevier, 2013). doi:10.1016/B978-0-12-415846-7.00010-X.
 115. Fàbrega, A. & Vila, J. *Salmonella enterica* serovar Typhimurium skills to succeed in the host: Virulence and regulation. *Clin. Microbiol. Rev.* **26**, 308–341 (2013).
 116. Kozak, G. K., Macdonald, D., Landry, L. & Farber, J. M. Foodborne outbreaks in Canada linked to produce: 2001 through 2009. *Journal of Food Protection* vol. 76 (2013).
 117. Vojdani, J. D., Beuchat, L. R. & Tauxe, R. V. Juice-associated outbreaks of human illness in the United States, 1995 through 2005. *J. Food Prot.* **71**, (2008).
 118. Kariuki, S. *et al.* Characterisation of community acquired non-typhoidal *Salmonella* from bacteraemia and diarrhoeal infections in children admitted to hospital in Nairobi, Kenya. *BMC Microbiol.* **6**, 1–10 (2006).
 119. Crim, S. *et al.* *Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Surveillance Report for 2012*.
 120. Waturangi, D. E., Wiratama, E. & Sabatini, A. Prevalence and molecular characterization of *Salmonella enterica* serovar Typhimurium from ice and beverages in Jakarta, Indonesia. *BMC Res. Notes* **12**, 10–15 (2019).
 121. Ahmed, A. O. *et al.* Salmonellosis: Serotypes, prevalence and multi-drug resistant profiles of *Salmonella enterica* in selected poultry farms, Kwara State, North Central Nigeria. *Onderstepoort J. Vet. Res.* **86**, 1–8 (2019).
 122. van Belkum, A. *et al.* Innovative and rapid antimicrobial susceptibility testing systems. *Nat. Rev. Microbiol.* **18**, 299–311 (2020).

123. Maugeri, G., Lychko, I., Sobral, R. & Roque, A. C. A. Identification and Antibiotic-Susceptibility Profiling of Infectious Bacterial Agents: A Review of Current and Future Trends. *Biotechnol. J.* **14**, (2019).
124. Denamur, E., Clermont, O., Bonacorsi, S. & Gordon, D. The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* **19**, 37–54 (2021).
125. Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
126. Bawn, M. *et al.* Evolution of *Salmonella enterica* serotype Typhimurium driven by anthropogenic selection and niche adaptation. *PLOS Genet.* **16**, e1008850 (2020).
127. Didelot, X. *et al.* Recombination and Population Structure in *Salmonella enterica*. *PLOS Genet.* **7**, e1002191 (2011).
128. Holt, K. E. *et al.* Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3574–E3581 (2015).
129. Illumina. An Introduction to Next-Generation Sequencing Technology. *Dev. Biol.* **276**, 391–402 (2017).
130. Signor, L. & Erba, E. B. Matrix-assisted Laser Desorption/Ionization Time of Flight (MALDI-TOF) Mass Spectrometric Analysis of Intact Proteins Larger than 100 kDa. *JoVE (Journal Vis. Exp.)* e50635 (2013) doi:10.3791/50635.
131. Webster, J. & Oxley, D. Protein Identification by Peptide Mass Fingerprinting using MALDI-TOF Mass Spectrometry. 1117–1129 (2009) doi:10.1007/978-1-59745-198-7_120.
132. Dimzon, I. K. D. & Knepper, T. P. MALDI-TOF MS for Characterization of Synthetic Polymers in Aqueous Environment. *Compr. Anal. Chem.* **58**, 307–338 (2012).
133. Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-Assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* **26**, 547–603 (2013).
134. Flores-Treviño, S. *et al.* Screening of biomarkers of drug resistance or virulence in ESCAPE pathogens by MALDI-TOF mass spectrometry. *Sci. Rep.* **9**, 1–10 (2019).
135. Wang, C. C., Lai, Y. H., Ou, Y. M., Chang, H. T. & Wang, Y. S. Critical factors determining the quantification capability of matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* vol. 374 (2016).
136. Weis, C. *et al.* Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. *Nat. Med.* **28**, 164–174 (2022).
137. Burckhardt, I. & Zimmermann, S. Using matrix-assisted laser desorption ionization-time of flight mass spectrometry to detect carbapenem resistance within 1 to 2.5 hours. *J. Clin. Microbiol.* **49**, 3321–3324 (2011).
138. Esener, N. *et al.* Mass spectrometry and machine learning for the accurate diagnosis of benzylpenicillin and multidrug resistance of *Staphylococcus aureus* in bovine mastitis. *PLoS Comput. Biol.* **17**, 1–28 (2021).
139. Bauman, R. W., Machunis-Masuoka, E. & Montgomery, J. E. *Microbiology: With Diseases by Body System*. (Pearson Education, Inc., 2012).
140. Schermelleh, L. *et al.* Super-resolution microscopy demystified. *Nat. Cell Biol.* **21**, 72–84 (2019).
141. Tavaddod, S. & Naderi-Manesh, H. Evidence of Multi-Domain Morphological Structures in Living *Escherichia coli*. *Sci. Rep.* **7**, 1–7 (2017).
142. Davey, M. E. *et al.* The nucleoid-associated protein HU β affects global gene expression in *Porphyromonas gingivalis*. *Microbiol. (United Kingdom)* **159**, 219–229 (2013).

143. Liu, B., Persons, L., Lee, L. & de Boer, P. A. J. Roles for both FtsA and the FtsBLQ subcomplex in FtsN-stimulated cell constriction in *Escherichia coli*. *Mol. Microbiol.* **95**, 945–970 (2015).
144. Zhao, T. *et al.* Super-resolution imaging reveals changes in *Escherichia coli* SSB localization in response to DNA damage. *Genes to Cells* **24**, 814–826 (2019).
145. Sridhar, S. *et al.* High-Content Imaging to Phenotype Antimicrobial Effects on Individual Bacteria at Scale. *mSystems* **6**, (2021).
146. Lv, Y. *et al.* Antimicrobial properties and membrane-active mechanism of a potential α -helical antimicrobial derived from cathelicidin PMAP-36. *PLoS One* **9**, (2014).
147. Liberali, P., Snijder, B. & Pelkmans, L. Single-cell and multivariate approaches in genetic perturbation screens. *Nat. Rev. Genet.* **16**, 18–32 (2015).
148. Boutros, M., Heigwer, F. & Laufer, C. Microscopy-Based High-Content Screening. *Cell* **163**, 1314–1325 (2015).
149. Chong, Y. T. *et al.* Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* **161**, 1413–1424 (2015).
150. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721–727 (2010).
151. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
152. Sato, S., Rancourt, A., Sato, Y. & Satoh, M. S. Single-cell lineage tracking analysis reveals that an established cell line comprises putative cancer stem cells and their heterogeneous progeny. *Sci. Rep.* **6**, 1–11 (2016).
153. Sridhar, S. *et al.* Inhibitory Concentrations of Ciprofloxacin Induce an Adaptive Response Promoting the Intracellular Survival of *Salmonella enterica* Serovar Typhimurium. *MBio* (2021) doi:10.1128/MBIO.01093-21.
154. Maes, M. *et al.* A novel therapeutic antibody screening method using bacterial high-content imaging reveals functional antibody binding phenotypes of *Escherichia coli* ST131. *Sci. Rep.* **10**, 1–13 (2020).
155. Zoffmann, S. *et al.* Machine learning-powered antibiotics phenotypic drug discovery. *Sci. Rep.* **9**, 1–14 (2019).
156. Quach, D. T., Sakoulas, G., Nizet, V., Pogliano, J. & Pogliano, K. Bacterial Cytological Profiling (BCP) as a Rapid and Accurate Antimicrobial Susceptibility Testing Method for *Staphylococcus aureus*. *EBioMedicine* **4**, 95–103 (2016).
157. Nonejuie, P., Burkart, M., Pogliano, K. & Pogliano, J. Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. *Proc. Natl. Acad. Sci.* **110**, 16169–16174 (2013).
158. Nonejuie, P. *et al.* Application of bacterial cytological profiling to crude natural product extracts reveals the antibacterial arsenal of *Bacillus subtilis*. *J. Antibiot. (Tokyo)*. **69**, 353–361 (2016).
159. Smith, T. C. *et al.* Morphological profiling of tubercle bacilli identifies drug pathways of action. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 18744–18753 (2020).
160. Htoo, H. H. *et al.* Bacterial Cytological Profiling as a Tool To Study Mechanisms of Action of Antibiotics That Are Active against *Acinetobacter baumannii*. *Antimicrob. Agents Chemother.* **63**, 1–11 (2019).
161. Rajput, A. *et al.* Identifying the effect of vancomycin on health care-associated methicillin-resistant *Staphylococcus aureus* strains using bacteriological and physiological media. *Gigascience* **10**, 1–8 (2021).
162. Petite, J. *et al.* Use of high-content analysis and machine learning to characterize complex microbial samples via morphological analysis. *PLoS One* **14**, 1–15 (2019).
163. de Wet, T. J., Winkler, K. R., Mhlana, M., Mizrahi, V. & Warner, D. F. Arrayed

- crispr and quantitative imaging describe the morphotypic landscape of essential mycobacterial genes. *Elife* **9**, 1–36 (2020).
164. Godinez, W. J. *et al.* Morphological Deconvolution of Beta-Lactam Polyspecificity in *E. coli*. *ACS Chem. Biol.* **14**, 1217–1226 (2019).
 165. French, S., Côté, J.-P., Stokes, J. M., Truant, R. & Brown, E. D. Bacteria Getting into Shape: Genetic Determinants of *E. coli* Morphology. *MBio* **8**, 6–11 (2017).
 166. Schneider, J. P. & Basler, M. Shedding light on biology of bacterial cells. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150499 (2016).
 167. Mitchell, T. M. *Machine Learning*. (McGraw-Hill, 1997).
 168. A.L.Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **3**, (1959).
 169. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, (1958).
 170. Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, (1986).
 171. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nat.* 1986 3236088 **323**, 533–536 (1986).
 172. Boser, B. E., Guyon, I. M. & Vapnik, V. N. Training algorithm for optimal margin classifiers. in *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory* (1992). doi:10.1145/130385.130401.
 173. Kan, A. Machine learning applications in cell image analysis. *Immunology and Cell Biology* (2017) doi:10.1038/icb.2017.16.
 174. The, H. C. *et al.* Assessing gut microbiota perturbations during the early phase of infectious diarrhea in Vietnamese children. *Gut Microbes* **0976**, (2017).
 175. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
 176. Lewin-Epstein, O., Baruch, S., Hadany, L., Stein, G. Y. & Obolski, U. Predicting Antibiotic Resistance in Hospitalized Patients by Applying Machine Learning to Electronic Medical Records. *Clin. Infect. Dis.* **72**, e848–e855 (2021).
 177. Luz, C. F. *et al.* Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. *Clin. Microbiol. Infect.* **26**, 1291–1299 (2020).
 178. Anahtar, M. N., Yang, J. H. & Kanjilal, S. Applications of Machine Learning to the Problem of Antimicrobial Resistance: an Emerging Model for Translational Research. *J. Clin. Microbiol.* **59**, e0126020 (2021).
 179. McGuire, R. J. *et al.* A pragmatic machine learning model to predict carbapenem resistance. *Antimicrob. Agents Chemother.* **65**, (2021).
 180. Langlois, R. N. Computers and Semiconductors. in *Technological Innovation and Economic Performance* (2021). doi:10.2307/j.ctv1mjqv2j.15.
 181. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science (80-.)*. **313**, (2006).
 182. Liu, Z. *et al.* Dynamic laser speckle imaging meets machine learning to enable rapid antibacterial susceptibility testing (DYRAST). *ACS Sensors* **5**, 3140–3149 (2020).
 183. Thrift, W. J. *et al.* Deep learning analysis of vibrational spectra of bacterial lysate for rapid antimicrobial susceptibility testing. *ACS Nano* **14**, 15336–15348 (2020).
 184. Weis, C. V., Jutzeler, C. R. & Borgwardt, K. Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. *Clin. Microbiol. Infect.* **26**, 1310–1317 (2020).
 185. Dave, P. *et al.* Has introduction of rapid drug susceptibility testing at diagnosis impacted treatment outcomes among previously treated tuberculosis patients in Gujarat, India? *PLoS One* **10**, e0121996 (2015).

