Applying at scale genomic and phenotypic approaches to study pathogens linked to nosocomial transmission and disease

Benjamin Michael Warne

Department of Medicine

University of Cambridge

This thesis is submitted for the degree of

Doctor of Philosophy

Clare College

October 2021

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the Clinical and Veterinary Medicine Degree Committee, being fewer than 60,000 words excluding appendices, references, tables and figure legends (final word count 57,671 words).

Data described in this thesis have been published in:

Sridhar, S., Forrest, S., Warne, B., *et al.* (2021) High-Content Imaging to Phenotype Antimicrobial Effects on Individual Bacteria at Scale. *mSystems*. **6**(3). (**Chapter 4**)

Reacher, M. Warne, B., Reeve, L. *et al.* (2019) Influenza-associated mortality in hospital care: a retrospective cohort study of risk factors and impact of oseltamivir in an English teaching hospital, 2016 to 2017. *Euro Surveill.* **24** (44) (**Chapter 6**)

Meredith, L. W., Hamilton, W., Warne, B., *et al.* (2020) Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis.* **20**, 1263-1271 (**Chapter 7**)

Warne, B., Enright, J., Metaxaki, M., *et al.* (2021) Efficacy of mass testing for SARS-CoV-2 in a UK university using swab pooling and PCR. *Research Square.* (Preprint) doi:10.21203/rs.3.rs-520626/v1 (**Chapter 8**)

Aggarwal, D., Warne, B., Jahun, A., *et al.* (2021) Genomic epidemiology of SARS-CoV-2 in the University of Cambridge identifies dynamics of transmission. *Nat Commun.* (In press). (**Chapter 8**)

Benjamin Michael Warne

October 2021

Abstract

Applying at scale genomic and phenotypic approaches to study pathogens linked to nosocomial transmission and disease

Benjamin Michael Warne

Infectious diseases continue to pose major threats to human health. This manifests in different ways, from the incessant rise of antimicrobial resistance, to seasonal epidemics of respiratory viruses and global pandemics, all of which cause high levels of morbidity, mortality and burden on healthcare systems. In this thesis I have investigated three pathogens that are archetypes of these threats: the Gram-negative bacterium *Klebsiella pneumoniae*, the seasonal respiratory virus influenza and the pandemic coronavirus SARS-CoV-2.

Using isolates prospectively collected at Cambridge University Hospitals, I have utilised a combination of genomic, epidemiological and clinical datasets to characterise these pathogens. Phylogenetic analysis of *K. pneumoniae* bacteraemia isolates from a four year period show considerable genomic diversity and possible cryptic transmission events involving a lineage associated with hospital outbreaks. Using a novel method of high content imaging using confocal microscopy, I have developed high throughput pipelines that can quantify morphological changes induced in multiple bacterial species by a range of clinically relevant antimicrobials. I have used this approach to determine the phenotypic variation in a collection of 175 *K. pneumoniae* isolates that represent the diversity of the species. In paired imaging and RNAseq experiments using clinical isolates, I have identified large numbers of genes that are differentially expressed in the presence of antimicrobials. Among them are

pathways associated with antimicrobial resistance, including mutagenesis, drug efflux and plasmid conjugation. Together, these data suggest that rather than passively selecting resistant isolates, antimicrobials may also actively drive resistance in *K. pneumoniae*.

Using large genomic epidemiology studies I have identified extensive networks of nosocomial transmission of influenza and SARS-CoV-2 in CUH, involving patients and healthcare workers with an associated high inpatient mortality. For hospitalised influenza patients, I have shown the protective effect of the antiviral oseltamivir in multivariable analyses, and how rates of hospital acquired infection and mortality have fallen in response to multidisciplinary interventions over three winters. During the COVID-19 pandemic, real-time genomic-epidemiology analyses have been integrated with clinical services to understand and reduce viral spread.

Finally, I have applied these approaches to a community setting, identifying SARS-CoV-2 transmission dynamics in the University of Cambridge and how they have been interrupted by local and national measures to control the COVID-19 pandemic. In particular, I have detailed the implementation and evaluation of a large regular screening programme for asymptomatic SARS-CoV-2 infection using pooled sampling, a programme that has provided insights into the efficacy of mass testing approaches.

Acknowledgements

I thank my supervisor, Gordon Dougan, for his support and mentoring throughout this work, as well as Estée Török and Stephen Baker who have provided supervision and guidance at varying times during my fellowship; Josefin Bartholdson Scott for her supervision and direction on the bacterial experimental work and imaging; Sally Forrest, Mailis Maes and Sushmita Sridhar for their contributions, assistance and comments in optimising imaging procedures and RNAseq data; Fahad Khokhar for his assistance with assembling isolate collections, long read sequence data and advice on laboratory techniques; Beth Blane and Del Pickard for advice on laboratory methods; Peter Gallagher for his work on phenotyping *K. pneumoniae* collections; Hayley Wilson, Zoe Dyson, Leah Roberts, Lindsay Pike, and the Wellcome Sanger Institute Pathogen Informatics team for advice and support with bioinformatic analyses; James Hutt and his colleagues at PerkinElmer for helping to develop the image analysis pipelines; Kat Holt, Kelly Wyres, Nick Thomson and Eva Heinz for useful discussions regarding *K. pneumoniae*; Catherine Ludden, the British Society for Antimicrobial Chemotherapy, Colin Macleod, Francesca Short and Sebastian Bruchmann for contributing collections of isolates; the PHE Clinical Microbiology and Public Health Laboratory staff for their efforts in processing and storing isolates; Avijit Dutta for providing provisional Fluidigm data; all members of the Dougan, Baker and Török groups, past and present, who have provided helpful discussion over the course of my fellowship; and Clare Harris for enabling all of this work. For the influenza and SARS-CoV-2 studies, I thank Hamid Jalal, Mark Reacher and Maria Zambon for their additional supervision and support; David Williams, Will Hamilton, Dinesh Aggarwal, Ian Goodfellow and his laboratory team, Luke Meredith, Ewan Harrison, Sharon Peacock and her laboratory team, and various contributors from COG-UK for their genomic expertise and sequencing efforts; Neville Verlander, Richard Samworth and Jessica Enright for contributions to statistical analyses;

Lucy Reeve, Ashley Popay and Iain Roddick for their epidemiological analyses; Elinor Moore, Isobel Ramsay, Kat Sharrocks, Theo Gouliouris, Clare Leong, David Enoch, Richard Smith and the infectious diseases, infection control, microbiology and patient safety teams for their work on nosocomial transmission; Peter Driscoll, Stuart Wills and Afzal Chaudhry for their assistance in developing of clinical data extracts; Rhys Tassell, Michelle Lineham and the CUH POC laboratory; Mike Weekes, Patrick Maxwell, Duncan MacFarlane, Rob Howes, Jon Holgate, Vijay Samtani, Gillian Weale and the members of the University of Cambridge Rapid Response Working Group to Test, Trace and Isolate, as well as teams from the Cambridge Lighthouse Laboratory, University Information Services, Institute for Manufacturing, the University of Cambridge and its constituent Colleges who have all dedicated much time and effort into ensuring the success of the asymptomatic COVID-19 screening programme; Marina Metaxaki, Stewart Fuller and the clinical nursing team for helping to run the clinical and research elements of the screening programme; Nick Matheson for leadership and supervision of all aspects of the University screening programme and associated projects; and finally to all patients and students who have contributed data to these studies, and the staff who have cared for them.

This research was primarily funded by the National Institute for Health Research, Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust. Additional salary support was provided from the University of Cambridge and Cambridge University Hospitals NHS Foundation Trust. Preliminary research was supported by a Clinical Research Fellowship from the Addenbrooke’s Charitable Trust.

Contributions to Chapters

All chapters	Study design, data collection, data analysis, figure preparation, manuscript writing	Ben Warne
	Supervisor	Gordon Dougan, Stephen Baker

The manuscript and figures in this thesis were produced by Ben Warne, except where explicitly stated in the text or figure legends. All experimental work, data collection and analyses were designed and performed by Ben Warne, unless stated in the list of contributors below.

Chapter 3	Additional supervision	Estée Török
	Isolate collection and DNA extraction	Ben Warne, Fahad Khokhar
	RNA extraction, sequence analysis, phenotyping	Ben Warne
	Plasmid curing and conjugation experiments	Ben Warne, Derek Pickard

All clinical data were extracted from the laboratory information management system and analysed by Ben Warne. The CUH bacteraemia isolate collection was collated from the CMPHL and stored by Ben Warne and Fahad Khokhar. DNA extraction was performed by Ben Warne and Fahad Khokhar, and sequenced by Fahad Khokhar and the Wellcome Sanger Institute. RNA extraction was performed by Ben Warne and sequencing performed at the Wellcome Sanger Institute. Experimental design, quality control, bioinformatic, phylogenetic and RNAseq analyses were performed by Ben Warne. Plasmid curing and conjugation experiments were performed by Ben Warne under the supervision of Derek Pickard.

Chapter 4	Additional supervision	Josefin Bartholdson Scott
	Plate coating experiments	Sally Forrest
	Imaging experiments	Ben Warne, Sushmita Sridhar, Sally Forrest, Mailis Maes
	Image analysis pipelines and data analysis	Ben Warne, Josefin Bartholdson Scott, James Hutt, Sally Forrest

	Clinical isolate investigation and imaging; live imaging experiments	Ben Warne

Imaging optimisation experiments were designed by Ben Warne, Josefin Bartholdson Scott, Sally Forrest and Sushmita Sridhar. Imaging of *Klebsiella pneumoniae* was performed by Ben Warne; experiments with other species were performed by Sally Forrest, Mailis Maes and Sushmita Sridhar. Experiments refining plate coatings were performed by Sally Forrest, and analysed by Ben Warne and Sally Forrest. Image analysis pipelines for all experiments, including quality control and downstream data analysis, were developed by Ben Warne, with input from Josefin Bartholdson Scott, James Hutt and the remainder of the team. All *Klebsiella* clinical isolate and live bacterial imaging experiments and analyses were designed and performed by Ben Warne.

Chapter 5	Isolate collection, sequence analysis, RNA extraction	Ben Warne
	Long read sequencing and hybrid assemblies	Fahad Khokhar
	Additional Vitek, MALDI-TOF and phenotyping	Peter Gallagher
	Bacterial phenotyping and imaging	Ben Warne, Peter Gallagher
	Preliminary Fluidigm experiment	Avijit Dutta, Ben Warne

Long read sequencing and assembly of *Klebsiella* genomes was performed by Fahad Khokhar, assisted by Ben Warne, with quality control of resultant sequences performed by Ben Warne. RNA extraction was performed by Ben Warne and sequencing performed at the Wellcome Sanger Institute. Transciptomic experimental design, quality control and RNAseq analyses were performed by Ben Warne. The *Klebsiella* isolate collection was compiled by Ben Warne. Antimicrobial susceptibility testing and basic bacterial phenotyping was performed by Ben Warne and Peter Gallagher. Isolate collection high content imaging experiments were designed and analysed by Ben Warne, and performed by Ben Warne and Peter Gallagher. Differential expression experiment using the Fluidigm platform was designed and analysed by Ben Warne, with the Fluidigm assay performed by Avijit Dutta

Chapter 6	Additional supervision	Hamid Jalal, Mark Reacher, Maria Zambon, Clare Sander, Chris Smith, Hongyi Zhang
	Clinical data cleaning and analysis, combined cluster analysis, service evaluation	Ben Warne

	Clinical data collection	Ben Warne, Nicholas Jones, Kyriaki Ranellou, Silvana Christou, Callum Wright, Nyaradzai Sithole, Rory Carpenter, Angharad Everden, Elizabeth Jarman, Ali Khalid, Kyle Lam, Chloe Myers, Shuhui Ren, Kathryn J Rolfe, Tommy Sutton, Saher Choudhry
	Genome sequencing and analysis	Monica Galiano, David Williams
	Epidemiological analysis	Iain Roddick, Ben Warne, Lucy Reeve, Ashley Popay
	Multivariable analysis	Neville Verlander

The clinical case report form was designed by Hamid Jalal, Ben Warne and Mark Reacher. Clinical data were manually abstracted by Ben Warne and the clinicians listed above. Data cleaning and quality control was performed by Ben Warne with further cleaning by Lucy Reeve. Multivariable analysis of clinical data was performed by Neville Verlander, with study design and assurance of the analysis performed by Ben Warne, Mark Reacher and Hamid Jalal. Design, data collection and analysis of the influenza clinical service evaluation was completed by Ben Warne. Influenza genome sequencing was performed by Monica Galiano and phylogenetic analysis by David Williams. Epidemiological analysis was performed by Iain Roddick, Ben Warne, Lucy Reeve and Ashley Popay. The analysis combining genomic, epidemiological and clinical datasets was designed and performed by Ben Warne.

Chapter 7	Additional supervision	Estée Török, Ian Goodfellow
	Genomic analysis	William Hamilton, Luke Meredith, Ben Warne
	Epidemiological analysis	Ashley Popay, Ben Warne, Lucy Reeve
	Clinical data extraction and analysis	Ben Warne, Peter Driscoll
	Hospital associated infection reviews	Ben Warne, Elinor Moore, Richard Smith, Theo Gouliouris, David Enoch, Isobel Ramsay, Kat Sharrocks and the infectious diseases team

Clinical data abstraction tool was developed by Ben Warne and Peter Driscoll. Clinical data quality control and analysis was performed by Ben Warne. Genome sequencing was performed by Ian Goodfellow and his laboratory, with phylogenetic analysis performed by William Hamilton, Luke Meredith and Ben Warne. Epidemiological analysis was performed by Ashley Popay, Ben Warne and Lucy Reeve. Analysis combining genomic, epidemiological and clinical datasets was designed and performed by Ben Warne. The hospital acquired infection review process was developed by

Ben Warne, Elinor Moore and Richard Smith, with meetings chaired by Ben Warne and supported by the clinical infection specialties.

Chapter 8	Additional supervision	Nicholas Matheson
	Data collection, cleaning and analysis	Ben Warne
	Statistical advice	Richard Samworth, Jessica Enright
	Genomic and epidemiological analysis	Dinesh Aggarwal, Ben Warne
	Telephone survey	Marina Metaxaki, University of Cambridge COVID-19 helpdesk

The asymptomatic screening programme was designed and led by Nicholas Matheson with Ben Warne and the University of Cambridge Rapid Response Working Group to Test, Trace and Isolate. Project ethics application, study protocol development, public engagement, liaison with University groups, consumables procurement, staff recruitment and clinical implementation of the programme was led by Ben Warne. Programme monitoring and evaluation was performed by Ben Warne. The symptom telephone survey was designed by Ben Warne, Nicholas Matheson and Marina Metaxaki, and performed by Marina Metaxaki and the University of Cambridge COVID-19 helpdesk. Survey results were analysed by Ben Warne. Multivariable analyses were performed by Ben Warne and reviewed by Richard Samworth and Jessica Enright. Whole genome sequencing was performed by the COG-UK collaborative. Phylogenetic analysis was led by Dinesh Aggarwal, with input from Ben Warne; epidemiological analysis was led by Ben Warne with input from Dinesh Aggarwal. Combined genomic and epidemiological analysis was designed and performed by both Ben Warne and Dinesh Aggarwal.

Table of contents

List of figures

List of tables

1. Introduction

1.1 The persistent threat of infectious diseases to global health

Over the past 100 years, enormous progress has been made in the treatment and prevention of infectious diseases. Interventions including programmes of vaccination against a range of pathogens[1], the development of numerous antimicrobial classes[2], improved diagnostics and research tools including the revolution in genome sequencing[3], combined with improved epidemiological surveillance and public health measures[4] have dramatically reduced global infectious disease burden.

However, communicable diseases continue to cause enormous morbidity and mortality and remain a persistent major threat to public health[5]. According to a global burden of disease study published in 2019, six infectious diseases were among the top ten causes of disability adjusted life years (DALYs) in young children[6]. While the incidence of some communicable diseases (including tuberculosis, HIV, neglected tropical diseases, diarrhoeal and lower respiratory tract infections) is falling, changes in patterns of human behaviour, including increasing international travel, urbanisation and climate change, have led to increased risks of infectious disease epidemics[7]. In 2019 the World Health Organization (WHO) published a list of what they consider to be the ten largest threats to global health[8]. Six of these specifically refer to communicable diseases: antimicrobial resistant pathogens, influenza, high-threat pathogens (including the coronaviruses severe acute respiratory syndrome

(SARS-CoV) and Middle East respiratory syndrome (MERS-CoV)), dengue, HIV, and vaccine hesitancy. Since the publication of this report, the emergence and rapid global dissemination of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to over 240 million confirmed cases by October 2021, with 4.9 million associated deaths[9].

While my PhD fellowship was initially focused on investigating the first of the WHO's identified threats to global health, antimicrobial resistance, the SARS-CoV-2 pandemic changed my circumstances dramatically. Like many clinicians and academics across the globe, my existing plans were put on hold and over a year of my fellowship has been dedicated to the study and management of coronavirus disease 2019 (COVID-19) in both secondary care and in the University of Cambridge. Much of my research on COVID-19 has arisen in the context of work undertaken early in my fellowship studying the impact of seasonal influenza in secondary care.

In this thesis I present work on all three of these areas, which have overlapping approaches to the study of major human pathogens. In particular, I demonstrate how genomic methods can be used in conjunction with novel diagnostic and imaging approaches which, when combined with extensive metadata from large isolate collections, can provide insight into pathogen diversity, transmission, and response to interventions to treat and prevent infection.

1.2 *Klebsiella pneumoniae* and AMR

Although hundreds of antimicrobials have been developed over the last century, with a range of chemical structures and mechanisms of action, resistance to each new antimicrobial class has emerged within a few years of their introduction into clinical practice[10]. The rise of antimicrobial resistance (AMR) in bacteria has been described as a "global health emergency" by the United Nations[11], with high levels of associated mortality[12], burden on healthcare[13] and threat to the global economy[14]. The EARS-Net study, for example, identified an estimated 671,000 infections with antimicrobial resistant bacteria in the European region in 2015 alone, associated with 33,100 attributable deaths and 874,541 DALYs[15]. This burden of disease disproportionately affects some of the most vulnerable members of society, including infants, the elderly and patients admitted to hospital. Secondary care providers and long term care facilities are recognised as reservoirs of AMR bacteria[16,17], with high rates of antimicrobial resistance in hospital acquired infections (HAIs), and AMR infections being associated with increased risk of mortality[18,19]. Recent studies have also demonstrated the key role of human community transmission of these pathogens[20,21]. There has been an exponential rise of AMR colonisation and infection in animals[22], disproportionately affecting low and middle income countries, that has also been implicated in rising rates of infection in humans[23].

One of the most important species in both the study and management of AMR is *Klebsiella pneumoniae*. It is a key microorganism, in which many AMR genes have first been identified before disseminating to other Gram-negative pathogens. These mechanisms include genes coding for extended-spectrum beta-lactamases[24], carbapenemases[25,26] and plasmid-mediated colistin resistance[27]. *K. pneumoniae* has caused multidrug-resistant outbreaks in hospitals across the globe[28-30], including Cambridge University Hospitals NHS Foundation

Trust (CUH)[31]. In recognition of this threat, the World Health Organization has identified antimicrobial-resistant *Enterobacteriaceae*, including *K. pneumoniae*, as one of three groups of pathogens in its highest priority category for the research and development of new antimicrobials[32].

1.2.1 Overview of *Klebsiella* biology and phylogeny

The genus *Klebsiella* comprises Gram-negative, non-motile, aerobic rod-shaped bacteria. These include a range of species commonly found in the environment, with variable propensity for human infection, including *K. pneumoniae*, *K. oxytoca*, *K. aerogenes* and *K. michiganensis*. The species most frequently associated with human colonisation and disease is *K. pneumoniae*, although others are increasingly recognised for their potential for invasive human diseases and AMR gene acquisition[33]. Genomic and proteomic analysis of the genus *Klebsiella* has identified that *K. pneumoniae* consists of several subspecies, including *K. pneumoniae* subsp. *pneumoniae*, *K. variicola*, *K. quasipneumoniae* subsp. *quasipneumoniae* and K. *quasipneumoniae* subsp. *similipneumoniae*[34,35]. Of these, *K. pneumoniae* subsp. *pneumoniae* comprises the majority of clinical isolates with the highest rates of resistance and virulence. Both species of *K. quasipneumoniae* are found most commonly in humans, and are associated with HAI but low levels of virulence and resistance, consistent with opportunistic infection. *K. variicola* is found in a range of plants and animals. It is distinguished by other species from its ability to fix nitrogen, enabling survival in the environment[36].

Lineages of *K. pneumoniae* are commonly described using a multilocus sequence typing (MLST) scheme using seven housekeeping genes[37]. This MLST approach has been extended to a core gene MLST (cgMLST) scheme targeting 694 core genes, which can be

used to define high-resolution sequence types (STs) and their aggregation into clonal groups (CGs)[38]. The mean *K. pneumoniae* genome is 5.5 Mb in size, encoding approximately 5,700 genes[34]. However, there is considerable genomic diversity across the species, with a large accessory genome. The size of the core genome (i.e. genes present in >99% of isolates) varies by study, depending on the sampling methods used, for example ranging from 1,888 genes in a global phylogeny to 2,958 genes in a phylogeny of isolates from the UK[34,39]. The non-core genome (genes shared by between 0 and 95% of isolates) is large, consisting of >25,000 genes. However, most non-core genes are rare, with two thirds found in ≤5% of *K. pneumoniae* genomes and one third found in only one isolate in the global phylogeny above. Analysis of the genetic diversity, taxonomy and G+C content of these genes shows that they have been derived from a wide variety of bacterial taxa, including *Enterobacteriaceae*, *Vibrio*, and *Acinetobacter*. Up to one third of protein-coding genes identified in the species have no known identified function[34].

Phylogenetic analysis of *K. pneumoniae* is characterised by a deep-branching structure, indicating early divergence of the species into hundreds of distinct lineages with approximately 0.5% nucleotide difference between their core genomes[35]. The estimated rate of mutation is a mean of 3.8 SNPs per genome per year[39]. In comparison to other species, where one lineage appears to have become dominant and is implicated in major global outbreaks[40,41], numerous lineages of *K. pneumoniae* have emerged as important causes of infection, transmission and the acquisition of antimicrobial resistance[42].

Additional typing methods are based on the two most common antigenic targets on the bacterial surface: the capsular (K) and O antigens[43]. There are 12 O antigen serotypes, shared by numerous lineages. The capsular antigen is more diverse, with 77 phenotypically defined variants identified and many more predicted by genetic analysis of the K locus.

Multiple K locus types can be identified within single STs and, as with the O antigen, the same K locus can be found across multiple lineages as a result of horizontal gene transfer and recombination[44]. Typing methods have additionally been used to classify plasmids in *K. pneumoniae* associated with AMR genes, in particular the FII and FIB incompatibility types that commonly carry beta-lactamases and can be shared with other members of the *Enterobacteriaceae* family[45].

1.2.2 Spectrum of clinical disease caused by *K. pneumoniae*

K. pneumoniae is globally distributed and a common commensal organism, found in the stool of 5-38% of individuals in the community, and in the nasopharynx of 1-6% of healthy individuals[46]. Increased rates of carriage have been identified in people who use excess alcohol[47] and those with Chinese ethinicity[48], though it is most commonly associated with antimicrobial use and hospitalisation, being detectable in the stool of 77% of inpatients, and the nasopharynx of 19%[46]. A number of host factors dispose for *K. pneumoniae* infection, most of which are associated with impaired host defences, including diabetes mellitus, malignancy, chronic kidney disease and long term steroid use[49,50]. Recent administration of antimicrobials has also been associated with invasive disease[50,51].

Infection with *K. pneumoniae* commonly presents in one of two ways. Firstly, it presents as HAIs, frequently related to medical devices, such as ventilator associated pneumonia associated with endotracheal tubes, urinary tract infections associated with catheters and bacteraemia associated with line infections. These are most common in patients with the underlying co-morbidities listed above[46]. Secondly, *K. pneumoniae* causes community acquired infections, which are less common and account for a minority of urinary tract infections and pneumonias, again associated with underlying host risk factors. An exception

is the syndrome of severe, disseminated *K. pneumoniae* causing liver abscesses, pneumonia, meningitis and endophthalmitis, commonly in previously healthy individuals and predominantly associated with hypervirulent *K. pneumoniae* strains in South East Asia[52].

There are a number of genes that are known to contribute to virulence in *K. pneumoniae*. These include the K1 or K2 capsular polysaccharide (which provides resistance to phagocytosis)[53,54], hypermucoviscosity related to extracellular polysaccharide (which has a role in resistance to antibody-mediated immunity)[55], lipopolysaccharides (that act as endotoxins and may enhance invasion)[56], siderophores (that enhance growth and invasion)[57] and a range of adhesins and pili that facilitate adhesion to both human tissues and synthetic materials[58,59]. Currently, the co-existence of these virulence determinants and AMR genes in the same lineage is relatively uncommon. However, given the high degree of horizontal gene transfer in this species, there is concern that a combined hypervirulent, multi-drug resistant lineage will arise. Over the last few years, there have been increasing reports of such strains [60-62], including worrying descriptions of single plasmids that contain both hypervirulence genes and carbapenemases[63]. The emergence of hypervirulent, multidrug resistant isolates of *K. pneumoniae* has major public health implications, especially if they spread globally in the same manner as other *K. pneumoniae* lineages[64].

1.2.3 Impact of *K. pneumoniae* in secondary care

K. pneumoniae causes a substantial burden on healthcare services. Over four years from 2017-21, there were 42,625 reported *Klebsiella* bacteraemias in England, of which over two thirds were caused by *K. pneumoniae* and 31% were hospital acquired[65]. There has been a gradual increase in the rate of *K. pneumoniae* bacteraemia over the last decade, rising from 7.8 per 100,000 population in 2009, to 12.5 per 100,000 population in 2018 in England,

Wales and Northern Ireland[66]. Current rates of antimicrobial resistance in the UK are low but rising for certain drug classes, standing at 31% resistance to amoxicillin-clavulanic acid, 14% to piperacillin-tazobactam and 13% to ceftazidime in 2018[66]. While resistance to carbapenems in the UK is low (<2%), it is considerably higher in other parts of Europe, and is frequently associated with nosocomial transmission[67]. While data on resistance rates in low and middle income countries are sparse, one study estimated that isolates of *E. coli* and *K. pneumoniae* resistant to carbapenems were responsible for 0.5 million bloodstream infections and 3.1 million serious infections in 2014[68].

The spread of AMR *K. pneumoniae* has important implications for healthcare. Patients with carbapenem resistant *K. pneumoniae* infections have higher total medical costs, longer length of inpatient stays and higher mortality than those with susceptible isolates[69]. A significant proportion of the cost associated with these admissions has been attributed to antimicrobial therapy[70]. Reported rates of mortality in patients with invasive carbapenemase-producing *K. pneumoniae* have been as high as 50%[71]. While hospital acquired infection is an important route of acquisition of these pathogens, in settings of high endemicity in the environment or local community, patients are often colonised before they enter the hospital and may serve as important sources of transmission to the hospital environment and other patients[72].

1.2.4 Relationship between *K. pneumoniae* biology and AMR

There are a number of reasons why *K. pneumoniae* presents such a large contribution to AMR. The species has a broader geographical and ecological distribution, a more varied pangenome, greater AMR gene diversity and a higher plasmid burden than many other Gram-negative opportunistic pathogens[73]. One of the major routes to AMR in *K. pneumoniae*

is through the acquisition of resistance genes via horizontal transfer, primarily by plasmids. Over 100 distinct AMR genes have been identified in *K. pneumoniae*, with hundreds of plasmids belonging to dozens of distinct incompatibility and replicon types. Evidence of the transfer of AMR plasmids between *Klebsiella*, and between *Klebsiella* and other *Enterobacteriaceae*, has been detected in studies of hospitalised patients and in hospital environments[35]. Both mutational and plasmid-mediated resistance to antimicrobials have emerged multiple times in different lineages[34,39]. However, a number of clades have become widely disseminated globally, associated with outbreaks in secondary care despite the observation that they are not generally associated with known *Klebsiella* virulence determinants[29,74].

Some of the most important resistance mechanisms, for both public health and individual clinical treatment, are the genes encoding the carbapenemases KPC, OXA-48, and NDM, and the ESBL CTX-M-15[35]. Each of these genes is associated with a specific transposon that mobilises it between different plasmid backbones (and on to different strains and species) and sometimes into the bacterial chromosome itself. All four of these beta-lactamases have been reported in a variety of *K. pneumoniae* lineages and are frequently associated with other resistance determinants. Their increased prevalence has necessitated different approaches to antimicrobial therapy, including novel beta lactam/beta lactamase inhibitor combinations and revisiting older classes of antimicrobial, either repurposed drugs such as colistin, or modifications of existing classes, such as the tetracycline derivative tigecycline[75]. However, there are a number of limitations with these drugs, including cost, suboptimal pharmacokinetic properties, side effects and the emergence of resistance.

While the majority of resistance to beta lactams is conferred by enzymatic destruction with beta lactamases, such as those listed above, for other antimicrobial classes the mechanisms

are often more diverse. For example, there are multiple means of resistance to fluoroquinolones, a class of antimicrobial that inhibits the topoisomerase II class of bacterial intracellular enzymes (most commonly DNA gyrase and topoisomerase IV) that have essential roles in DNA replication. This leads to bacterial DNA damage and cell death. Resistance mechanisms include: 1) mutational resistance in *gyrA* and *parC*, the genes encoding DNA gyrase and topoisomerase IV; 2) reduced membrane permeability, through a reduction in porin expression or altered outer membrane organisation; 3) efflux pump overexpression; 4) target protection by the DNA-binding protein Qnr; and 5) enzymatic modification of antimicrobials[76]. The last three mechanisms are commonly plasmid-encoded[77]. There are a number of bacterial species where the development of mutational quinolone resistance has been associated with the emergence of dominant clades that disseminate widely in both community[78] and hospital settings[41,79], but in *K. pneumoniae* the relative contributions and impact of these quinolone resistance mechanisms is unclear. Although outbreak strains are often associated with carbapenem resistance, these are normally plasmid-encoded alongside other resistance determinants, including those targeting ciprofloxacin.

1.3 Genomic approaches to investigating AMR

The last decade has seen increasing recognition of AMR as a global health concern[8,12]. Multiple UK and international action plans have been created, using a multidisciplinary, multifaceted approach to tackle this complex problem[80,81]. They identify a knowledge gap in the mechanisms by which antimicrobial resistance emerges and is propagated within human populations, which limits approaches to reducing transmission and developing novel

therapeutics. One of the major contributors to AMR research over the last two decades has been whole genome sequencing. Genomics has played a major role in characterising the diversity and evolution of resistance mechanisms in *K. pneumoniae*[34,39,63,67]. The exponential rise of genome sequences submitted to public databases, combined with automated bioinformatics tools for analysing *K. pneumoniae* genomes[42,44], has enabled the rapid implementation of genomics in the study of AMR pathogens and large platforms for international surveillance of outbreak strains[82]. At a local level, genomics approaches are increasingly being used as part of AMR pathogen outbreak investigations, including at CUH[31,83].

However, there remain considerable limitations of genomics approaches when used in isolation. As outlined above, up to one third of the genes in *K. pneumoniae* encode proteins of unknown function; many other genes are poorly characterised with unclear mechanisms of regulation. In studies of the correlation between the measured AMR phenotype and genotype in *K. pneumoniae*, attempts to predict susceptibility from whole genome sequence data have yielded varying results. While there are increasingly reliable predictors of resistance to the beta-lactam classes of antimicrobials, there remains poor sensitivity and specificity for the prediction of resistance to fluoroquinolones. In one study, 7/69 *K. pneumoniae* isolates were resistant to ciprofloxacin, but no known resistance gene could be identified[84]. Although mutations in *gyrA* and *parC* are known to be major determinants of resistance, there is wide variation in fluoroquinolone minimum inhibitory concentrations (MICs) for isolates exhibiting the same mutations, indicating that other determinants contribute to resistance[85].

Other genomic approaches have been used to improve our understanding of AMR mechanisms. Bacterial RNA sequencing (RNAseq) has been employed to identify novel

colistin resistance mechanisms in *K. pneumoniae* and to further elucidate candidate pathways of ciprofloxacin resistance in other bacterial species[86-90]. Although RNAseq has both reproduced known genes of significance and identified potential targets for future research, the approach has been criticised for over-reliance on changes in gene expression that may not translate to important phenotypic differences[91].

Alternative approaches include transposon directed insertion-site sequencing (TraDIS). For example, in a 2017 paper Jana *et al.* were able to identify a gene associated with colistin resistance using a TraDIS library generated from a ST258 *K. pneumoniae* isolate. However, they were unable to determine any novel genes associated with ciprofloxacin resistance[92]. The approach has been criticised for its inability to implicate essential genes that may be important in resistance, as well as practical limitations in scaling to large isolate collections. Another approach is genome-wide association studies (GWAS) which, in an era of rapid expansion in the number of available genomes, may identify a variety of biological mechanisms through correlation between phenotype and genotype[42]. However, previous GWAS attempting to identify ciprofloxacin resistance mechanisms have only found associations with the known resistance gene *qnr* or mutations in chromosomal *gyrA* and *parC* genes[93], with only 90% accuracy in predicting the crude phenotype of susceptibility or resistance. A recent study using machine learning approaches to predict MIC based on genotype claims better accuracy, but is limited by using data from a single centre and bacteria with a narrow range of ciprofloxacin MICs[94].