186. Banerjee, R. & Humphries, R. Rapid Antimicrobial Susceptibility Testing Methods for Blood Cultures and Their Clinical Impact. *Front. Med.* **8**, 167 (2021).
187. Li, Y. *et al.* HMD-ARG: hierarchical multi-task deep learning for annotating antibiotic resistance genes. *Microbiome* **9**, 1–12 (2021).
188. Valizadehaslani, T., Zhao, Z., Sokhansanj, B. A. & Rosen, G. L. Amino Acid k-mer Feature Extraction for Quantitative Antimicrobial Resistance (AMR) Prediction by Machine Learning and Model Interpretation for Biological Insights. *Biology (Basel)*. **9**, 365 (2020).
189. Maguire, F., Rehman, M. A., Carrillo, C., Diarra, M. S. & Beiko, R. G. Identification of Primary Antimicrobial Resistance Drivers in Agricultural Nontyphoidal Salmonella enterica Serovars by Using Machine Learning. *mSystems* **4**, (2019).
190. Toan, N. D. *et al.* Clinical features, antimicrobial susceptibility patterns and genomics of bacteria causing neonatal sepsis in a children's hospital in Vietnam: Protocol for a prospective observational study. *BMJ Open* **8**, (2018).
191. Thompson, C. N. *et al.* A cohort study to define the age-specific incidence and risk factors of Shigella diarrhoeal infections in Vietnamese children: A study protocol. *BMC Public Health* **14**, 1–7 (2014).
192. Sridhar, S. An integrated approach to ciprofloxacin susceptibility analysis and high-throughput bacterial phenotyping in Salmonella. (University of Cambridge, 2020). doi:10.17863/CAM.64658.
193. CLSI. *Performance Standards for Antimicrobial Susceptibility Testing*. (Clinical And Laboratory Standards Institute, 2018). doi:10.1108/08876049410065598.
194. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 1–16 (2014).
195. Alcock, B. P. *et al.* CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).
196. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* (2012) doi:10.1089/cmb.2012.0021.
197. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu153.
198. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
199. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
200. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
201. Hoiseth, S. K. & Stocker, B. A. D. Aromatic-dependent Salmonella typhimurium are non-virulent and effective as live vaccines. *Nature* **291**, 238–239 (1981).
202. Kröger, C. *et al.* The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1277–E1286 (2012).
203. SMALT – Wellcome Sanger Institute. <https://www.sanger.ac.uk/tool/smalt-0/>.
204. Hunt, M. *et al.* ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb. Genomics* **3**, e000131 (2017).
205. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.* **57**, 3348–3357 (2013).
206. R Core Team. R: A language and environment for statistical computing. (2014).
207. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* vol. 14 (2003).
208. Raivo, K. & Kolde, R. pheatmap: Pretty Heatmaps. *R Pacakage Version* vol. 61

- <https://cran.r-project.org/package=pheatmap> (2019).
209. Nakazawa, M. fmsb: Functions for Medical Statistics Book with some Demographic Data. *R package version 0.7.1* <https://cran.r-project.org/package=fmsb> (2021).
 210. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer New York, 2009). doi:10.1007/978-0-387-98141-3.
 211. Xiao, N. ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for ‘ggplot2’. *R package version 2.9* <https://cran.r-project.org/package=ggsci> (2018).
 212. Kassambara, A. ggpubr: ‘ggplot2’ Based Publication Ready Plots. *R package version 0.4.0* <https://github.com/kassambara/ggpubr> (2020).
 213. Ishwaran, H., Kogalur, U. B., Blackstone, E. H. & Lauer, M. S. Random survival forests. *Ann. Appl. Stat.* **2**, (2008).
 214. Gilpin, W., Huang, Y. & Forger, D. B. Learning dynamics from large biological data sets: Machine learning meets systems biology. *Curr. Opin. Syst. Biol.* **22**, 1–7 (2020).
 215. Hahn, S. J., Kim, S., Choi, Y. S., Lee, J. & Kang, J. Prediction of type 2 diabetes using genome-wide polygenic risk score and metabolic profiles: A machine learning analysis of population-based 10-year prospective cohort study. *eBioMedicine* **86**, 104383 (2022).
 216. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* (2007) doi:10.1128/AEM.00062-07.
 217. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. Catboost: Unbiased boosting with categorical features. in *Advances in Neural Information Processing Systems* vols 2018-December (2018).
 218. Christoph, M. *Interpretable Machine Learning A Guide for Making Black Box Models Explainable. Book* (2020).
 219. Zhao, Q. & Hastie, T. Causal Interpretations of Black-Box Models. *Journal of Business and Economic Statistics* vol. 39 (2021).
 220. Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *J. Comput. Graph. Stat.* **24**, (2015).
 221. Covert, I. & Lee, S.-I. Improving KernelSHAP: Practical Shapley Value Estimation via Linear Regression. (2020) doi:10.48550/arxiv.2012.01536.
 222. Li, B., Qiu, Y., Shi, H. & Yin, H. The importance of lag time extension in determining bacterial resistance to antibiotics. *Analyst* **141**, 3059–3067 (2016).
 223. Tjørve, K. M. C. & Tjørve, E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS One* **12**, 1–17 (2017).
 224. Buchanan, R. L., Whiting, R. C. & Damert, W. C. When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiol.* **14**, 313–326 (1997).
 225. Chatterjee, T., Chatterjee, B. K., Majumdar, D. & Chakrabarti, P. Antibacterial effect of silver nanoparticles and the modeling of bacterial growth kinetics using a modified Gompertz model. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 299–306 (2015).
 226. Griggs, D. J., Gensberg, K. & Piddock, L. J. V. Mutations in gyrA Gene of Quinolone-Resistant Salmonella Serotypes Isolated from Humans and Animals. *Antimicrob. Agents Chemother.* **40**, 1009–1013 (1996).
 227. Kingsley, R. A. *et al.* Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res.* **19**, 2279–2287 (2009).
 228. Mather, A. E. *et al.* New variant of multidrug-resistant Salmonella enterica serovar

- typhimurium associated with invasive disease in immunocompromised patients in Vietnam. *MBio* **9**, (2018).
229. Van Puyvelde, S. *et al.* An African Salmonella Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. *Nat. Commun.* **10**, 1–12 (2019).
 230. van Oosten, L. N. & Klein, C. D. Machine Learning in Mass Spectrometry: A MALDI-TOF MS Approach to Phenotypic Antibacterial Screening. *J. Med. Chem.* **63**, 8849–8856 (2020).
 231. Lasch, P. *et al.* MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. *Anal. Chem.* **80**, 2026–2034 (2008).
 232. Hernychova, L. *et al.* Detection and identification of *Coxiella burnetii* based on the mass spectrometric analyses of the extracted proteins. *Anal. Chem.* **80**, 7097–7104 (2008).
 233. Marklein, G. *et al.* Matrix-assisted laser desorption ionization-time of flight mass spectrometry for fast and reliable identification of clinical yeast isolates. *J. Clin. Microbiol.* **47**, 2912–2917 (2009).
 234. Gibb, S. & Strimmer, K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics* **28**, 2270–2271 (2012).
 235. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
 236. Gligorijević, V. *et al.* Structure-based protein function prediction using graph convolutional networks. *Nat. Commun.* **12**, (2021).
 237. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, (2021).
 238. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, 10800 (2021).
 239. Otero, F. *et al.* Rapid Detection of Antibiotic Resistance in Gram-Negative Bacteria Through Assessment of Changes in Cellular Morphology. *Microb. Drug Resist.* **23**, 157–162 (2017).
 240. Choi, J. *et al.* A rapid antimicrobial susceptibility test based on single-cell morphological analysis. *Sci. Transl. Med.* **6**, 267ra174-267ra174 (2014).
 241. Spalding, C., Keen, E., Smith, D. J., Krachler, A. & Jabbari, S. Mathematical modelling of the antibiotic- induced morphological transition of *Pseudomonas aeruginosa*. *PLoS Comput. Biol.* 1–28 (2018).
 242. Magiorakos, A. P. *et al.* Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* **18**, 268–281 (2012).
 243. Blázquez, J. *et al.* PBP3 inhibition elicits adaptive responses in *Pseudomonas aeruginosa*. *Mol. Microbiol.* (2006) doi:10.1111/j.1365-2958.2006.05366.x.
 244. Yao, Z., Kahne, D. & Kishony, R. Distinct Single-Cell Morphological Dynamics under Beta-Lactam Antibiotics. *Mol. Cell* **48**, 705–712 (2012).
 245. Buijs, J., Dofferhoff, A. S. M., Mouton, J. W., Wagenvoort, J. H. T. & Van Der Meer, J. W. M. Concentration-dependency of β -lactam-induced filament formation in Gram-negative bacteria. *Clin. Microbiol. Infect.* **14**, 344–349 (2008).
 246. Miller, C. *et al.* SOS response induction by β -lactams and bacterial defense against antibiotic lethality. *Science* (80-.). (2004) doi:10.1126/science.1101630.
 247. Ryan, D. M. & Monsey, D. Bacterial filamentation and in vivo efficacy: A comparison of several cephalosporins. *J. Antimicrob. Chemother.* (1981) doi:10.1093/jac/7.1.57.

248. Rolinson, G. N. Effect of β -lactam antibiotics on bacterial cell growth rate. *J. Gen. Microbiol.* (1980) doi:10.1099/00221287-120-2-317.
249. Baltekin, Ö., Boucharin, A., Tano, E., Andersson, D. I. & Elf, J. Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. *Proc. Natl. Acad. Sci.* **114**, 9170–9175 (2017).
250. Levy, S. B. *The Antibiotic Paradox*. (Perseus Publishing, 2002).
251. Karami, N., Nowrouzian, F., Adlerberth, I. & Wold, A. E. Tetracycline resistance in escherichia coli and persistence in the infantile colonic microbiota. *Antimicrob. Agents Chemother.* **50**, 156–161 (2006).
252. Ahmad, A. *et al.* Pharmacodynamic modelling of in vitro activity of tetracycline against a representative, naturally occurring population of porcine Escherichia coli. *Acta Vet. Scand.* **57**, 1–8 (2015).
253. Dunai, A. *et al.* Rapid decline of bacterial drug-resistance in an antibiotic-free environment through phenotypic reversion. *Elife* **8**, 1–20 (2019).
254. WHO. *Critically Important Antimicrobials for Human Medicine*. (2019).
255. WHO. Presumptive use of azithromycin.
https://www.who.int/maternal_child_adolescent/guidelines/development/provision-of-azithromycin-to-infants/en/.
256. Bergman, M. *et al.* Macrolide and azithromycin use are linked to increased macrolide resistance in Streptococcus pneumoniae. *Antimicrob. Agents Chemother.* **50**, 3646–3650 (2006).
257. Darton, T. C. *et al.* Azithromycin resistance in Shigella spp. in Southeast Asia. *Antimicrob. Agents Chemother.* **62**, (2018).
258. Lin, L. *et al.* Azithromycin Synergizes with Cationic Antimicrobial Peptides to Exert Bactericidal and Therapeutic Activity Against Highly Multidrug-Resistant Gram-Negative Bacterial Pathogens. *EBioMedicine* **2**, 690–698 (2015).
259. Gomes, C., Ruiz-Roldán, L., Mateu, J., Ochoa, T. J. & Ruiz, J. Azithromycin resistance levels and mechanisms in Escherichia coli. *Sci. Rep.* **9**, 1–10 (2019).
260. Horner, P. J. Azithromycin antimicrobial resistance and genital Chlamydia trachomatis infection: Duration of therapy may be the key to improving efficacy. *Sex. Transm. Infect.* **88**, 154–156 (2012).
261. Levin, B. R. *et al.* A Numbers Game: Ribosome Densities, Bacterial Growth, and Antibiotic-Mediated Stasis and Death. *MBio* **8**, 1–13 (2017).
262. Tadmor, A. D. & Tlusty, T. A coarse-grained biophysical model of E. coli and its application to perturbation of the rRNA operon copy number. *PLoS Comput. Biol.* **4**, (2008).
263. Greulich, P., Scott, M., Evans, M. R. & Allen, R. J. Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics. *Mol. Syst. Biol.* **11**, 796–796 (2015).
264. Garneau-Tsodikova, S. & Labby, K. J. Mechanisms of resistance to aminoglycoside antibiotics: Overview and perspectives. *Medchemcomm* **7**, 11–27 (2016).
265. Cushnie, T. P. T., O'Driscoll, N. H. & Lamb, A. J. Morphological and ultrastructural changes in bacterial cells as an indicator of antibacterial mechanism of action. *Cell. Mol. Life Sci.* **73**, 4471–4492 (2016).
266. Dunne, W. M., Jaillard, M., Rochas, O. & Van Belkum, A. Microbial genomics and antimicrobial susceptibility testing. <https://doi.org/10.1080/14737159.2017.1283220> **17**, 257–269 (2017).
267. Schoepp, N. G. *et al.* Rapid pathogen-specific phenotypic antibiotic susceptibility testing using digital LAMP quantification in clinical samples. *Sci. Transl. Med.* **9**, (2017).
268. Liu, Y. Y. *et al.* Emergence of plasmid-mediated colistin resistance mechanism MCR-