All of these approaches may be limited by the restrictions of bacterial phenotyping *in vitro* using basic microbiological antimicrobial susceptibility testing, or the relatively small number of bacterial isolates used in experimental design. To address these problems requires

improved phenotypic techniques and approaches that can study larger collections of isolates, representing a broader range of diversity in species such as *K. pneumoniae*.

1.4 Novel phenotyping approaches using high content imaging

Advances in bacterial genomics have transformed the field of AMR research, enabling fine resolution investigations of genomic diversity. However, it remains critical to link genotypic findings with bacterial phenotype to interpret the biological and clinical relevance of these results. Current approaches to resistance phenotyping are constrained by the use of traditional microbiological techniques, for example relying on simple antimicrobial gradients to determine a single concentration that inhibits bacterial growth. Such methods are often low throughput and aggregate data from mixed populations of bacterial cells, preventing analysis of sub-populations that may be important in the development of resistance, including tolerant and persistent bacteria whose phenotypes and molecular mechanisms remain poorly understood[95].

Recent advances in high content imaging (HCI) platforms present an opportunity to dramatically improve bacterial phenotyping[96]. Combining automated microscopy and image analysis enables detailed morphological studies of individual bacteria under a variety of conditions that cannot be captured using traditional screening approaches[97]. Such image-based profiling has great potential in a range of applications. These include high-throughput drug screening, one of the main applications of HCI applied to eukaryotic cells and tissues[98,99]. Studies in the field of bacteriology have often focused on a limited assessment

of intracellular organisms, for example simple enumeration of bacterial load of *Mycobacterium tuberculosis* or *Salmonella* in infected eukaryotic cell lines[100-102].

By contrast, Nonejuie *et al.* have used HCI to make detailed morphological studies of *E. coli* growing in liquid culture, utilising developments in image analysis software to provide detailed descriptions of individual bacteria[103]. They have demonstrated clear phenotypic changes that are characteristic for bacteria incubated with a range of antimicrobial classes and have used this approach to predict the mechanism of action of a novel antimicrobial compound. More recently, Zoffmann *et al.* have demonstrated the utility of HCI in screening large compound libraries for antimicrobial activity against a number of species, including *K. pneumoniae*[104]. Similar approaches have been used to phenotype an increasing breadth of bacterial species in different experimental conditions[105-108].

However, the full potential of this approach has yet to be explored. The above studies are often limited by a small number of bacterial isolates or growth conditions. Further developments in automated imaging hardware and software have produced microscopes such as the Opera Phenix, that are capable of producing HCI in high throughput using 96/384 well plates[109,110]. In principle, this enables detailed phenotyping of large numbers of cells in controlled conditions in the same experiment, though methods have yet to be optimised for large bacterial collections.

Additionally, there is the prospect of combining HCI data with genomic approaches to produce comprehensive comparisons of genotype and phenotype. Combinations of genomic and imaging approaches have successfully been used to study bacterial responses to antimicrobials[111]. For example van Laar *et al.* used sub-lethal concentrations of carbapenems to determine their effect on *K. pneumoniae* using both RNAseq and scanning

electron microscopy (SEM)[112]. They identified a quantifiable change in the size and shape of bacteria in the presence of carbapenems, which was associated with differential gene expression that contributed to this phenotype. They also identified changes in expression of a range of genes associated with peptidoglycan processing (the drug target of carbapenems) and AMR mechanisms to other antimicrobial classes. Although this study was limited in scale, there is clear potential in correlating multiple experimental approaches to address the complex challenge of AMR.

1.5 Influenza in secondary care

While the biology and pathology of influenza is strikingly different from AMR bacteria, there are a number of features that they have in common. Like *K. pneumoniae*, influenza has a number of genomically distinct, globally disseminated lineages associated with a high burden of morbidity and mortality, with the potential to acquire virulence leading to large epidemics of public health significance[113,114]. While the role of community transmission of influenza has long been established, there is an emerging recognition of influenza as a hospital associated pathogen and genomics is playing a key role in its surveillance[115].

1.5.1 Overview of influenza biology and genomics

Influenza viruses are enveloped, single-stranded, negative-sense RNA viruses of the family *Orthomyxoviridae*. They are zoonotic infections, causing infections in humans, other mammalian species and birds. There are three viral types, based on phylogenetic and antigenic variation: influenza A, B and C. Influenza A viruses have the greatest genetic

diversity, infecting the widest range of host species and causing a substantial majority of severe disease in humans[113]. However, there is increasing recognition of the potential virulence of influenza B lineages, which have caused a considerable burden of infection and hospitalisation in the UK in recent winters[116]. Influenza C, by contrast, is associated only with mild disease.

The influenza A and B virus genome is approximately 14 kb in size and consists of eight segments, which have the potential to be exchanged through reassortment. Influenza A is further categorised into subtypes based on the two main viral antigens, the haemagglutinin (HA) and neuraminidase (NA) surface glycoproteins[117]. There is a relatively high mutation rate for the virus, associated with the phenomenon of antigenic drift, by which structural changes in HA and NA are associated with incomplete immune protection against viral lineages over time. The most common circulating lineages in humans over the last decade have been H3N2 and H1N1ndm09[116].

1.5.2 Spectrum of clinical disease caused by influenza A and B

Influenza is predominantly spread by droplets from the respiratory tract. Classical presentations consist of a syndrome of sudden onset fever, myalgia, malaise, dry cough, sore throat and nasal congestion, although atypical presentations, such as nausea, vomiting and diarrhoea, are more common in infants, the elderly and the immunocompromised[118]. The presentation is non-specific and the infection can mimic a range of other infectious and non-communicable diseases. The mean incubation period of influenza is 2 days (range 1 to 4 days). Influenza can cause serious disease and death, commonly associated with bacterial superinfection in the lower respiratory tract, and a number of high risk populations have been

identified including patients who are immunocompromised, elderly, pregnant or morbidly obese, and those with a range of underlying co-morbidities[119,120].

A more substantial burden of morbidity and mortality is associated with influenza A(H3N2), in comparison to other commonly circulating lineages such as A(H1N1) or influenza B[121]. The average rates of excess pneumonia and influenza hospitalization are reported to be twice as high during seasons characterised by high prevalence of H3N2 infection, disproportionately affecting infants and young children, the elderly and other high-risk populations[122].

1.5.3 Impact of influenza in secondary care

There is considerable concern for the pandemic potential of the virus, with the H1N1 Spanish influenza pandemic of 1918–1919 having caused acute illness in an estimated 25–30% of the world's population and the death of 40 million people[123]. Certain viral lineages, such as influenza A(H5N1), are predicted to be a major public health threat and a potential cause of future pandemics[124].

However, the majority of the burden of influenza in the UK comes from seasonal epidemics. Influenza is associated with 14.1% of all acute respiratory admissions to secondary care in adults, contributing an estimated 5.7 million admissions worldwide each year, and between 291,243 and 645,832 deaths[125-127]. The highest rates of mortality are seen in sub-Saharan Africa, South East Asia and in people aged over 75[127]. This high burden of disease translates to seasonal peaks in acute admissions to secondary care and influenza has been attributed as one of the major causes of winter bed crises in the NHS over the last two decades[128-130].

Understanding risk factors for poor outcome is important in developing strategies in secondary care to reduce the burden of disease and improve clinical outcomes. However, the majority of data on such risk comes from studies dating from the H1N1pdm09 pandemic, with few studies in the context of seasonal epidemics, especially with H3N2 circulating lineages[120,131-134].

One of the key controversies in the management of hospitalised influenza patients is the use of neuraminidase inhibitors such as oseltamivir. Although recommended by a number of national and international bodies, including the National Institute for Health and Care Excellence (NICE), the US Centers for Disease Control and Prevention (CDC), the Infectious Diseases Society of America (IDSA) and the WHO[135-138], their use remains contentious. Concerns have been raised about data withheld from clinical trials and the appropriateness of oseltamivir stockpiling for influenza pandemic preparedness[139,140]. While meta-analyses of randomised controlled trials have demonstrated that oseltamivir reduces lower respiratory tract complications, there was no evidence of a benefit on mortality[141,142]. However, these studies were predominantly performed in younger adults with few co-morbidities, where mortality was a rare event; trial data in patients hospitalised with influenza is lacking. Observational studies conducted during the H1N1pdm09 pandemic have demonstrated that oseltamivir reduces length of stay, symptom duration, admission to intensive care, progression to respiratory failure and mortality[131,143-145]. However, there is little observational data from inpatients with other viral strains, including influenza A(H3N2) and influenza B.

1.5.4 Role of influenza in nosocomial transmission

Many influenza outbreaks have been reported from acute hospital settings[146-149]. A review of 12 nosocomial influenza outbreaks reported patient attack rates that ranged from 3% to

50%, with a median mortality of 16%[146]. Influenza genome sequencing has been used to investigate viral linkage among infected patients; however, these studies are limited to small outbreaks occurring in a single clinical area in secondary care[148,150-156]. A combined use of epidemiological and molecular techniques has demonstrated limited nosocomial transmission of influenza cases on a small scale[153,157,158], but these studies have been impacted by poor case ascertainment and the quality of associated epidemiological data.

Previous studies have shown that combating nosocomial influenza transmission is complex and requires multidisciplinary interventions[159]. In particular, there are challenges in the recognition and diagnosis of mild or asymptomatic influenza cases, disproportionate impact on the frail elderly and the potential role of healthcare worker (HCW) and hospital visitors in transmission networks. There is therefore limited current knowledge regarding the extent of nosocomial transmission of influenza and the most effective measures to limit its spread in secondary care.

1.6 COVID-19 in secondary care

Although the public health importance of influenza pandemics has been recognised for over a century, there has been increasing recognition of the pandemic potential for coronaviruses since the emergence of SARS in 2002, a disease that caused infection in more than 8,000 individuals and 774 deaths in 26 countries on 5 continents[160]. The subsequent identification of MERS-CoV in 2012, with associated 40% mortality, was also concerning for further uncontrolled transmission[161], although extensive human-to-human spread has not been identified. Of note, both SARS and MERS-CoV were associated with a high risk of

nosocomial transmission. However, it was a third coronavirus, SARS-CoV-2, that has fulfilled the potential of its predecessors and caused the largest global pandemic in over 100 years.

1.6.1 Overview of SARS-CoV-2 biology and genomics

Coronaviruses are enveloped, single-stranded positive-sense RNA viruses, in the *Coronaviridae* subfamily *Orthocoronavirinae*. They have a genome 26–32 kb in length, in a single segment. They are characterised by relatively high mutation rates and are able to adapt to a range of animal hosts, including birds and a variety of mammals[162]. Seven coronaviruses have been identified in humans. Four of these (coronaviruses 229E, OC43, NL63 and HKU1) are responsible for 10–30% of upper respiratory tract infections in humans, the majority of which cause mild illness[163]. As shown above, SARS and MERS-CoV can be associated with severe human respiratory diseases. The genome of SARS-CoV-2 shares 80% sequence homology with SARS-CoV[164]. There are multiple schemes for the nomenclature of viral lineages, but the most commonly used are Pango lineages (e.g. A, B.1.1.7, etc.)[165] or the WHO classification system of variants based on letters of the Greek alphabet (alpha, beta, delta, etc.)[166].

The mean incubation period of SARS-CoV-2 infection is approximately 5 days (range 2-14 days)[167]. In comparison to SARS-CoV, peak infectiousness occurs earlier in the natural history of infection, often coinciding with, or slightly preceding, the onset of symptoms[168,169].

The major antigen of coronaviruses is the spike glycoprotein, found on the surface of the virion (providing it with the characteristic “corona” appearance on electron microscopy). It is responsible for mediating virus entry into cells and is the primary determinant of cell tropism, pathogenicity and immunity[170]. Mutations in the S gene have been associated with increased

virulence and transmissibility and reductions in vaccine efficacy and antibody-mediated immunity[171]. They therefore confer some of the major characteristics of SARS-CoV-2 variants of concern.

1.6.2 Spectrum of clinical disease caused by SARS-CoV-2

As with influenza, COVID-19 can present with a range of non-specific symptoms, most commonly fever, shortness of breath and cough[172]. More specific symptoms, such as dysgeusia and anosmia, are present in approximately half of infected individuals. The disease is associated with a wide range of extrapulmonary symptoms occurring at reduced frequency, including a number of cardiac, neurological, gastrointestinal, thromboembolic, endocrine and dermatological presentations[173]. The proportion of asymptomatic infections varies depending on the immune status of the individual, but a meta-analysis of published studies estimates this figure at approximately one third of all cases in unvaccinated populations[174]. Higher rates of asymptomatic infection are associated with individuals who have received a complete course of vaccination[175].

Unlike influenza, the main cause of mortality in these patients is a hyperinflammatory response in the host, most commonly manifest as acute respiratory distress syndrome[176]. The overall mortality associated with SARS-CoV-2 rises exponentially with age. In one study this was shown to be 0.4% at age 55, rising to >25% at age ≥90 years[177]. Other associations with severe disease include male sex, non-white ethnicity, obesity, social deprivation and a number of co-morbidities, including diabetes mellitus, haematological malignancy, chronic kidney disease, chronic obstructive pulmonary disease, chronic liver disease, stroke and solid organ transplantation[178].

An increasingly diverse spectrum of chronic symptoms (lasting for more than 90 days from the onset of infection) has been described, grouped under the title “post-COVID syndrome” but known more commonly as “long COVID”[179]. These affect between 4.7-80% of all cases, depending on the study and the definitions used[180]. Various pathophysiological hypotheses have been proposed for the mechanisms underlying this condition, but the diversity of symptoms, clinical and biochemical findings (which are frequently unremarkable) make this an area of uncertainty and intense research interest.

1.6.3 Impact of SARS-CoV-2 in secondary care

National public health interventions to combat the pandemic, including major social restrictions enforced in the UK, have explicitly stated that they were being employed to maintain a manageable burden of disease on secondary care services. Nevertheless, the pandemic has still placed an unprecedented demand on hospitals.

By October 23 2021, there had been 564,131 admissions to hospital related to COVID-19 in the UK[181]. The peak bed occupancy in England was in January 2021, when 34,336 beds in England were occupied by patients with COVID-19 (representing a third of the country’s pre-pandemic acute bed capacity)[182]. The peak demand for mechanical ventilation beds was 3,736 in England in January 2021. Prior to the pandemic, there was a baseline of 4,123 mechanical ventilation beds in England, a capacity that was increased by 53% in the first wave in an effort to address the demand of the pandemic[183].

This re-direction of acute secondary services to combating COVID-19 has led to concerns for the “collateral damage” of the pandemic on medical conditions, with a fall in all referrals to secondary care for the management of a range of chronic diseases and a rise in acute

conditions such as substance abuse and the effects of domestic violence[184]. The indirect consequences of the pandemic are broad and still being measured, extending beyond hospitals into areas as diverse as mental health[185], education[186] and academic research[187].

1.6.4 Role of SARS-CoV-2 in nosocomial transmission

The role of hospital transmission of SARS-CoV-2 early in the pandemic was unclear. While subsequent data have been published, at the time of the first wave in the UK there was a paucity of evidence on HAIs and optimal infection control procedures related to SARS-CoV-2. As noted above, SARS-CoV and MERS-CoV have both been associated with high rates of nosocomial infection. The first study to report from China during the COVID-19 pandemic stated that 41% of all cases identified in patients and HCWs were HAIs[188]. A subsequent study from the UK estimated that 9.1% of all infections diagnosed in secondary care were HAIs[189], but the authors acknowledge the difficulties in attributing the location of infection due to the comparably long incubation period of the virus.

The implications of nosocomial transmission are considerable. Many inpatients are vulnerable to developing severe infection, with case fatality rates in HAIs approaching 30%[190,191]. Infections in HCWs may dispose to significant viral propagation[192] and staff sickness, with associated absences that may affect patient safety. It was therefore important to rapidly understand and learn from HAIs in order to introduce measures to prevent transmission. Infection prevention & control (IPC) teams had to adapt rapidly to ensure patient and staff safety, taking into account national policies that were frequently introduced and revised after local policies had been put in place[193].

Previous implementation of rapid whole genome sequencing in pandemic settings has provided valuable data on circulating lineages, viral introductions into geographic regions, transmission networks and how these networks could be impacted by infection control interventions[3,194-196]. Building on local experience, the potential for the exploitation of genomics to study nosocomial transmission of the pandemic in CUH was considerable. As with influenza, there is a relatively high rate of mutation in SARS-CoV-2, with initial estimates of approximately 2.5 nucleotides per month[197]. Despite initial concerns for the low diversity of a virus so recently introduced into the human population, early sequencing of SARS-CoV-2 provided valuable information on virus biology and evolution, making this a feasible approach for outbreak investigation during the first wave[198].

1.7 COVID-19 in higher education institutions

While early data on infection control practices for COVID-19 in secondary care was limited, there was a dearth of any information for community settings. While there is an extensive literature on the management of nosocomial pathogens in secondary care, in community settings, specifically non-health and social care settings such as education, there is a paucity of recent evidence on infectious disease control for any communicable disease. While many case reports of outbreaks in universities have been published, their focus is frequently on vaccination to impact transmission of diseases such as meningitis B, mumps and measles[199-202]. In summer 2020, when attention was being given to re-opening higher education settings for the autumn term, vaccines were not available and so alternative methods of pandemic control were being considered.

The focus on universities reflects concerns that higher education institutions (HEIs) may act as reservoirs of infection that could spill over into the local community. University students are at higher risk for viral transmission, due to their large social networks, frequent social contacts, group teaching and dense student accommodation[203]. Young adults themselves are unlikely to develop severe disease and may be at higher risk of asymptomatic or minimally symptomatic infection, but retain the potential to propagate SARS-CoV-2 spread[204,205].

1.7.1 Importance of universities in the COVID-19 pandemic

Early in the COVID-19 pandemic in the UK, the majority of university students were sent home until the start of the autumn term. The first insights into the transmission potential of SARS-CoV-2 in HEIs therefore came from other countries. Outbreaks were first reported during the summer of 2020, especially from the United States where transmission was identified in association with large social gatherings, student accommodation and athletics events, in spite of some (admittedly varied) infection control measures introduced to limit viral spread[206-209]. However, the underlying dynamics of transmission were poorly described and their application to UK settings unclear[210]. Understanding these dynamics is important in order to design and implement effective infection control strategies, while minimising disruption to education, research and the wellbeing of students and staff[210,211]. Additionally, there was concern that outbreaks in HEIs could seed infections in more vulnerable populations, such as staff, local communities, and household contacts upon returning to permanent residences at the end of term[212,213]. Identifying and controlling possible sources of transmission in universities is therefore highly desirable.

1.7.2 Mass testing for SARS-CoV-2 in at-risk populations

In the absence of vaccination, the control of COVID-19 relies on a combination of case ascertainment and isolation, and tracing and quarantining high risk contacts[214-216]. However, high rates of asymptomatic, paucisymptomatic and presymptomatic transmission, thought to be more common in young adults[168,169], mean that interventions based on symptomatic case ascertainment alone are unlikely to be effective at controlling the spread of the virus.

To overcome this, attention had turned to the use of regular screening of asymptomatic individuals. Such approaches have the potential to identify presymptomatic and asymptomatic individuals with SARS-CoV-2 infection. Modelling studies have shown that regular (weekly or twice-weekly) screening, when combined with isolation of cases and their contacts, has the potential to substantially reduce transmission[217-219]. Screening approaches were routinely deployed early in the pandemic in settings associated with a high risk (or high consequence) of infection, such as health and social care[220-223], and later in other settings with high risk of transmission such as schools, prisons and professional sports[224-226]. In 2021, regular mass testing was offered to all adults in the community[227]. However, few of these programmes have been evaluated for their efficacy or cost-benefit, and the approach has drawn considerable criticism[228,229].

Nevertheless, screening of asymptomatic students was one of the approaches used by a number of universities in North America and the UK during the autumn of 2020, often using individual molecular diagnostic tests performed in university laboratories[209,230]. Variable levels of detail have been provided on the implementation and evaluation of these programmes and the impact that they had on limiting transmission to surrounding communities. Available reports have focussed on the practicability and acceptability of

screening[231,232] or enhanced case ascertainment[209,230,233], but not the effect on transmission, either within the university campus or in the community. While some modelling studies have been published about the theoretical benefit of mass testing in HEIs[234,235], the real-world impact on SARS-CoV-2 transmission remains unclear.

Additionally, regular screening of large student populations requires high numbers of tests and their associated consumables and laboratory capacity. Different approaches have been used as alternatives to individual molecular tests. The first is antigen detection using lateral flow tests (LFTs)[236,237]. However, aside from the lack of availability of LFTs in universities until the end of 2020, concerns have been raised about their accuracy, especially at low viral loads and when performed by untrained individuals[238,239].

An alternative approach is to use laboratory polymerase chain reaction (PCR) tests, but screen individuals in pools. Pooling has the potential to increase efficiency of a range of resources that may be limited in pandemic settings, including consumables and testing capacity. Most published studies have combined multiple samples into pools in the laboratory at the time of testing (sample pooling)[240-245]. However, this method requires additional sample manipulation with associated increases in hands-on time for laboratory staff, and leads to dilution of viral RNA present in individual samples. Alternatively, tests can be pooled at the time of sample collection, by combining multiple swabs from individuals in the same sample tube (swab pooling, also known as cohort pooling, or pooling at source)[246-248]. This does not impact laboratory workflow or test sensitivity caused by sample dilution. However, its application in a university setting is untested and there are clear logistical, clinical and even ethical considerations to such approaches.

A further advantage of laboratory based PCR testing is the availability of samples for subsequent whole genome sequencing. At the time of this study, no large scale genomic studies of SARS-CoV-2 had been conducted in higher education settings. However, the combination of sequences from asymptomatic and symptomatic testing pathways suggested that case ascertainment would be high. Alongside contextual genomic data from community isolates identified through surveillance by COG-UK[249] and contact tracing information from NHS test and trace, this would yield detailed resolution of transmission in this cohort.

1.8 Summary and aims

In this introduction I have outlined the burden of disease associated with three major infectious threats to human health: the rising trend of antimicrobial resistance associated with *K. pneumoniae*, seasonal epidemics of influenza and the global pandemic of SARS-CoV-2. While these three pathogens have markedly different biology and pathogenicity, they all cause substantial morbidity and mortality in secondary care and are important sources of nosocomial transmission. Despite extensive study, there remain large gaps in our understanding of these organisms and the impact that common treatments and infection control interventions have on their ability to cause disease.

The overall aim of this PhD was to combine genomic and phenotypic approaches with epidemiological and clinical data to study common human pathogens associated with nosocomial transmission and disease.

In this thesis, I will assemble and sequence large isolate collections from patients treated at CUH, and use a combination of genomics approaches to characterise the diversity and phylogeny of these three pathogens (**Chapters 3**, **6** and **7**). I will combine genomic, epidemiological and clinical data derived from the patient medical record to address the burden of nosocomial infection at CUH, including dynamics of transmission and the impact of clinical and infection control interventions to reduce the burden of disease (**Chapters 3**, **6** and **7**). I will use a novel method of high content imaging combined with RNAseq for detailed phenotypic and transcriptomic analysis of *K. pneumoniae* under antimicrobial pressure (**Chapters 4** and **5**). Finally, I will use a combination of genomic and epidemiological analyses to understand transmission dynamics in a university setting, and detail the impact of a mass testing programme for asymptomatic COVID-19 at the University of Cambridge (**Chapter 8**).

The specific aims of this thesis are:

1. To characterise the burden of *K. pneumoniae* bacteraemia at CUH over a four-year period, using a combination of clinical and laboratory data in conjunction with phylogenetic analysis of a large prospective isolate collection (**Chapter 3**);
2. To study the effect of various antimicrobial classes on clinical isolates of *K. pneumoniae*, using RNAseq to determine differential expression of genes associated with drug resistance (**Chapters 3** and **5**);
3. To develop a method for high content imaging of bacteria at scale, and to apply this to quantify phenotypic variation in individual bacteria exposed to a range of antimicrobial classes (**Chapter 4**);
4. To produce a collection of *K. pneumoniae* isolates, reflecting the diversity of the species, that has been extensively genotyped (using whole genome sequencing and

phylogenetic analysis) and phenotyped (using antimicrobial susceptibility testing and high content imaging) (**Chapter 5**);

5. To use this collection to study the response of *K. pneumoniae* isolates to antimicrobials, using a combination of high content imaging and differential gene expression using qPCR (**Chapter 5**);
6. To characterise the burden and mortality associated with influenza A and B infection at CUH over three winters using a variety of clinical datasets, and to evaluate the impact of control measures to reduce transmission and mortality (**Chapter 6**);
7. To characterise the burden of SARS-CoV-2 infection in CUH during the first year of the COVID-19 pandemic, using clinical data automatically extracted from the medical record, and to evaluate the impact of control measures to reduce transmission (**Chapter 7**);
8. To describe the transmission of influenza A and SARS-CoV-2 within CUH using a combination of phylogenetic and epidemiological analyses (**Chapters 6** and **7**);
9. To describe the implementation and evaluation of an asymptomatic COVID-19 screening programme at the University of Cambridge (**Chapter 8**);
10. To use a combination of genomic and epidemiologic approaches to understand the introduction of SARS-CoV-2 into the University, the dynamics of transmission once established, and the extent of onward transmission to the local community (**Chapter 8**).

2. Methods

2.1 Clinical data

2.1.1 Study setting and population

Cambridge University Hospitals (CUH) NHS Foundation Trust includes both Addenbrooke's Hospital and the Rosie Hospital in Cambridge, in the East of England. The Trust provides secondary care services for Cambridge and the surrounding area (a population of approximately 580,000), as well as tertiary referral services (including infectious diseases, oncology, haematology, solid organ transplantation, neurosurgery, paediatrics and neonatology) for the East of England (a population of approximately 2.5 million). CUH employs approximately 10,100 members of staff, in addition to hosting several thousand staff on honorary contracts employed through other agencies. Prior to the COVID-19 pandemic, CUH had 1,165 inpatient beds across 54 wards and 103 day care beds[250]. CUH has a completely electronic medical record using Epic (Epic Systems).

2.1.2 Clinical specimen processing

The on-site Public Health England Clinical Microbiology and Public Health Laboratory (CMPHL) forms part of the national PHE network of regional reference laboratories and provides bacteriology and virology diagnostic services for CUH. Blood culture sets consisting of two bottles (aerobic and anaerobic, BacT/ALERT, bioMérieux) were obtained from

patients as clinically indicated and cultured by the diagnostic laboratory. Species identification was made using mass spectrometry (MALDI-TOF MS, Bruker Daltonics). Antimicrobial susceptibility testing was determined using the disc diffusion method[251], with breakpoints determined by BSAC or EUCAST[252,253].

In cases of suspected respiratory viral infection, a nose and throat swab was collected into a single tube of viral transport medium for each patient. Samples were either tested directly using point of care diagnostics (Cepheid GeneXpert for influenza A and B, SAMBA-II for SARS-CoV-2[254,255]) or sent to the CMPHL for nucleotide extraction and PCR testing. RNA extraction was performed using the NucliSENS easyMag (bioMérieux) according to the manufacturer's recommendations. Detection of respiratory viruses was performed using a multiplex real-time PCR assay based on the Rotor-Gene Q platform (Qiagen) using primers and probes developed in the laboratory[256,257]. Tests were reported positive at cycle threshold (CT) values ≤36. In summer 2020, a second method for molecular SARS-CoV-2 diagnostics was introduced to increase capacity. This automated, proprietary PCR based assay (Panther, Hologic) was validated against the CMPHL in-house PCR assay. Samples were considered positive for SARS-CoV-2 with a relative light unit value ≥600.

As part of routine laboratory standard operating procedures, all blood culture and respiratory virus isolates were prospectively stored at -80°C.

Additional SARS-CoV-2 testing capacity for healthcare worker (HCW) screening at CUH was developed in the Cambridge Institute for Therapeutic Immunology and Infectious Disease (CITIID) on the Cambridge Biomedical Campus and was available between April and June 2020. Similar protocols were used as those in CMPHL, adapted for a different qPCR platform (QuantStudio, Thermo Fisher)[258].

HCW screens from July 2020 onwards, as well as the University asymptomatic screens, were tested by PCR in the University of Cambridge/AstraZeneca COVID-19 Testing Centre (Anne McLaren Building, Cambridge Biomedical Campus), part of the UK Lighthouse Laboratories Network. RNA was extracted from heat-inactivated samples using the Beckman RNA Advance Viral Kit, then tested by RT-qPCR for the presence of the orf1ab gene of SARS-CoV-2 using the Genesig Real Time PCR COVID-19 High Throughput (HT) Assay Kit.

2.1.3 Case ascertainment

Bacteraemia isolates of interest were identified by filtering a database of all positive culture results exported from the laboratory information system, based on species. This provided data on sample collection and result dates, patient identifiers and antimicrobial susceptibility results determined by the clinical laboratory. All stored *Klebsiella* isolates collected between December 1 2016 and December 31 2019 were retrieved from the CMPHL freezer for whole genome sequencing and long term storage.

Cases of influenza and SARS-CoV-2 were identified by running searches for all laboratory investigations performed either at the CMPHL or the CUH point of care (POC) laboratories using the reporting function in Epic. Results were exported into Stata 14.1 for cleaning and to exclude void, negative or duplicate results. Samples identified by respiratory microarray were excluded. Case numbers for other respiratory viruses were determined using the respiratory virus trackboard, using data extracted from the laboratory information management system and visualised in Qlikview. Influenza isolates included in the service evaluation were collected between September 1 and the final Sunday of the influenza

season, as determined locally by the clinical virology department (March 19 in 2017, April 30 in 2018 and April 16 in 2019). Influenza A(H3N2) cases from the 2016/17 winter were identified for sequencing. For the multivariable analysis of risk factors associated with mortality from influenza, all cases identified from August 20 2016 and May 12 2018 were included.

All SARS-CoV-2 diagnoses confirmed in CUH from the first case on March 10 2020 until April 7 2021 were identified for further analysis. For the purposes of the phylogenetic analysis presented in **Chapter 7**, cases between March 10 and April 24 2020 were included.

2.1.4 Clinical data collection

Clinical metadata for the service evaluation of influenza management were obtained from a number of sources. Prescribing data for antivirals were obtained from a dataset of all drug prescriptions in CUH, derived from the Epic patient record and visualised using the Qlikview platform. A separate Epic report was compiled of all patient admissions over each influenza season, providing data on admission and discharge dates. A database of all bed moves for each patient with an admission associated with a PCR-confirmed influenza result was compiled by the Information Management Team at CUH. This was filtered to determine the first bed to which a patient was admitted after transfer from the Emergency Department, and compared to a master list of all side rooms and cohort facilities in the Trust.

The medical records of all patients with influenza identified in the 2016/17 and 2017/18 winters were additionally reviewed by a clinician to obtain demographic data, clinical information regarding their presentation, co-morbidities, smoking and alcohol history, key laboratory and imaging results and outcome (including death). Data were abstracted to a

structured case report form (**Appendix F.2**). The Charlson co-morbidity index was calculated[259]. Hospital and primary care records were reviewed for evidence of seasonal influenza vaccination. Oseltamivir dose, frequency and duration were placed in three categories: i) standard course of 75mg two times daily for five days; ii) non-standard course, i.e. less than the standard course of 75mg two times daily for five days, by dose, frequency, or duration; or iii) oseltamivir not given. Where oseltamivir was not given, the medical record was reviewed to determine any documented reason for omitting treatment.

Data for COVID-19 patients were derived automatically from the patient medical record, using an extract produced by the CUH Data Management Team based on criteria used in the influenza case report in **Appendix F.2**. This included demographic, clinical, laboratory, ward movement and outcome data from all patients identified as having a confirmed SARS-CoV-2 result within the laboratory information management system. Variables for the Charlson co-morbidity index were derived from ICD-10 codes from the patient's active problem list or past medical history. Additional data extracts from Epic were used to produce daily updates of case numbers and bed occupancy.

2.1.5 Definitions of hospital acquired infection

For influenza, cases were provisionally assigned as being community acquired infection (CAI) if swabs were taken within 48 hours of inpatient admission and hospital acquired (HAI) if they were taken more than 48 hours from admission. The medical record of each case was reviewed manually to independently assess the validity of this approach, determining the onset of symptoms consistent with influenza in relation to the time of their hospital admission.

Due to the variable incubation period of SARS-CoV-2, separate criteria were established to reflect increasing certainty of HAI with longer delays from admission to symptom onset or sample collection. These are detailed in **Chapter 7**.

2.1.6 Risk factors for inpatient influenza mortality

Associations between inpatient influenza mortality and individual risk factors were examined by logistic regression using purposeful selection[260,261]. Variables with $P<0.2$ by logistic regression tests (LRT) in the single variable analysis and gender were considered in multivariable logistic regression analysis. Appropriate functional forms for continuous variables were determined by fitting cubic, quadratic and linear functions and selecting the simplest for which no improvement in fit was observed by LRT in both single and multivariable analyses. The initial model consisted of gender and variables with $P\leq0.01$ by LRT from the single variable analysis. A backwards stepwise procedure was undertaken dropping non-confounding variables with the largest LRT $P>0.1$ one at a time. A variable was confounding if its omission resulted in a change of more than 20% in the odds ratio (OR) in one or more of the parameters still in the model. The process continued until no further variables could be removed. Variables with $P\leq0.1$ in the single variable analysis were then added in a forward procedure and the above procedure repeated, at the end of which the remaining variables with $P>0.1$ were added and the backwards stepwise procedure implemented.

A second multivariable model was fitted replacing the variable for the category of oseltamivir treatment with the variable for the delay between the date of onset of influenza illness and the date of commencing oseltamivir treatment.

2.2 Bacteriological techniques

2.2.1 Isolate selection

As outlined in **Chapters 3** and **4**, provisional bacterial phenotypic characterisation was performed using the reference isolates ATCC 43816 and NCTC 13438, representing commonly studied *K. pneumoniae* isolates with varying patterns of antimicrobial susceptibility. The method of selecting isolates to form a larger collection representing the diversity of the species is described in **Chapter 5**.

Additional clinical isolates derived from a CUH collection of bacteraemia isolates from 2016 were selected as described in **Chapter 3**. These isolates were obtained through enhanced surveillance of resistant *Enterobacteriaceae* in CUH and had previously been sequenced at the Wellcome Sanger Institute[31]. Antimicrobial susceptibility data for all 127 isolates were previously obtained using the Vitek-2 antimicrobial susceptibility testing system (*Enterobacteriaceae* AST-N350 cards, bioMérieux). Genome assemblies were interrogated for resistance genes and sequence type using Kleborate (v0.3.0)[42]. Pilot RNAseq studies were planned on two closely genomically related clinical isolates with varying ciprofloxacin susceptibilities. In this collection, only one sequence type, ST45, contained isolates with disparate ciprofloxacin minimum inhibitory concentrations (MICs): KC006 (MIC≥4) and KC036 (MIC≤0.25). These two isolates were therefore chosen for ongoing experimental work.

When required, species identification was confirmed using mass spectrometry (MALDI-TOF MS, Bruker Daltonics), following the manufacturer's instructions.

2.2.2 Growth medium and culture techniques

All isolates were streaked from frozen stocks onto Iso-Sensitest or Luria Bertani (LB) agar (Oxoid), and incubated overnight at 37°C in a static incubator. Isolates were re-streaked twice before use in any experimental protocol. When overnight cultures were required in liquid media, single colonies were picked and inoculated into 10 mL of Iso-Sensitest or LB broth and incubated overnight at 37°C in a shaking incubator.

2.2.3 Antimicrobial susceptibility testing

MICs were initially obtained using Etests (bioMérieux) according to the manufacturer's instructions. In brief, bacterial isolates were streaked onto Iso-Sensitest agar (Oxoid) plates as described above, before single colony picks were diluted into phosphate buffer solution (PBS) until the solution approximated the 0.5 MacFarland standard. 100 µL of solution was spread evenly over an agar plate and an Etest applied using sterile forceps. The plate was incubated overnight at 37°C and read at 16-18 hours. Where the MIC fell between two values, the higher value was recorded[262].

The upper limit MIC of the ciprofloxacin Etest is 32 µg/mL. As the MIC of KC006 exceeded this limit, an accurate MIC was determined using agar and broth dilution. For agar dilution, serial two-fold dilutions of 1 mL ciprofloxacin were added to 19 mL fresh Iso-Sensitest agar, to make up a final concentration series from 8 to 500 µg/mL. The isolate was prepared as above to approximate the MacFarland standard. Three spots of 2 µL were placed on each antimicrobial-infused plate and incubated overnight at 37°C. The MIC value was interpreted as the lowest concentration of an antimicrobial where all visible growth was inhibited[263].

For broth macrodilution, 100 µL of overnight culture was added to 9.9 mL Iso-Sensitest broth (Oxoid), inoculated with serial two-fold dilutions of ciprofloxacin making a final concentration series from 8 to 500 µg/mL. The samples were incubated overnight at 37°C. From each clear tube, 10 µL were inoculated onto an Iso-Sensitest agar plate and incubated for a further 24 hours. The MIC value was interpreted as the lowest concentration of ciprofloxacin where all visible growth was inhibited[263].

For the large isolate collection in **Chapter 5**, all isolates were tested using the Vitek-2 antimicrobial susceptibility system (using *Enterobacteriaceae* AST-N350 cards, bioMérieux). Etest results were compared to Vitek results to ensure consistency. Where there was a discrepancy in the results between methods (greater than a fourfold difference in MIC), the Etest was repeated. Because of the wide range of MICs present across the collection and the need to rationalise the approach used for high throughput imaging, for final imaging experiments MICs falling between serial doubling values of concentration (e.g. 0.25, 0.5, 1, 2, 4 µg/mL, etc.) were rounded up to the nearest used concentration (e.g. an MIC of 0.375 µg/mL was rounded up to 0.5 µg/mL).

All antimicrobial solutions and dilutions were prepared according to the manufacturer's recommendations.

2.2.4 Time kill assays

The killing rate of bacterial isolates by ciprofloxacin was determined using a time-kill assay[263,264]. In brief, 100 µL of overnight culture was inoculated into each of 7 tubes containing Iso-Sensitest broth, with each tube containing serial two-fold dilutions of ciprofloxacin around the isolate MIC (i.e. 4x, 2x, 1x, 0.5x, 0.25x, 0.125xMIC and an

untreated control). The samples were incubated at 37°C. At intervals from 0 to 24 hours, 200 µL of each tube was pipetted into the first row of a 96 well plate and diluted in 10-fold series with PBS. 100 µL of appropriate wells were spread evenly over an agar plate. After overnight incubation at 37°C, the plate showing 10 to 200 colonies was used to calculate the final inoculum size. For example, if there were 40 colonies on the plate inoculated with 0.1 mL of a 1:1000 dilution (10^{-4} final dilution), then the inoculum contained 4 x 10^5 CFU/mL.

2.2.5 Growth curves

Growth curves of bacterial isolates in the presence of antimicrobials were determined in an automated plate reader (FLUOStar Omega microplate reader, BMG Labtech). For each sample, 100 µL of an overnight culture was inoculated into 10 mL of Iso-Sensitest broth to create a 1:100 dilution, then 90 µL of this dilution was inoculated into wells of a 96 well plate. The plate was transferred to the plate reader, incubated at 37°C at 200 rpm, and absorbance at 600 nm measured every 10 minutes. At 90 minutes of incubation, 10 µL of ciprofloxacin solution was added to each well, to generate 2-fold dilutions around the isolate's MIC (i.e. 4x, 2x, 1x, 0.5x, 0.25x, 0.125xMIC and an untreated control). Incubation was continued for a further 24 hours. All samples were performed in triplicate and the mean and standard deviation of the absorbance value at each time point used to plot growth curves.

2.2.6 Basic phenotyping methods

A hypermucoid phenotype was determined using string test[55]. Isolates were incubated overnight at 37°C before a standard inoculation loop was used to stretch a mucoviscous string from a single colony. Hypermucoidy was defined by the formation of viscous strings >5 mm in length. The experiment was repeated with overnight incubation at 4°C.

2.2.7 Plasmid curing by electroporation

Attempts were made to cure the KC006 isolate of its plasmid through electroporation[265]. Isolates were streaked three times from frozen and then grown overnight in culture as above. 50 µL of overnight culture was inoculated into 100 mL LB broth and incubated for 37°C in a shaking incubator to an optical density of 0.6. Samples were split into 4x 25 mL tubes and then centrifuged for 10 min at 4,000 RPM. Excess fluid was removed and 40 mL of ice-cold glycerol added per sample. The samples were again split into 4 further tubes (making a total of 16) and centrifuged at 4,000 RPM. Excess fluid was pipetted and 15 mL ice-cold glycerol added to each sample. Each sample was centrifuged for 40 min at 4000 RPM and the supernatant removed. All samples were pooled and then separated into 16x 1.5 mL Eppendorf tubes, before being centrifuged at 14,000 RPM at 4°C for 4 minutes. The first tube was re-suspended In 30 µL glycerol and the samples pooled by pipetting all of the sample from one tube and inoculating into the next. Samples were aliquoted into 6x 60 µL volumes and frozen at -80°C.

Samples were thawed and transferred into 700 µL SOC media and transferred to an electroporation device. Three samples were tested, pulsed once, twice or three times respectively for 5.4 ms at 1.8 kV. Samples were then incubated for 1 hour at 37°C. 100 µL from each sample was added to 900 µL of LB broth and serial 1:10 dilutions made, to a minimum of $1:10^9$. 100 µL from the $1:10^6$ to $1:10^9$ dilutions was streaked across LB plates and incubated overnight at 37°C. The following day, 48 individual colonies were picked from plates from each electroporation sample and simultaneously placed on both an LB plate and an Iso-Sensitest plate infused with gentamicin at 8 µg/mL, using a grid system to locate each isolate. The gentamicin concentration reflects the MIC of isolate KC006 (32 µg/mL), so that if the plasmid was cured growth would only occur on the plate without gentamicin. Single

colonies from a positive control (isolate KC006) and negative control (isolate KC036) were included in the same grid. Plates were incubated overnight at 37°C. Any colonies that grew on the LB plate but not on the gentamicin plate were picked and streaked onto both LB and gentamicin plates to confirm that they had lost their gentamicin resistant phenotype.

2.2.8 Plasmid conjugation using sodium azide

Azide and amoxicillin infused LB plates were prepared to generate a final concentration of 100 µg/mL of each, either individually or combined. The donor (KC006) and recipient (J53 – an azide resistant *E. coli* isolate) were streaked three times from frozen and then grown overnight in LB broth at 37°C in a shaking incubator. Bacterial cultures were then diluted 1:100 in separate aliquots of 5mL fresh LB broth and incubated for 2 hours. 0.5 mL of each of the donor and recipient culture were combined with 4 mL of fresh LB broth and incubated overnight at 37°C without shaking to aid conjugation. The sample was then vortexed to stop conjugation and 100 µL of the resultant culture spread across the azide and amoxicillin-infused plates. Growth on the combined amoxicillin-azide plate represented transconjugants. The species of these plates was confirmed using MALDI-TOF, as above.

Due to poor yields, the experiment was repeated using tetracycline in place of amoxicillin (again at 100 µg/mL), otherwise using the method above. It was further repeated using tetracycline plates, but incubating the initial combined isolate culture at 30°C overnight to aid conjugation.

After this failed to yield any transconjugants, the experiment was repeated using a cross-streaking approach on solid media. Single colony picks were taken from the donor isolate and streaked across an LB agar plate. A single colony pick was then taken from the recipient

and cross-streaked across the same plate. Duplicate plates were then incubated overnight at either 30°C or 37°C. The following day a streak of the plate was taken, suspended in PBS, and re-streaked on the azide-tetracycline infused plate before a further overnight incubation at 37°C. Growth on this plate should represent transconjugants. The species of these plates was confirmed using MALDI-TOF, as above.

2.3 High content imaging of bacteria

2.3.1 Preparation of plate coatings

Coating matrices were prepared in sterile conditions according to the manufacturer's instructions, as summarised in **Appendix B.2**. All coatings, with the exception of poly-L-lysine, were incubated in ultra-thin 96 well plates (PerkinElmer CellCarrier Ultra) overnight at 37°C and were rinsed 1-3 times with wash buffer prior to use. For poly-L-lysine, wells were coated for 5 minutes. The solution was aspirated, and wells were left to dry overnight at 37°C.

2.3.2 Experimental growth conditions

Isolates were passaged at least twice on solid media from frozen stock, before single colonies were picked and grown to stationary phase overnight in LB broth at 37°C. For the optimisation experiments described in **Chapter 4**, bacterial cultures were then diluted in LB broth and mixed with antimicrobials to reach a final antimicrobial concentration of 5x the MIC of that isolate. If an MIC could not be determined using Etests (i.e. if bacterial growth

continued beyond the upper limit of the MIC strip), the upper limit of the Etest was arbitrarily used in place of the MIC.

For subsequent experiments, overnight bacterial cultures were diluted in LB broth and incubated for 90 minutes in a shaking incubator at 37°C, so that they were in early exponential growth phase at the time of inoculation with antimicrobials. Varying dilutions of antimicrobial were used, as specified in **Chapters 4** and **5**.

In all experiments, 100 µL from each sample was inoculated into a well of a CellCarrier-96 Ultra 96 well plate (PerkinElmer), and incubated in a static incubator at 37°C for 2 hours. The culture was then aspirated and the remaining adherent bacteria fixed using 50 µL 4% paraformaldehyde (Alfa Aesar) added to all wells and left for 10 minutes at room temperature. After a single wash with PBS, fixed cells were stained using the following dilutions: FM4-64 (Thermo Fisher), 2-4 µg/mL; 4′,6-Diamidino-2-phenylindole dihydrochloride (DAPI, Sigma) 20 µg/mL; SYTOX green (Thermo Fisher) 0.125 µM. Concentrations of FM4-64 were adjusted depending on stain intensity, due to batch-to-batch variation and deterioration of stocks over time. All stains were diluted in HBSS buffer (Thermo Fisher) and 50 µL added to each well and left in the dark at room temperature for 20 minutes. After a further PBS wash, 50 µL PBS was added to each well and plates were imaged within 24 hours of staining.

2.3.3 Imaging using the Opera Phenix

All plates were imaged using the Opera Phenix confocal microscope (PerkinElmer), using a 63x water immersion lens. Ten fields of view, each covering an area of 0.2 x 0.2 mm, were imaged for each well, with 3-6 z stacks per field at 0.5 µm intervals to ensure comprehensive

imaging of the bacterial monolayer. Exposure time at each wavelength was consistent between all experiments. Optimisation experiments were performed in biological triplicate, with 3 technical repeats (i.e. 3 wells from the same source culture) per plate.

2.3.4 Image analysis and feature detection

Image analysis was performed using Harmony (v4.9) based on Acapella software (v5.0.0.122842). The full analysis pipelines for Gram negative and Gram positive bacteria are included in **Appendix B**. All 6 z stacks for each field were concatenated into a single image using maximum projection. Optical correction was performed using flatfield and brightfield correction. Bacteria were identified and segmented using the SER ridge feature using the FM4-64 channel and filtered based on intensity to remove artefacts. Basic morphological features (area, length, width and roundness) were calculated for each object, while advanced features (symmetry, threshold compactness, axial, radial and profile measurements) were calculated for each of the FM4-64, DAPI and SYTOX green channels for each object. These features were used to filter artefacts based on size (excluding objects ≤1.5 μm²) and a linear classifier machine learning function (PhenoLogic, PerkinElmer). Objects were classified as either single cells, dividing cells or artefacts, with training including over 100 objects manually classified into each group using the KC006 isolate in the presence of a range of antimicrobials across a single plate. This approach was manually verified by visualising the classification of wells from other plates imaged using the same experimental protocol.

2.3.5 Phenotype evaluation

The mean and standard deviation of each morphological feature for objects classified as single cells in each well were calculated in Harmony. These were used to calculate Z' statistics to determine the relative importance of each morphological feature in distinguishing between untreated bacteria and those treated with antimicrobials[266]. The same data were exported into TIBCO Spotfire Analyst (v7.11.1.0.26, Tibco Software) for further analysis and quality control using the customised High Content Profiler module (PerkinElmer). This enables rapid visualisation of the data in graphical and tabular form, to identify and remove outlying wells. The same module was used to determine principal components of this dataset comparing treated and untreated wells and to visualise their distribution using 3D principal component analysis (PCA) plots. Final results were plotted in GraphPad Prism version 9.0.0 (GraphPad Software) and R using the *ggplot2* and *scatter3d* packages[267]. A full list of morphological properties used in this analysis is included in **Appendix B.5.**

2.3.6 Live bacterial imaging

Experimental methods were replicated from those outlined above wherever possible. After inoculation of bacteria and antimicrobials into 96 well plates and incubation for 90 minutes, wells were aspirated and stains added, either alone or in combination, at 0.1x, 1x or 10x the recommended concentration for imaging fixed specimens. The plates were sealed with a gas permeable membrane and placed into the Opera Phenix, pre-heated to 37°C, where fields in each well were imaged at baseline and at 2 hours, to determine whether the stains interfered with bacterial growth or imaging. Further experimental details, including the stains used and adaptations using GFP and mCherry expressing bacteria, are included in **Chapter 4**.

Additional image analysis using the time series function in Harmony was employed to identify and track individual bacteria over the course of the experiment.

2.4 Molecular techniques

2.4.1 Bacterial DNA extraction

Bacterial isolate collections were grown on agar plates as described above. Single colonies were picked into a single well of a 96 deep well plate containing Iso-Sensitest broth, before overnight incubation. Bacteria were pelleted and re-suspended in PBS and 20 µL of 100 mg/mL lysozyme solution. DNA extraction was performed using either the MagNAPure small volume nucleic acid extraction kit (Roche) on the FLOW flex platform (Roche), or using the QIAamp 96 DNA QIAcube HT Kit (Qiagen), according to the manufacturer's instructions. Resultant DNA concentration and purity was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific) and a Tapestation 4200 using D1000 ScreenTapes (Agilent Technologies).

For long read sequence data (generated by PacBIO) DNA was extracted using the Epicentre MasterPure Complete DNA and RNA Purification Kit. Bacterial streaks from each isolate were suspended in 300 µL tissue and cell lysis solution with 1 µL proteinase K and incubated in a heating block at 65°C for 40 minutes. Samples were incubated with RNAse A at 37°C for 30 minutes, before the addition of 200 µL MPC protein precipitation reagent and centrifugation at 4°C for 10 minutes. Isopropanolol was used to precipitate the DNA, which was washed with 70% ethanol. DNA quantity and quality were determined using a NanoDrop

spectrophotometer (Thermo Fisher Scientific) and Qubit 2.0 fluorometer (Invitrogen, Life Technologies).

2.4.2 Bacterial RNA extraction

Single colonies were picked and added to 10 mL Iso-Sensitest broth and incubated overnight at 37°C. The following day, overnight culture was inoculated into a number of tubes containing Iso-Sensitest broth to a dilution of 1:100, and were grown in a shaking incubator at 37°C. At 90 minutes of incubation, antimicrobial solution was added to the culture, to make a final concentration related to the isolate's MIC, varying by experiment. One tube was left as an untreated control. At predetermined time points, 1 mL of sample was taken from each tube and mixed with 2 mL RNAprotect Bacteria Reagent (Qiagen), vortexed and incubated for 5 minutes at room temperature. The samples were centrifuged for 10 min at 3,000g, before the supernatant was decanted and the remaining pellet frozen immediately at -80°C. All samples were thawed and extracted within 4 weeks using the RNEasy mini kit (Qiagen), as per the manufacturer's instructions. RNA quantity and integrity were determined using the Agilent RNA 6000 Pico kit and the Agilent 2100 Bioanalyzer (Agilent Technologies). All experiments were repeated in triplicate.

2.4.3 Quantitative PCR using the Fluidigm platform

Using RNA extracted as above, 4 µL sample was mixed with 1 µL of Reverse Transcription Master Mix (Fluidigm) and run in a Thermocycler using the following protocol: 25°C for 5 minutes, 42°C for 30 minutes and 85°C for 5 minutes, as per the manufacturer's instructions. Samples were stored at 4°C prior to further processing. cDNA quantification was performed using a NanoDrop spectrophotometer (Thermo Fisher Scientific). qPCR analysis was

performed using the BioMark HD qPCR system (Fluidigm), following the manufacturer's instructions[268]. In brief, samples were pre-amplified then prepared for IFC (integrated fluidics circuit) loading, with EvaGreen and DNA Binding Dye Sample Loading Reagent. 5 µl of each assay mix and sample mix was loaded into inlets of a 96.96 IFC before loading using the Fluidigm Juno. Once complete, the IFC was moved to the BioMark HD for qPCR. Preliminary thresholding of the amplification data was completed using the Fluidigm Real-Time PCR Analysis Software, before raw data were exported to apply manually defined melting curve peak thresholds.

Quantification of differential gene expression was calculated using the steps[269]:

1. Normalisation of the sample Ct value in relation to the housekeeping gene:

$$\Delta CT \text{ values} = \text{sample CT value} - \text{housekeeping gene CT value}$$

2. The sample CT value can then be compared to a related control sample:

$$\Delta\Delta CT \text{ values} = \text{sample } \Delta CT \text{ value} - \text{control } \Delta CT \text{ value}$$

3. Fold difference values can be obtained using the following formula

$$\text{Fold difference} = 2^{-(\Delta\Delta CT \text{ values})}$$

2.5 Genome sequencing and assembly

2.5.1 Bacterial DNA – short read

Bacterial isolates collected in 2016 and 2017 were sent to the Wellcome Sanger Institute for sequencing. Library preparation was undertaken with automated systems using the IHTP WGS NEB Ultra II library kit. Libraries were sequenced on an Illumina HiSeq platform (Illumina) using standard protocols. Annotated assemblies were produced using the pipeline described in Page *et al*[270]. For each sample, sequence reads were used to create multiple assemblies using VelvetOptimiser (v2.2.5)[271] and Velvet (v1.2.10)[272]. An assembly improvement step was applied to the assembly with the best N50 and contigs were scaffolded using SSPACE (v2.0)[273] and sequence gaps filled using GapFiller (v1.11)[274]. Automated annotation was performed using PROKKA (v1.5)[275] and a genus specific database from RefSeq[276]. Taxonomic identity was assigned to all short reads and assemblies using Kraken[277]. The quality of the sequences was checked by determining the length of each genome, the number of contigs in the assembly and depth of coverage against the mapped reference.

Bacterial isolates in 2018 and 2019 were sequenced in-house. Library preparation was undertaken using NEBNext Ultra II FS DNA library prep kit (New England BioLabs Inc.). Sequencing was performed using the Illumina MiSeq platform (Illumina). With the resultant fastq file, read filtering for length and quality was performed with fastp (v0.21.0) using the default settings[278]. Assemblies were produced in Unicycler (v0.4.9b)[279] and the assemblies annotated using PROKKA (v1.5) as above. Assembly quality was checked as above.

2.5.2 Bacterial DNA – long read using Nanopore sequencing

DNA was extracted using the RNeasy method from Qiagen above. Library preparation was performed using the Nanopore Protocol for Rapid Barcoding Sequencing (SQK-RBK004). Sequencing was performed using a MinION SpotON flow cell (Oxford Nanopore Technologies), following the manufacturer's instructions. Basecalling was performed using Guppy (v5.0.11)[280] with read filtering by fastp (v0.21.0)[278]. Additional bar code and adapter trimming was performed using porechop (v0.2.4)[281], and subsequent short read filtering with filtlong (v0.2.0) (https://github.com/rrwick/Filtlong).

Hybrid assemblies were produced for clinical isolates used as references for RNAseq mapping. Short read sequence data produced by Illumina sequencing at Wellcome Sanger Institute was combined with the trimmed and filtered long read data in Unicycler (v0.4.9b)[279] using bold mode.

2.5.3 Bacterial DNA – long read using PacBio sequencing

DNA was sequenced at the Wellcome Sanger Institute using the PacBio RS II instrument (Pacific Biosciences Inc.). PacBIO reads were converted from BAM to FASTQ format using Samtools (v1.6)[282], corrected and assembled using CANU (v1.6)[283] and circularised using circlator (v1.5.3)[284]. The assembly was then annotated using PROKKA (v1.5)[270,275].

2.5.4 Bacterial RNA

Total RNA was reverse transcribed using SuperScript III reverse transcriptase (Invitrogen). Library construction, fragmentation, 1st and 2nd strand synthesis, end preparation and

ligation was performed using the NEB Ultra II RNA custom kit (New England BioLabs) on a Bravo WS automation system (Agilent Technologies). Sequencing was performed at the Wellcome Sanger Institute using the HiSeq platform (Illumina).

2.5.5 Viral RNA – short read (influenza)

All samples were transferred from the CMPHL to the PHE central sequencing laboratory in London for sequencing. Multi-segment reverse transcription-PCR (MRT-PCR) was used to amplify H3N2 RNA[285]. Reverse-transcription and amplification were performed using One-step RT-PCR system with Superscript III and Platinum Taq HiFi polymerase (Life Technologies). The MRT-PCR products were cleaned, diluted to the required concentration and submitted for Nextera library preparation for Illumina short-read sequencing with a MiSeq instrument (Illumina).

2.5.6 Viral RNA – long read (SARS-CoV-2)

All PCR-positive, pre-extracted diagnostic samples were identified and transferred from diagnostic laboratories (CMPHL for clinical samples, the CITIID laboratory for HCW samples or the Cambridge lighthouse laboratory for University samples), to the University of Cambridge Department of Virology for sequencing. Where CT values were available prior to sample selection, positive samples with a CT value ≤33 were sequenced using a multiplex PCR based approach according to the modified ARTIC v2 protocol v3 primer set[286]. Amplicon libraries were sequenced using MinION flow cells (v9.4.1) (Oxford Nanopore Technologies). Genomes were assembled using reference-based assembly and a bioinformatics pipeline using 20x minimum coverage cut-off for any region of the genome and 50.1% cut-off for calling single nucleotide polymorphisms. Genomic data were filtered to

exclude sequences with >5% 'N's and those of spuriously low file sizes (<29KB). Genomes were aligned with minimap2[287] to the Wuhan Hu-1 reference genome (MN908947.3). All samples were processed through COVID-CLIMB pipelines[288,289]. Samples with ≥95% genome coverage were included for phylogenetic analysis. Protocols are available at https://github.com/COG-UK.

Individual samples from students identified as being SARS-CoV-2 positive through asymptomatic screening were not available for the first week of the University term (week commencing October 5 2020). Given the potential importance of identifying lineages present in the University in the first week of term, attempts were made to sequence all RNA extracts from pooled samples where an individual positive student had been identified.

2.6 Bioinformatic analysis of bacterial sequences

2.6.1 Phylogenetic analysis

Reads were mapped to the reference genome *Klebsiella pneumoniae* NTUH-K2044. Annotation files produced by PROKKA were used as the input for the pangenome package Roary (v3.13.0) to produce a core genome alignment[290]. This was used to produce a maximum likelihood, midpoint rooted tree using IQTREE (v1.6.10)[291]. Phylogenetic trees were visualised and annotated in the Interactive Tree Of Life (iTOL, v5)[292]. SNP distance matrices for closely related lineages were produced using snp-dists (v0.6.2) (https://github.com/tseemann/snp-dists).

2.6.2 Genotyping

Sequence type, subspecies and genes associated with antimicrobial resistance and virulence were determined using Kleborate (v2.0.1)[42]. Kleborate definitions of sequence type are based on the 7-locus scheme from the BIGSdb database hosted at the Institut Pasteur[37]. Capsule and O locus typing was assigned using Kaptive (v0.7.0)[44]. Plasmid types were identified by searching assemblies using SRST2[293] to report against the PlasmidFinder database[294]. Where specific contigs needed to be associated with plasmid types, the hybrid assembly was visualised in Bandage (v0.8.1)[295] and contigs copied as fasta files into the online version of PlasmidFinder (v2.1) (https://cge.cbs.dtu.dk/services/PlasmidFinder/)[294,296].

2.6.3 RNA sequencing and differential gene expression analysis

For each sample, sequence reads were mapped against the reference genome for that isolate using BWA (v0.7.17)[297] to produce a BAM file. BWA was used to index the reference and the reads were aligned using default parameters but with the quality threshold for read trimming set to 15 (q=15) and maximum insert size (a) set as the maximum requested fragment size of the sequencing library. Gene expression values were computed from the read alignments to the coding sequencing to generate the number of reads mapping and RPKM (reads per kilobase per million)[298,299]. Only reads with a mapping quality score of 10 were included in the count. Differential expression analysis was conducted using DESeq2[300], with data visualisation and quality checked using DEAGO[301] . PCA plots of normalised counts were used to identify samples that did not cluster with their replicates. Genes encoding hypothetical proteins of unknown function with substantial changes in expression were checked using the Uniprot database[302].

All of the software developed by the Pathogen Informatics team at the WSI is freely available for download from GitHub [299,301,303,304] under an open source license, GNU GPL 3.

2.7 Genomic epidemiology analysis of respiratory viruses

2.7.1 Spatio-temporal analysis in CUH

Ward movement data was extracted for each hospital admission associated with a confirmed case of influenza or SARS-CoV-2, as detailed above. An algorithm was written in Structured Query Language (SQL) by the PHE Field Service to detect all instances of co-admission to a single ward of two or more cases with overlap between the infectious interval of one case and the susceptible interval of another. For example, for influenza, the susceptible interval was taken to be from between 96 to 24 hours before the first positive specimen; and the contagious interval from 24 hours before to 168 hours after the first positive specimen, based on the CDC description of influenza infections[305].

The SQL algorithm generated a list of co-admissions that fit these parameters, detailing which patients were co-located together, on which ward, for what period, and where the risk of transmission was present. A possible between-ward seeding event was defined as a network in which a primary case had been transferred from one ward to another during the case's infectious interval.

2.7.2 Phylogenetic analysis

For influenza, multiple sequence alignment was performed using BWA MEM (v0.7.12-r1039)[297]. Maximum likelihood phylogenies were constructed using IQ-TREE (v1.6.1)[291]. Cluster membership was assigned using the DendroPy package (v4.3.0)[306] in Python (v3.5.2). A nosocomial person-to-person transmission was considered plausible where fewer substitutions than expected within 168 hours at a rate of 0.006 substitutions/site/year[307] were observed since their most recent common ancestor. Phylogenetic plots were created using MatPlotLib (v2.1.2)[308].

For SARS-CoV-2, multiple sequence alignment was performed using MAFFT (v7.458)[309]. Phylogenetic trees were produced using IQ-TREE (v2.1.2)[291] and rooted using Wuhan Hu-1 (MN908947.3) as an outgroup. They were visualised in Microreact[310] and the R package *ggtree* (v2.0.4)[311]. Viral lineages were assigned with the PANGOLIN package (v1.07 for CUH analyses, v2.1.6 for University analyses)[312]. A SNP difference matrix was produced from the multiple sequence alignment using the *snp-dists* package (v0.7.0, as above).

2.7.3 SARS-CoV-2 transmission networks in Cambridge University Hospitals

Clusters were defined by genomically identical isolates (0 SNP differences). They were reviewed alongside the spatio-temporal clusters produced by the PHE Field Service. Cases within both the same space-time cluster and the same genomic cluster were deemed to be supportive of a recent transmission event. Where a spatio-temporal association was not identified, the patient records for each case within a genomic cluster were manually reviewed and assigned a score of 1 (strong evidence supporting a recent linked transmission

chain, e.g. patients co-located in another location (such as pre-admission) and becoming positive within the incubation period of the virus), 2 (a plausible transmission chain was present e.g. patients becoming positive while located in nearby wards within the hospital but who did not appear to be in direct contact), and 3 (no evidence of any epidemiological associations between cases).

Genomic and epidemiological data were reviewed in a weekly working group comprising infectious disease physicians, microbiologists, infection control nurses and the CUH patient safety team. Representatives from occupational health, PHE Field Service and SARS-CoV-2 whole genome sequencing teams (University of Cambridge, COG-UK) were also invited. All potential HAIs were reviewed within one week of a positive test by an infection clinician to confirm whether cases were more likely to be community or hospital acquired. They then used a structured case report form to extract information from Epic. Cases were presented at the next working group meeting to obtain a consensus view on whether the balance of probability favoured hospital or community acquisition. The clinical teams directly involved in patient care were not involved in data collection or analysis.

2.7.4 SARS-CoV-2 transmission networks in the University of Cambridge

SARS-CoV-2 sequences were derived from the University and compared to isolates from the local community, identified from the COG-UK dataset for Cambridgeshire (https://www.cogconsortium.uk/data). These samples were derived from NHS local community testing from non-hospitalised, symptomatic individuals. Other samples were derived from patients treated at three Cambridgeshire hospital trusts: CUH; Royal Papworth Hospital NHS Foundation Trust (specialist heart and lung hospital, providing surge capacity

for COVID-19); Cambridgeshire and Peterborough NHS Foundation Trust (provider of community, mental health and learning disability services in Cambridgeshire). Hospital samples were obtained from both asymptomatic screening and those exhibiting COVID-19 symptoms. Finally, samples were also derived from the asymptomatic HCW screening programme at CUH[220,258].

Possible transmission clusters were defined using CIVET v2.1.0 (https://github.com/artic-network/civet). Where indicated, collapsed nodes from trees generated from CIVET were inspected to visualise data in the context of the COG-UK national database (https://www.cogconsortium.uk/). Each genomic cluster was manually reviewed with data available from the University and local and national contact tracing to determine epidemiological associations between members (see below).

2.8 COVID-19 at the University of Cambridge

2.8.1 Study setting and population

The University of Cambridge has approximately 23,000 students and 12,600 staff. It is divided into 31 colleges and 150 departments, faculties and other institutions. Students belong to a College community, as well as being members of the University and an academic faculty/department. Colleges provide residential accommodation for approximately two thirds of all University students, either on campuses or in off-site housing, and offer pastoral and academic support for each individual as well as social and sports activities. All Colleges have membership from students across multiple courses and departments; conversely, departments are comprised of students and staff affiliated with multiple Colleges.

2.8.2 Asymptomatic screening programme

The University of Cambridge Asymptomatic COVID-19 Screening Programme was a weekly screening programme for SARS-CoV-2 infection for students living in College owned or maintained accommodation. During Michaelmas term 2020, screening took place over a 9 week period from October 5 2020 to December 6 2020. Students were screened using swab pooling and two-step confirmatory PCR testing as detailed below.