- 1 in animals and human beings in China: A microbiological and molecular biological study. *Lancet Infect. Dis.* **16**, 161–168 (2016).
269. Ivonina, M. V., Orimoto, Y. & Aoki, Y. Quantum chemistry-machine learning approach for predicting and elucidating molecular hyperpolarizability: Application to [2.2]paracyclophane-containing push-pull polymers. *J. Chem. Phys.* **154**, (2021).
 270. TI=((‘antimicrobial resistance’ OR ‘antibiotic resistance’ OR resistan*) AND (‘MALDI-TOF’ OR ‘MALDI-TOF MS’ OR ‘matrix-assisted laser desorption ioni*’ OR ‘matrix-assisted laser desorption ioni*mass spectrometry’) AND (detect* OR identif* OR screen*)) – 111 – Web of Science Core Collection.
<https://www.webofscience.com/wos/woscc/summary/0583dc10-d618-44a0-9b23-c720998943d4-4c2eaed9/relevance/1>.
 271. TI=((‘high content screen*’ OR ‘high content imag*’ OR ‘bacteri* cytologi* profil*’) AND (((antimicrobial OR antibiotic) AND resistan*) OR antimicrobial OR antibiotic OR antibacteri* OR bactericid*)) – 6 – Web of Science Core Collection.
<https://www.webofscience.com/wos/woscc/summary/dd859749-5a3f-400c-8413-53e2a76086b1-4c2fbbf8/relevance/1>.
 272. Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).
 273. Duong, V. T. *et al.* Evaluation of luminex xTAG gastrointestinal pathogen panel assay for detection of multiple diarrheal pathogens in fecal samples in Vietnam. *J. Clin. Microbiol.* **54**, 1094–1100 (2016).
 274. Jernigan, J. A. *et al.* Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012–2017. *N. Engl. J. Med.* **382**, 1309–1319 (2020).
 275. Nkansa-Gyamfi, N. A., Kazibwe, J., Traore, D. A. K. & Nji, E. Prevalence of multidrug-, extensive drug-, and pandrug-resistant commensal *Escherichia coli* isolated from healthy humans in community settings in low- and middle-income countries: a systematic review and meta-analysis. *Glob. Health Action* **12**, (2019).
 276. Šámal, V., Paldus, V., Fáčková, D., Mečl, J. & Šrám, J. The prevalence of antibiotic-resistant and multidrug-resistant bacteria in urine cultures from inpatients with spinal cord injuries and disorders: an 8-year, single-center study. *BMC Infect. Dis.* **22**, 1–11 (2022).
 277. Mirzaei, B. *et al.* Prevalence of multi-drug resistant (MDR) and extensively drug-resistant (XDR) phenotypes of *Pseudomonas aeruginosa* and *Acinetobacter baumannii* isolated in clinical samples from Northeast of Iran. *BMC Res. Notes* **13**, 4–9 (2020).
 278. Gnimatin, J. P., Weyori, E. W., Agossou, S. M. & Adokiya, M. N. Bacterial infections epidemiology and factors associated with multidrug resistance in the northern region of Ghana. *Sci. Rep.* **12**, 1–12 (2022).
 279. Hemm, M. R., Weaver, J. & Storz, G. *Escherichia coli* Small Proteome. *EcoSal Plus* **9**, (2020).
 280. Small, T. *et al.* The Small Protein CydX Is Required for Cytochrome bd Quinol Oxidase Stability and Function in *Salmonella enterica* Serovar Typhimurium: a Phenotypic Study. *J. Bacteriol.* **202**, 1–15 (2020).
 281. Lescat, M., Magnan, M., Kenmoe, S., Nordmann, P. & Poirel, L. Co-Lateral Effect of Octenidine, Chlorhexidine and Colistin Selective Pressures on Four Enterobacterial Species: A Comparative Genomic Analysis. *Antibiotics* **11**, (2022).
 282. Shi, D. *et al.* Combined exposure to non-antibiotic pharmaceuticals and antibiotics in the gut synergistically promote the development of multi-drug-resistance in *Escherichia coli*. *Gut Microbes* **14**, (2022).
 283. Huang, C. S., Pedersen, B. P. & Stokes, D. L. Crystal structure of the potassium-importing KdpFABC membrane complex. *Nature* **546**, 681–685 (2017).

284. Delgado, M. A., Vincent, P. A., Farías, R. N. & Salomón, R. A. YojI of *Escherichia coli* functions as a microcin J25 efflux pump. *J. Bacteriol.* **187**, 3465–3470 (2005).
285. Cava, F., De Pedro, M. A., Lam, H., Davis, B. M. & Waldor, M. K. Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids. *EMBO J.* **30**, 3442–3453 (2011).
286. Khasheii, B., Mahmoodi, P. & Mohammadzadeh, A. Siderophores: Importance in bacterial pathogenesis and applications in medicine and industry. *Microbiol. Res.* **250**, 126790 (2021).
287. Misra, R. & Benson, S. A. A novel mutation, cog, which results in production of a new porin protein (OmpG) of *Escherichia coli* K-12. *J. Bacteriol.* **171**, 4105–4111 (1989).
288. Zheng, F. *et al.* Interpretation of cancer mutations using a multiscale map of protein systems. *Science (80-.)*. **374**, (2021).
289. Bourgeais, V., Zehraoui, F., Ben Hamdoune, M. & Hanczar, B. Deep GONet: self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data. *BMC Bioinformatics* **22**, 1–24 (2021).
290. Yang, J. H. *et al.* A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **177**, 1649-1661.e9 (2019).
291. Xie, Y. R., Castro, D. C., Bell, S. E., Rubakhin, S. S. & Sweedler, J. V. Single-Cell Classification Using Mass Spectrometry through Interpretable Machine Learning. *Anal. Chem.* **92**, 9338–9347 (2020).
292. Seethi, V. D. R., LaCasse, Z., Chivte, P., Gaillard, E. R. & Bharti, P. An Explainable-AI approach for Diagnosis of {COVID-19} using MALDI-ToF Mass Spectrometry. *CoRR abs/2109.1*, 1–9 (2021).
293. Yu, F., Wei, C., Deng, P., Peng, T. & Hu, X. Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. *Sci. Adv.* **7**, (2021).
294. Goisauf, M., Akyüz, K. & Martin, G. M. Moving back to the future of big data-driven research: reflecting on the social in genomics. *Humanit. Soc. Sci. Commun.* **7**, 1–9 (2020).
295. Janion, C. Some aspects of the SOS response system - A critical survey. *Acta Biochimica Polonica* vol. 48 (2001).
296. Diver, J. M. & Wise, R. Morphological and biochemical changes in *Escherichia coli* after exposure to ciprofloxacin. *J. Antimicrob. Chemother.* **18**, 31–41 (1986).
297. Barrett, T. C., Mok, W. W. K., Murawski, A. M. & Brynildsen, M. P. Enhanced antibiotic resistance development from fluoroquinolone persists after a single exposure to antibiotic. *Nat. Commun.* **10**, 1–11 (2019).
298. Ponmalar, I. I., Swain, J. & Basu, J. K. Modification of bacterial cell membrane dynamics and morphology upon exposure to sub inhibitory concentrations of ciprofloxacin. *Biochim. Biophys. Acta - Biomembr.* **1864**, 183935 (2022).
299. Wong, F. *et al.* Cytoplasmic condensation induced by membrane damage is associated with antibiotic lethality. *Nat. Commun.* **12**, (2021).
300. Drlica, K., Malik, M., Kerns, R. J. & Zhao, X. Quinolone-mediated bacterial death. *Antimicrob. Agents Chemother.* **52**, 385–392 (2008).
301. Pitruzzello, G. *et al.* Multiparameter antibiotic resistance detection based on hydrodynamic trapping of individual: *E. coli*. *Lab Chip* **19**, 1417–1426 (2019).
302. Zahir, T. *et al.* High-throughput time-resolved morphology screening in bacteria reveals phenotypic responses to antibiotics. *Commun. Biol.* **2**, 1–13 (2019).
303. Zahir, T. *et al.* Image-Based Dynamic Phenotyping Reveals Genetic Determinants of Filamentation-Mediated β -Lactam Tolerance. *Front. Microbiol.* **11**, (2020).
304. Milo, R. & Phillips, R. *Cell Biology by the Numbers. Cell Biology by the Numbers*

- (2015). doi:10.1201/9780429258770.
305. Ge, M. C. *et al.* Routine identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Success rate, economic analysis, and clinical outcome. *J. Microbiol. Immunol. Infect.* **50**, 662–668 (2017).
 306. Peach, K. C., Bray, W. M., Winslow, D., Linington, P. F. & Linington, R. G. Mechanism of action-based classification of antibiotics using high-content bacterial image analysis. *Mol. Biosyst.* **9**, 1837–1848 (2013).
 307. Xavier, J. B. *et al.* Mathematical models to study the biology of pathogens and the infectious diseases they cause. *iScience* **25**, 104079 (2022).
 308. Zhu, Y., Zhao, J. & Li, J. Genome-scale metabolic modeling in antimicrobial pharmacology. *Eng. Microbiol.* **2**, 100021 (2022).
 309. Tomlin, C. J. & Axelrod, J. D. Biology by numbers: Mathematical modelling in developmental biology. *Nat. Rev. Genet.* **8**, 331–340 (2007).
 310. Banwarth-Kuhn, M. & Sindi, S. How and why to build a mathematical model: A case study using prion aggregation. *J. Biol. Chem.* **295**, 5022–5035 (2020).
 311. Cranmer, M. *et al.* Discovering symbolic models from deep learning with inductive biases. *Adv. Neural Inf. Process. Syst.* **2020-Decem**, 1–14 (2020).
 312. Cuomo, S. *et al.* Scientific Machine Learning Through Physics-Informed Neural Networks: Where we are and What's Next. *J. Sci. Comput.* **92**, 1–62 (2022).
 313. B. R. Nogueira, I., V. Santana, V., Ribeiro, A. M. & Rodrigues, A. E. Using scientific machine learning to develop universal differential equation for multicomponent adsorption separation systems. *Can. J. Chem. Eng.* **100**, 2279–2290 (2022).
 314. Rackauckas, C. *et al.* Universal Differential Equations for Scientific Machine Learning. *arXiv* 1–55 (2020).

Appendix

Publications during the DPhil candidature

Check for updates

NEWS & ANALYSIS

GENOME WATCH

Seeing the bigger picture

Caroline Zellmer¹, Tuan Anh Tran and Sushmita Sridhar

This month's Genome Watch highlights the need for improved phenotypic approaches to complement the power of genomics.

In recent years, we have observed a phenomenal genomics revolution, which has transformed our understanding of microbiology and facilitated the characterization of bacteria, including deep exploration of the human microbiome. Whole-genome sequencing has elucidated larger-scale trends, such as species-level relatedness, presence in various geographies and trends of antimicrobial resistance (AMR), as well as the diversity of microbial communities in the gut and environment. However, other highly relevant and information-rich bacterial profiling technologies have lagged behind. In particular, bacterial phenotyping remains largely unimproved beyond conventional confocal microscopy, which generates a limited number of representative images.