The programme was announced to students on September 9 2020. Formal invitations to participate were emailed to all eligible students on September 26 2020 and weekly thereafter. These emails contained links to the programme website, which contained the consent form, student information sheet and privacy notice. Additional communications were sent from Colleges and the programme was promoted on social media by student representatives.

The screening programme was introduced in phases, to ensure that all aspects of the programme operated effectively and had capacity to manage uncertain numbers of positive cases. Initially, two students from each testing pool were randomly selected to contribute swabs for pooled sample collection each week (weeks 1-2). Then, half of the students in each testing pool were asked to contribute swabs each week (weeks 3-7, such that every student was tested every two weeks). Finally, all students in each testing pool were asked to contribute swabs each week (weeks 8-9, ensuring every student was tested weekly).

Students were excluded from participating in pooled sample collection if: 1) they had tested positive for SARS-CoV-2 within the preceding 8 weeks (later 90 days, in line with national guidance); 2) they had symptoms of COVID-19 or were awaiting an individual symptomatic

test result; 3) they were self-isolating for any other reason, for example due to contact with a confirmed or suspected COVID-19 case or international travel from a high prevalence area.

College accommodation was divided into households based on shared facilities such as a bathroom or kitchen. For the purposes of screening, some households were split or merged with neighbouring households to ensure testing pools of 6-10 students. Household and testing pool membership were determined by each College prior to the start of term (based on guidance provided by the University and the screening programme) and recorded in a central database.

Customised testing kits were prepared within the University. Each kit was labelled with the names of students belonging to a particular testing pool, and contained packaging and cleaning materials, a number of nasopharyngeal flocked swabs (A04-96000P, Trafalgar Scientific, or MSC-96000, Sterilab) corresponding to the number of students within that pool, and one tube of viral transport medium (HiViral Transport Medium, Trafalgar Scientific) labelled with a pseudonymised code associated with that testing pool. Screening kits were delivered to each College weekly for distribution to testing pools. Each College was assigned a testing day, with all students from that College invited to participate on the same day. Drop-off points were allocated in every College, with same-day sample collection and delivery to the testing facility. Screening took place from Monday to Thursday, with individual confirmatory tests from Tuesday to Friday.

Combined nose and throat swabs were obtained by self-administration in the students' own accommodation. All swabs from members of a pool were immediately placed in the same tube of viral transport medium. The physical dimensions of the tube limited capacity to 10 swabs, therefore determining the maximum number of students in a pool. To confirm the

number of students contributing swabs, each pool was asked to complete a simple electronic "swab count" at the time of sample collection, indicating the number of swabs present in the sample tube.

All samples were tested by PCR in the established University of Cambridge/AstraZeneca Cambridge COVID-19 Testing Centre, as detailed above. Pseudonymised results were transferred to the secure data hosting service (SDHS) provided by the University of Cambridge School of Clinical Medicine, where they were associated with members of each testing pool. Results were communicated to students by text message, and shared with Colleges and the University COVID-19 Operations Centre via the SDHS. Results of individual confirmatory tests were also reported to Public Health England using the Second Generation Surveillance System (SGSS), with automated notification to NHS Test and Trace for parallel local and national contact tracing efforts. Additionally, cases were asked to complete a monitoring form held centrally by the University, which requested students to document contacts who were at risk from transmission.

Students contributing swabs to positive pooled tests were instructed to self-isolate, and either invited for same-day individual confirmatory tests at a University of Cambridge/CUH testing facility (weeks 1-6) or provided with a test kit for individual self-swabbing in their own accommodation (weeks 7-9). Students with positive individual confirmatory tests were treated in the same manner as students with positive results from NHS pathways for symptomatic testing.

Isolation of cases and contacts was in accordance with UK national guidance, instructions from NHS Test and Trace and the local public health teams. In brief, cases were required to self-isolate for 10 days from their date of symptom onset (or test date, if asymptomatic).

Households of confirmed cases were quarantined for 14 days. Colleges provided support for student isolation; the nature of this support varied between Colleges and households, but included provisions such as food and drink, educational and psychological support where required.

If a positive student could not be identified by individual testing, the CT value of the positive pooled screening test was reviewed: if >36, the pooled screening test was considered to be an operational false positive; if ≤36, a further round of individual tests was performed, to minimise the risk of a false negative individual test result. If all individual confirmatory tests were negative, students were released from self-isolation/household quarantine.

2.8.3 Other University control measures

In addition to asymptomatic screening, the University provided a testing service for students and staff with the cardinal symptoms of COVID-19 (fever, cough, and/or anosmia/ageusia). Combined nose and throat swabs were obtained by self-administration at a University testing facility in the city centre, Monday to Friday during the study period. Ascertainment of students testing positive by NHS pathways was derived from self-reporting to the University COVID-19 Operations Centre, completion of the University's COVID-19 monitoring form, or information provided by the Local Authority or Public Health England.

In addition to offering symptomatic testing and asymptomatic screening, the University supported a number of COVID-19 reduction measures for the duration of the study period under the banner of their "StaySafeCambridgeUni" campaign. These measures reflected contemporary national policy in the UK. They were communicated to students prior to the start of term, periodically as part of routine email communications from the University and

Colleges or Departments during the term, and supported by posters across the University campus. All policies and a range of supporting materials were made available online. Additional support and information was provided by Colleges, departments and through a network of collegiate student representatives and the University-wide student union.

COVID-19 reduction measures included the mandatory use of face coverings in work or study settings (unless the individual concerned had a medical exemption), widespread promotion of hand washing and availability of virucidal hand sanitisers, and guidance on social distancing to a minimum of 2 metres. Lectures and other large group teaching was moved online. Small group teaching and practical classes were continued in accordance with the policies above. All international students were asked to self-isolate immediately on arriving in Cambridge, for a period of 14 days.

From the start of term until November 4 2020, extra-curricular, sporting and social activities were permitted in line with the guidance above. There were no additional restrictions placed on student movement between the University and the local community. Individual Colleges made policies on the admission of the general public and members of other Colleges into their premises.

Overall support and coordination for the University's COVID-19 response was provided by a dedicated Operations Centre, providing telephone and e-mail advice and decision-making for students and staff. Oversight was provided by senior members of the University in a gold-silver-bronze command structure.

Representatives of the University and the COVID-19 Operations Centre met with members of the local public health authorities at least twice-weekly during the study period to monitor

surveillance activities and coordinate interventions to reduce transmission. Where high rates of transmission were identified in a particular College, additional meetings were held with representatives from all parties. Numbers of newly confirmed cases were shared daily between the screening programme, symptomatic testing, COVID Operations Centre, public health authorities and members of the University's gold command.

Announced on October 31 2020, a national lockdown was implemented by the UK government on November 5 which lasted until December 1. Stricter restrictions were put in place during this time, including the closure of all hospitality venues, limitations on mixing between households (unless students were part of an existing social bubble) and movements outside the home unless for essential activities (such as shopping or medical care) or physical exercise. All group sporting activities were cancelled, as were social activities involving multiple households. The majority of in-person teaching was either postponed or moved online, with the exception of students on vocational training programmes such as clinical medical students. Further pastoral support was provided through the Colleges.

2.8.4 Retrospective telephone symptom survey

To investigate the symptoms experienced by students with SARS-CoV-2 infection identified by asymptomatic screening, a retrospective telephone survey was conducted of all students with positive individual confirmatory tests. Students were contacted between November 24 2020 and December 17 2020 by a clinical research nurse or members of the University COVID-19 Operations Centre, and assessed using a set questionnaire.

To estimate the proportion of all students identified by asymptomatic screening who experienced symptoms, the sample of students responding to the telephone survey was assumed to be representative and independent of the manifestation of symptoms.

2.8.5 Epidemiologic and NHS contact tracing data

Student demographic data were derived from the University of Cambridge student electronic record system CamSIS, and household structure and membership data from the asymptomatic screening programme. The Second Generation Surveillance System (SGSS) and contact-tracing data provided by NHS Test and Trace were interrogated to identify additional University affiliated cases (students and staff) and hospital staff accessing the national community testing programme.

Epidemiologically linked ‘common exposures’ for students, University staff and the local community were identified through Test and Trace data. Common exposures were defined as locations or events that two or more people testing positive for SARS-CoV-2 visited in the same two to seven day period before symptom onset or positive test. Additional contact tracing information was also provided by the University COVID-19 Operations Centre.

2.8.6 Additional statistical analyses

Participation rates were estimated by comparing the number of samples received and reported by the testing facility with the mean number of students participating per testing pool, reported through “swab counts” detailed above. Student characteristics associated with

programme participation and positive test results were assessed using unpaired 2 sample 2-tailed *t*-tests (continuous variables) or chi-square tests (categorical variables).

To evaluate the secondary attack rate amongst household contacts, we first identified all index cases (the first case diagnosed in each household). For each index case, we then compared the number of household contacts testing positive within 14 days of the index, compared with the total number of household contacts. Results were stratified according to the characteristics of the index case (testing route, symptoms and CT values of their individual sample). Where multiple index cases were identified in the same household on the same day, they were either reported as a separate “mixed” category or the household was excluded from analysis. Secondary attack rates were compared using Fisher’s exact test. Associations between testing positive for SARS-CoV-2 (by any testing route at any time during the study) and individual risk factors were examined by logistic regression and expressed as odds ratios with 95% confidence intervals and associated *P* values. All variables were treated as categorical, with the exception of household size, which was treated as continuous.

A multivariable logistic regression analysis was fitted with six covariates, namely gender, ethnicity, residence (out of term time), year group, course and household size. All variables were significant, so were retained in the final model. Pairwise interaction terms were also considered, but did not lead to an improved model.

2.9 Statistical analyses

Comparisons between categories were made using chi-square tests, while distributions of continuous variables were compared using paired or unpaired 2 sample 2-tailed *t*-tests (2 distributions) or ordinary one-way ANOVA (3 or more distributions), as specified in the text and figure legends.

Datasets were variously compiled in Azure (Microsoft) and Excel 2013 (Microsoft) and analysed in STATA 14.1 (College Station) and GraphPad Prism version 9.0.0 (GraphPad Software) unless otherwise stated.

2.10 Ethics

Investigation of *K. pneumoniae* bacteraemia isolates from CUH was conducted under the terms of ethical approval granted by the NHS Research Ethics Committee for the project "Whole genome sequencing of bacterial pathogens" (reference: 12/EE/0439).

On advice from the chair of the East of England – Cambridge Central Research Ethics committee, the influenza study was exempted from ethical review. The work has been registered with the Safety and Quality Support department at CUH under a service evaluation/outbreak investigation remit (project numbers: 6905, 7053, 7787).

Investigations of SARS-CoV-2 transmission in CUH and the University of Cambridge were conducted as part of surveillance for COVID-19 infections under the auspices of Section 251 of the NHS Act 2006 and therefore did not require individual patient consent or ethical

approval. The COG-UK study protocol was approved by the Public Health England Research Ethics Governance Group (reference: R&D NR0195).

Ethical approval for the University of Cambridge asymptomatic screening programme was granted by the University of Cambridge Human Biology Research Ethics Committee (reference: HBREC.2020.35). Participation was voluntary, with no coercion or incentives. Informed consent was obtained from all participants, who were free to withdraw from the programme at any time. All data was processed fairly, lawfully and transparently, in accordance with the EU General Data Protection Regulation (GDPR) 2016/679 and the UK Data Protection Act 2018.

3. Characterising invasive *K. pneumoniae* at Cambridge University Hospitals

3.1 Introduction

Despite advances in genomic sequencing and phenotype-genotype correlations for antimicrobial susceptibility, there is still much that is unknown about the mechanism and drivers that lead to the development of antimicrobial resistance. This is particularly true of certain antimicrobial classes such as the fluoroquinolones, where the acquisition of resistance to drugs such as ciprofloxacin has been associated with the global dissemination of resistant clades of public health importance[41,78,79].

To illustrate this problem, **Figure 3.1** plots the minimum inhibitory concentration of ciprofloxacin for 129 *K. pneumoniae* isolates in the British Society for Antimicrobial Chemotherapy (BSAC) bacteraemia collection[39], grouped by known fluoroquinolone resistance determinants identified using Kleborate[42]. Three groups of isolates are included: those with no known genes or mutations associated with resistance; those with only a single mutation in the key fluoroquinolone resistance gene *gyrA*; and those with a common combination of mutations in *gyrA* and *parC*. While each mutation in this example is associated with increased resistance, there is clearly a wide distribution of MICs for each group, suggesting that there are other important determinants of resistance. Of note, the distribution of MICs for isolates with no known resistance mutations crosses the recognised breakpoint of antimicrobial susceptibility[253]. Not only does this demonstrate that there are

unknown resistance determinants in *Klebsiella pneumoniae*, they are also sufficient to impact clinical outcomes in patients treated empirically with ciprofloxacin.

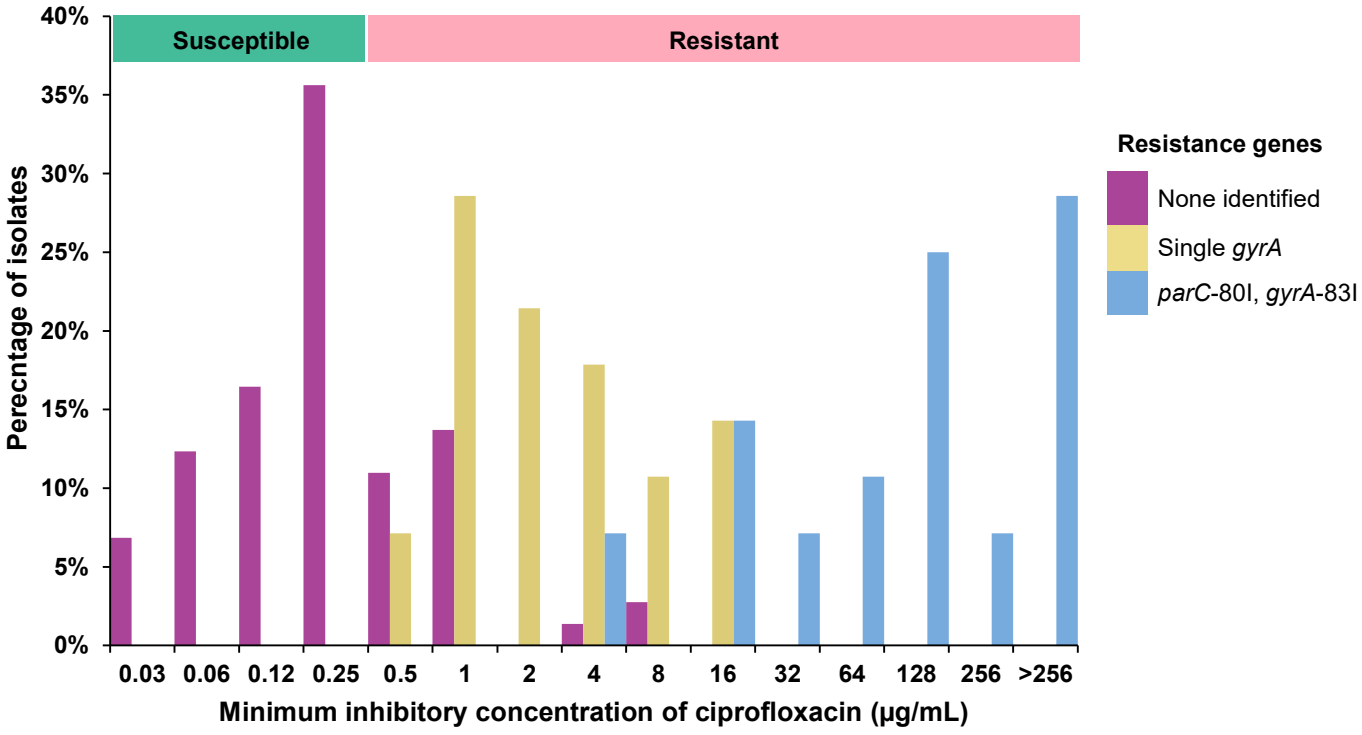


Figure 3.1. Distribution of minimum inhibitory concentrations of ciprofloxacin for 129 *K. pneumoniae* isolates in the BSAC bacteraemia collection.
Isolates are grouped by known fluoroquinolone resistance mutations as determined by Kleborate. MICs were determined by broth microdilution. As defined by the European Committee on Antimicrobial Susceptibility Testing (EUCAST), a breakpoint of ≥0.5 confers resistance[253].

Existing datasets of genome sequenced *Klebsiella* isolates frequently focus either on outbreaks or longitudinal surveillance where little metadata is available for these isolates[34,39]. Using a detailed sentinel site approach has the potential to provide a wealth of additional

clinical metadata that can be used to both study the burden and development of antimicrobial resistance, as well as the evaluation of novel antimicrobial therapies.
In this chapter I demonstrate the burden of *K. pneumoniae* bacteraemia in the context of other major pathogens at Cambridge University Hospitals, a large secondary care provider in the East of England. I will describe the key clinical, phenotypic and genotypic features of invasive *K. pneumoniae* from this hospital, using isolates collected over a four-year period that have been whole genome sequenced. I place these organisms into a contextual phylogeny with a global collection of *K. pneumoniae* and a historic collection of bacteraemia isolates from the same hospital. Finally, I use a single bacteraemia isolate to address the impact of ciprofloxacin on differential gene expression using RNAseq and demonstrate a range of transcriptomic changes that may influence antimicrobial resistance.

3.2 Burden of key pathogens at Cambridge University Hospitals

Cambridge University Hospitals National Health Service Foundation Trust (CUH) is a 1,160 bed teaching hospital in the United Kingdom. It provides a combination of routine secondary care services for the population of Cambridge and the surrounding area and tertiary referral services (including infectious diseases, oncology, haematology, transplantation, neurosurgery, paediatrics and neonatology) for the East of England. The on-site diagnostic microbiology laboratory forms part of the national Public Health England network of regional reference laboratories and provides bacteriology and virology diagnostic services for CUH.

The burden of key infectious diseases of clinical importance over a five year period from January 2016 to March 2021 is shown in **Figure 3.2**. Panel **a** demonstrates the incidence of bacteraemia episodes illustrated using the ESKAPEE group of pathogens, which represent some of the most pathogenic bacterial species associated with antimicrobial resistance. While there is a relatively consistent burden of disease throughout the calendar year, there is some stochastic and seasonal variation in incidence. For *E. coli*, bloodstream infections are more common in summer (mean 32.3 bacteraemias per month in Jun-Sep compared with 29.1 for the rest of the year, P=0.049, unpaired 2-tail t-test), while *Enterococcus faecium* are less common (mean 5.5 v 7.6 bacteraemias per month, P=0.014). Over the study period there was no significant difference in *K. pneumoniae* bloodstream infections by season (7.2 v 6.6 infections per month, P=0.19), or year (P=0.21, ordinary one-way ANOVA).

By contrast, there is marked seasonal variation in the frequency of common respiratory viral infections (panel **b**) with the considerable majority of cases occurring in the winter months. In addition, there is year-to-year variation in viral incidence, such as the rise in influenza B cases over the 2017-18 winter or the spike in respiratory syncytial virus in December 2019. The impact of influenza A and B on CUH, and factors that may impact the frequency of cases observed in **Figure 3.2b**, are described in detail in **Chapter 6**.

Panel **c** shows the frequency of laboratory confirmed SARS-CoV-2 cases at CUH over the first 15 months of the COVID-19 pandemic. This is described in detail in **Chapter 7**. It is worth noting here the considerable impact that the pandemic, and measures to control COVID-19, have had on reducing cases of other respiratory viral infection over the 2020-21 winter to negligible levels.

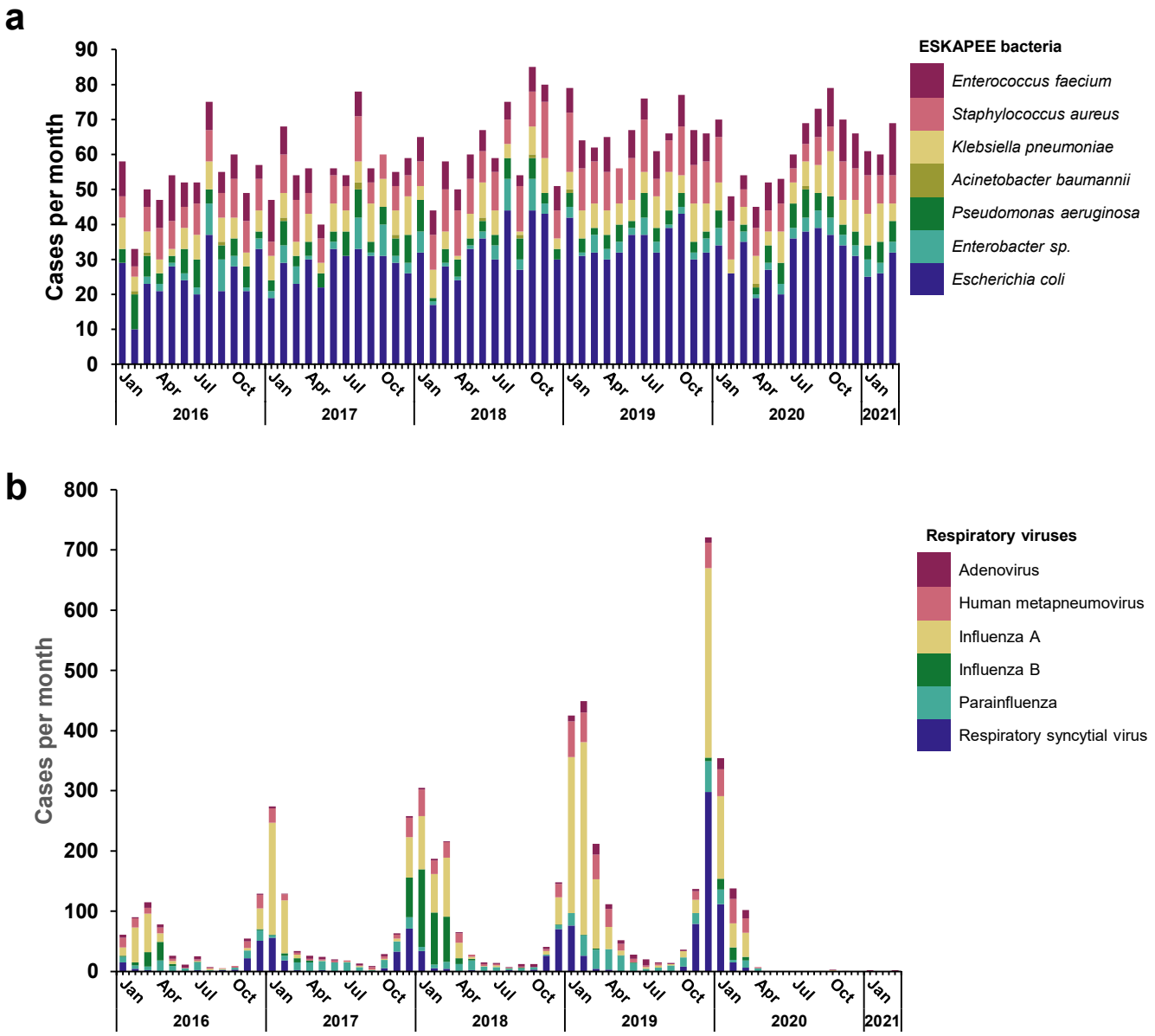


Figure 3.2 Burden of common infectious diseases at CUH, 2016-21

Monthly incidence of: **a**) confirmed bloodstream infections with ESKAPEE group of bacteria; **b**) confirmed respiratory virus infection; or **c**) confirmed SARS-CoV-2 infections. Bacteraemia episodes are defined as first positive blood culture with that organism over a 14 day period. Bacterial species were identified from positive blood culture specimens by MALDI-TOF. Respiratory viruses were diagnosed from nose and/or throat swabs using PCR or equivalent molecular point of care test. Each virus is counted once per season and only inpatient admissions are included. Note that the y axes are not comparable between plots to aid with visualisation.

3.3 *K. pneumoniae* bacteraemia in CUH, 2016-19

In total, there were 515 positive blood cultures for *Klebsiella* species in CUH between January 1 2016 and December 31 2019, reflecting 406 unique bacteraemia episodes (**Figure 3.3**). *K. pneumoniae* caused the largest number of infections (308/406, 75.9%), with *K. oxytoca* (80/406, 19.7%) and *K. aerogenes* (13/406, 3.2%) causing the majority of the remainder. The clinical features of *K. pneumoniae* bacteraemic patients, alongside antimicrobial susceptibility testing of the associated isolates, are shown in **Table 3.1**.

There is a relatively even distribution of cases between male and female patients, across a wide range of ages though disproportionately affecting older adults. Antimicrobial susceptibility testing was performed using disc diffusion at the diagnostic microbiology laboratory. Direct comparison of antimicrobial resistance between years is not possible, as the breakpoints changed during the study period when the laboratory moved from BSAC to EUCAST guidelines in May 2018[252]. For example, the breakpoint for ciprofloxacin resistance increased from 16 mm (BSAC) to 22 mm (EUCAST), causing a reported rise in resistance between 2016/17 and 2018/19 that may not represent a true change in susceptibility.

To investigate the change in resistance patterns over time, the original zone diameters measured using disc diffusion for six antimicrobials were obtained from the laboratory information management system (LIMS). A one-way ANOVA was conducted to associate the year of isolation to antimicrobial resistance (as determined by disc diffusion testing). The results are presented in **Table 3.2**.

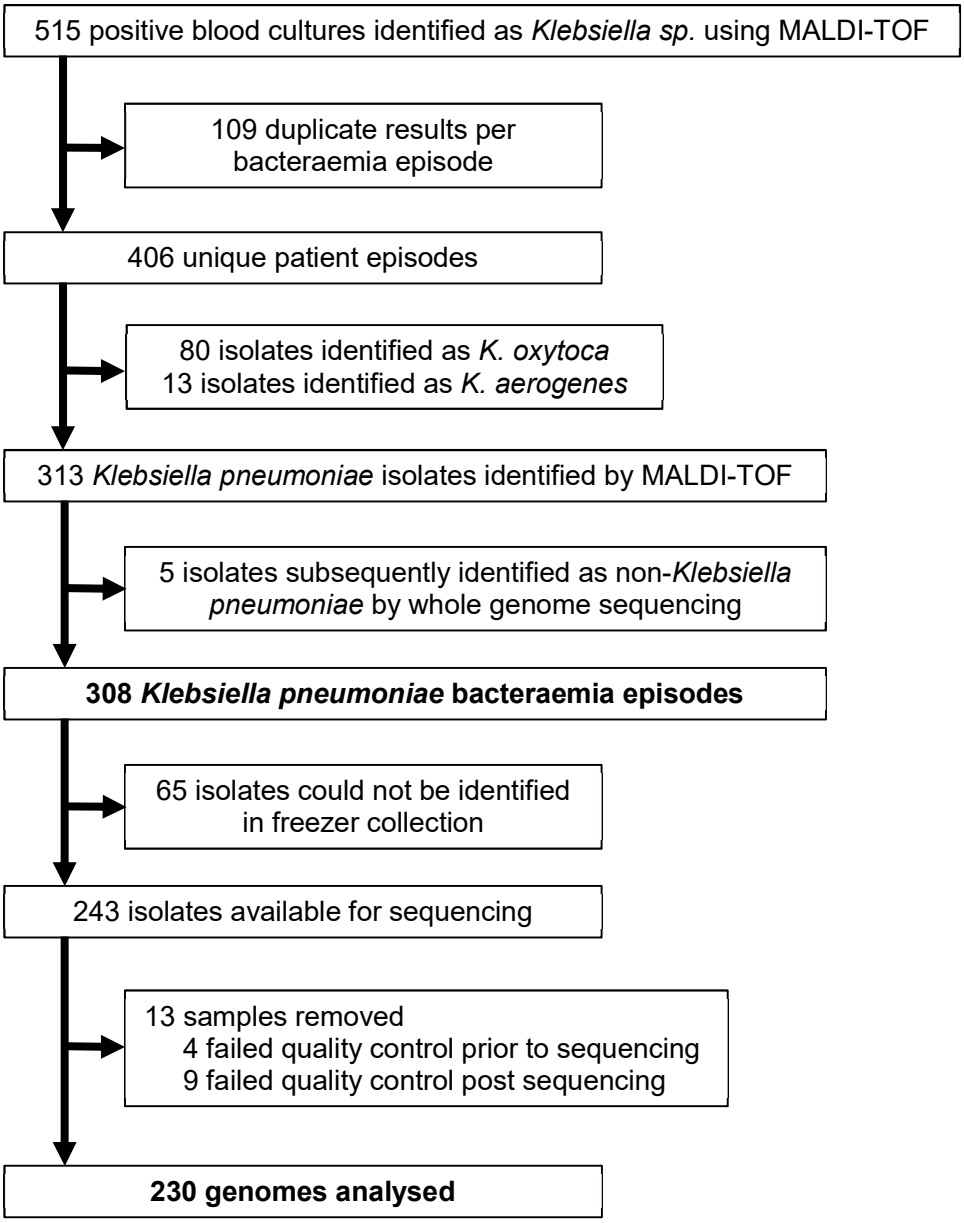


Figure 3.3 Case ascertainment of *K. pneumoniae* bacteraemia isolates from CUH, 2016-19.

Initial case ascertainment was determined using a comprehensive list of all microbiological samples available from the hospital laboratory information management system.

Table 3.1 – Patient and isolate characteristics of *K. pneumoniae* bacteraemia at Cambridge University Hospitals, 2016-2019

		All (n=308*)	Sequenced (n=230)
Sex	Female	143 (46.3%)	102 (44.3%)
	Male	166 (53.7%)	128 (55.7%)
Age	Mean (range)	63.1 (0-102)	63.6 (0-102)
	Median (IQR)	69 (52-80)	69 (50-81)
Year of isolation	2016	69	60 (87.0%)
	2017	86	76 (88.4%)
	2018	67	41 (61.2%)
	2019	86	53 (61.6%)
	Total	308	230 (74.7%)
Antimicrobial resistance**	Ampicillin	301 (97.7%)	224 (97.4%)
	Amoxicillin-clavulanate	88 (28.6%)	66 (28.7%)
	Piperacillin-tazobactam	39 (12.7%)	29 (12.6%)
	Cefpodoxime	38 (12.3%)	27 (11.7%)
	Cefotaxime	37 (12.0%)	26 (11.3%)
	Ceftazidime	41 (13.3%)	30 (13.0%)
	Ertapenem	8 (2.6%)	5 (2.2%)
	Meropenem	5 (1.6%)	3 (1.3%)
	Aztreonam	36 (11.7%)	26 (11.3%)
	Ciprofloxacin	53 (17.2%)	34 (14.8%)
	Gentamicin	32 (10.4%)	21 (9.1%)
	Amikacin	3 (1.0%)	1 (0.4%)

*Unique bacteraemic episodes with *K. pneumoniae* identified by MALDI-TOF, excluding 5 isolates identified as non-*K. pneumoniae* species by genome sequencing.

**Tested using disc diffusion with breakpoints defined by BSAC (Jan 2016-April 2018) or EUCAST (May 2018-Dec 2019).

There was a statistically significant difference in susceptibility across the time period as determined by one-way ANOVA for piperacillin-tazobactam, cefotaxime and gentamicin. A Tukey post hoc test for each of these three antimicrobials confirmed that susceptibility was

most significantly reduced between 2016 and 2019 for cefotaxime (*P*=0.003) and piperacillin-tazobactam (*P*<0.001), and between 2017 and 2019 for gentamicin (*P*=0.001).

Table 3.2 – Trends in antimicrobial resistance of *K. pneumoniae* bacteraemia at Cambridge University Hospitals, 2016-2019

Antimicrobial susceptibility determined by disc diffusion at the diagnostic laboratory. All values for the mean and standard deviation are in millimetres.

	2016		2017		2018		2019		One-way ANOVA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	*F* (3,304)	*p value*
Amoxicillin-clavulanate	20.5	6.4	20.8	5.1	21.2	3.9	20.4	4.0	0.40	0.75
Piperacillin-tazobactam	23.7	5.1	24.8	3.7	21.4	4.6	19.3	3.9	26.3	**<0.0001**
Cefotaxime	28.9	10.0	28.6	8.0	25.6	8.1	24.3	6.6	6.02	**0.0005**
Meropenem	30.6	3.7	30.4	2.6	30.6	3.9	29.5	3.1	2.14	0.095
Ciprofloxacin	22.4	7.8	23.2	6.5	23.3	7.7	24.6	6.9	1.27	0.28
Gentamicin	20.1	5.7	21.3	4.3	18.5	4.8	18.7	3.4	6.61	**0.0002**

SD = standard deviation.

3.4 Genomic insights into *K. pneumoniae* bacteraemia at CUH

To further characterise *K. pneumoniae* bacteraemia at CUH, whole genome sequencing was attempted on at least one isolate from each bacteraemia episode. Blood culture isolates were prospectively stored by the diagnostic laboratory and were made available for this study for sequencing. Isolates from January to November 2016 had already been sequenced as part of an investigation into transmission of multidrug resistant organisms in CUH[31]. The remainder were identified, stored, extracted and sequenced as described in **Chapter 2**. In

total, sequences from 230/313 (73.7%) bacteraemia episodes were available for further analysis (**Figure 3.3**). These 230 isolates were obtained from 217 patients; 11 patients had two separate bacteraemia episodes, while one patient had three episodes over the study period.