Phenotypic profiling through high-content imaging (HCI) leverages automated high-throughput, high-resolution microscopy followed by bespoke analysis pipelines to measure morphological and fluorescence

intensity parameters. These can be deconvoluted using machine learning algorithms to identify relevant features that distinguish bacteria from one another. Such a method creates the opportunity to combine genomic data with extensive single-cell-level phenotypic data to infer what genetic information means at scale in bacterial populations. HCI can help complete our understanding of microorganisms and should be exploited to address the dire AMR crisis. Specifically, HCI can contribute to our understanding of how antimicrobials affect individual cells within communities. As Montaña and colleagues¹ outline, previously unappreciated morphological parameters have helped to uncover valuable information about bacteria that can be useful to differentiate isolates, including non-uniform interactions with antimicrobials.

A HCI approach can expose further variation between drug susceptibility profiles of individual cells within a population, as demonstrated by Pettit et al.². This includes response evaluation of perturbed isolates of interest, measurement of heterogeneity in populations and temporal tracking of phenotypic changes in situ. This approach has great potential for antimicrobial resistance profiling. Conventional AMR evaluation typically assesses the presence of genes encoding β -lactamases and measures disk diffusion or e-test strip results. Unfortunately, these methods fail to consider differentiation within bacterial communities of interest and suitably bridge the genotype–phenotype gap.

Given the unmitigated AMR crisis, our field requires new tools to understand how AMR is leveraged by microorganisms. To accurately and precisely measure the biological impact of a chemical against individual cells and populations, a comprehensive assessment, such as HCI, is needed. With this exhaustive approach, terabytes of high-resolution data are generated, which must be efficiently processed by machine learning-trained models for subsequent analyses and otherwise obscured

pattern identification. These models can identify morphological parameters relevant to bacterial resistance, and these morphologies can be evaluated during antimicrobial treatment and exposure. HCI combined with post-hoc machine learning analysis enables us to combine large amounts of genetic and morphological data to generate unprecedented insight, as Kim et al.³ elegantly demonstrated. Lee and collaborators⁴ also recently outlined how machine learning algorithms can be used to reveal hidden information about bacteria by relying on high-resolution microscopy with large quantities of quality morphological data.

Although these papers are useful and evaluate distinct aspects of microorganism phenotypic variation, including heterogeneity within populations, they do not address the urgent need of improved antimicrobial surveillance and understanding of genome–phenome interactions. Ultimately, there is a strong need to create new phenotypic approaches to complement existing genomic approaches. Early establishment of HCI frameworks could yield actionable and cumulative insights across the range of actors interested in resistance profiling data, from policymakers to patients.

Caroline Zellmer¹✉, Tuan Anh Tran^{1,2} and Sushmita Sridhar^{1,2}

¹Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK

²Baker Lab, Department of Medicine, University of Cambridge, Cambridge, UK

✉e-mail: cz2@sanger.ac.uk

<https://doi.org/10.1038/s41579-021-00640-6>

1. Montaña, E. T. et al. Bacterial cytological profiling identifies rhodamine-containing PAINS analogs as specific inhibitors of *Escherichia coli* thymidylate kinase in vivo. *J. Bacteriol.* **203**, e01010521 (2021).
2. Pettit, J. et al. Use of high-content analysis and machine learning to characterize complex microbial samples via morphological analysis. *PLoS ONE* **14**, e0222528 (2019).
3. Kim, H. et al. Deep learning frameworks for rapid Gram stain image data interpretation: protocol for a retrospective data analysis. *JMIR Res. Protoc.* **9**, e16843 (2020).
4. Lee, J.-Y. et al. Deep learning predicts microbial interactions from self-organized spatiotemporal patterns. *Comput. Struct. Biotechnol. J.* **18**, 1259–1269 (2020).



Credit: Philip Paton/Springer Nature Limited

Article

Pathogenic *Escherichia coli* Possess Elevated Growth Rates under Exposure to Sub-Inhibitory Concentrations of Azithromycin

Tran Tuan-Anh ¹, Ha Thanh Tuyen ¹, Nguyen Ngoc Minh Chau ¹, Nguyen Duc Toan ¹, Tran Hanh Triet ², Le Minh Triet ¹, Nguyen Hoang Thu Trang ¹, Nguyen Thi Nguyen To ¹, Josefin Bartholdson Scott ³, Hao Chung The ¹ , Duy Pham Thanh ¹, Hannah Clapham ⁴ and Stephen Baker ^{3,*}

¹ The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City Q5, Vietnam; anhtt@oucru.org (T.T.-A.); tuyenht@oucru.org (H.T.T.); chaunnm@oucru.org (N.N.M.C.); toannnd@oucru.org (N.D.T.); trietlm@oucru.org (L.M.T.); trangnht@oucru.org (N.H.T.T.); tontn@oucru.org (N.T.N.T.); haoct@oucru.org (H.C.T.); duypt@oucru.org (D.P.T.)

² Division of Aquacultural Biotechnology, Biotechnology Center of Ho Chi Minh City, Ho Chi Minh City Q5, Vietnam; trieth@dac.vn.org

³ Department of Medicine, Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), University of Cambridge, Cambridge CB2 0AW, UK; jb2143@medschl.cam.ac.uk

⁴ Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549, Singapore; hannah.clapham@nus.edu.sg

* Correspondence: sgb47@medschl.cam.ac.uk

Received: 26 August 2020; Accepted: 26 October 2020; Published: 26 October 2020



Abstract: Antimicrobial resistance (AMR) has been identified by the World Health Organization (WHO) as one of the ten major threats to global health. Advances in technology, including whole-genome sequencing, have provided new insights into the origin and mechanisms of AMR. However, our understanding of the short-term impact of antimicrobial pressure and resistance on the physiology of bacterial populations is limited. We aimed to investigate morphological and physiological responses of clinical isolates of *E. coli* under short-term exposure to key antimicrobials. We performed whole-genome sequencing on twenty-seven *E. coli* isolates isolated from children with sepsis to evaluate their AMR gene content. We assessed their antimicrobial susceptibility profile and measured their growth dynamics and morphological characteristics under exposure to varying concentrations of ciprofloxacin, ceftriaxone, tetracycline, gentamicin, and azithromycin. AMR was common, with all organisms resistant to at least one antimicrobial; a total of 81.5% were multi-drug-resistant (MDR). We observed an association between resistance profile and morphological characteristics of the *E. coli* over a three-hour exposure to antimicrobials. Growth dynamics experiments demonstrated that resistance to tetracycline promoted the growth of *E. coli* under antimicrobial-free conditions, while resistance to the other antimicrobials incurred a fitness cost. Notably, antimicrobial exposure heterogeneously suppressed bacterial growth, but sub-MIC concentrations of azithromycin increased the maximum growth rate of the clinical isolates. Our results outline complex interactions between organism and antimicrobials and raise clinical concerns regarding exposure of sub-MIC concentrations of specific antimicrobials.

Keywords: *Escherichia coli*; azithromycin; tetracycline; growth rate; fitness cost; antimicrobial resistance; resistome; morphology; physiology

1. Introduction

Antimicrobial resistance (AMR) in bacterial pathogens is a well-recognized global health problem, causing significant mortality and morbidity. In 2016, 58.6% of *Escherichia coli* (*E. coli*), 34.5% of *Klebsiella pneumoniae* (*K. pneumoniae*), 33.9% *Pseudomonas aeruginosa*, 55.4% of *Acinetobacter spp.*, and 19.8% of *Staphylococcus aureus* isolates in Europe were resistant to at least one antimicrobial [1]. The situation in low- and middle-income countries (LMICs), such as Vietnam, may be worse, where approximately 50% of *E. coli* [2] and 91% of *Shigella flexneri* [3] isolates from clinical samples were resistant to at least three antimicrobial classes (multi-drug-resistant (MDR)). According to WHO, hospitalization costs for patients infected with drug-resistant bacteria are significantly higher than that for patients infected with susceptible bacteria [4]. Indeed, it is estimated that resistant pathogens may result in a total economic cost of USD 0.5 billion and USD 2.8 billion in Thailand and the US, respectively [5]. The outlook is bleak, as the discovery and development of new antimicrobials have declined dramatically [6]. Hence, there is an urgent need for novel approaches to comprehend how AMR organisms emerge and spread and understand how antimicrobials impact on bacterial cells.

E. coli is a member of the Enterobacteriaceae and a component of the healthy human gut microbiota. *E. coli* has been reported as the most prevalent pathogen associated with bacteraemia in Asia (43.8%) [2] and urinary tract infections in Africa (41.9%) [7]. In South America, diarrheagenic *E. coli* was responsible for 45.2% of diarrheal cases in children under five years [8]. Particularly, ETEC (enterotoxigenic *E. coli*) is estimated to cause approximately 222 million diarrheal cases and 51,200 deaths globally in 2016 [9]. The high prevalence and burden of *E. coli* has made it one of the most important pathogens and one of the best studied group of microorganisms in science. However, we have limited contemporary data on the impact of antimicrobials on MDR *E. coli* associated with disease. A scenario may be all too familiar in LMICs where empirical antimicrobials are administered to patients without any microbiological diagnostics or antimicrobial susceptibility testing.

Although antimicrobial susceptibility testing is the gold standard approach to identify and describe resistant bacteria, it provides no insight into the molecular and cellular response triggered by antimicrobial exposure. There have been previous investigations into the molecular mechanisms of AMR in specific bacteria [2] and their morphological [10–12] and physiological characteristics [13] under antimicrobial treatments. However, there are few studies directly linking molecular mechanisms of AMR with the morphology and physiology of antimicrobial-resistant organisms. Given the high burden of resistant *E. coli* in LMICs, we designed a holistic investigation to provide an in-depth understanding of AMR in bloodborne *E. coli* from both a cellular and molecular perspective using microscopic techniques and whole-genome sequencing. Additionally, we studied the growth dynamics of pathogenic *E. coli* to measure the physiological changes of *E. coli* with a diverse compendium of AMR genes upon short-term exposure to different antimicrobial classes.

2. Results

2.1. Resistome and Morphology Associated with AMR

The 27 clinical *E. coli* isolates exhibited a high prevalence of AMR phenotypes, with all isolates resistant to at least one antimicrobial (Table 1). Among those, 70.4%, 70.4%, 88.9%, 74.1%, and 64.0% were resistant to ciprofloxacin, azithromycin, ceftriaxone, tetracycline, and gentamicin, respectively (Table 1). Additionally, 22/27 (81.5%) isolates were resistant to at least three different classes of antimicrobials and defined as MDR [14].

Table 1. Minimum inhibition concentrations (MICs) and antimicrobial resistance of ATCC 25922 and 27 clinical isolates used in this study.

Isolate ID	Minimum Inhibition Concentrations (mg/L)				
	CIP ¹	AZI ¹	CRO ¹	TE ¹	CN ¹
E3585	32 (R)	48 (R)	256 (R)	256 (R)	256 (R)
E3823	32 (R)	48 (R)	256 (R)	256 (R)	256 (R)
E4569	0.064 (S)	12 (S)	32 (R)	4 (S)	1 (S)
E6348	0.38 (S)	8 (S)	256 (R)	4 (S)	256 (R)
E8964	0.38 (S)	64 (R)	256 (R)	256 (R)	64 (R)
E9833	0.38 (S)	16 (S)	256 (R)	256 (R)	1 (S)
E10085	32 (R)	48 (R)	256 (R)	16 (R)	256 (R)
E10487	32 (R)	256 (R)	256 (R)	256 (R)	96 (R)
E10996	0.094 (S)	16 (S)	256 (R)	256 (R)	1 (S)
E11030	32 (R)	4 (S)	0.094 (S)	256 (R)	64 (R)
E12236	0.25 (R)	256 (R)	1.5 (I)	256 (R)	1 (S)
E12241	0.094 (S)	8 (S)	0.064 (S)	128 (R)	1 (S)
E12674	32 (R)	6 (S)	64 (R)	128 (R)	1.5 (S)
E14252	12 (R)	16 (S)	256 (R)	256 (R)	32 (R)
E14488	32 (R)	256 (R)	256 (R)	4 (S)	16 (R)
E15475	32 (R)	256 (R)	256 (R)	256 (R)	1 (S)
E15476	32 (R)	256 (R)	256 (R)	256 (R)	1.5 (S)
E15583	32 (R)	256 (R)	256 (R)	256 (R)	1 (S)
E1456	0.25 (S)	48 (R)	256 (R)	64 (R)	256 (R)
E1965	32 (R)	256 (R)	256 (R)	4 (S)	256 (R)
E2408	32 (R)	256 (R)	256 (R)	4 (S)	64 (R)
E2542	32 (R)	256 (R)	256 (R)	4 (S)	256 (R)
E4751	32 (R)	256 (R)	256 (R)	256 (R)	256 (R)
E5306	0.38 (S)	48 (R)	256 (R)	256 (R)	256 (R)
E5610	32 (R)	256 (R)	256 (R)	4 (S)	256 (R)
E5896	0.38 (S)	256 (R)	128 (R)	256 (R)	256 (R)
E6227	32 (R)	256 (R)	256 (R)	256 (R)	1.5 (S)
ATCC25922	0.094 (S)	4 (S)	0.125 (S)	4 (S)	1.5 (S)
Susceptible	8/27 (29.6%)	8/27 (29.6%)	2/27 (7.4%)	7/27 (25.9%)	10/27 (37.0%)
Intermediate	0/27 (0.0%)	0/27 (0.0%)	1/27 (3.7%)	0/27 (0.0%)	0/27 (0.0%)
Resistant	19/27 (70.4%)	19/27 (70.4%)	24/27 (88.9%)	20/27 (74.1%)	17/27 (63.0%)

¹ CIP: ciprofloxacin, AZI: azithromycin, CRO: ceftiozone, TE: tetracycline, CN: gentamicin, S: susceptible, I: intermediate, R: resistant.

To investigate the AMR gene profiles of the 28-isolate collection, WGS was performed on the 27 clinical *E. coli* isolates and a whole-genome sequence of *E. coli* ATCC 25922 was obtained from the NCBI database. The most prevalent sequence type (ST) was ST1193, which accounted for 8/27 isolates (29.6%); the remainder fell into 11 different STs (Figure 1b). A resistome analysis demonstrated that resistance to azithromycin and tetracycline was associated with the presence of *mphA* and the *tet* gene variants (*tetA/tetR* or *tetB*), respectively. All isolates harboring *aac(3)-IIa* were resistant to gentamicin. In contrast, the resistome data were insufficient for explaining ciprofloxacin resistance, which is commonly attributed to mutations in the Quinolone resistance determining region (QRDR) [15]. Eleven and eight of 27 clinical isolates, respectively, harbored *bla*_{CTX-M-1} and *bla*_{CTX-M-9}, which, as extended-spectrum beta-lactamases, confer resistance to third-generation cephalosporins. Ceftriaxone MICs were positively correlated with the number of acquired resistance genes to beta-lactams in these isolates (i.e., the more *bla* resistance genes, the higher the MIC; $\rho = 0.66$, $p < 0.001$). Moreover, three isolates (ST69 and ST410) carried *bla*_{NDM-1}, which is responsible for carbapenem resistance.

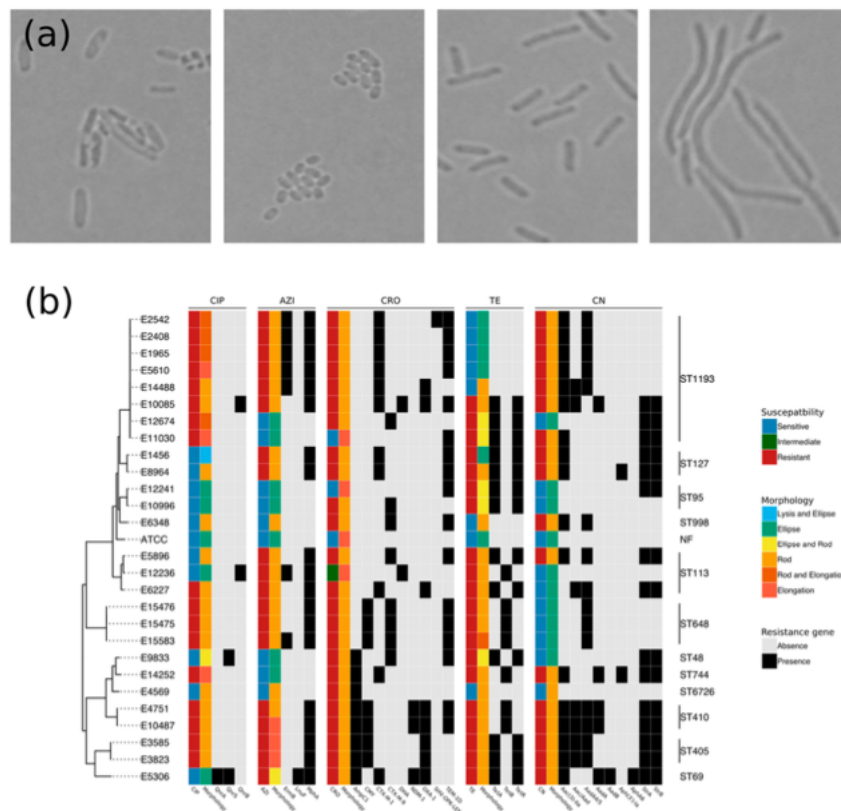


Figure 1. Bacterial morphology associated with antimicrobial resistance (AMR) and resistome. (a) The morphological features of *E. coli* after 3 h of being treated with ciprofloxacin. Brightfield images of *E. coli* in different morphologies, including, from left to right, cell lysis, ellipsoid, rod-like shape, and elongation, were acquired with a Nikon Ni-E upright microscope and a 100× objective lens. Morphology of the reference strain is in rod-like shape in drug-free condition. (b) The phylogenetic structure of *E. coli* with corresponding susceptibility, AMR genes, and morphology. (Left) The unrooted phylogenetic tree constructed from core genomes of the 28 *E. coli* isolates. (Right) Sequence type (ST) of those isolates obtained by mapping the whole-genome reads to a database of seven housekeeping genes. (Middle) Five heatmap blocks, each of which demonstrates (first column) antimicrobial susceptibility, (second column) morphology under exposure to antimicrobials, and (remaining columns) the presence (black) and absence (gray) of resistance genes against quinolone (CIP: ciprofloxacin), macrolide (AZI: azithromycin), β -lactam (CRO: ceftriaxone), tetracycline (TE), and aminoglycoside (CN: gentamicin). NF: non-identified ST.

Aiming to explore the impacts of AMR and AMR genes on morphological changes, the 28 *E. coli* isolates (27 clinical isolates and the reference strain ATCC 25922) were subjected to antimicrobial exposure for three hours at CLSI breakpoint concentrations [16] and examined by light microscopy. The microscopy images were qualitatively classified into four principal categories: lysis, ellipsoid, rod-like shape, and elongation (Figure 1a). Figure 1b demonstrates that *E. coli* cells responded to antimicrobial exposure with differential modifications in their morphology, with some showing an

association with a particular AMR phenotype. Specifically, ceftriaxone-susceptible and -intermediate resistant isolates were elongated, while resistant isolates retained a rod-like shape when exposed to ceftriaxone (Figure 1b). Similarly, the majority of *E. coli* that were resistant to azithromycin, tetracycline, and gentamicin retained a rod morphology, while the majority of susceptible organisms were ellipsoid when exposed to these antimicrobials (Figure 1b).

Upon exposed to azithromycin and tetracycline, morphological changes were more heterogeneous and not consistent with the respective susceptibility profiles of the organisms. For example, of eight tetracycline-resistant *E. coli* isolates, five were ellipsoid and three were rod-like shape. Ciprofloxacin exposure induced the highest level of morphological heterogeneity, in which resistant organisms were rod-like shape, elongated, or a rod–elongated mixture; ciprofloxacin-susceptible organisms were ellipsoid, rod-shaped, or lysed when exposed to ciprofloxacin. We did not observe any obvious association between AMR gene content and bacterial morphological changes under exposure to antimicrobials (Figure 1b), with the exception of the aforementioned rod morphology with *mphA*, *aac(3)-IIa* gene, and *tet* gene variants conferring resistance to azithromycin, gentamicin, and tetracycline, respectively.

2.2. Assessing the Growth Dynamics of *E. coli* under Exposure to Antimicrobials

The 28 *E. coli* isolates were exposed to different concentrations (0 to 256 mg/L) of ciprofloxacin, azithromycin, ceftriaxone, tetracycline, and gentamicin; growth curves were performed over 24 h. One isolate was chosen to perform three technical replicates (Figure S1) and the pair-wise correlation coefficients among three replicates were calculated. An average correlation coefficient of 0.89 was determined. Additionally, each isolate was exposed to azithromycin and tetracycline with an additional one–two biological replicates to confirm the observed effects of resistance on their growth dynamics.

More than 2500 growth curves from the 28 *E. coli* isolates in different antimicrobial exposure conditions were generated, and Gompertz's, Baranyi's, and the three-phase linear models were fit to the growth curves. Gompertz's and Baranyi's models fit with the lowest MSE (mean square error) and MAE (mean absolute error), and the highest ρ (correlation coefficient) and R^2 (correlation of determination) (Figure S2). However, the performance of Gompertz's model was superior to that of Baranyi's model, which indicates that Gompertz's (with three parameters) was the most suitable model for these data. Consequently, we employed Gompertz's model to estimate the growth parameters of each curve for further analysis.

The growth features were plotted separately for five experimental conditions, and against different antimicrobial concentrations as relative to the isolate's respective MIC (Figure 2). Generally, antimicrobials suppressed the growth of *E. coli*, which could be observed as the inferred A reduced gradually and the area under the curve (AUC) reduced rapidly with increasing drug concentrations. Alternatively, increased antimicrobial concentrations induced a prolonged λ in all cases. The μ against relative drug concentration was more heterogeneous and largely dependent on antimicrobial class. The μ of *E. coli* treated with ciprofloxacin and tetracycline declined inversely with respect to an increase in the respective antimicrobial concentrations. However, when organisms were exposed to ceftriaxone, the μ remained constant, with some organisms showing increased μ at higher concentrations. Notably, under azithromycin exposure, the general trend of the 28 *E. coli* isolates was that μ increased in lower concentrations of the macrolide and then decreased gradually at higher drug concentrations. These data suggest that sub-inhibitory concentrations of azithromycin induce an increased rate of *E. coli* replication.

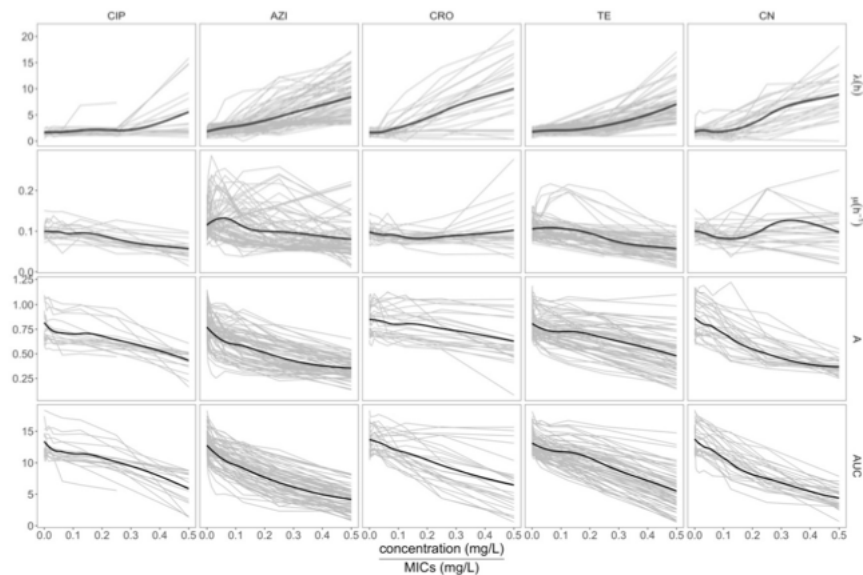


Figure 2. The dynamics of the growth features. From top to bottom, lag-phase period (λ), maximum growth rate (μ), maximum cell density (A), and area under the curve (AUC) corresponding to antimicrobial concentration relative to MICs of 28 *E. coli* isolates from clinical samples. From left to right, ciprofloxacin (CIP), azithromycin (AZI), ceftriaxone (CRO), tetracycline (TE), and gentamicin (CN). The horizontal axis is antimicrobial concentrations relative to MICs of each isolate. The vertical axis represents the values of growth features. Bold lines are loess smooth functions and thin lines are the growth features of each isolate.