The presence of genes or mutations associated with antimicrobial resistance was identified using Kleborate[42]. Of the 189 isolates determined to be ciprofloxacin susceptible, 5 (2.6%) possessed the *qnrS* gene. No genetic determinants of resistance were identified in the 7 isolates with intermediate ciprofloxacin resistance and in 8/34 (23.5%) resistant isolates. The remaining ciprofloxacin resistant isolates had a combination of *parC* and *gyrA* mutations and *qnrB* genes. One resistant isolate also possessed *qepA*, while two possessed only *qnrS* and 7 possessed only *qnrB*. Only one carbapenemase-producing isolate was identified, which was an NDM-1 carrying isolate identified during an outbreak of ST78 at CUH in 2016[31].

The core genome, reflecting genes in ≥99% of all isolates in this collection, was determined using Roary[290]. An alignment of the 1,732 core genes identified was used to produce a maximum likelihood phylogenetic tree in IQTREE[291], shown in **Figure 3.4**. The majority of isolates (187/230, 81.3%) and the burden of resistance genes and phenotypes were from the subspecies *K. pneumoniae* sensu stricto. Of the remaining isolates, 28/230 (12.2%) were *Klebsiella variicola,* 8/230 (3.5%) *Klebsiella quasipneumoniae* subsp. *similipneumoniae* and 7/230 (3.0%) *Klebsiella quasipneumoniae* subsp. *quasipneumoniae*.

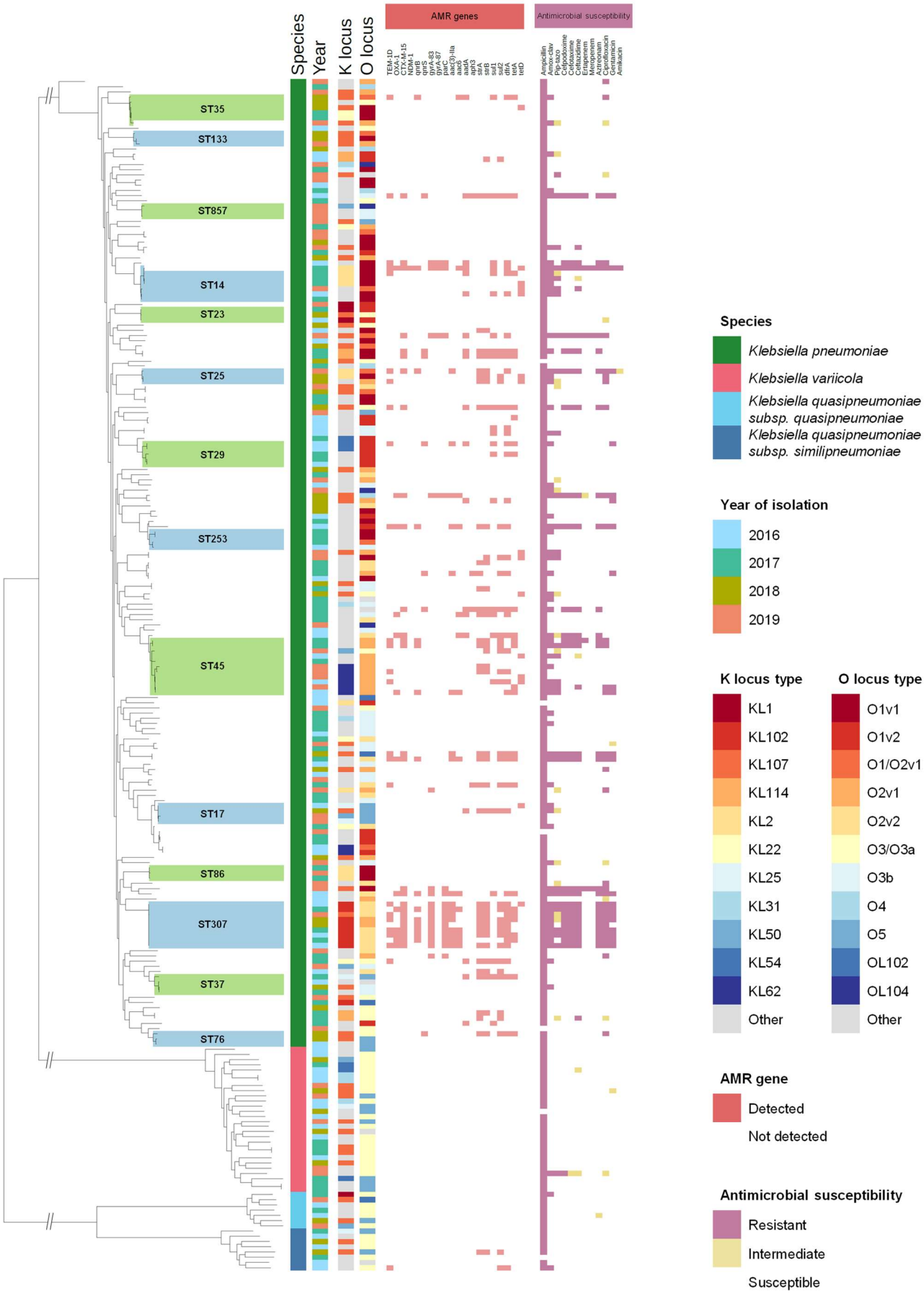


Figure 3.4 Phylogenetic tree of *K. pneumoniae* bacteraemia isolates from CUH, 2016-19.

Maximum likelihood phylogenetic tree produced by IQTREE, based on SNPs in the core genome determined by Roary. The tree is midpoint rooted. The branches to *K. pneumoniae*, *K. variicola* and *K. quasipneumoniae* are not to scale to facilitate visualisation. Species, MLST and AMR gene profile were determined using Kleborate. K and O locus types were determined using kaptive. Phenotypic susceptibilities were determined using disc diffusion. Common and/or clinically important sequence types represented by 3 or more isolates are highlighted in alternating colours and annotated.

Over the four-year study period there were a large number of sequence types identified at low frequency (**Figure 3.5a**). The highest proportion of isolates belonged to the ST45 group, but these account for fewer than 5% of all bloodstream infections, demonstrating that CUH has a genetically diverse range of *K. pneumoniae* causing invasive disease. Similarly, there was no single dominant O antigen type and although KL107 formed 20.9% of all capsule types, the remainder were present at frequencies of less than 5%.

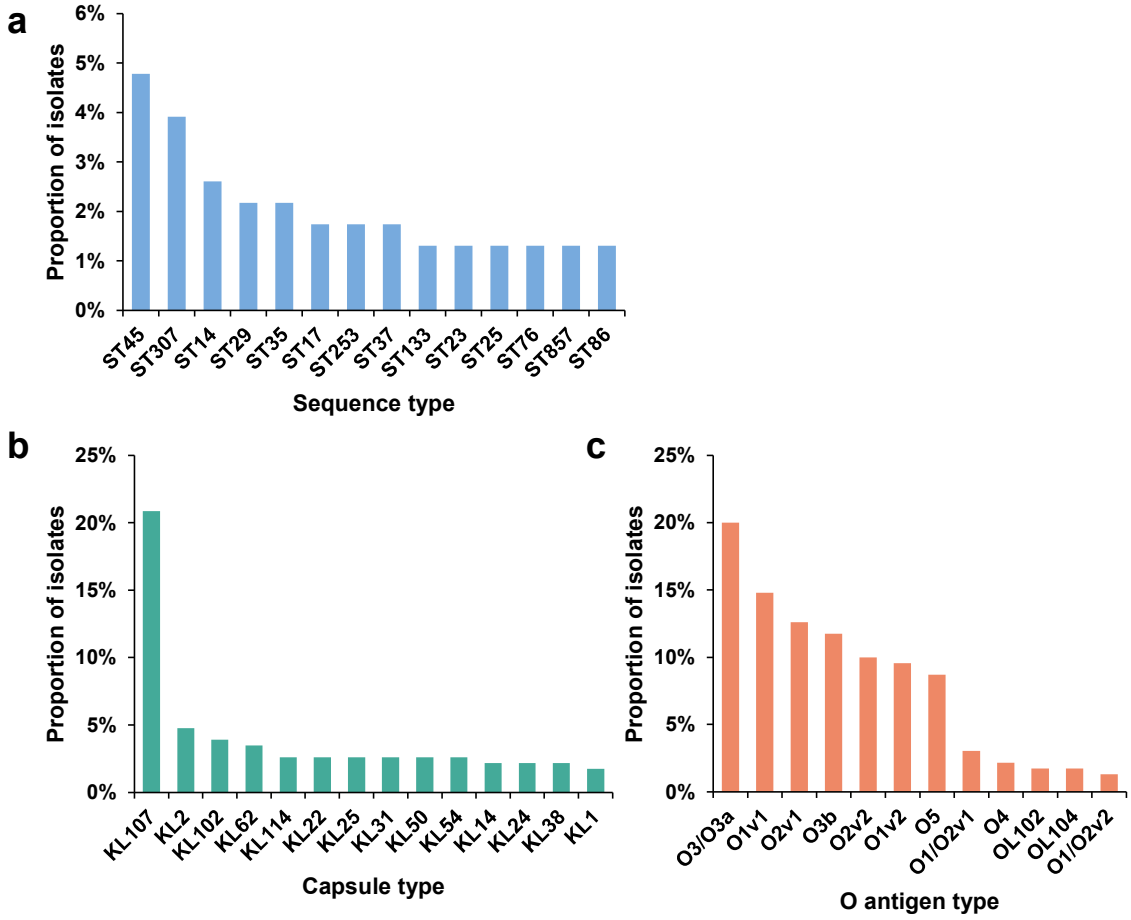


Figure 3.5 Common sequence types, capsule and O locus types in *K. pneumoniae* bacteraemia isolates from CUH, 2016-19.

Panel **a** shows the frequency distribution of the most common multilocus sequence types for isolates in this study, as determined by Kleborate. All STs with 3 or more isolates represented are included. Single locus variants of any sequence type are included with that sequence type. Panel **b** shows the distribution of capsular types, and panel **c** the distribution of O antigen types, as determined by Kaptive.

Figure 3.6 places the sequences in this study in the context of a global phylogeny of *K. pneumoniae* isolates published by Holt *et al.* in 2015 [34]. Only isolates sequenced at the Wellcome Sanger Institute from the global collection are included in this tree. This demonstrates that the proportion of each subspecies and the variety of sequence types isolated in CUH is broadly representative of the species as a whole, with the majority of major sequence types represented in both collections.

Figure 3.7 shows the sequences in this study in the context of a collection of 106 blood culture isolates from CUH collected between 2006 and 2012 and published by Ellington *et al.* [313]. The proportion of *K. pneumoniae* sensu stricto isolates formed a smaller number of isolates in the 2016-19 collection in comparison to the historic CUH collection (81.3% v 87.7%, P=0.14, chi^2 test). Additionally, the diversity of sequence types in CUH has increased between the two studies. While no single ST represented more than 5% of isolates in the 2016-19 collection, 35% of sequences in the previous CUH study could be attributed to two sequence types (ST874 and ST101), neither of which were identified in the more recent collection. The phenomenon of successive replacement of dominant sequence types in *K. pneumoniae* observed by Ellington *et al*. was not present over the 2016-19 study: not only was there a more diverse range of sequence types, those that were present at higher frequency were identified throughout the 4 year study period, indicating that no single clade has become dominant in the study population over this time.

To examine the relationship between isolates in the current study, the core genome alignment produced by Roary was used to produce a SNP difference matrix using the snp-dists package. This demonstrated a number of clusters of closely genomically related isolates.

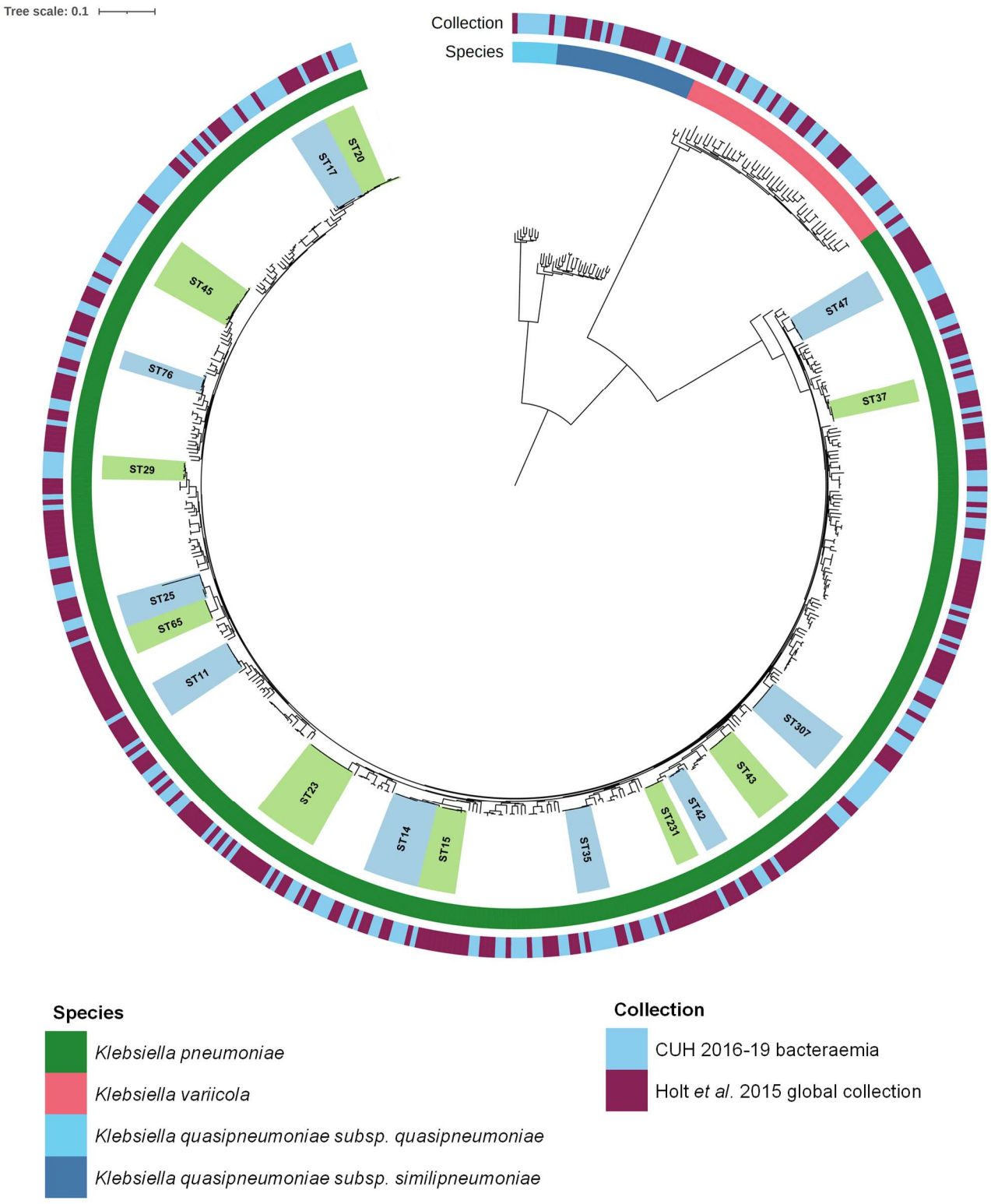


Figure 3.6 Phylogenetic tree of *K. pneumoniae* bacteraemia isolates from CUH, 2016-19, in context of a global *K. pneumoniae* collection.

Maximum likelihood phylogenetic tree produced by IQTREE, based on SNPs in the core genome determined by Roary (n=1,390 genes present in >99% isolates). The tree is midpoint rooted. Species and MLST were determined using Kleborate. Common and/or clinically important sequence types represented by 4 or more isolates are highlighted in alternating colours and annotated.

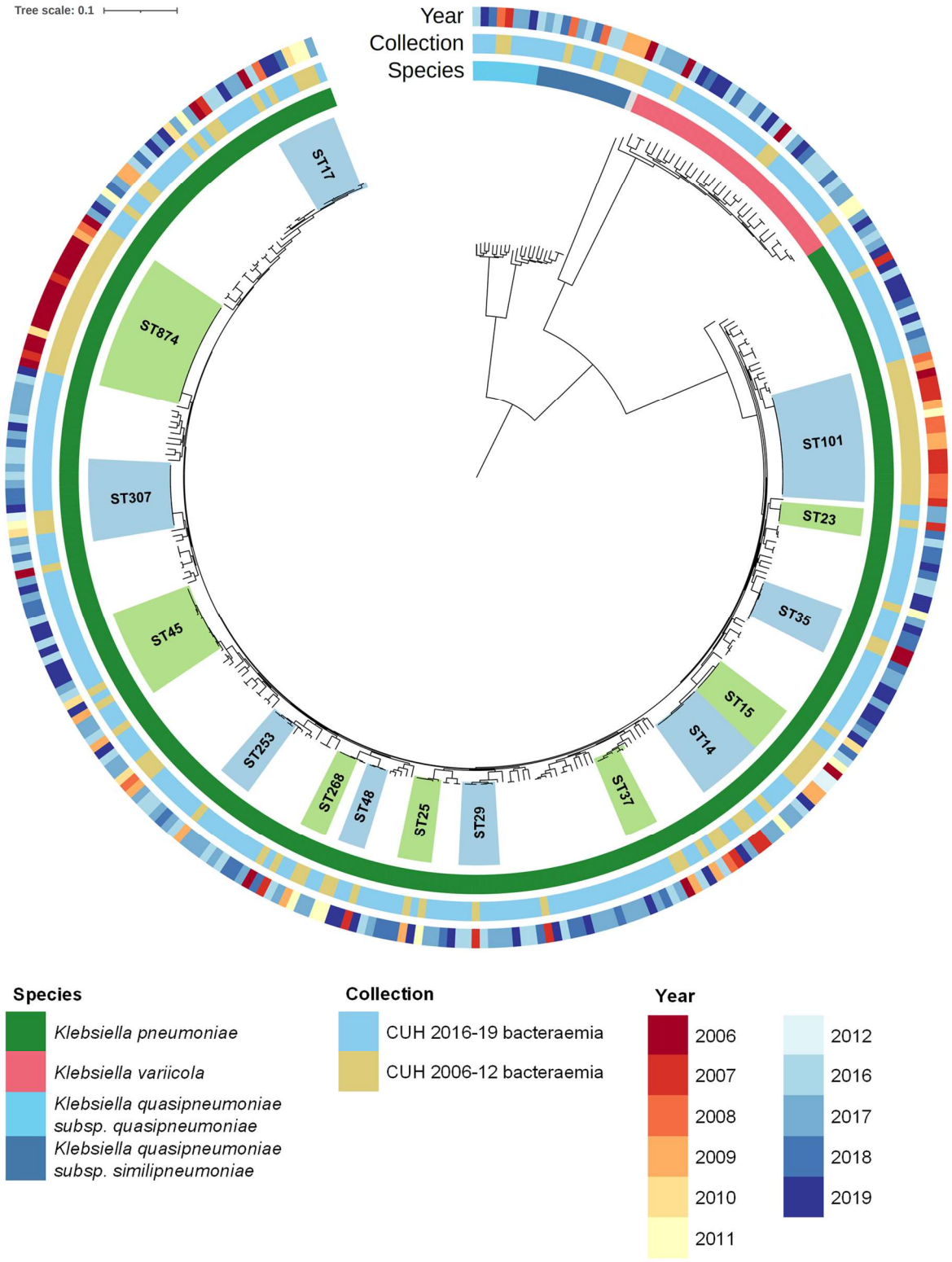


Figure 3.7 Phylogenetic tree of *K. pneumoniae* bacteraemia isolates from CUH, 2016-19 in context of isolates from 2006-12

Maximum likelihood phylogenetic tree produced by IQTREE, based on SNPs in the core genome determined by Roary (n=2,166 genes present in >99% isolates). The tree is midpoint rooted. Species and MLST were determined using Kleborate. Common and/or clinically important sequence types represented by 4 or more isolates are highlighted in alternating colours and annotated.

Of the 12 patients with recurrent bacteraemia during the study period, all of which were from the sub-species *K. pneumoniae* sensu stricto, 10 grew isolates with identical sequence types on two separate occasions. The median delay between the start of each bacteraemia episode for these 10 patients was 54.7 days (IQR 16.7-160.9, range 15.4-310.0). The core genome SNP difference between each of these 10 pairs of isolates was 0-11 SNPs in 8 cases. For the remaining 2 pairs of isolates, despite being of the same sequence type, there was a difference of 352 and 396 SNPs, isolated 69 days and 310 days apart respectively.

There were two clusters of genomically related sequences with more than two isolates. The first was a cluster of seven ST45 *K. pneumoniae* isolates, with a range of 0-30 SNPs between isolates collected across the four year study period. There was no obvious epidemiological association between these isolates. The second cluster was comprised of six ST307 isolates, range of 0-9 SNPs between isolates. Four of these isolates were hospital acquired. Of the two community acquired cases, one had recently been discharged from a prolonged hospital admission; the second was a resident in the same care home as one of the four hospital acquired infections (HAIs). Three further ST307 isolates had a minimum of 17-30 SNP differences with the larger cluster. Two of these were HAIs; the third was an immunocompromised patient with recurrent outpatient hospital contact and previous hospital admissions, all at CUH.

There were 5 further pairs of sequences with fewer than 20 core SNP differences, from separate patients. There was no obvious associations in the available demographic information for these patients.

3.5 Characterisation of blood culture isolates KC006 and KC036

To further investigate the effect of antimicrobials on clinical isolates of *K. pneumoniae*, a provisional phylogeny of bacteraemia isolates from the above CUH collection isolated in 2016 was analysed to identify two closely genomically related isolates with varying ciprofloxacin susceptibilities. Only one sequence type, ST45, contained isolates with disparate ciprofloxacin susceptibilities: KC006 (resistant) and KC019 and KC036 (susceptible). KC006 and KC019 differ by 35 SNPs in the core genome alignment used to produce Figure 3.4; KC006 and KC036 differ by 481 SNPs. By contrast, the mean difference between any two genomes across the 2016-19 CUH bacteraemia isolate collection was approximately 22,300 SNPs (range 1 to 65,149 SNPs). KC006 and KC019 share the same predicted capsule type (KL62), while KC036 has a different type (KL24). All three isolates share the same predicted O antigen type (O2v1). Given the divergent ciprofloxacin MICs despite similar genomic backgrounds, the isolates KC006 and KC036 were selected for ongoing experimental work.

KC006 was isolated from a blood culture specimen from an elderly male patient admitted to hospital with sepsis secondary to a presumed catheter-associated urinary tract infection. The patient died as a result of their infection. KC036 was isolated from an elderly male patient with recurrent aspiration pneumonia secondary to a previous stroke. The ciprofloxacin MIC of KC036 was 0.03 µg/mL determined by E-test. Growth of KC006 extended just beyond the highest concentration marker on the ciprofloxacin E-test (32 µg/mL); broth and agar dilution tests both confirmed an MIC of 64 µg/mL. MICs determined by Etest for a range of antimicrobials are listed in **Appendix C**.

Examining the resistance profiles of these two isolates using Kleborate, they both possessed *tet*, *shv* and *ampH* resistance genes. The KC006 genome additionally contained *aac3-IIa* (aminoglycoside resistance), *qnrB1* (fluoroquinolone resistance), *drfA14* (trimethoprim resistance), *catB4* (chloramphenicol resistance) and the OXA-1 beta-lactamase genes. The *gyrA* and *parC* genes in both isolates contained no known resistance mutations, nor the quinolone resistance gene *qepA*. KC006 had two circular plasmids: one 198 kbp plasmid (IncFIB(K)/IncFII(K)) encoding 210 predicted protein coding sequences including a number of known plasmid-mediated resistance genes; and one 3.9 kbp plasmid encoding 4 predicted protein coding sequences. KC036 had three circular plasmids: one 308 kbp plasmid (IncFIB(K)/IncFII(K)) encoding 319 predicted protein coding sequences, including a smaller repertoire of antimicrobial resistance genes; one 34.6 kbp plasmid encoding 45 predicted protein coding sequences including the virB virulence regulon transcriptional activator and a type IV secretion system; and one 3.4 kbp plasmid encoding 5 proteins, including a homologous protein to the S-type pyocin suggesting that this is in the Col plasmid family.

To determine the growth characteristics of KC006 and KC036, growth curves of each isolate were performed using an automated plate reader over a 24 hour period. This was performed in liquid media in the presence of varying concentrations of ciprofloxacin, related to the MIC of each isolate (**Figure 3.8a,b**). Time-kill assays were additionally performed to determine viable bacterial cell counts over the same time course and antimicrobial concentrations (**Figure 3.8c,d**). Increasing concentrations of ciprofloxacin were associated with a dose-dependent inhibition of growth and a reduction in the number of viable bacteria up to 4 hours for both isolates. After this point, there was a divergence in the pattern of growth seen. For KC036 there was complete recovery in growth and an increase in the number of viable bacteria for all ciprofloxacin concentrations, approaching untreated levels within 24 hours.

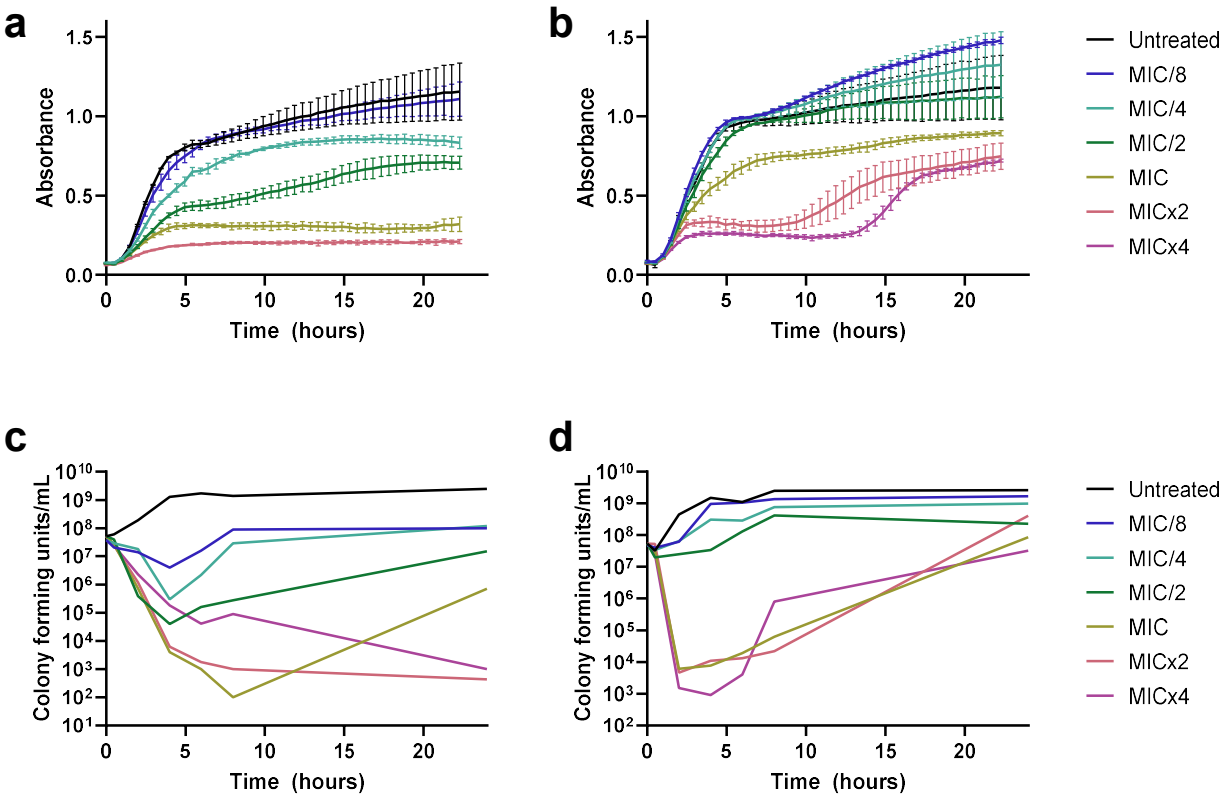


Figure 3.8 Growth characteristics of isolates KC006 and KC036

Growth curves for **a)** bacteraemia isolate KC006, and **b)** isolate KC036, performed in triplicate in a plate reader. Mean values and standard error plots are plotted for every 30 minutes of incubation. Panels **c)** and **d)** show the time kill assays for isolate KC006 and KC036 respectively. Bacteria were isolated after 0, 0.5, 2, 4, 6, 8 and 24 hours incubation. In all experiments bacteria were grown to early exponential growth phase before inoculation with ciprofloxacin at different concentrations based on their MIC.

For KC006, there was partial recovery of growth in the presence of sub-inhibitory concentrations (0.125 to 0.5x MIC) of ciprofloxacin, but which plateaued before reaching the level seen with untreated samples. At MIC, a similar effect was seen but bacterial growth was suppressed until at least 8 hours after inoculation. At higher concentrations of ciprofloxacin (2 to 4x MIC) there was a persistent reduction in viable bacteria, but at 24 hours there were still approximately 1,000 CFU/mL.

3.6 Differential gene expression of isolate KC006 in response to ciprofloxacin

To further investigate the effect of fluoroquinolones on clinical isolates of *K. pneumoniae*, RNAseq was used to identify genes that are significantly differentially expressed following exposure to ciprofloxacin. Given the effect of both antimicrobial concentration and length of incubation on bacterial growth demonstrated by the growth curves and time kill assays above, provisional experiments were conducted by sampling the isolate KC006 30 minutes, 2 hours, 4 hours and 8 hours after inoculation with ciprofloxacin at varying concentrations (untreated, 0.25x MIC, 1x MIC and 4x MIC). Reads were mapped to an assembly of KC006 produced using PacBio sequencing. Further RNAseq experiments involving isolate KC036 are described in **Chapter 5**.

The total number of genes significantly up- and down-regulated for each concentration of ciprofloxacin in comparison to an untreated control are shown in **Figure 3.9a**. While there was modest transcriptional change at sub-inhibitory concentrations of ciprofloxacin, at MIC and 4x MIC there was a linear rise in differentially expressed genes, highest at 4 hours by which time over 1000 genes were affected. By 8 hours the differential expression remained elevated for 4x MIC in comparison to the untreated control, but began to fall again at MIC of ciprofloxacin. Using the results obtained at MIC as an example, **Figure 3.9b** shows a heatmap of differentially expressed genes at 4 different time points. Genes are ordered based on the $\log_2$ fold change observed at 2 hours. In comparison to this reference, a smaller number of the same genes are also differentially expressed at 30 minutes, but the majority remain persistently up- or down-regulated at 4 and 8 hours.

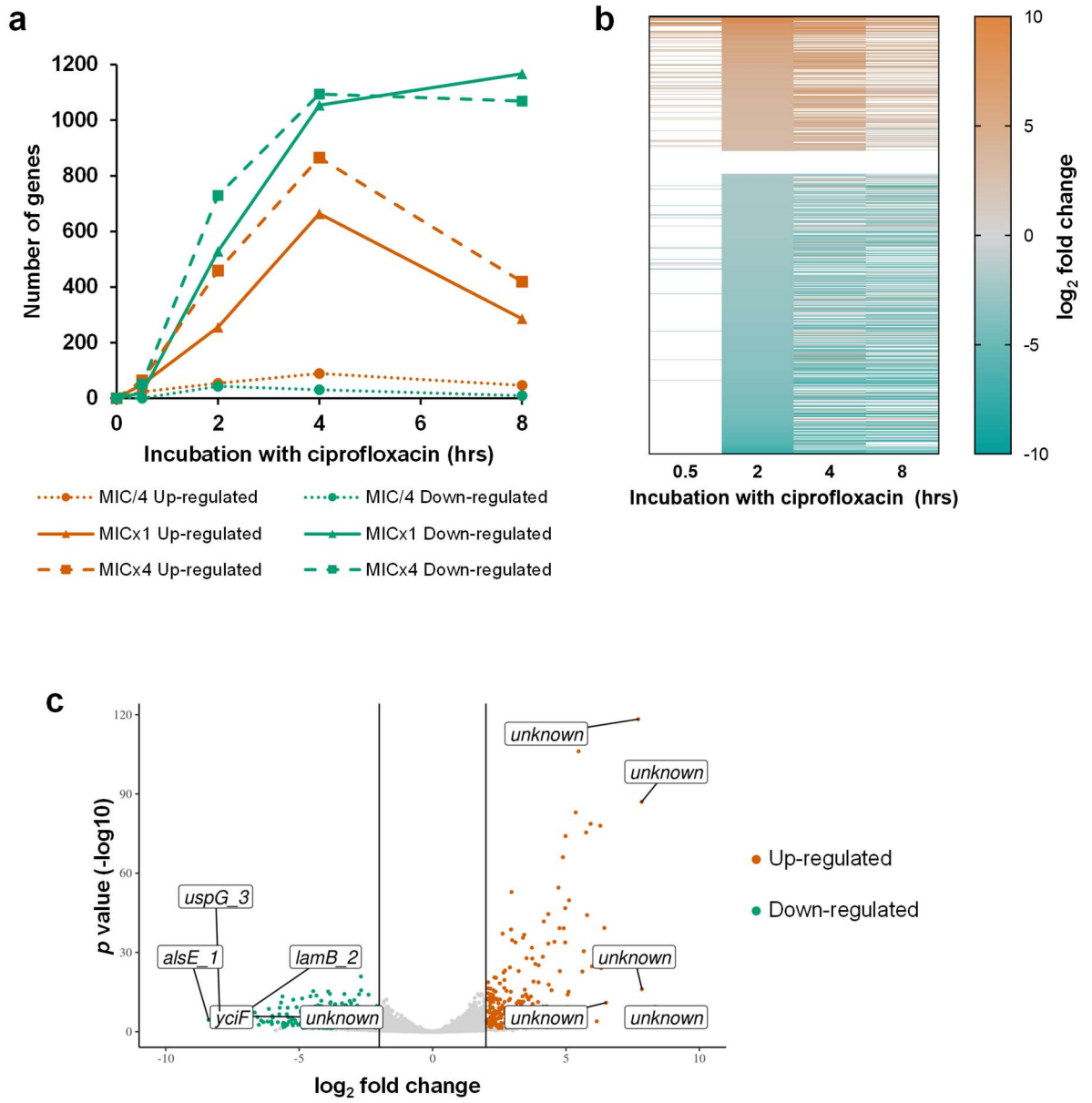


Figure 3.9 Differential gene expression of isolate KC006 in the presence of ciprofloxacin

a) The overall transcriptional changes in *K. pneumoniae* isolate KC006 over time after incubation with varying concentrations of ciprofloxacin. The isolate was in early exponential growth phase at the time of antibiotic inoculation. The number of genes significantly up-or down-regulated is defined as a minimum 2x $\log_2$ fold change in gene expression, with an adjusted *p* value <0.05. **b)** Heat map demonstrating significantly differentially expressed genes at 2 hrs in the presence of ciprofloxacin at MIC relative to an untreated control, and how their expression changes over time. Genes that are not significantly differentially expressed (adjusted *P* <0.05) are coloured white. **c)** Volcano plot comparing ciprofloxacin at MIC with an untreated control at 2 hours incubation. Points are coloured where the adjusted *p* value <0.05 and $\log_2$ fold change in expression is ≥2 or ≤-2. The 5 genes with the greatest $\log_2$ fold change in expression are annotated. Where genes are labelled as "unknown", they encode hypothetical proteins of unknown function.