On further investigation, we also found that tetracycline-resistant isolates possess a fitness advantage over susceptible isolates in both antimicrobial-free conditions and at $0.5 \times \text{MIC}$ (Figure 3). Specifically, at $0.5 \times \text{MICs}$ of tetracycline, the μ ($p < 0.0001$), A ($p < 0.001$), and AUC ($p < 0.0001$) of the tetracycline-resistant *E. coli* were all significantly higher than those of the susceptible organism (Figure 3). Without antimicrobial exposure, the AUC of tetracycline-resistant isolates was significantly higher than that of susceptible isolates ($p < 0.05$). Conversely, azithromycin-susceptible *E. coli* in drug-free conditions had a fitness advantage over azithromycin-resistant *E. coli* with significantly higher AUCs ($p < 0.05$) (Figure 3).

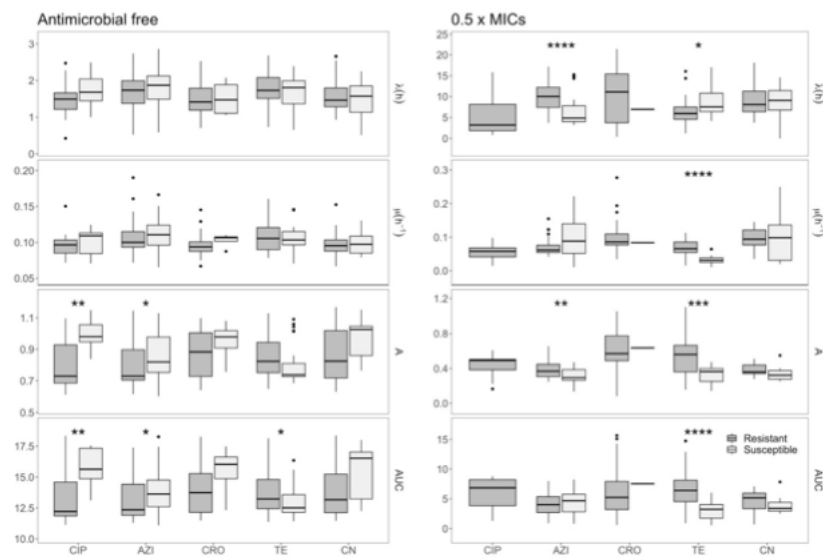


Figure 3. A comparison between the growth features for susceptible isolates and those for resistant isolates. From top to bottom, maximum cell density (A), maximum growth rate (μ), lag-phase period (λ), and area under the curve (AUC) of resistant isolates (dark gray) and those of susceptible isolates (light gray) without antimicrobial (left) and at half MICs (right). *: p -value < 0.05 , **: p -value < 0.01 , ***: p -value < 0.001 , ****: p -value < 0.0001 .

2.3. Effects of Azithromycin and Tetracycline Resistance on Growth Dynamics of *E. coli*

To further investigate the impact of azithromycin and tetracycline resistance on the growth dynamics of *E. coli*, we plotted the growth features of resistant and susceptible isolates at different relative concentrations of these antimicrobials. Figure 4 shows a comparison between trends of growth features (as inferred by Gompertz's model) on data from three replicates of resistant and susceptible organisms grown in increasing concentrations of azithromycin and tetracycline. We found that the λ of the tetracycline-resistant organism was lower than that of the susceptible organisms, while the μ and AUC of organisms resistant to tetracycline were higher than those of susceptible organisms. Conversely, the azithromycin-resistant organisms exhibited a longer λ , a higher A , and a lower μ than the azithromycin-susceptible organisms. Additionally, we found that the μ of both azithromycin-resistant and -susceptible *E. coli* peaked at low antimicrobial concentrations relative to MIC, which was supported by both Gompertz's, Baranyi's, and the three-phase linear models (Figure S3). These data suggest that sub-inhibitory concentrations of azithromycin actually accelerate the growth of both susceptible and resistant *E. coli*.

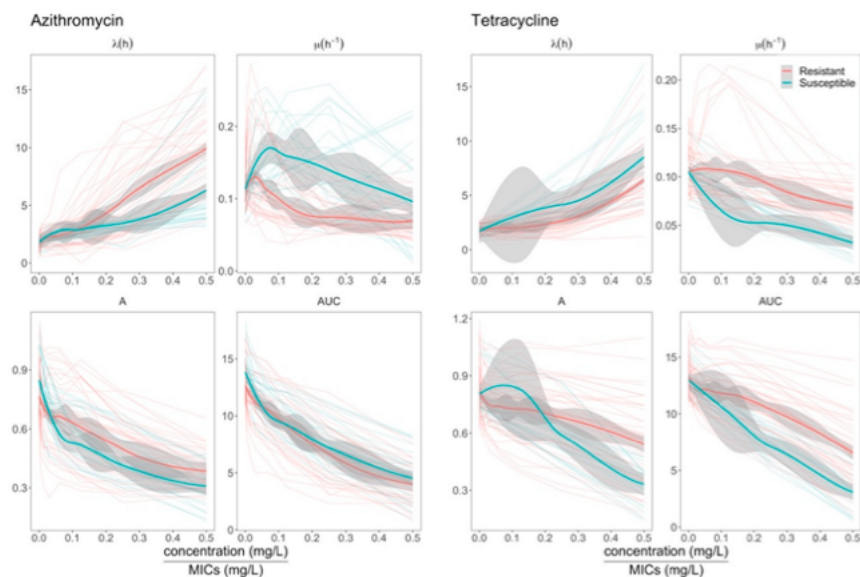


Figure 4. A comparison between growth features of resistant and susceptible isolates under exposure to azithromycin and tetracycline. The horizontal axis is antimicrobial concentrations relative to MICs. The vertical axis is values of the four growth features, including lag time (λ), maximum growth rate (μ), maximum cell density (A), and area under the curve (AUC). For each isolate, growth features were plotted against relative concentrations of antimicrobials as thin lines. Bold red (resistant isolates) and blue (susceptible isolates) lines indicate loess smooth regression. Gray shading represents 95% confident intervals.

2.4. Elevation of Growth Rate in *ermB*⁺ *E. coli* at Sub-MIC Concentrations of Azithromycin

All azithromycin-resistant *E. coli* isolates ($n = 19$) carried *mphA*; additionally, seven harbored *ermB* and one isolate harbored *lnuF*. In order to investigate the influence of *ermB*, which was prevalent among *E. coli* resistant to azithromycin, on the growth features of *E. coli*, we compared the growth features of the *E. coli* with the presence (*ermB*⁺) and absence (*ermB*[−]) of this gene. The μ , A , and AUC decreased in both *ermB*⁺ and *ermB*[−] isolates, while λ was elevated with increased relative concentrations of azithromycin (Figure 5a). We found that *ermB*⁺ *E. coli* exhibited a higher A and AUC , shorter λ , and lower μ than those of *ermB*[−] *E. coli*. At low concentrations of azithromycin relative to MICs, μ in *ermB*[−] isolates increased rapidly (peaked at $\sim 0.03 \times \text{MIC}$) and then gradually decreased. The highest μ value from the *ermB*⁺ organisms was lower than that from *ermB*[−] organisms, which peaked at a greater relative concentration of azithromycin. We further found that at lower concentrations of azithromycin (absolute concentrations of 1 and 2 mg/L), the *ermB*[−] organisms exhibited significantly higher μ than that of *ermB*⁺ organisms (Figure 5b), which was in concordance with data generated with relative azithromycin concentrations. Lastly, we found that MICs of *ermB*⁺ isolates were significantly higher than those of *ermB*[−] isolates ($p < 0.05$) (Figure 5c). Our data show that MICs against azithromycin are associated with the accumulation of the macrolide resistance gene.

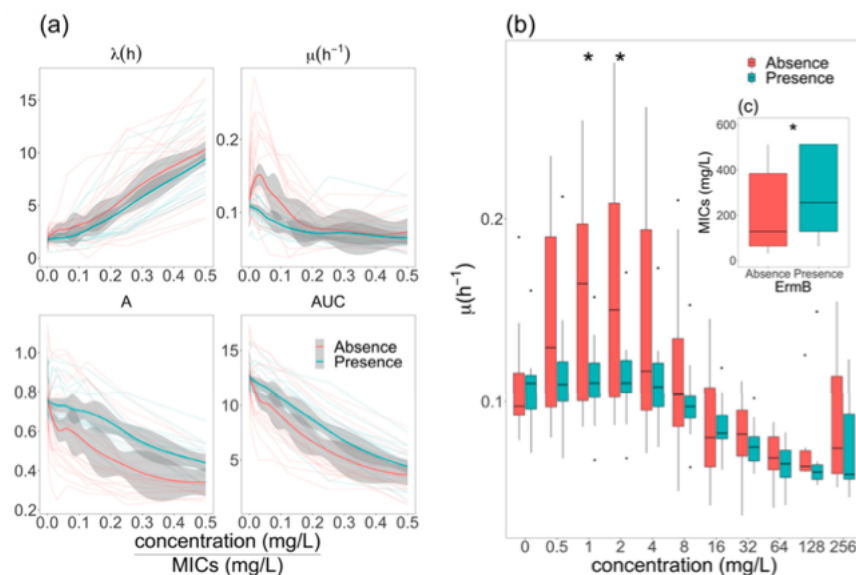


Figure 5. The impact of the *ermB* gene on growth dynamics of azithromycin-resistant *E. coli*. (a) Growth features of *E. coli* in the presence (blue thin lines) and absence (red thin lines) of *ermB*. Vertical axis is values of λ : lag time, μ : maximum growth rate, A: maximum cell density, and AUC: area under the curve. Horizontal axis is antimicrobial concentration relative to MICs. Bold red and blue lines are loess smooth regression. Gray shading is 95% confident intervals. (b) Maximum growth rate (μ) of *ermB*⁻ *E. coli* (red boxes) compared to that of *ermB*⁺ *E. coli* (blue boxes). Horizontal axis is the absolute concentration of azithromycin (mg/L). (c) Comparison MICs of *ermB*⁺ and *ermB*⁻ *E. coli*. *: p -value < 0.05.

3. Discussion

This study systematically assessed the physiological and morphological responses of AMR *E. coli* to clinically relevant antimicrobials. Regarding morphology, we found a relationship between resistance to ceftriaxone and cell elongation, which has been previously reported [17–22] and also utilized to rapidly detect resistant isolates [10,11]. However, we found that the molecular basis of AMR was not associated with antimicrobial-induced morphological responses. This observation suggests that morphology-based rapid antimicrobial susceptibility testing may be performed without prior knowledge of resistance mechanisms, which is a drawback of nucleic acid-based and protein-based methods. A limitation of morphology-based testing is the highly diverse dependence of morphological changes on the types of antimicrobial, even for antimicrobials belonging to the same family. Considering the β -lactams as an example, meropenem induces a spindle morphology in *E. coli*, while moxalactam induces filamentous cells [23]. Thus, for each antimicrobial, a prior imaging dataset of both resistant and susceptible organisms is required to infer the susceptibility of newly isolated organisms. For other antimicrobials used in this study, the morphology–resistance correlation was less consistent, which may be addressed by additional time data provided by live-cell imaging, which may be required to build highly accurate antimicrobial susceptibility tests [11,24]. The heterogeneous responses to CIP, AZI, TE, and CN might be explained by the highly diverse genetic background of the *E. coli* collection comprising of 11 different STs. Our results showed antimicrobial-induced morphological changes in both resistant and susceptible isolates, suggesting that bacterial cells respond to drug pressure despite the resistance to antimicrobials. This is probably due to the residue, in the cytoplasm, of active antimicrobials which is not yet extruded by efflux pumps (e.g., TetA and TetB) or inactivated by enzymes

(e.g., β -lactamases). This could also be explained by the weak interaction between antimicrobials (e.g., CIP) and modified target proteins (e.g., mutated GyrA).