Further differential expression analysis was conducted at a single time point and ciprofloxacin concentration: 2 hours at MIC. This was to facilitate comparison with planned transcriptomic studies in the same isolate in the presence of different antimicrobials. This is supported by my observation that growth curves and imaging experiments (shown in **Chapters 4 and 5**) have demonstrated impaired cell growth and/or death in the presence of beta lactam antimicrobials after 2 hours of incubation or at concentrations greater than MIC, limiting both transcriptomic and phenotypic studies. Given the limited transcriptomic data at sub-inhibitory concentrations and in an attempt to maintain uniform methods wherever possible, I decided to focus on the 2 hour time point and 1x MIC.

A volcano plot of the effect of ciprofloxacin at MIC after two hours of incubation in comparison to an untreated control is shown in **Figure 3.9c**. A large number of genes were significantly differentially expressed, including 255 up-regulated and 528 down-regulated genes with a minimum fourfold change in expression. The 60 coding sequences with the largest increase in expression at 2 hours at MIC are shown in **Table 3.3**.

Table 3.3 – Differentially expressed genes of KC006 exposed to ciprofloxacin at MIC after 2 hours incubation

Results are ordered by log_2 fold change in expression. Only upregulated transcripts encoding predicted coding regions are included. Genes encoded on a plasmid are highlighted in red. Gene product details are derived from annotations from Prokka and, for genes of unknown function, Uniprot.

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
05199	8.34	$3.8x10^{-10}$	unknown	hypothetical protein
05213	7.85	$7.93x10^{-17}$	unknown	hypothetical protein
01153	7.84	$1.04x10^{-87}$	unknown	toxic peptide TisB
05198	6.50	$1.05x10^{-11}$	unknown	hypothetical protein
05200	6.48	$1.64x10^{-31}$	unknown	putative dopa decarboxylase protein remnant
05195	6.44	$5.28x10^{-40}$	unknown	hypothetical protein
05214	6.32	$9.15x10^{-25}$	unknown	domain of unknown function (DUF932)

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01152	6.29	$1.07x10^{-78}$	unknown	putative acyltransferase (GNAT) family
04905	6.15	0.000119	unknown	hypothetical protein
00718	5.97	$1.83x10^{-25}$	unknown	alkylphosphonate utilization operon protein PhnA
05104	5.92	$2.04x10^{-79}$	*dinI_2*	DNA-damage-inducible protein I
05105	5.79	$7.29x10^{-45}$	*umuD_1*	error-prone repair protein UmuD, DNA polymerase V subunit
05106	5.75	$3.71x10^{-76}$	*umuC_1*	error-prone repair protein UmuC, DNA polymerase V subunit
05212	5.62	$1.65x10^{-23}$	unknown	domain of unknown function (DUF3560)
02336	5.47	$7.3x10^{-107}$	*recN*	DNA repair protein RecN (Recombination protein N)
05222	5.36	$9.8x10^{-84}$	*umuC_2*	error-prone repair protein UmuC, DNA polymerase V subunit
05209	5.27	$3.43x10^{-8}$	*ykfF*	hypothetical protein
05206	5.21	$2.69x10^{-8}$	unknown	PsiA protein, plasmid SOS inhibition protein
05210	5.10	$7.34x10^{-16}$	*ssb_2*	single-stranded DNA-binding protein
00315	5.07	$5.19x10^{-10}$	*nhaA*	sodium/proton antiporter nhaA
04909	5.02	0.000262	unknown	hypothetical protein
04860	4.98	$8.68x10^{-75}$	*sulA*	SOS-response cell division inhibitor SulA
00574	4.97	$1.69x10^{-47}$	*nrdD*	anaerobic ribonucleoside-triphosphate reductase
05221	4.92	$6.84x10^{-40}$	*umuD_2*	error-prone repair protein UmuD, DNA polymerase V subunit
03927	4.89	$8.35x10^{-67}$	*yebE*	inner membrane protein YebE
04906	4.85	$3.46x10^{-5}$	unknown	hypothetical protein
00575	4.76	$6.26x10^{-40}$	*pflA_1*	anaerobic ribonucleoside-triphosphate reductase activating protein
03928	4.72	$2.91x10^{-55}$	unknown	DNA damage-inducible protein YebG (LexA family transcriptional regulator)
00515	4.69	0.009717	unknown	hypothetical protein
01238	4.46	$1.95x10^{-16}$	*yicG*	predicted membrane protein, UPF0126 domain
05208	4.35	$1.53x10^{-18}$	unknown	DNA-binding protein, ParB/RepB/Spo0J family partition protein
05159	4.35	0.018844	unknown	hypothetical protein
04992	4.34	$4.4x10^{-34}$	*dinI_1*	DNA-damage-inducible protein I
02257	4.34	$3.14x10^{-45}$	*recA*	recombinase A, bacterial DNA recombination protein
00533	4.29	0.023846	*ycjP_1*	glycerol-3-phosphate transporter membrane protein
04907	4.28	$2.43x10^{-10}$	unknown	exodeoxyribonuclease VIII
04902	4.18	0.00111	unknown	hypothetical protein
04908	4.14	$3.18x10^{-10}$	*dam_2*	adenine-specific DNA methyltransferase
02258	4.13	$4.2x10^{-29}$	*recX*	recombination regulatory protein RecX
00793	3.97	$7.06x10^{-26}$	*lexA*	SOS-response repressor and protease LexA
00576	3.97	$1.97x10^{-19}$	*relE*	RelE/StbE plasmid replicon stabilization toxin,mRNA interferase RelE,
05192	3.97	0.022623	unknown	conjugal transfer transcriptional regulator TraJ
01237	3.93	$3.56x10^{-9}$	*ligB_3*	NAD-dependent DNA ligase LigB
01137	3.87	$1.81x10^{-6}$	*ibpB*	16 kDa heat shock protein B
04321	3.87	0.001591	*cadA*	lysine decarboxylase
05107	3.85	$2.2x10^{-26}$	*dmlR_29*	LysR family transcriptional regulator
04903	3.74	0.000994	unknown	hypothetical protein
01248	3.74	$4.66x10^{-8}$	*pyrE*	orotate phosphoribosyltransferase
00792	3.74	$2.02x10^{-28}$	*dinF*	DNA-damage-inducible SOS response protein, MATE efflux family protein
01048	3.72	$1.58x10^{-32}$	*aes*	acetyl esterase
05109	3.68	$6.46x10^{-7}$	*ydeA*	putative arabinose (sugar) efflux transporter
01418	3.68	$4.52x10^{-10}$	*livK*	high-affinity leucine-specific transport system
01244	3.66	$6.16x10^{-5}$	unknown	HlyD family efflux pump
05108	3.60	$6.43x10^{-13}$	*butA*	diacetyl reductase ,short chain dehydrogenase
01874	3.59	$7.15x10^{-8}$	*endA*	DNA-specific endonuclease I
05193	3.56	$3.7x10^{-8}$	*traM*	conjugal transfer protein TraM
05207	3.55	$6.32x10^{-6}$	*psiB*	plasmid SOS inhibition protein B PsiB
05223	3.54	$5.8x10^{-14}$	unknown	RelB/StbD, putative antitoxin of the YafO-YafN toxin-antitoxin system
04904	3.52	0.000844	unknown	hypothetical protein
01243	3.47	$5.58x10^{-9}$	unknown	putative acetyltransferase

As anticipated from a drug with a mechanism of action that causes double stranded DNA breaks, ciprofloxacin induced upregulation of genes associated with DNA damage and repair, including *dinI*, *recN* and *yebG*. However, there were also a number of differentially expressed genes that may be related to antimicrobial resistance. Some of the most substantially upregulated were *umuC* and *umuD*, which encode the low-fidelity DNA polymerase V, and *ssb*, which encodes a single stranded DNA-binding protein which is required for the polymerase to function. There were two different copies of these three genes in the KC006 genome: one chromosome- and the other plasmid-encoded. All 6 genes were highly upregulated in the presence of ciprofloxacin and may therefore contribute to higher rates of mutagenesis. Additionally, *lexA*, the main repressor of *umuC* and *umuD* was also significantly upregulated, as were the other LexA-repressed genes *polB* and *dinB*.

Other upregulated genes included *dinF*, which encodes a DNA-damage inducible efflux pump implicated in resistance, and an analogue of the putative efflux pump gene *hlyD*, which itself has similarities to the AcrAB efflux pump which confers resistance to fluoroquinolones. One of the most substantially downregulated genes was *lamB*, a porin whose loss has been associated with antimicrobial resistance. A number of genes associated with plasmid replication and transfer, and therefore implicated in the spread of plasmid-mediated resistance, were differentially expressed in the presence of ciprofloxacin. These included significant upregulation of *traM* and *traJ*, key proteins in plasmid replication and conjugation, and downregulation of the conjugation suppressor *traR*. Of note, the gene *traM* was one of the small number of genes significantly upregulated at 2 hours in the presence of sub-inhibitory concentrations of ciprofloxacin (0.25x MIC).

A region of the larger KC006 plasmid encoded a number of known antimicrobial resistance genes and which were upregulated (albeit it to a lesser extent in comparison to the genes above) in the presence of ciprofloxacin. These included genes encoding resistance mechanisms for tetracycline (*tetA*, $\log_2$ increase 1.27, P=0.01), chloramphenicol (*cat*, $\log_2$ increase 2.05, P=2.6x10^{-5}), aminoglycosides (*aacA-aphD*, $\log_2$ increase 1.37, P=6.6x10^{-5}) and beta lactams (*bla*, $\log_2$ increase 2.23, P=2.3x10^{-6}). The ciprofloxacin gene *qnrS1* is also encoded in this region, and showed a 2-fold increase in expression (P=0.00078). Other, non-plasmid encoded genes associated with ciprofloxacin resistance include the AcrAB efflux pump: the two variants of *acrA* in the KC006 genome were not significantly upregulated; of the five variants of *acrB*, only one had a significant change in expression at two hours, being downregulated ($\log_2$ decrease 1.31, P=0.027); the acrAB transcription repressor, *acrR*, was also downregulated ($\log_2$ decrease 1.31, P=5.0x10^{-5}).

A substantial proportion of differentially expressed reads map to genes of unknown function. Of the down-regulated genes, 85/528 (16.1%) were predicted to be non-protein coding, while 14/528 (2.7%) were annotated as being hypothetical proteins by Prokka. Similarly, of the up-regulated genes, 61/255 (23.9%) were predicted as non-protein coding, while 22/255 (8.6%) were identified as hypothetical proteins. Of the 10 most up-regulated genes, 8 had no confirmed function from their initial annotation; 6 of these were plasmid-encoded. **Figure 3.10** shows a linear map of the 198 kbp plasmid in the KC006 genome. It demonstrates the number of hypothetical proteins of unknown function and intergenic regions that were differentially expressed in the presence of ciprofloxacin, as well as their co-location with genes known to be involved with antimicrobial resistance and plasmid conjugation. Using the predicted amino acid sequences for these hypothetical proteins, the UniProt database revealed some limited information about a small number genes which is included in **Table 3.3**. One predicted protein had sequence homology to PsiA, which inhibits plasmid SOS

responses. Another was predicted to belong to the ParB/RepB/Spo0J family of plasmid partition proteins. Two predicted proteins had recognised domains of unknown function (DUF): one was a putative dopa decarboxylase while another was a putative antitoxin of the YafO-YafN toxin-antitoxin system. Six plasmid-encoded hypothetical genes had no further annotation in UniProt.

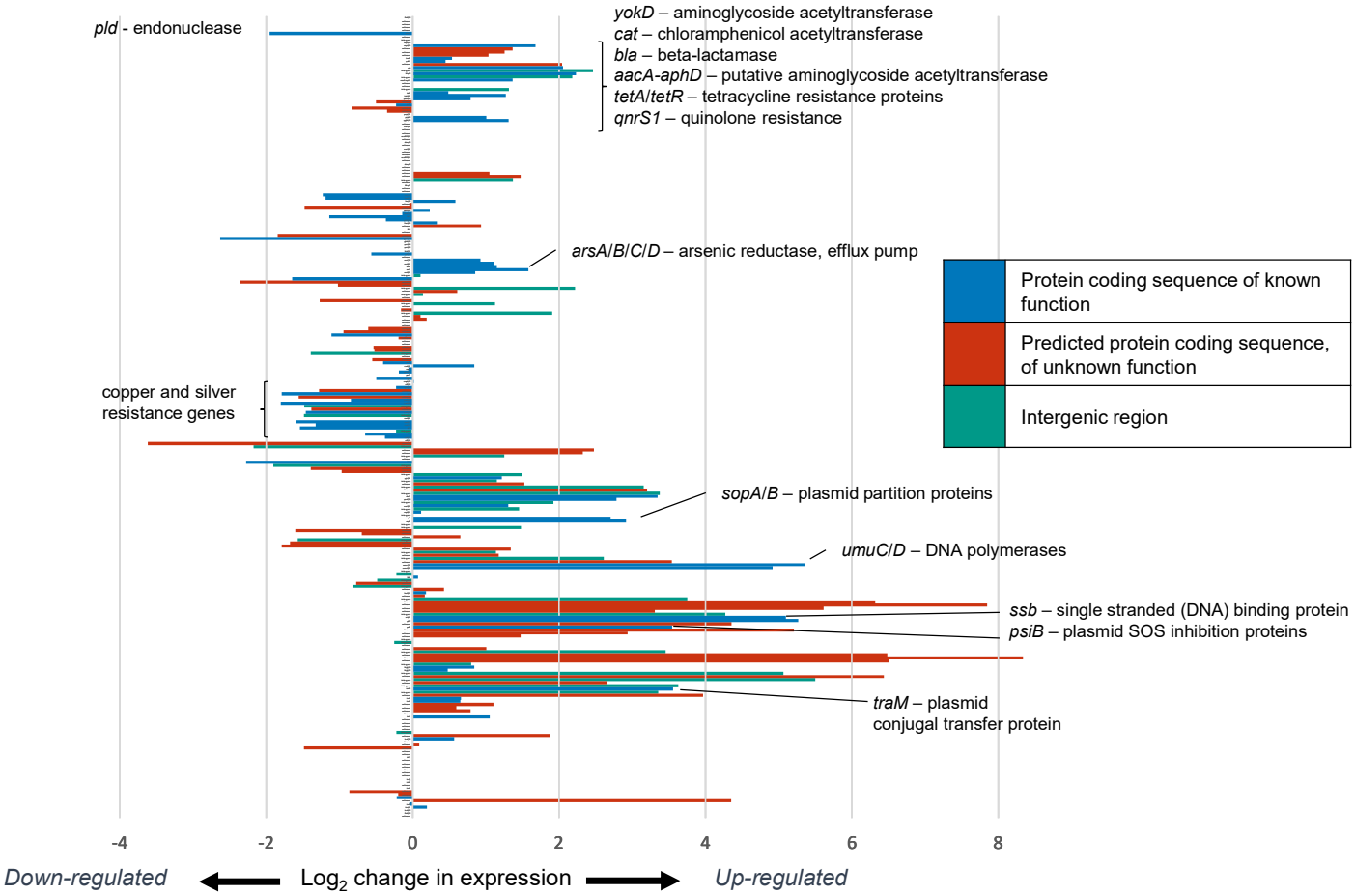


Figure 3.10 Linear map of the 198 kbp plasmid in isolate KC006, with associated differential gene expression results in the presence of ciprofloxacin

Differential gene expression results compare ciprofloxacin treatment at MIC with an untreated control after 2 hours of incubation. The length of the coloured bars represent the log_2 fold change in expression, with no limits on significance. Blue bars represent genes encoding proteins of known function; red bars represent genes encoding hypothetical proteins of unknown function; green bars represent intergenic areas. Note that the y axis represents position relative to other genes and is not to scale. Genes of interest are annotated.

3.7 Identifying the role of the KC006 plasmid in resistance

The results of my RNAseq analysis showed that the 198 kbp plasmid identified in isolate KC006 was likely to be important in driving its antimicrobial resistant phenotype, coding both known and putative resistance loci. To evaluate its role further, two attempts were made to remove the plasmid from KC006 and to evaluate any impact on this phenotype.

First, attempts were made to cure the plasmid using electroporation. In brief, KC006 was repeatedly pelleted and re-suspended in ice-cold glycerol before being transferred to SOC media. Three samples were electroporated for 5.4 ms at 1.8 kV, with each sample receiving one, two or three pulses. Samples were streaked on LB agar and the resultant colonies streaked on both normal LB and gentamicin infused plates. Gentamicin was used as an antimicrobial for counter-selection of resistance because aminoglycosides are, with the exception of fluoroquinolones, the only antimicrobial class with resistance genes identified on the plasmid but not the chromosome. However, all of the ~140 colonies identified grew on the gentamicin plate, indicating that the isolate had not been cured of the plasmid.

Second, plasmid conjugation was attempted using sodium azide counter-selection. While the majority of plasmid conjugation experiments rely on selection of isolates using antimicrobials, isolates that are resistant to a large number of antimicrobial classes (such as KC006) necessitate the use of alternative selection techniques. Sodium azide is a toxin that can be used to select for antimicrobial resistance plasmids. Resistance to azide is uncommon in most bacteria, but some isolates (such as *E. coli* J53) have developed resistance due to mutations in the *secA* gene[314]. Because these isolates are not resistant to other antimicrobials, they can be used as recipients in plasmid conjugation experiments.

Multiple attempts were made to transconjugate the plasmid from KC006 into the *E. coli* recipient J53, as detailed in **Chapter 2.1.9**. Different growth conditions were used on liquid and solid media at both 30°C and 37°C, and counter-selection using sodium azide and either ampicillin or tetracycline. In the most successful of these experiments, in which donor and recipient were combined in liquid culture, 4 colonies derived from 30°C incubation and a further 4 from 37°C incubation grew on combined ampicillin and azide infused plates. However, of these 8 colonies, 1 was confirmed to be *E. coli* and 7 to be *Klebsiella pneumoniae* using MALDI-TOF. Subsequent evaluation of the single *E. coli* colony using E-tests showed that it had a susceptibility profile that was highly similar to J53, including ampicillin (MIC = 2ug/mL) despite initial cross-selection on a plate on which the untreated J53 negative control did not grow. The reason for the growth of *K. pneumoniae* colonies on these plates is unclear, but may reflect *de novo* mutations in the *secA* gene, or alternative mechanisms of azide resistance.

Further work on plasmid curing, conjugation and quantification of phenotypic effects of ciprofloxacin on KC006 was impacted by the COVID-19 pandemic. Work that was planned but not possible to contribute to this thesis is discussed below.

3.8 Discussion

In this chapter I have demonstrated the considerable burden of *K. pneumoniae* bacteraemia in a single hospital over a four year period, re-iterating its importance as a clinically important pathogen in secondary care. I have used a range of bioinformatics tools to characterise these isolates and produced a phylogeny demonstrating the marked genetic diversity of

sequences. The wide range of sequence types broadly reflects that seen globally and demonstrates a relative increase in diversity since a comparable historic study of bacteraemia isolates in the same hospital. Using this collection I have identified a single isolate for further detailed characterisation of the effects of fluoroquinolones on transcription, and have demonstrated that a number of genes associated with antimicrobial resistance were significantly differentially expressed. These included genes encoding a range of resistance mechanisms including efflux, porin loss, mutagenesis, enzymatic alteration of antimicrobials and increased plasmid mobilisation and conjugation. Additionally, I have identified a number of genes encoding hypothetical proteins, which were up-regulated under ciprofloxacin pressure and were present on a plasmid in close proximity to known resistance determinants. These genes may therefore play a role in bacterial virulence and/or antimicrobial resistance but have not previously been described.

The clinical characteristics of *Klebsiella* bacteraemia in this study are comparable to contemporaneous national datasets. The proportion of all *Klebsiella* bacteraemia caused by *K. pneumoniae* is similar (75.9% in this study compared to 73% across England, Wales and Northern Ireland), as is the disposition for older individuals[65]. Although they constitute the minority of *Klebsiella* bacteraemias, *K. oxytoca* has an increasingly recognised capability to acquire resistance and virulence factor genes from *K. pneumoniae*, with the potential to become an opportunistic or hospital associated pathogen warranting further surveillance and investigation[33].

Rates of antimicrobial resistance are comparable to contemporaneous national bacteraemia datasets from England, Wales and Northern Ireland for 2018, which demonstrate 31% amoxicillin-clavulanate resistance, 14% piperacillin-tazobactam resistance and 13% ceftazidime resistance, but low levels of carbapenem resistance (<2%), similar to this

study[66]. The rise in resistance to beta lactams observed here also reflects changes at a national level over the same period[66].

The clinical features of *Klebsiella* bacteraemia, including the demographic features of affected patients and their outcomes, are generally poorly described in the literature in comparison to the detailed genomic characterisation now routinely widely available using bioinformatics tools such as Kleborate and Kaptive. One of the original intentions of this PhD was to produce detailed extracts of routinely collected clinical and demographic data from patients with bacteraemia from the electronic health record system at CUH. In principle, the advent of comprehensive electronic documentation of all clinical data provides an opportunity to extract a wealth of information that can be paired with other datasets, such as whole genome sequencing, for detailed characterisation of infectious diseases. However, the process of defining the required variables and constructing these extracts is not straightforward and relies on individuals with expertise in both clinical data handling and the underlying structure of the Epic health record. Following initial discussions with the CUH data management team to assist in the creation of this database at the end of 2019, all work on the project was re-directed towards producing extracts of patients admitted with COVID-19 at the start of 2020 to assist in local and national efforts to combat the pandemic. That work is described in **Chapter 7**. An application has been made to adapt the COVID-19 data extracts for bacteraemia patients; while these may provide a great insight into invasive bacterial diseases, including this dataset, unfortunately the data were not available in time to contribute to this thesis.

There are a number of strengths to the genomic analyses in this study. First, prospective storage of bacteraemia isolates in the diagnostic laboratory that were made available to this

study have ensured a high case ascertainment per bacteraemia episode. Secondly, this study was able to take advantage of genomic investigations of Gram-negative bacteraemia at CUH, as well as a global study of *K. pneumoniae*, to provide both local and national context to these sequences[34,313]. Both studies support my finding that the genetic diversity of isolates at CUH is both broad and appears to be diversifying over time, reflected by a wide range of sequence types and major antigenic variants. This is important when considering the development of novel antimicrobial therapies which target these antigens including vaccination, monoclonal antibodies and phage therapies[43,315-317]. Physical isolate collections that have been genetically and phenotypically characterised are likely to be important in determining the efficacy of novel antimicrobials across the *Klebsiella* species *in vitro* prior to clinical trials. The addition of detailed clinical metadata on patients with invasive *Klebsiella* infections will enable better understanding of when and how these antimicrobials may be used as part of clinical management, and how additional clinical diagnostics may need to be developed to guide the use of narrow spectrum therapeutics such as phage and monoclonal antibodies.

Despite the genetic diversity of isolates in this study, I have identified two clusters of closely related sequences that may form part of a transmission network within CUH. This is concerning in a study focused on bacteraemia isolates, that likely represent a small fraction of transmission leading to clinical infection or colonisation with pathogenic strains of *K. pneumoniae*. The largest of these clusters, from a ST307 clade, are among the most phenotypically resistant in the study. ST307 isolates are strongly associated with a conserved plasmid possessing the bla-CTX-M-15 ESBL gene[318], possess a number of other genetic factors that aid its survival in the hospital environment, including fimbriae and siderophores[319], and have been associated with a number of hospital-associated outbreaks[320-322]. This study also demonstrates potential transmission between residents in

the same care home; such transmission of ST307 has previously been described in a rehabilitation centre[323], and in other *K. pneumoniae* strains between hospital and long term care facilities[324]. CUH itself has previously experienced outbreaks of *K.* pneumoniae; a surveillance study of the prevalence of multidrug resistant organisms in CUH in 2016 identified two outbreaks of the same ST78 clone five months apart[31]. However, only one of these patients developed invasive disease. The finding of a hospital lineage that has caused multiple hospital-associated invasive infections requires further investigation and strengthens the case for routine genomic surveillance of high impact pathogens to better understand transmission dynamics.

The RNAseq data presented here has identified a range of resistance mechanisms with a significant change in gene expression. These include known resistance genes, including plasmid-mediated resistance via *qnr*[325], the efflux pump *acrAB*[326] and low-fidelity polymerase genes repressed by LexA and associated with mutagenesis, a key feature in the development of resistance-conferring mutations to fluoroquinolones in *Enterobacteriaceae*[327,328]. Many of these genes have previously been identified through RNAseq studies in other species, including *qnr*[90] and efflux pumps such as *acrAB*[89,329]. Additionally I have identified substantial downregulation of the porin *lamB*, supporting the hypothesis that it may play a role in inducible resistance[330]. I have also identified a number of resistance genes for non-quinolone antimicrobials, suggesting a generalised response that may be common to a number of antimicrobial stresses. Of interest, I have found significant up-regulation of genes associated with plasmid mobilisation and conjugation, including sub-inhibitory concentrations of ciprofloxacin. This phenomenon has previously been described by Thanh Duy *et al*, who quantified the conjugation efficiency of ESBL-encoding plasmids from human commensal *E. coli* to ciprofloxacin-resistant *Shigella sonnei*

and found that it was increased following supplementation with ciprofloxacin[331]. This presents a potentially important mechanism for the spread of resistance; an unintended consequence of antimicrobial therapy may be the increased spread of plasmids encoding resistance and virulence determinants to other bacterial species.

Also of interest are the number of hypothetical proteins that were significantly upregulated in the presence of ciprofloxacin. Many of these were encoded in regions of the genome that are adjacent to other known resistance determinants and demonstrate differential gene expression of similar orders of magnitude, suggesting they may be under the same mechanism of regulation. A number of upregulated hypothetical proteins were also identified by previous RNAseq studies, such as Molina-Santiago *et al.* when performing transcriptomic studies of *Pseudomonas putida* in the presence of ciprofloxacin[329]. This raises an opportunity to identify important resistance determinants that have not previously been described.

A key limitation of this study is the extensive investigation of one antimicrobial in a single bacterial isolate, especially when a number of differentially expressed genes are located on a plasmid. It is unclear whether this response can be generalised to other *K. pneumoniae* isolates, especially given the genetic diversity of the species, and to other antimicrobial classes that may induce different transcriptional responses. These questions are addressed further in **Chapter 5**.

Although I have demonstrated significant differential gene expression in this study, my original intention was to further characterise these findings by quantifying phenotypic changes associated with these transcriptomic findings. Unfortunately, planned experiments were not possible as a result of my work on the COVID-19 pandemic, but are described here

in outline. Firstly, I would have evaluated the rate of mutagenesis of isolate KC006 in the presence of ciprofloxacin in comparison to an untreated control by using a mutation accumulation/whole genome sequencing approach. Similar experiments by Long *et al.* in *E. coli* have demonstrated a dose-dependent increase in rates of mutation with fluoroquinolones that directly contributes to the acquisition of resistance through mutations in *gyrA*[328].

Secondly, given the number of differentially expressed efflux pumps and porins identified, I planned to measure the intracellular antimicrobial concentrations in the presence and absence of ciprofloxacin. The methods for doing so were not determined by the start of the pandemic, but will either have employed traditional techniques, such as those using Hoechst H33342 or ethidium bromide as a proxy for intracellular antimicrobial concentrations, or fluorescently labelled antimicrobials themselves[332]. Both approaches would have used the high content imaging system described in **Chapter 4**.

Finally, I intended to determine the effect of ciprofloxacin on plasmid conjugation frequency between KC006 and a recipient isolate[331]. However, given the AMR profile of KC006 there were difficulties finding appropriate methods for cross-selection. There was very low (if any) conjugation with *E. coli* isolate J53 in all experimental conditions used, certainly insufficient to quantify frequency. Further work was required to determine the ideal growth conditions and recipients for plasmid conjugation with isolate KC006, or use an alternative donor in whom similar transcriptomic effects had been identified (such as those in **Chapter 5**).

4. High content imaging for bacterial phenotyping

4.1 Introduction

In the previous chapter I illustrated how genomic approaches can be used to characterise large bacterial collections. Freely available bioinformatics tools are becoming increasingly widely available, making detailed genomic analysis of pathogens possible for a variety of research and clinical purposes. However, detailed phenotypic evaluation of bacterial pathogens is comparably lacking, relying on microbiological techniques that have remained largely unchanged for decades.

However, recent developments in high throughput confocal imaging and improved software tools for image analysis have resulted in the development of high content imaging (HCI). Used primarily in the field of eukaryotic cell research, HCI can determine subtle differences in cellular structure and shape not discernible to the human eye or conventional phenotypic approaches[96]. Importantly, the technique can be used to screen large numbers of cells grown simultaneously in controlled growth conditions in conventional formats, such as 96/384 well plates. While recent studies have demonstrated the potential utility of high content imaging for bacteria growing in liquid culture[103,106], the methods have not yet been optimised on a wide variety of species, nor a diverse range of lineages within species.

In this chapter I describe the development and optimisation of a high content imaging method using the Opera Phenix confocal microscope (PerkinElmer) to systematically phenotype individual bacteria under standardised growth conditions. I initially focus on a small number of reference isolates from three important species in the field of AMR, namely *Klebsiella pneumoniae*, *Salmonella enterica* serovar Typhimurium and *Staphylococcus aureus*. These organisms have distinct phenotypes (Gram-positive v Gram-negative, adherent v non-adherent) that necessitate three distinct imaging and analysis pipelines. Finally, I describe provisional experiments to characterise individual live bacteria in liquid culture.

4.2 High content imaging of bacteria

Initial optimisation experiments were performed using two reference isolates from each of the species *Klebsiella pneumoniae*, *Staphylococcus aureus* and *Salmonella enterica* serovar Typhimurium. Pairs of isolates were selected for their different antimicrobial susceptibility profiles (one susceptible to the majority of antimicrobial classes, the other resistant). The MICs of these isolates in relation to a range of antimicrobials were determined by Etest and are listed in **Appendix B.1**.

Bacterial cultures in stationary phase were diluted into fresh LB broth in ultra-thin 96-well microtitre plates. The resulting plates were incubated for 2 hours to enable sufficient time for morphological changes in their respective growth conditions. Bacteria were fixed and stained *in situ* using three commonly used dyes: FM4-64, a lipophilic dye that stains cell membranes and bacterial cell walls; DAPI, a cell permeable dye that stains nucleic acids; and SYTOX green, which also stains nucleic acids but only enters bacteria with permeabilised cell membranes and is therefore used as a marker of cell death[103]. Imaging was conducted

within 24 hours of staining using the Opera Phenix confocal microscope. Image analysis was performed using the associated software package Harmony. An overview of the imaging pipeline and stains used is shown in **Figure 4.1.**

Both the Opera Phenix and Harmony have been developed and largely utilised to study eukaryotic cells. To investigate morphological changes in individual bacteria, some orders of magnitude smaller than human cell lines, bespoke image analysis pipelines necessitated development for this study. Additionally, due to marked phenotypic differences, parallel pipelines were necessary for Gram-positive cocci (*S. aureus*) and Gram-negative rods (*K. pneumoniae* and *S.* Typhimurium). Each pipeline consisted of two key stages: i) segmentation of the image to identify individual bacteria; ii) measuring morphological properties of each bacterium using individual and combined measurements of the distribution of the three dyes above. For the Gram-negative pipeline there was an additional third stage required to detect (and exclude) overlapping or dividing bacteria, which need to be distinguished from individual cells to provide reliable phenotypic measurements. Multiple approaches were applied using the image analysis “blocks” in Harmony, with iterative improvements made through trial and error with the additional support of the manufacturers and software developers from PerkinElmer. However, no new software was specifically designed for this study. Illustrative images of the final analysis pipelines are shown in **Figure 4.2**. The full final analysis pipelines are included in **Appendices B.3** and **B.4**.