Due to improved manufacturing technology, the price of tetracyclines has decreased dramatically, to the extent where they are the cheapest antimicrobials available. This has made tetracycline the second most widely used antimicrobial worldwide [25]. Although the usage of tetracyclines in human health is becoming less common, they are still important antimicrobials disease prevention and growth promotion in animals [26]. Our data demonstrated a high prevalence of tetracycline-resistant *E. coli* via the *tet* gene, which is in accordance with previous studies [27]. By comparing growth features of *E. coli* treated with tetracycline, we found that tetracycline-resistant isolates possessed a fitness advantage over susceptible isolates. This fitness advantage in the absence of the antimicrobial may explain the ubiquity of *tet* gene carriage in the gut microbiota of children and the sustained persistence of tetracycline resistance [27]. Our observations concur with a pharmacodynamic model [28], in which the fitness of resistant *E. coli* did not differ from susceptible *E. coli* in antimicrobial-free conditions. However, a recent study using growth curve analysis found that tetracycline resistance may exert a fitness cost of up to 10% on *E. coli* compared with a wild-type strain [29].

Azithromycin has been listed by WHO as one of the critically important antimicrobial reagents for human medicine [30]. Moreover, WHO have been developing guidelines to prophylactically prescribe azithromycin to infants and children to reduce child mortality [31], which might increase the resistance against macrolide antimicrobials globally [32]. Therefore, the molecular mechanisms and prevalence of azithromycin resistance have been studied extensively [3,31,33–35]. However, the influence of azithromycin exposure on the physiology of bacteria is not well understood. Here, we observed an accelerated μ of both resistant and susceptible isolates under exposure to low concentrations of azithromycin, raising a concern about potential impacts of this antimicrobial on developing new resistant mutants. These results suggest further investigations are required to evaluate the potential benefits of presumptive antimicrobial treatments compared to the risk of AMR development. We speculate that at sub-MIC concentration, azithromycin molecules inhibit protein synthesis by binding to the P site on the 50S subunit of the bacterial ribosome [36], inducing up-regulation of ribosome synthesis which is associated with higher μ [37,38]. With increased concentration of the macrolide, more ribosomal molecules are bound by azithromycin, which leads to insufficient protein synthesis and reduction in μ .

Notably, the trend of μ in azithromycin-resistant *E. coli* was lower than the trend in azithromycin-susceptible isolates. This observation may be experimental evidence supporting a mathematical model for ribosome-targeting antimicrobials, which showed that faster growth induces greater drug susceptibility [39]. Moreover, amongst the azithromycin-resistant *E. coli* harboring the *mphA* gene, the *ermB*⁺ isolates exhibited lower μ than that of *ermB*[−]. We hypothesize two feasible causal pathways of how *ermB* influences μ : (1) the presence/absence of *ermB* directly impacts on the μ trend or (2) the presence of *ermB* increases the MICs of *E. coli* and consequently changes the trend of μ against the relative concentration of azithromycin. The lower μ in *ermB*⁺ may be due to reduced ribosome synthesis as *ermB* allows constitutive methylations of 23S rRNA [40], which protects the complexes from azithromycin binding and potentially interferes with ribosome assembly. This suggests that the presence/absence of *ermB* is likely to have a direct impact on μ .

This study has limitations. First, using these approaches, we were unable to develop an algorithmic tool to quantitatively measure morphological changes and to categorize cell shapes. Therefore, the cell shape classification may be subjective. Second, the relationship between morphology, growth dynamics, and cell viability was not investigated. Third, the correlation between the live cell counts, measured CFUs, and OD values, representing the density of both live and dead cells, was not verified. Although bacterial enumeration provides information on the actual growth rates of the cultured bacteria, we opted to analyze the growth of *E. coli* by monitoring OD as it enables high-throughput experiments with biological replications. Lastly, the 27 *E. coli* isolates possessed potential cross-interaction among AMR genes and high diversity in genetic background, which we were not able to address.

4. Materials and Methods

4.1. Organisms and Whole Genome Sequencing

Twenty-seven *E. coli* isolates were isolated from the blood samples of Vietnamese children hospitalized with sepsis in a previous study [41], which was approved by the Oxford Tropical Research Ethics Committee (OxTREC, reference number 35–16) and the Ethic Committee of Children's Hospital 1 (reference number 73/GCN/BVND1). *E. coli* ATCC 25922 was included as a control in all analyses. Minimum inhibitory concentrations (MICs) against ciprofloxacin, azithromycin, ceftriaxone, tetracycline, and gentamicin were determined for all isolates by E-test (bioMérieux, Ivry-sur-Seine, France) and serial broth dilution. Whole-genome sequencing was performed on the 27 clinical isolates. DNA was extracted and purified using the Wizard Kit (Promega, Madison, WI, USA); the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) was used to fragment DNA with adapter ligation. DNA sequencing was performed on a MiSeq platform, using MiSeq Reagent Kit V2 (Illumina, San Diego, CA, USA) to produce 250-bp paired-end reads. Sequencing data are available at www.ebi.ac.uk/ena (accession numbers from ERR4342165 to ERR4342191).

4.2. Monitoring and Modeling Bacterial Growth

For growth dynamics experiments, two-fold serial broth dilutions were conducted in 96-well U-bottom plates (Thermo Fisher Scientific, Waltham, MA, USA) to obtain differing concentrations of antimicrobials (from 0 to 512 mg/L) in 100 µL cation-adjusted Mueller-Hilton (MH) broth (Oxoid, Basingstoke, UK). The selected antimicrobials were ciprofloxacin, azithromycin, ceftriaxone, tetracycline, and gentamicin (Sigma Aldrich and AK Scientific, MO, USA). For each of the 28 *E. coli* isolates, one colony of overnight-incubated *E. coli* was inoculated into 5 mL Luria-Bertani (LB) broth (Basingstoke, UK) for 3 h at 37 °C with shaking to reach OD₆₀₀ ~ 0.3. The bacterial suspension was diluted 1/1000 in MH broth, and 100 µL of the diluted suspension was mixed into each well of the 96-well plates. The final concentration of antimicrobials in each well varied from 0 to 256 mg/mL (two-fold intervals, i.e., 0, 0.5, 1, 2, ..., 256). OD₆₀₀ was measured on the plates every 15 min for 24 h at 37 °C, with shaking before each measurement, in a FLUOstar OPTIMA microplate reader (BMG Labtech, Ortenberg, Germany). The output data for each condition were curated and modified Gompertz's (1), Baranyi's (2), and three-phase linear (3) models [13] were fit to the data by minimizing the mean square error (MSE).

$$N(t) = N_0 + A \exp \left\{ - \exp \left[\frac{\mu e}{A} (\lambda - t) + 1 \right] \right\} \quad (1)$$

$$N(t) = N_0 + \mu C(t) - \ln \left\{ 1 + \frac{e^{\mu C(t)} - 1}{e^A} \right\} \quad (2)$$

$$N(t) = \begin{cases} N_0, & t \leq \lambda \\ N_0 + \mu(t - \lambda), & \lambda < t < t_s \\ A, & t \geq t_s \end{cases} \quad (3)$$

where $N(t)$ is the cell density, measured in OD₆₀₀, at timepoint t , N_0 is the cell density at $t = 0$, A is the maximum cell density, μ is the maximum growth rate (per hour), λ is the lag time (in hours), t_s is the time to reach stationary phase (in hours), and $C(t) = t + \frac{1}{\mu} \ln [e^{-\mu t} + e^{-\mu \lambda} - e^{-\mu(t+\lambda)}]$. Area under the curve (AUC) was approximately estimated from the growth curves using the Reimann summation method. We termed the A , μ , λ , and AUC as growth features of a growth curve.

4.3. Bright-Field Microscopy

In order to investigate the morphology of *E. coli* in response to antimicrobial exposure, isolates were inoculated on a nutrient agar (NA) plate and incubated overnight at 37 °C. For each isolate, <10 colonies were inoculated into saline buffer (0.85% NaCl) to reach OD₆₀₀ ~ 0.6. Resistant isolates were diluted 1/20 to reduce cell density in a field of view under microscopic examination. The selected antimicrobial

(ciprofloxacin, azithromycin, ceftriaxone, tetracycline, and gentamicin) of its corresponding CLSI break-point concentration [16] (i.e., ceftriaxone at 4 µg/mL) was added to the bacterial suspension and to the thin NA plate (8 mL). Then, 100 µL of bacterial suspension was inoculated onto NA plates and incubated for 3 h at 37 °C. Subsequently, small pieces of agarose pad were excised from the NA plate, placed onto microscope slides, and examined under a Nikon Ni-e upright microscope with a 100× objective lens.

4.4. Data Analysis

Data analysis was performed in R (version 3.5.3) programming language [42]. To estimate the parameters from the models, the “optim” function (the R Stats package) was used to constrainedly minimize the MSE between the actual and predicted OD values. For each isolate, the growth features were plotted against relative MICs using the “ggplot2” and “gridExtra” packages. Wilcoxon rank-sum test on boxplots was carried out on the “ggpubr” package. Multilocus sequence typing (MLST) of *E. coli* was inferred from the whole-genome sequencing data using the SRST2 pipeline [43] on the seven housekeeping genes, including *adhA*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA*. Resistome analysis with AMR gene identification was performed on the SRST2 pipeline. De novo assembly of sequencing data was performed using SPAdes [44], gene annotation was completed using Prokka [45], and the pan- and core genomes were deduced using Roary [46] with a minimum 99% blastp identity. Finally, a phylogenetic tree was constructed on RAxML [47] using the alignment of the core genome from Roary’s output and was visualized using the “ggtree” package.

5. Conclusions

In conclusion, we exploited growth dynamics and bacterial morphology to study the impact of AMR in clinical *E. coli* isolates. The morphological responses to ceftriaxone were consistently associated with the resistance to the antimicrobial, while the responses to other antimicrobials were more heterogeneous. In the absence of the drug, tetracycline resistance confers a fitness advantage on *E. coli*, while the other antimicrobials exert a measurable fitness cost (lower AUC and μ) on these bacteria. At low concentrations of azithromycin, μ was accelerated and azithromycin-susceptible isolates possessed higher μ than resistant isolates. We additionally observed an impact of AMR genes on the growth dynamics of bacteria, in which *E. coli* harboring the *ermB* gene exhibited lower μ than *ermB*[−] *E. coli*. Our study raises a concern regarding the persistence of AMR and the development of new resistance phenotypes under antimicrobial exposure, showing the need for a better understanding of how morphological and physiological responses contribute to the AMR and tolerance in bacterial pathogens.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2079-6382/9/11/735/s1>. Figure S1. Three replicates of growth dynamics monitoring. A clinical *E. coli* isolate was treated, three times, with ciprofloxacin (CIP), azithromycin (AZI), ceftriaxone (CRO), tetracycline (TE), and gentamicin (CN) at different concentrations. Red dots: replicate #1, green dots: replicate #2, and blue dots: replicate #3. Figure S2. Comparison between performance of Gompertz’s, Baranyi’s and three-phase linear model. For each growth curve, the three models were utilized to predict OD values at every timepoints from 0 to 24h. The actual and predicted values were used to calculate the mean square error (MSE), mean absolute error (MAE), correlation coefficient (ρ), and correlation of determination (R^2). The lower MSE and MAE; higher ρ and R^2 , the better a model is. Figure S3. Comparison between growth parameters estimated from Gompertz’s, Baranyi’s and three-phase linear model. Parameters include lag-phase period (λ), maximum growth rate (μ), and maximum cell density (A). The horizontal and vertical axes are relative concentration of antimicrobials and values of the growth parameters, respectively.

Author Contributions: Conceptualization, S.B.; methodology, T.T.-A., H.T.T., N.N.M.C., T.H.T., L.M.T., N.H.T.T., N.T.N.T.; formal analysis, T.T.-A., H.C.T., D.P.T.; investigation, N.D.T.; writing—original draft preparation, T.T.-A., S.B.; writing—review and editing, T.T.-A., S.B., H.C., D.P.T., J.B.S., H.C.T., H.C.; supervision, S.B.; project administration, N.D.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a Wellcome senior research fellowship to Stephen Baker (215515/Z/19/Z). HCT is a Wellcome International Training Fellow (218726/Z/19/Z). DPT is funded as a leadership fellow through the Oak Foundation.