In the Gram-negative rod pipeline (**Figure 4.2a**), images were initially filtered using the FM4-64 channel. Due to the lipophilic nature of the dye, FM4-64 provides the most reliable visualisation of the external border of the cell, the bacterial cell wall, which was detected using Harmony’s ridge filter. Objects determined by this process were then measured and filtered in terms of stain intensity and size, to remove staining and other artefacts.

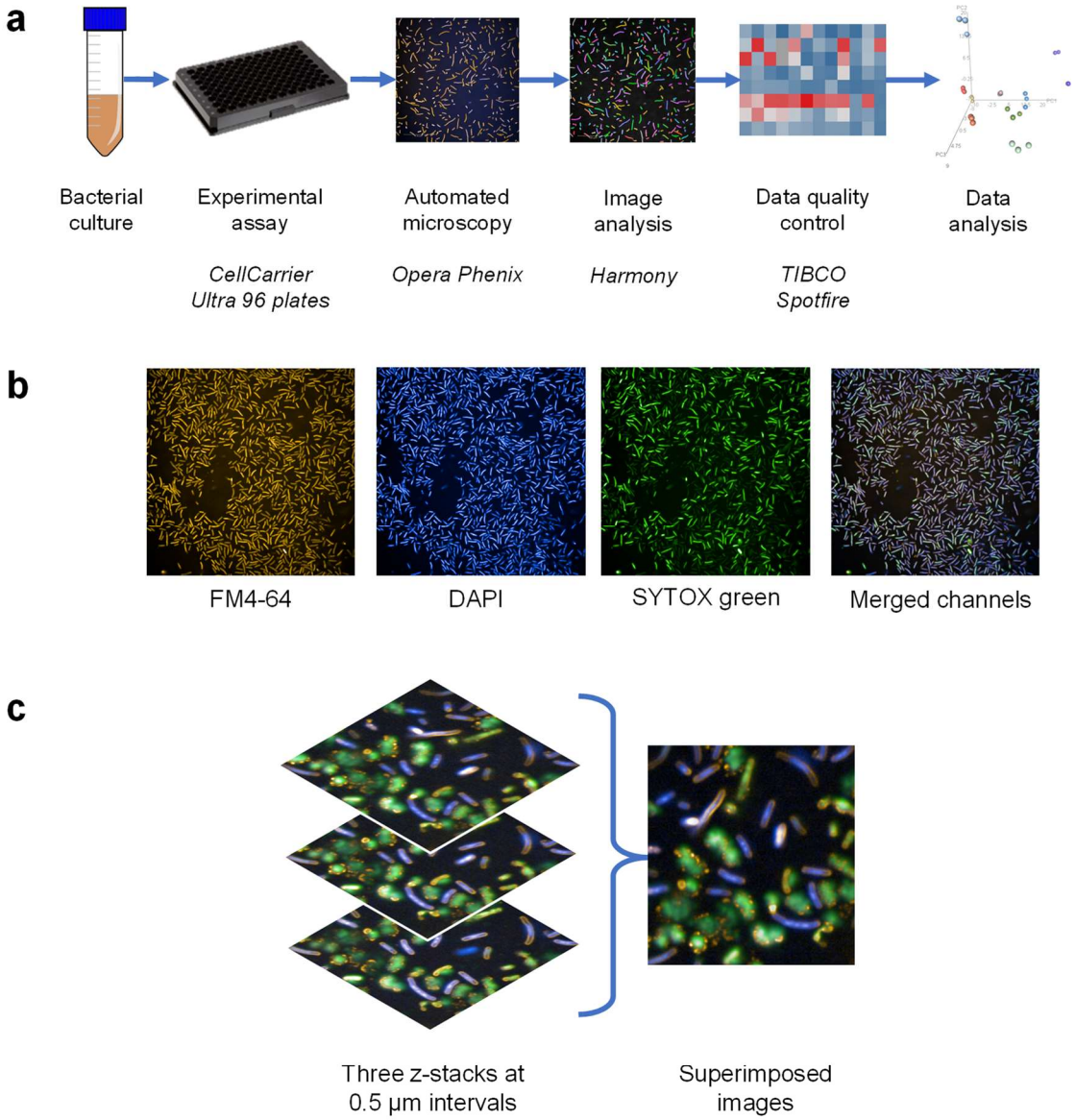


Figure 4.1 Bacterial high content imaging overview

(**a**) Schematic of the high-content imaging workflow. Bacterial cultures were added to ultra-thin bottom plates and incubated, with or without antimicrobial compounds. After 2 hours the contents of each well were aspirated and adherent bacteria fixed and stained before being imaged on an Opera Phenix confocal microscope using a 63x water immersion objective. Images were analysed using Harmony, with data quality controls performed in Spotfire prior to data analysis. (**b**) Representative whole field images of channels within Harmony for FM4-64 (cell membrane), DAPI (nucleic acid, membrane permeable) and SYTOX green (nucleic acid, membrane impermeable) and the 3 merged channels. Each field covers 200 x 200 µm, or 2,000 x 2,000 pixels. (**c**) Three composite images were taken at different depths of focus (z-stacks) for each field, at intervals of 0.5 µm, to capture the full thickness of the bacterial monolayer. These were combined (maximum projection) for further analysis.

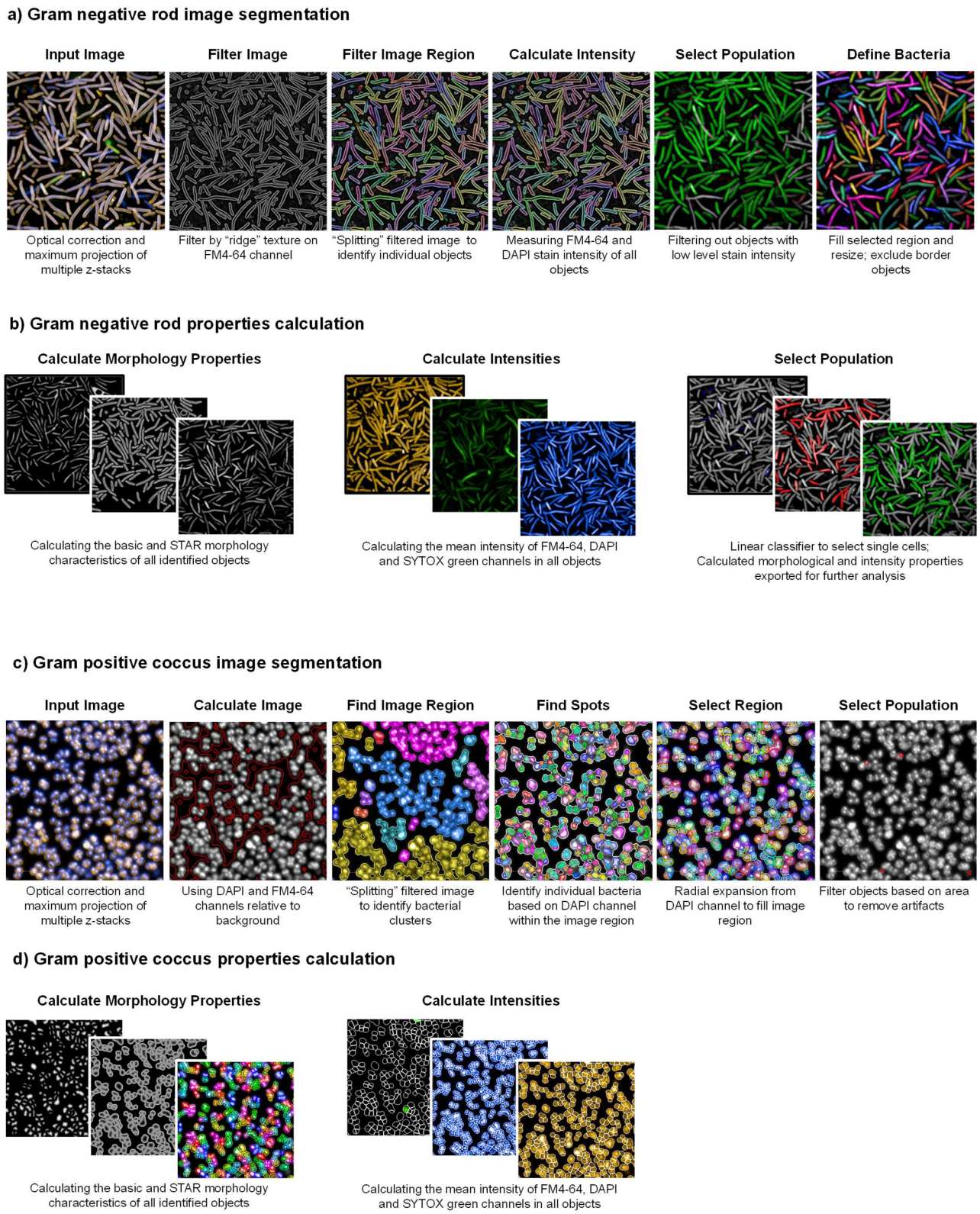


Figure 4.2 Bacterial image analysis workflow for Gram-negative rods and Gram-positive cocci

(**a**) Using basic brightfield correction and maximum projection, images were segmented by filtering the images using texture properties based on the FM4-64 channel to remove any background. The image region was filled and resized and border objects were excluded to include only whole objects. The image region was further calculated using FM4-64 and DAPI fluorescence and individual bacteria were defined. (**b**) Bacterial morphology and stain intensity properties were calculated using DAPI, SYTOX green and FM4-64 channels. Finally, a linear classifier was used to train the software to define single bacterial cells and exclude any artefacts. (**c**) Using basic brightfield correction and maximum projection, the bacterial region was defined using a calculated image based on DAPI and FM4-64 channels. Individual bacteria were identified within the image region using the DAPI channel, and the bacterial regions were defined and resized into individual bacterial cells. Any artefacts were removed using size filters. (**d**) Bacterial morphology and stain intensity properties were calculated using DAPI, SYTOX Green and FM4-64 fluorescence.

A combination of DAPI and FM4-64 stain intensities were used to determine the full extent of the bacteria within the detected cell border, resizing the object to include both cytosolic and cell wall regions. Objects at the image border (i.e. bacteria that were only partly imaged) were excluded and basic size filters applied to remove objects that were too small to be bacteria. Basic morphological features of each object were then calculated, including area, roundness, width and length, followed by more complex features such as the symmetry, intensity and distribution of each of the FM4-64, DAPI and SYTOX green channels within each object. A full list of the measured properties is shown in **Appendix B.5**.

All of these properties were then utilised by a linear classifier using Harmony PhenoLOGIC. Over one hundred representative objects were manually identified to define three classes of object, using images taken of *K. pneumoniae* during this experiment: single bacterial cells, dividing/overlapping cells and artefacts (such as stain deposits or fragments of dead bacterial cells). These were manually entered into the linear classifier, which used the above morphological properties to determine the key features in distinguishing each category of objects and applying them across all objects in this (and any subsequent) experiment. Data on over 60 morphological properties measured for single bacterial cells, in addition to aggregate well-based mean and standard deviation valuations of the same properties, were exported for further analysis. The number of single bacterial cells imaged per well varied depending on experimental conditions and isolate, but was a minimum of 2,000 bacteria per untreated well for all plates in these experiments.

The Gram-positive cocci pipeline (**Figure 4.2b**) is comparable to that above, but relies on different methods for initial segmentation. Due to the tight clustering of neighbouring bacteria, the ridge detection function is unable to reliably identify and segment individual cells. Instead, the region of interest (i.e. an aggregate of all bacteria relative to the background) is determined using a combination of the FM4-64 and DAPI channel intensities.

Individual bacteria are determined using the DAPI channel. Because DAPI stains nucleic acid, the channel forms a “spot” within each object that can be identified by Harmony and used as the basis of individual bacterial identification. Each spot is then expanded radially to the edge of the region of interest, or until it abuts a neighbouring cell, to determine the area of each object. Filters are applied, as for the Gram-negative rod pathway, to remove artefacts and objects along the image border. The same morphological and intensity properties are calculated, but a linear classifier is not necessary as the cells form a more cohesive monolayer without overlapping cells and with fewer artefacts.

Due to the large number of bacterial cells assessed, unless otherwise stated the exported data is limited to well-based mean and standard deviation values for the parameters measured on single cells. All plates were reviewed using a HCI module developed by PerkinElmer and available in TIBCO Spotfire to check the quality of results and exclude outlying wells. Data was then exported for further analysis.

4.3 Optimisation of experimental conditions

An early observation encountered in provisional experiments was that different bacterial isolates displayed varying levels of adhesion to the 96 well plates. Poor adhesion leads to small movements of bacteria across the image field as it is being captured, both between the sequence of channels for each dye and the z stacks to provide additional depth to the image. This substantially reduces the image quality in downstream analysis. The two *K. pneumoniae* reference isolates adhered strongly to the bottom of the well, providing clear images in all experiments; by contrast isolates of *S.* Typhimurium and *E. coli* adhered weakly, leading to blurred, poorly aligned images in the flattened z-stack, as shown in **Figure 4.3**.

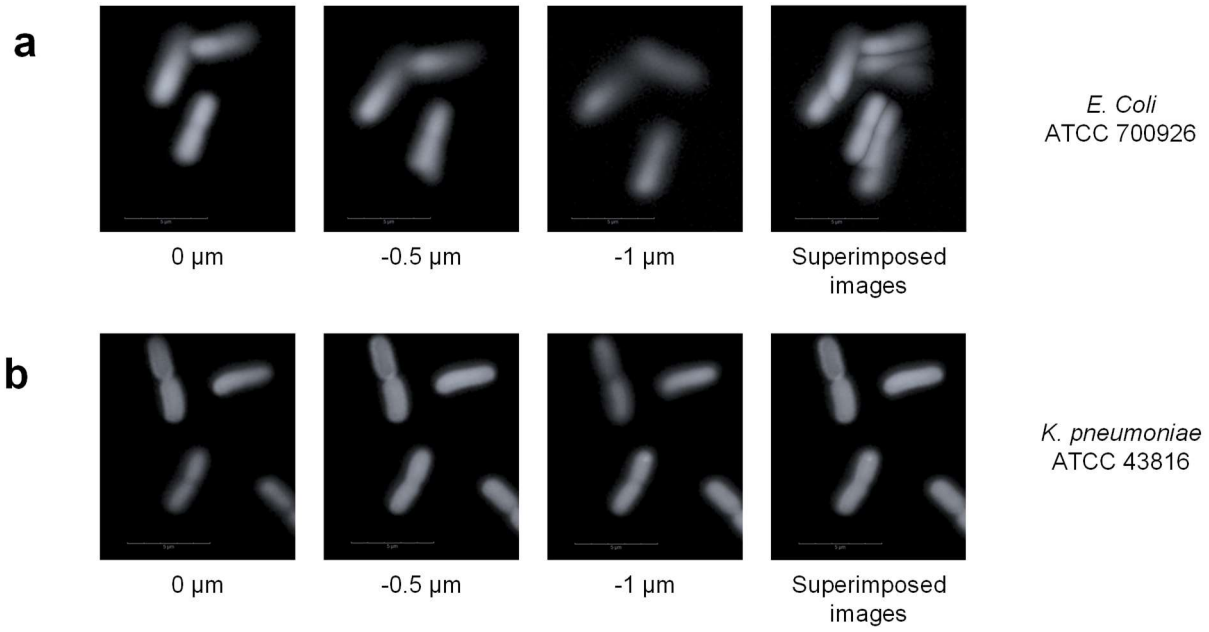


Figure 4.3 Species variation in adhesion and resulting image quality

Confocal images, using the DAPI and FM4-64 channels, of *E. coli* isolate ATCC 700926 (**a**) and *K. pneumoniae* isolate ATCC 43816 (**b**) on uncoated plates. The first 3 images in each row represent different z-stacks of the same field at 0.5 µm intervals, with the composite maximum projection image in the final column.

In an attempt to improve cellular adhesion for non-adherent bacteria, 11 coating matrices (thick and thin rat tail collagen, Matrigel, vitronectin, fibronectin, Cell-Tak, laminin, wheat germ agglutinin (WGA), poly-L-lysine, gelatine, and mouse collagen, detailed in **Appendix B.2**) were studied to identify the best imaging conditions in comparison to an untreated control. The analysis pipelines above were used to quantify the number of single bacterial cells per field as a proxy for cellular adhesion. Images were also manually reviewed to support this finding and identify any additional morphological differences associated with the plate coatings. All six reference isolates were studied in all conditions to determine the effect on adherent as well as non-adherent bacteria, and assess the impact of plate coatings in future HCI studies.

The number of bacteria identified on each plate coating are presented in **Figure 4.4**. The optimal coating varied between species and, to a lesser extent, between isolates of the same species. As described above, *K. pneumoniae* isolates displayed the best adhesion in uncoated wells, with comparable levels of bacteria imaged on WGA and fibronectin coatings (**Figure 4.4a**). In addition to the reduced cell numbers identified on other coatings, there was some associated change in bacterial morphology, with laminin, gelatin and mouse collagen all associated with bacterial aggregation that was not present on the uncoated well (**Figure 4.5a**). Given the uncoated plate is the fastest and simplest to prepare, and yields the best results, all subsequent experiments with these isolates were conducted in uncoated wells. This provides the default experimental conditions for further experiments involving *K. pneumoniae* described below and in **Chapter 5**.

By contrast, *S.* Typhimurium displayed poor adhesion to uncoated plates, with varying levels of adhesion to wells coated with thick rat tail collagen, Matrigel, and vitronectin (**Figure 4.4b**). Bacterial aggregation was again noted with some coatings, including mouse collagen and thin collagen (**Figure 4.5b**). Vitronectin and collagen were identified as the optimal conditions for *S.* Typhimurium NCTC 13347 and NCTC 13348 respectively. While the number of bacteria identified clearly improved with plate coatings, movement artefact could not be entirely eliminated. Further image analyses on *S.* Typhimurium were therefore conducted on single planes rather than a maximum projected image of multiple z-stacks.

Similarly, the two *S. aureus* isolates displayed different adhesion properties, with ATCC 29213 poorly adherent to uncoated wells in comparison to vitronectin, fibronectin, Cell-Tak, WGA and gelatin-coated wells, whereas NCTC 6571 only had sufficient cell counts on thin collagen and fibronectin coated wells (**Figure 4.4c**). Thin collagen and vitronectin were therefore selected for optimal adhesion of NCTC 6571 and ATCC 29213, respectively.

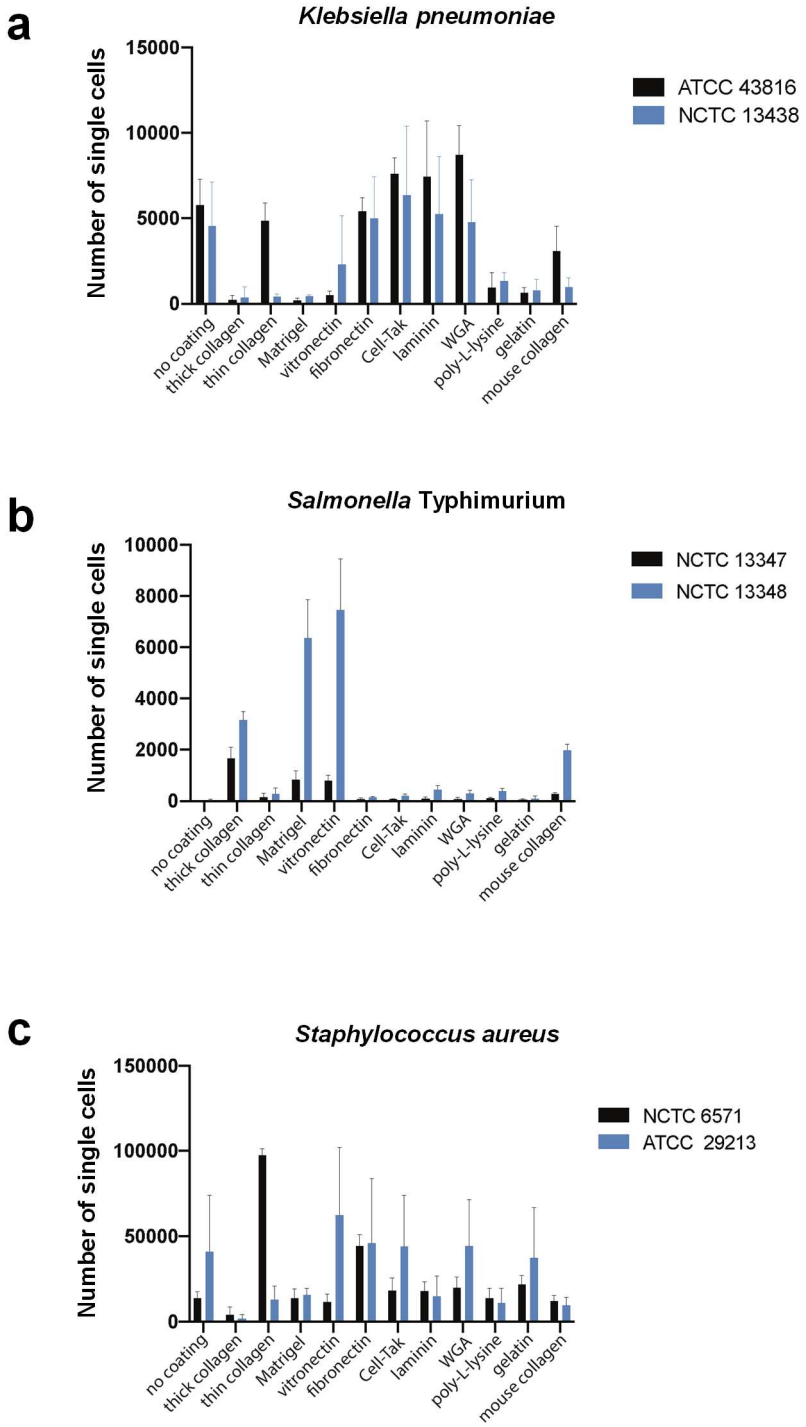


Figure 4.4 Optimising plate coating for bacterial adhesion

Isolates were grown in ultra-thin 96 well plates on different coating matrices. Harmony analysis pipelines were used to count the number of adherent bacteria after fixing, washing and staining. Graphs compare the adhesion of two representative isolates of *K. pneumoniae* (**a**), *S.* Typhimurium (**b**) and *S. aureus* (**c**) on each substrate. Error bars represent standard deviation of three biological replicates. Experiments conducted by Sally Forrest. Image analysis pipelines developed by Ben Warne. Figures prepared by Sally Forrest.

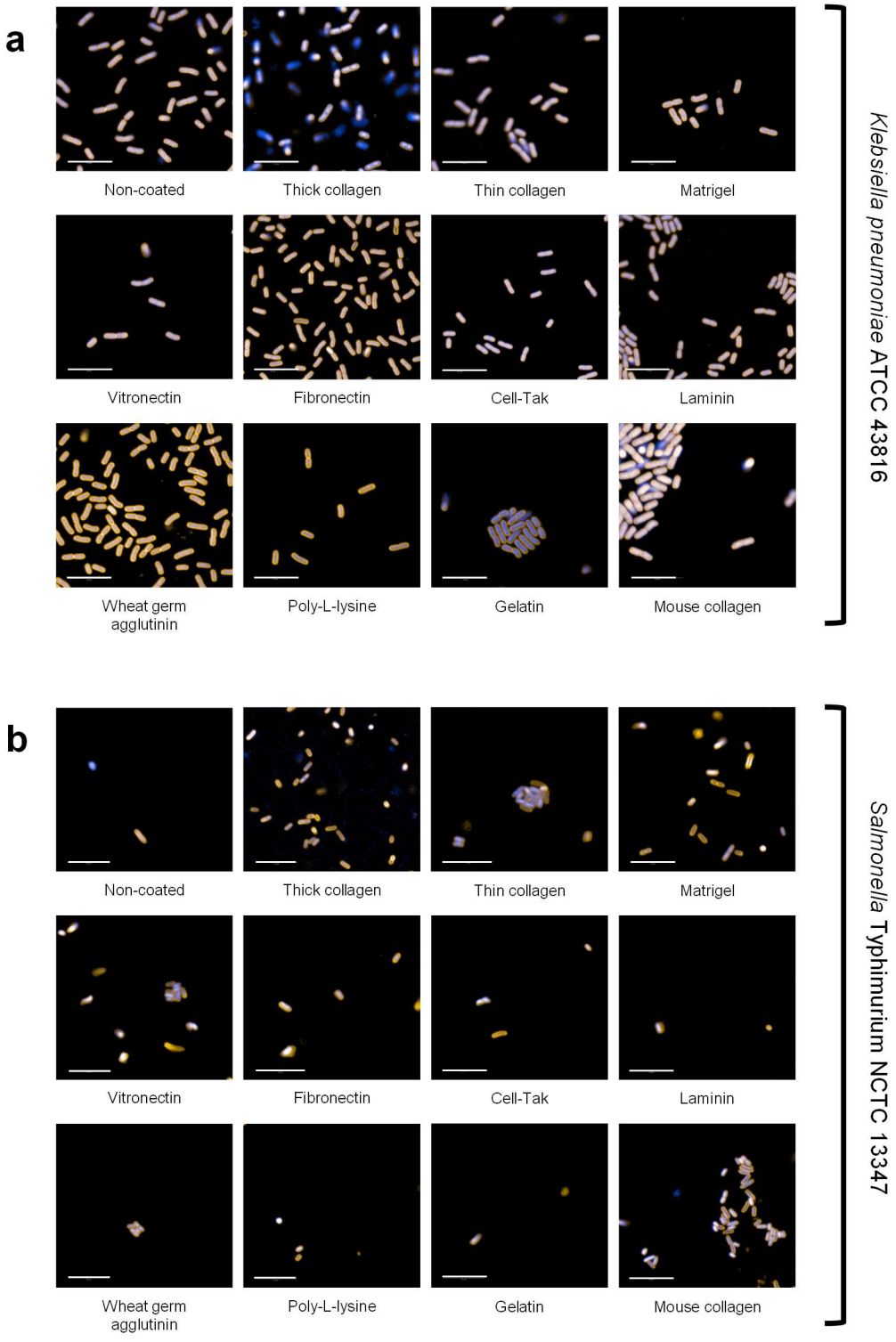


Figure 4.5 Representative images of isolates on different well coating matrices

Isolates *K. pneumoniae* ATCC 43816 (**a**) and *S.* Typhimurium NCTC 13347 (**b**) were incubated in non-coated wells or wells containing collagen, Matrigel, vitronectin, fibronectin, Cell-Tak, laminin wheat germ agglutinin poly-L-lysine, gelatin or mouse collagen for 2 hours, then fixed and stained with FM4-64 (membrane) and DAPI (nucleic acid). Images were acquired on the Opera Phenix using a 63x water immersion lens.

4.4 Morphological changes associated with exposure to antimicrobials

Having optimised the experimental conditions and image analysis pipelines, this high content imaging system was then used to compare the morphological changes that occur when bacteria are exposed to antimicrobials. All six reference isolates were exposed to a variety of antimicrobials (10 for the Gram-negative organisms; 5 for the *S. aureus isolates*) for 2 hours at 5x MIC, or at 5x the upper limit of the Etest used if bacterial growth occurred along the whole length of the MIC strip. Representative images of the observed phenotypic changes are shown for *K. pneumoniae* isolate ATCC 43816 (**Figure 4.6a**) and *S. aureus* ATCC 29213 (**Figure 4.7a**).

The image analysis pipelines calculated 62 morphological properties and channel intensities for each single bacterium identified. Well-based mean and standard deviation values were combined from the 3 technical and biological replicates (3 wells per plate; 3 plates per isolate) for each growth condition and analysed using principal component analysis (PCA), as shown in **Figures 4.6b** and **4.7b**. For each antimicrobial, the three technical replicate wells from each plate cluster tightly together, with antimicrobials that share a common class or related mechanism of action forming distinct clusters. For example, *K. pneumoniae* isolates treated with gentamicin and tigecycline, which both inhibit protein synthesis by binding the 30S ribosomal subunit at different locations, cluster together, as does rifampicin, which inhibits RNA synthesis (and, by extension, protein synthesis). The beta lactam antimicrobials cefuroxime and amoxicillin form a separate cluster, as do ciprofloxacin and trimethoprim-sulphamethoxazole, which both impact DNA replication. Where an isolate was resistant to an antimicrobial, such as ampicillin, trimethoprim-sulphamethoxazole and cefuroxime resistance in the *K. pneumoniae* isolate NCTC 13438, wells treated with these antimicrobials cluster with the untreated control.

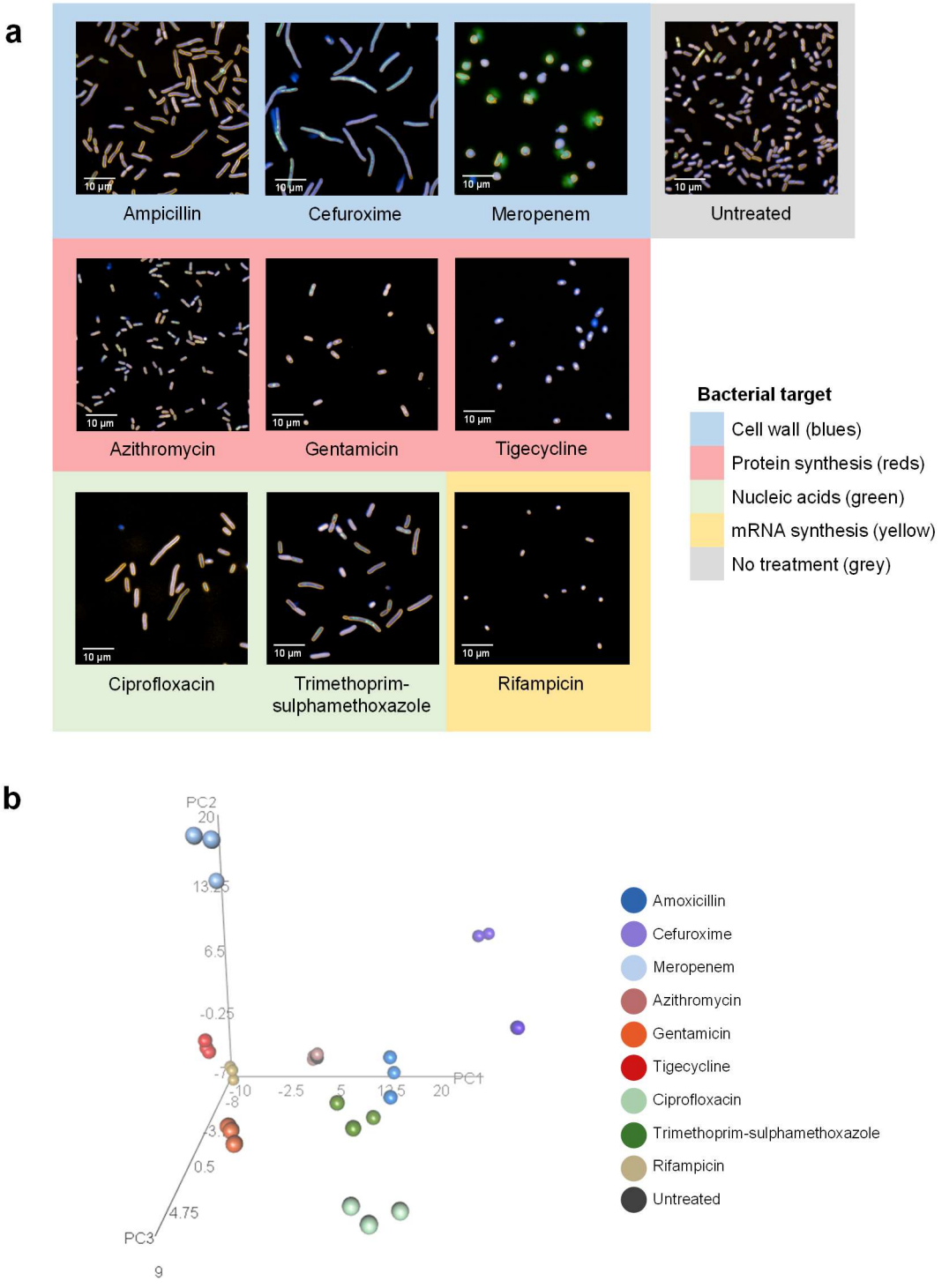


Figure 4.6 Phenotypic effects of antimicrobial pressure on *K. pneumoniae* ATCC 43816

(**a**) Representative images of the effect of different antimicrobials on the *K. pneumoniae* isolate ATCC 43816 in early exponential growth phase after 2 hours of incubation. Antimicrobials are grouped by similar cellular targets. Bacteria were stained with FM4-64, DAPI and SYTOX green. Images were acquired on an Opera Phenix using a 63x water immersion lens. (**b**) Three-dimensional principal component analysis of the mean and standard deviation values of 62 morphological properties measured for single bacterial cells in each well. Technical triplicate repeats are shown.