Acknowledgments: We wish to acknowledge all members of the molecular epidemiology group at Oxford University Clinical Research Unit (OUCRU) in Vietnam and the study team at Children's Hospital 1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. European Centre for Disease Prevention and Control Surveillance of antimicrobial resistance in Europe 2016. In *Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net)*; European Centre for Disease Prevention and Control: Stockholm, Sweden, 2017.
2. Lan, N.P.H.; Hien, N.H.; Le Thi Phuong, T.; Thanh, D.P.; Thieu, N.T.V.; Ngoc, D.T.T.; Tuyen, H.T.; Vinh, P.V.; Ellington, M.J.; Thwaites, G.E.; et al. Phenotypic and genotypic characteristics of ESBL and AmpC producing organisms associated with bacteraemia in Ho Chi Minh City, Vietnam. *Antimicrob. Resist. Infect. Control.* **2017**, *6*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Darton, T.C.; Thanh Tuyen, H.; Chung The, H.; Newton, P.N.; Dance, D.A.B.; Phetsouvanh, R.; Davong, V.; Campbell, J.I.; Van Minh Hoang, N.; Thwaites, G.E.; et al. Azithromycin resistance in *Shigella* spp. in Southeast Asia. *Antimicrob. Agents Chemother.* **2018**, *62*, e01748–17. [\[CrossRef\]](#) [\[PubMed\]](#)
4. World Health Organization. *Antimicrobial Resistance: Global Report on Surveillance 2014*; World Health Organization: Geneva, Switzerland, 2014.
5. Shrestha, P.; Cooper, B.S.; Coast, J.; Oppong, R.; Do Thi Thuy, N.; Phodha, T.; Celhay, O.; Guerin, P.J.; Wertheim, H.; Lubell, Y. Enumerating the economic cost of antimicrobial resistance per antibiotic consumed to inform the evaluation of interventions affecting their use. *Antimicrob. Resist. Infect. Control.* **2018**, *7*, 1–9. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Hamad, B. The antibiotics market. *Nat. Rev. Drug Discov.* **2010**, *9*, 675–676. [\[CrossRef\]](#)
7. Odoki, M.; Aliero, A.A.; Tibyangye, J.; Nyabayo Maniga, J.; Wampande, E.; Kato, C.D.; Agwu, E.; Bazira, J. Prevalence of Bacterial Urinary Tract Infections and Associated Factors among Patients Attending Hospitals in Bushenyi District, Uganda. *Int. J. Microbiol.* **2019**, *1*, 1–8. [\[CrossRef\]](#)
8. Gonzales, L.; Joffe, E.; Rivera, R.; Sjöling, Å.; Svennerholm, A.M.; Ifúñez, V. Prevalence, seasonality and severity of disease caused by pathogenic *Escherichia coli* in children with diarrhoea in Bolivia. *J. Med. Microbiol.* **2013**, *62*, 1697–1706. [\[CrossRef\]](#)
9. Khalil, I.A.; Troeger, C.; Blacker, B.F.; Rao, P.C.; Brown, A.; Atherly, D.E.; Brewer, T.G.; Engmann, C.M.; Houpt, E.R.; Kang, G.; et al. Morbidity and mortality due to shigella and enterotoxigenic *Escherichia coli* diarrhoea: The Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* **2018**, *18*, 1229–1240. [\[CrossRef\]](#)
10. Otero, F.; Santiso, R.; Tamayo, M.; Fernández, J.L.; Bou, G.; Lepe, J.A.; McConnell, M.J.; Gosálvez, J.; Cisneros, J.M. Rapid Detection of Antibiotic Resistance in Gram-Negative Bacteria Through Assessment of Changes in Cellular Morphology. *Microb. Drug Resist.* **2017**, *23*, 157–162. [\[CrossRef\]](#)
11. Choi, J.; Yoo, J.; Lee, M.; Kim, E.-G.; Lee, J.S.; Lee, S.; Joo, S.; Song, S.H.; Kim, E.-C.; Lee, J.C.; et al. A rapid antimicrobial susceptibility test based on single-cell morphological analysis. *Sci. Transl. Med.* **2014**, *6*, 267ra174. [\[CrossRef\]](#)
12. Spalding, C.; Keen, E.; Smith, D.J.; Krachler, A.; Jabbari, S. Mathematical modelling of the antibiotic-induced morphological transition of *Pseudomonas aeruginosa*. *PLoS Comput. Biol.* **2018**, *14*, e1006012. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Li, B.; Qiu, Y.; Shi, H.; Yin, H. The importance of lag time extension in determining bacterial resistance to antibiotics. *Analyst* **2016**, *141*, 3059–3067. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Magiorakos, A.P.; Srinivasan, A.; Carey, R.B.; Carmeli, Y.; Falagas, M.E.; Giske, C.G.; Harbarth, S.; Hindler, J.F.; Kahlmeter, G.; Olsson-Liljequist, B.; et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: An international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.* **2012**, *18*, 268–281. [\[CrossRef\]](#)
15. Aldred, K.J.; Kerns, R.J.; Oshero, N. Mechanism of quinolone action and resistance. *Biochemistry* **2014**, *53*, 1565–1574. [\[CrossRef\]](#) [\[PubMed\]](#)
16. *Performance Standards for Antimicrobial Susceptibility Testing*, 28th ed.; Clinical And Laboratory Standards Institute: Wayne, PA, USA, 2018.
17. Blázquez, J.; Gómez-Gómez, J.M.; Oliver, A.; Juan, C.; Kapur, V.; Martín, S. PBP3 inhibition elicits adaptive responses in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **2006**, *62*, 84–99. [\[CrossRef\]](#) [\[PubMed\]](#)

18. Yao, Z.; Kahne, D.; Kishony, R. Distinct Single-Cell Morphological Dynamics under Beta-Lactam Antibiotics. *Mol. Cell* **2012**, *48*, 705–712. [CrossRef] [PubMed]
19. Buijs, J.; Dofferhoff, A.S.M.; Mouton, J.W.; Wagenvoort, J.H.T.; Van Der Meer, J.W.M. Concentration-dependency of β -lactam-induced filament formation in Gram-negative bacteria. *Clin. Microbiol. Infect.* **2008**, *14*, 344–349. [CrossRef]
20. Miller, C.; Thomsen, L.E.; Gaggero, C.; Mosseri, R.; Ingmer, H.; Cohen, S.N. SOS response induction by β -lactams and bacterial defense against antibiotic lethality. *Science* **2004**, *305*, 1629–1631. [CrossRef]
21. Ryan, D.M.; Monsey, D. Bacterial filamentation and in vivo efficacy: A comparison of several cephalosporins. *J. Antimicrob. Chemother.* **1981**, *7*, 57–63. [CrossRef]
22. Rolinson, G.N. Effect of β -lactam antibiotics on bacterial cell growth rate. *J. Gen. Microbiol.* **1980**, *120*, 317–323. [CrossRef]
23. Godinez, W.J.; Chan, H.; Hossain, I.; Li, C.; Ranjitkar, S.; Rasper, D.; Simmons, R.L.; Zhang, X.; Feng, B.Y. Morphological Deconvolution of Beta-Lactam Polyspecificity in *E. coli*. *ACS Chem. Biol.* **2019**, *14*, 1217–1226. [CrossRef]
24. Baltekin, Ö.; Boucharin, A.; Tano, E.; Andersson, D.I.; Elf, J. Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9170–9175. [CrossRef] [PubMed]
25. Levy, S.B. *The Antibiotic Paradox*; Perseus Publishing: Cambridge, MA, USA, 2002.
26. Daghrir, R.; Drogui, P. Tetracycline antibiotics in the environment: A review. *Environ. Chem. Lett.* **2013**, *11*, 209–227. [CrossRef]
27. Karami, N.; Nowrouzian, F.; Adlerberth, I.; Wold, A.E. Tetracycline resistance in *Escherichia coli* and persistence in the infantile colonic microbiota. *Antimicrob. Agents Chemother.* **2006**, *50*, 156–161. [CrossRef] [PubMed]
28. Ahmad, A.; Zachariassen, C.; Christiansen, L.E.; Græsboell, K.; Toft, N.; Matthews, L.; Damborg, P.; Agersø, Y.; Olsen, J.E.; Nielsen, S.S. Pharmacodynamic modelling of in vitro activity of tetracycline against a representative, naturally occurring population of porcine *Escherichia coli*. *Acta Vet. Scand.* **2015**, *57*, 1–8. [CrossRef] [PubMed]
29. Dunai, A.; Spohn, R.; Farkas, Z.; Lázár, V.; Györkei, Á.; Apjok, G.; Boross, G.; Szappanos, B.; Grézal, G.; Faragó, A.; et al. Rapid decline of bacterial drug-resistance in an antibiotic-free environment through phenotypic reversion. *Elife* **2019**, *8*, 1–20. [CrossRef] [PubMed]
30. World Health Organization. *Critically Important Antimicrobials for Human Medicine*, 6th ed.; World Health Organization: Geneva, Switzerland, 2019; ISBN 9789241515528.
31. Presumptive Use of Azithromycin. Available online: https://www.who.int/maternal_child_adolescent/guidelines/development/provision-of-azithromycin-to-infants/en/ (accessed on 7 December 2019).
32. Bergman, M.; Huikko, S.; Huovinen, P.; Paakkari, P.; Seppälä, H.; Thompson, A.K.; Rautio, M.; Renkonen, R.; Muotiala, A.; Vaara, M.; et al. Macrolide and azithromycin use are linked to increased macrolide resistance in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **2006**, *50*, 3646–3650. [CrossRef]
33. Lin, L.; Nonejuie, P.; Munguia, J.; Hollands, A.; Olson, J.; Dam, Q.; Kumaraswamy, M.; Rivera, H.; Corriden, R.; Rohde, M.; et al. Azithromycin Synergizes with Cationic Antimicrobial Peptides to Exert Bactericidal and Therapeutic Activity Against Highly Multidrug-Resistant Gram-Negative Bacterial Pathogens. *EBioMedicine* **2015**, *2*, 690–698. [CrossRef]
34. Gomes, C.; Ruiz-Roldán, L.; Mateu, J.; Ochoa, T.J.; Ruiz, J. Azithromycin resistance levels and mechanisms in *Escherichia coli*. *Sci. Rep.* **2019**, *9*, 1–10. [CrossRef]
35. Horner, P.J. Azithromycin antimicrobial resistance and genital Chlamydia trachomatis infection: Duration of therapy may be the key to improving efficacy. *Sex. Transm. Infect.* **2012**, *88*, 154–156. [CrossRef]
36. Gomes, C.; Martínez-Puchol, S.; Palma, N.; Horna, G.; Ruiz-Roldán, L.; Pons, M.J.; Ruiz, J. Macrolide resistance mechanisms in Enterobacteriaceae: Focus on azithromycin. *Crit. Rev. Microbiol.* **2017**, *43*, 1–30. [CrossRef]
37. Levin, B.R.; McCall, I.C.; Perrot, V.; Weiss, H.; Ovesepian, A.; Baquero, F. A Numbers Game: Ribosome Densities, Bacterial Growth, and Antibiotic-Mediated Stasis and Death. *MBio* **2017**, *8*, e02253-16. [CrossRef] [PubMed]
38. Tadmor, A.D.; Tlusty, T. A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLoS Comput. Biol.* **2008**, *4*, e1000038. [CrossRef] [PubMed]

39. Greulich, P.; Scott, M.; Evans, M.R.; Allen, R.J. Growth-dependent bacterial susceptibility to ribosome-targeting antibiotics. *Mol. Syst. Biol.* **2015**, *11*, 796. [[CrossRef](#)] [[PubMed](#)]
40. Garneau-Tsodikova, S.; Labby, K.J. Mechanisms of resistance to aminoglycoside antibiotics: Overview and perspectives. *Medchemcomm* **2016**, *7*, 11–27. [[CrossRef](#)] [[PubMed](#)]
41. Toan, N.D.; Darton, T.C.; Boinett, C.J.; Campbell, J.I.; Karkey, A.; Kestelyn, E.; Thinh, L.Q.; Mau, N.K.; Thanh Tam, P.T.; Nhan, L.N.T.; et al. Clinical features, antimicrobial susceptibility patterns and genomics of bacteria causing neonatal sepsis in a children's hospital in Vietnam: Protocol for a prospective observational study. *BMJ Open* **2018**, *8*, e019611. [[CrossRef](#)]
42. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.
43. Inouye, M.; Dashnow, H.; Raven, L.A.; Schultz, M.B.; Pope, B.J.; Tomita, T.; Zobel, J.; Holt, K.E. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **2014**, *6*, 1–16. [[CrossRef](#)]
44. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [[CrossRef](#)]
45. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
46. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.G.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)]
47. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).