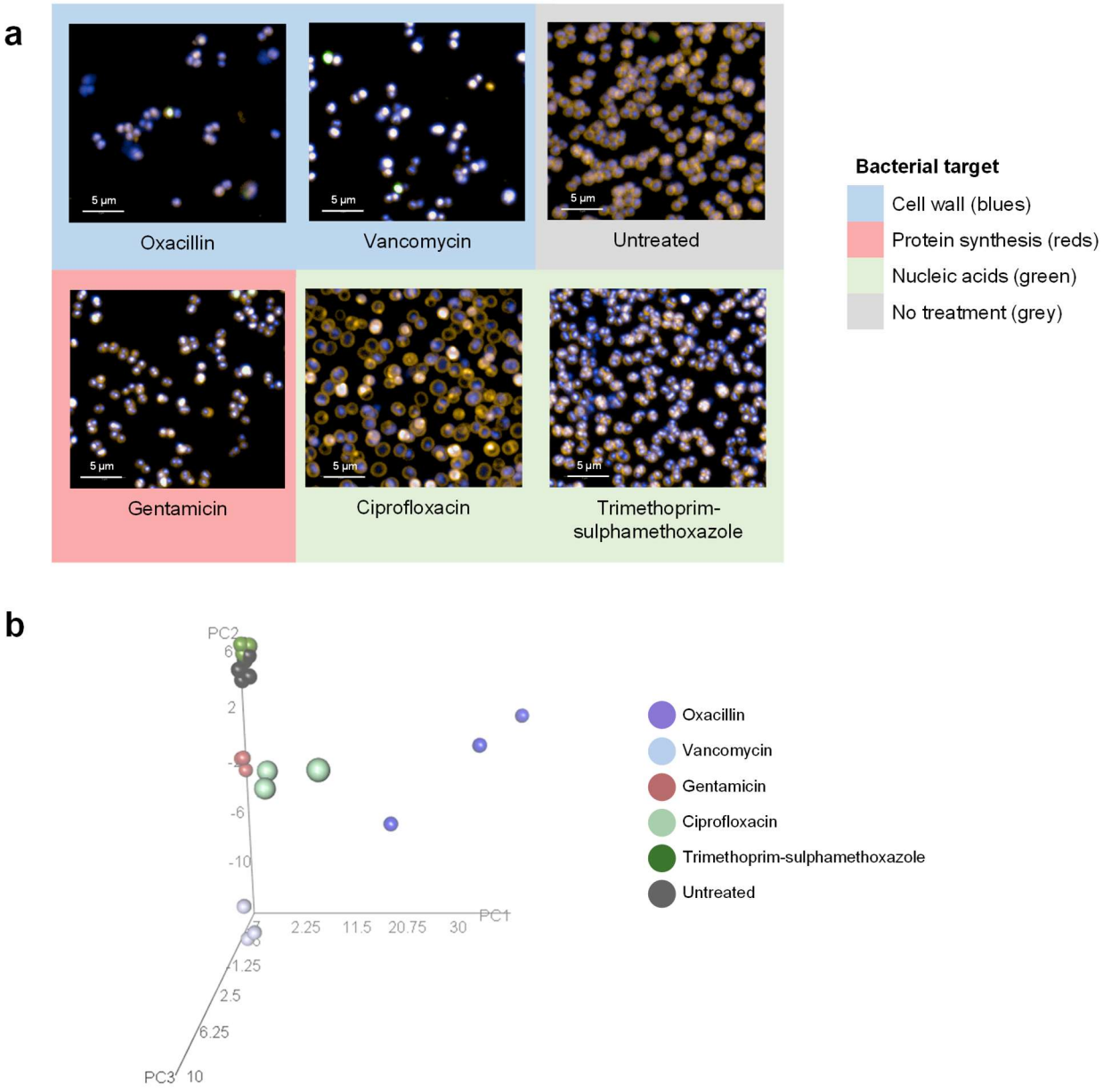


Figure 4.7 Phenotypic effects of antimicrobial pressure on *S. aureus* ATCC 29213

(**a**) Representative images of the effect of different antimicrobials on the *S. aureus* isolate ATCC 29213 in early exponential growth phase after 2 hours of incubation. Antimicrobials are grouped by similar cellular targets. Bacteria were stained with FM4-64, DAPI and SYTOX green. Images were acquired on an Opera Phenix using a 63x water immersion lens. (**b**) Three-dimensional principal component analysis of the mean and standard deviation values of 62 morphological properties measured for single bacterial cells in each well. Technical triplicate repeats are shown.

Biological replicates, from experiments performed on different days, form similar clusters, demonstrating that the assay is reproducible (**Figure 4.8a,b**). Although separate image analysis pipelines were developed for coccus- and rod-shaped bacteria, this demonstrates that each pipeline can reliably distinguish a wide variety of phenotypic changes induced by antimicrobial treatment.

The distribution of variance contributing to principal components 1, 2 and 3 in **Figure 4.6a** are shown in **Figure 4.8c**. The top 50 individual measurements contributing to principal component 1 in the same figure are shown in **Figure 4.8d**. These illustrate the relatively small individual contribution of each variable to the principal components, further stressing the power of combining multiple measurements for phenotypic characterisation of bacteria.

The Z' statistic is used in high throughput microscopy to determine key morphological features from the wide range of variables measured, that can distinguish positive and negative controls[266]. Values between 0.6 and 1 are considered reliable features on which to base this distinction. **Appendix B.6** shows the top 15 Z' values for distinguishing untreated *K. pneumoniae* from those treated with ciprofloxacin, meropenem and tigecycline. Some of the key phenotypic changes in the presence of ciprofloxacin were that the bacteria were longer with larger and more varied area (as determined by width-to-length ratio, area and axial length ratio). Such appearances may be intuitive for a drug that inhibits DNA replication and cell division, but does not inhibit bacterial protein synthesis *per se* – i.e. the bacteria can grow, but cannot divide, producing long filamentous bacterial phenotypes. However, there are more subtle morphological changes that may not be evident to the human eye. For example, ciprofloxacin treated cells had an apparently wider distribution of nuclear material spread within the cell (DAPI threshold compactness) and had a more homogeneously stained cell wall (FM4-64 radial relative deviation).

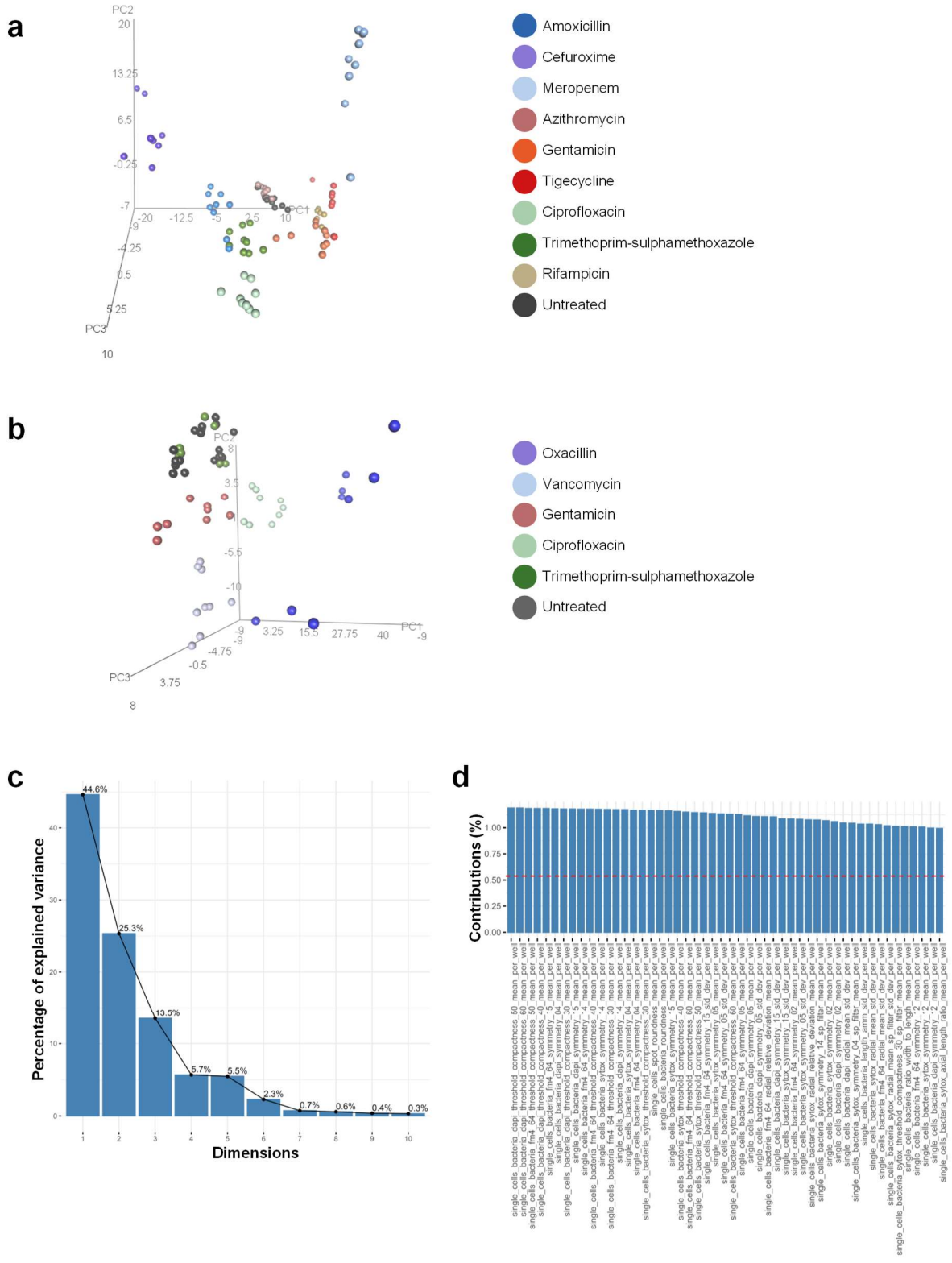


Figure 4.8 Assay reproducibility and the contribution of morphological features in PCA

Three-dimensional principal component analysis of the mean and standard deviation values of 62 morphological properties measured for *K. pneumoniae* ATCC 43816 (**a**) and *S. aureus* ATCC 29213 (b) after 2 hours of exposure to antimicrobials. Three biological replicates with three technical replicates per assay are shown. (**c**) Scree plot demonstrating the proportion of variation captured by each principal component for the isolate *K. pneumoniae* ATCC 43816. (**d**) The top 50 variables for principal component 1 for the same isolate, ranked in descending order by their contribution.

By contrast, meropenem induced morphological changes were characterised by increased cell area, width and roundness. This may be expected from a beta lactam antimicrobial that causes damage to the bacterial cell wall, leading to the cell taking on fluid by osmosis, swelling and forming larger, rounder cells. There is also a wider distribution of nuclear material spread through the cell (DAPI profile). Bacteria treated with tigecycline were again morphologically distinct, with more compact nuclear material at the centre of the cell (DAPI profile and threshold compactness) and more heterogeneous staining of the cell wall (FM4-64 threshold compactness) in comparison to untreated controls.

The distribution of individual bacterial measurements for a selection of variables with high Z' scores for a range of antimicrobials are shown in **Figure 4.9**. In addition to demonstrating how individual morphological features differ between antimicrobial exposures, they also demonstrate the heterogeneity of bacterial populations of one isolate grown from the same culture under identical growth conditions. For example, ciprofloxacin induces a significant increase in mean bacterial area and length-to-width ratio, but there is a wide distribution of both parameters, even in the same well. Some ciprofloxacin-treated bacteria retain morphological appearances that are consistent with untreated bacteria. Observing such differences within a population of bacteria is of potential utility in identifying persistence or emerging resistance in the face of antimicrobial pressure.

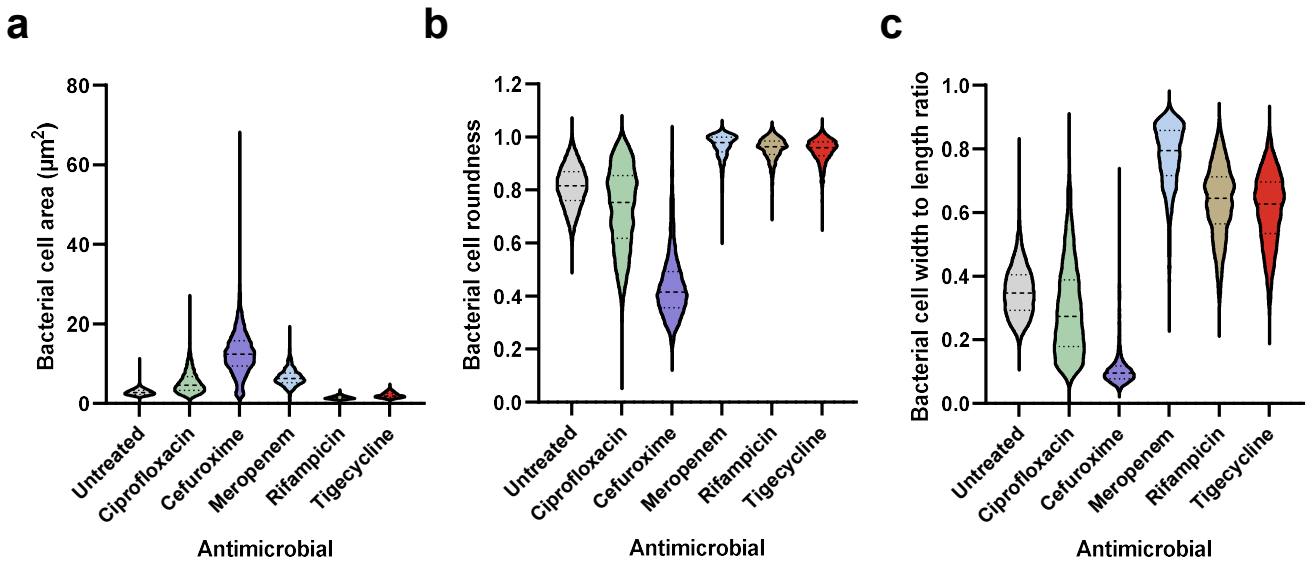


Figure 4.9 Comparison of individual morphological measurements

Violin plots of bacterial area (**a**), roundness (**b**) and width-to-length ratio (**c**) comparing *K. pneumoniae* ATCC 43816 treated with ciprofloxacin, cefuroxime, meropenem, rifampicin, and tigecycline, with untreated controls. Dashed lines represent the median and interquartile range.

4.5 Imaging clinical *K. pneumoniae* isolate KC006

Having optimised the high content imaging pipeline using two reference *K. pneumoniae* isolates, I next applied the same methodology to the blood culture isolate KC006, characterised in **Chapter 3**. Harmonising methods used for HCI and RNASeq, the isolate was grown to stationary phase in overnight culture, diluted into fresh LB and incubated at 37°C for 90 minutes. Samples were then inoculated with a range of antimicrobials with differing modes of action, at 1x MIC. In comparison to previous experiments, amoxicillin, rifampicin and azithromycin (to which KC006 is resistant) were substituted for chloramphenicol and colistin. MICs are detailed in **Appendix C**. All wells were fixed, stained and imaged after 2 hours of incubation. Representative images and a principal component analysis of one of the three replicates are shown in **Figure 4.10.**

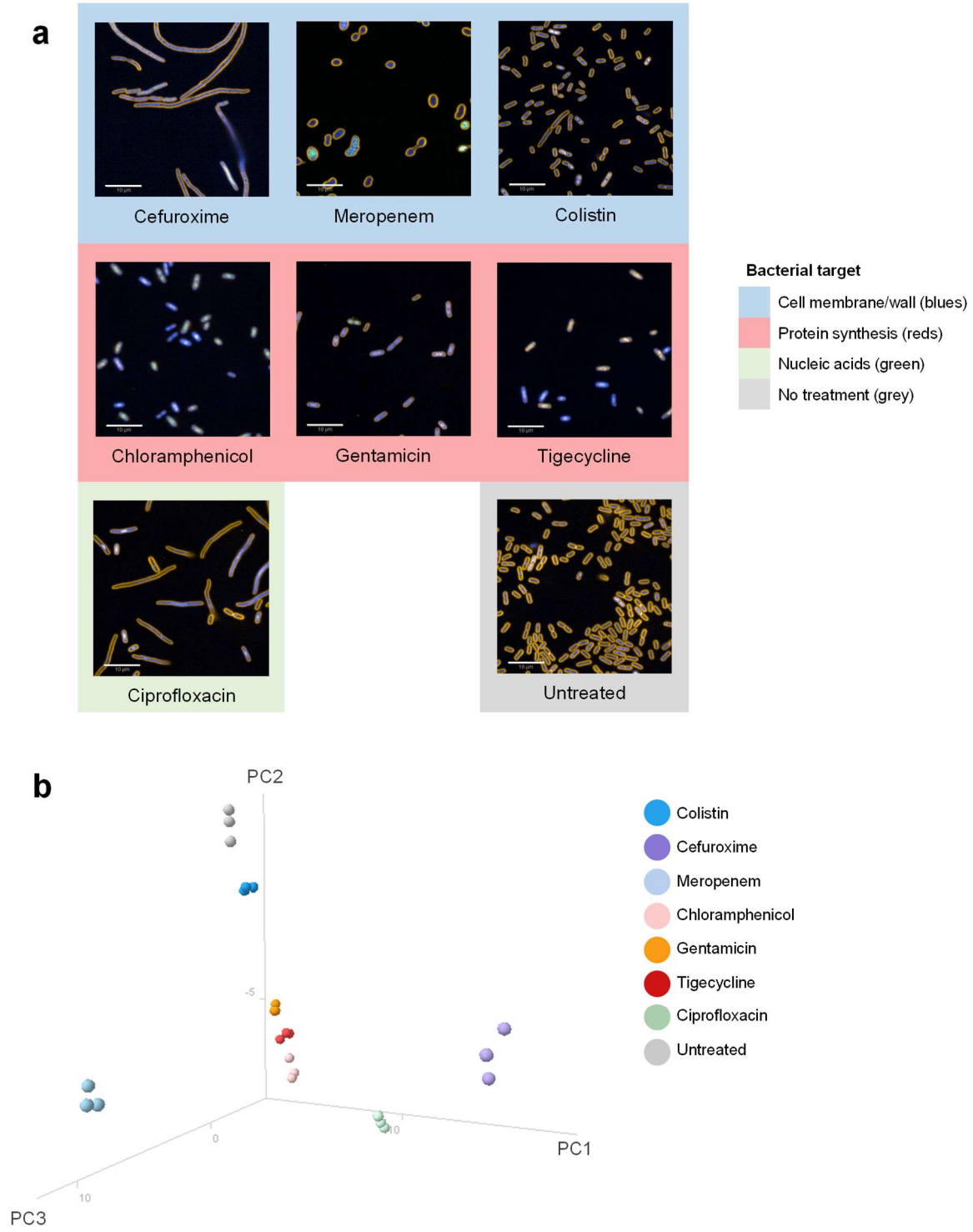


Figure 4.10 Phenotypic effects of antimicrobial pressure on clinical isolate KC006

(**a**) Representative images of the effect of different antimicrobials on the *K. pneumoniae* bacteraemia isolate KC006 in early exponential growth phase after 2 hours of incubation. Antimicrobials are grouped by similar cellular targets. Bacteria were stained with FM4-64, DAPI and SYTOX green. Images were acquired on an Opera Phenix using a 63x water immersion lens. (**b**) Three-dimensional principal component analysis of the mean and standard deviation values of 62 morphological properties measured for single bacterial cells in each well. Technical triplicate repeats are shown.

In keeping with previous experiments, there is a clear morphological difference and distinct clustering with all antimicrobials tested. Chloramphenicol, which targets the 50S subunit of the bacterial ribosome and therefore inhibits protein synthesis, induces a comparable morphological change to gentamicin and tigecycline. The effect of colistin was more subtle and distinct from other antimicrobial classes, but could still be differentiated using PCA.

4.6 Live bacterial imaging

All imaging demonstrated so far has been performed on fixed bacterial cells after a 2 hour incubation period. However, such assays are likely to provide greater detail through using serial measurements of bacteria over time. As shown in **Figure 4.9**, although antimicrobials cause a large morphological change in bacteria when aggregated at the population level within a well, reviewing data from individual bacteria demonstrates that a small sub-population are not substantially phenotypically altered in the presence of some antimicrobials. To investigate these patterns further, trial experiments were performed using live bacterial imaging.

Experimental conditions were kept as similar to those above as possible, with provisional experiments being conducted using the isolate KC006. After inoculation of bacteria and antimicrobials into 96 well plates and incubation for 90 minutes, wells were aspirated and the stains listed in **Table 4.1** were added directly to wells, either alone or in combination, at 0.1x, 1x and 10x the recommended concentration for imaging fixed specimens. The plates were sealed with a gas permeable membrane and placed into the Opera Phenix, pre-heated to 37°C, where fields in each well were imaged at baseline and at 2 hours, to determine whether the stains interfered with bacterial growth or imaging. In parallel, bacterial growth in

the presence of the stains at the same concentration was measured in an automated plate reader.

Table 4.1 – Stains used in live bacterial imaging trials

Stain	Spectrum	Cellular target	Cell membrane permeability
FM4-64	**Red**	Cell wall/membrane	Impermeable
DAPI	**Blue**	Nucleic acid	Permeable
SYTOX green	**Green**	Nucleic acid	Impermeable
SYTO-9	**Green**	Nucleic acid	Permeable
PI	**Red**	Nucleic acid	Impermeable
FM5-95	**Red**	Cell wall/membrane	Impermeable
Sytox blue	**Blue**	Nucleic acid	Impermeable
Hoechst	**Blue**	Nucleic acid	Permeable
HI	**Red**	Cell wall/membrane	Permeable
DiL	**Red**	Cell wall/membrane	Impermeable

All provisional experiments using stain at 0.1x the usual concentration did not yield any images that could be used for analysis, as there was insufficient contrast between the bacteria and the background stain intensity. Some stains were found to be bactericidal at concentrations sufficient for imaging (Hoescht, DiL). The remaining stains did not appear to have an effect on bacterial growth in the plate reader at any concentration and produced reasonable quality images using the Opera Phenix. However, for most stains the imaged bacterial appearances did not change substantially between baseline imaging and that at 2 hours. An example, using the regular combination of FM4-64, DAPI and SYTOX green, is shown in **Figure 4.11a**. Although there are a larger number of bacteria at the 2 hr time point,

there is no apparent bacterial replication of those cells present at baseline. Serial imaging of these wells at 4 minute intervals confirmed that these additional bacteria were present due to cells in free culture subsequently adhering to the plate. The one exception is the combination of SYTO-9 and the live/dead differentiating stain PI. As shown in **Figure 4.11b**, after 2 hours of incubation there appeared to be bacterial replication, seeded by cells present at baseline.

The reason for impaired bacterial replication is unclear, but may relate to a direct cytotoxic effect of the stains at the plate surface, or the impact of high energy laser excitation used in the imaging process. To circumvent both of these issues, further trials were conducted using *K. pneumoniae* reference strains that harboured a GFP or mCherry expressing plasmid. Eight bacterial isolates were provided by Sebastian Bruchmann and Francesca Short for use in this study: ATCC 43816 (GFP, mCherry), MKP103 (GFP, mCherry), RH201207 (GFP, mCherry), Ecl8 (mCherry) and NCTC 13438 (mCherry). Experimental conditions were the same as those used above, but isolates were also incubated in the presence and absence of ciprofloxacin at MIC (**Appendix B.1**).

Provisional experiments showed that the intensity of mCherry deteriorated rapidly over time. **Figure 4.11c** shows the reduction in mCherry intensity over approximately 20 minutes, with the field being imaged approximately every 3 minutes. Similar appearances were present for all mCherry expressing isolates. Of the GFP expressing isolates, MKP103 and RH201207 did not adhere to the plate surface, leaving only ATCC 43816. While GFP intensity in this isolate also deteriorated over time, the impact was much less substantial and Harmony analysis pipelines could identify individual bacteria over an 80 minute time course (**Figure 4.12**). Additionally, bacterial replication on the plate surface could clearly be visualised, showing that imaged bacteria were truly alive despite regular imaging (**Figure 4.12a**).

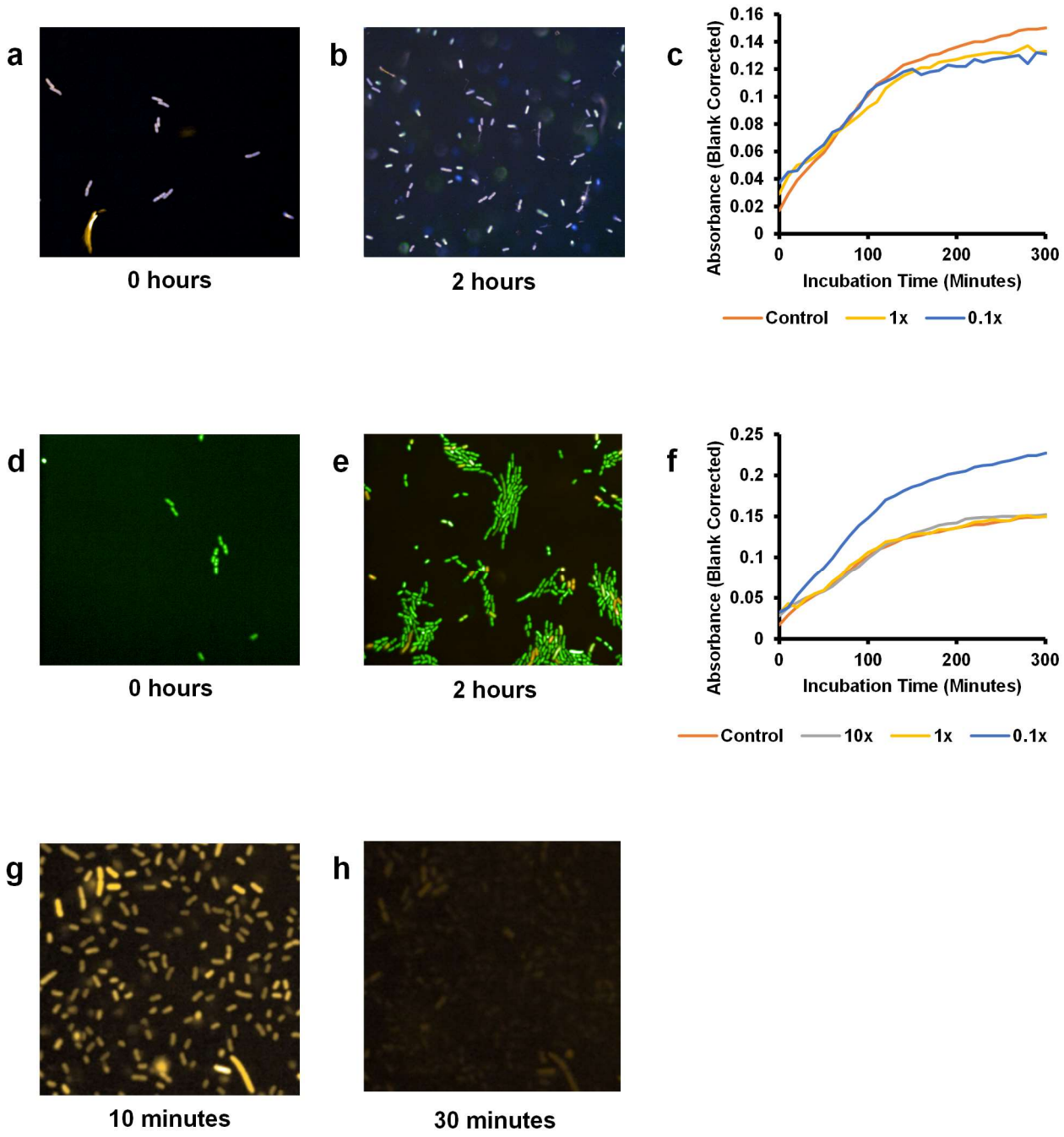


Figure 4.11 Live imaging using stains or fluorescent protein expressing *K. pneumoniae*

(**a-b**) *K. pneumoniae* isolate KC006 in live culture in the presence of the stains DAPI, FM4-64 and SYTOX green, with the same partial field imaged at 0 hours (**a**) and 2 hours (**b**) of incubation and demonstrating no bacterial replication. (**c**) Growth curve of KC006 in the same growth conditions with varying concentrations of stains, relative to that used in previous experiments.

(**d-e**) KC006 in live culture in the presence of the stains SYTO-9 and PI, with the same partial field imaged at 0 hours (**d**) and 2 hours (**e**) of incubation, with appearances strongly suggestive of bacterial replication. (**f**) Growth curve of KC006 in the presence of SYTO-9 and PI at different concentrations relative to that recommended by the manufacturer for confocal imaging of fixed cells.

(**g-h**) *K. pneumoniae* isolate ATCC 43816 with a plasmid encoding constitutively expressed mCherry, in live culture. Images taken at 10 minutes (**g**) and 30 minutes (**h**) with the same exposure and image settings. All images were acquired on an Opera Phenix using a 63x water immersion lens.

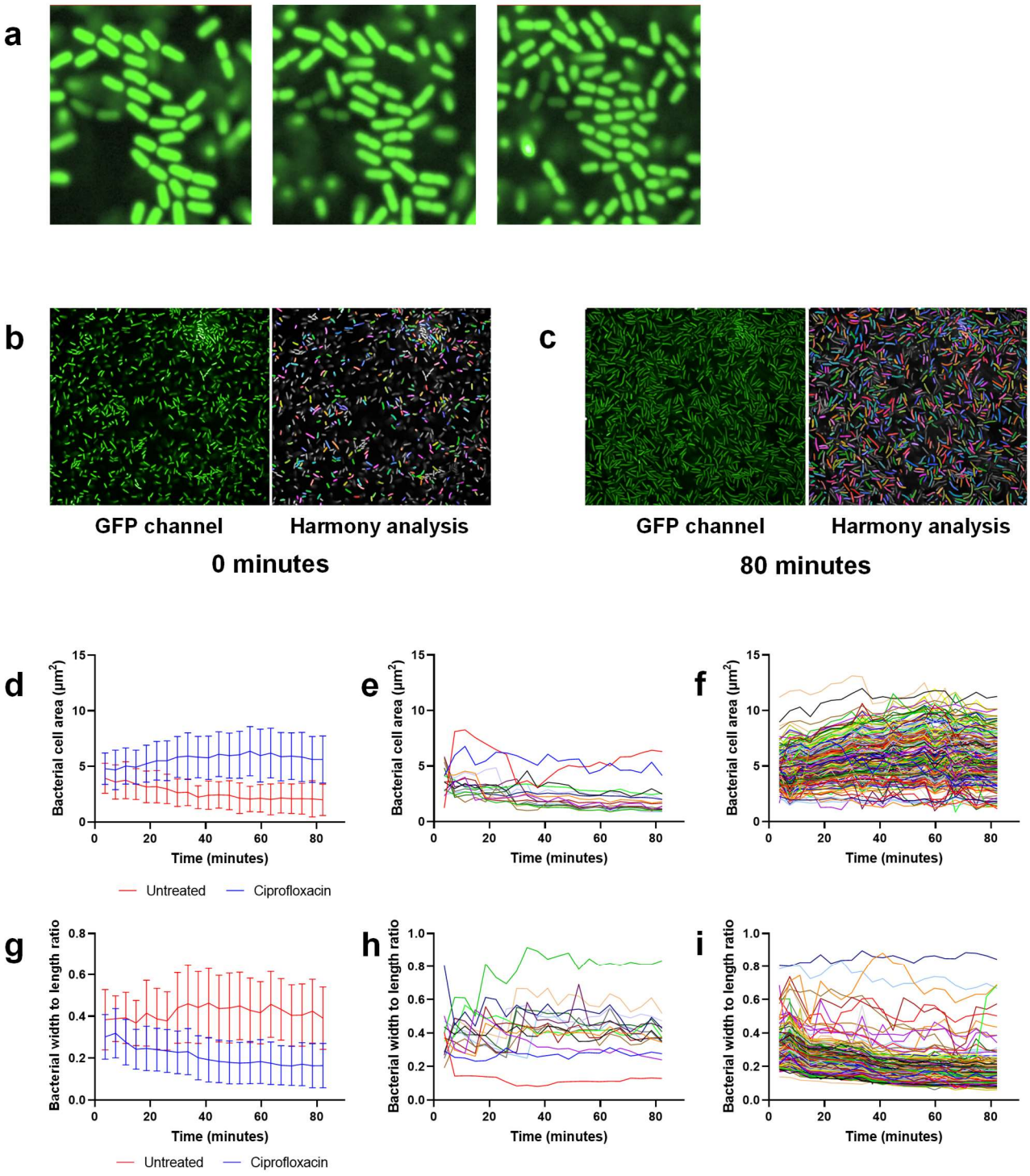


Figure 4.12 Individual GFP-producing ATCC 43816 *K. pneumoniae* imaged in live culture

(**a**) Serial cropped field images of GFP-expressing ATCC 43816 over approximately 20 minutes, demonstrating cell division. (**b-c**) Whole field images (left image in each panel) and the segmented bacteria identified by Harmony (right image) of the same isolate treated with ciprofloxacin at baseline (**b**) and after 80 minutes incubation (**c**).

(**d-i**) The mean and standard deviation of the bacterial cell area (**d**) and width-to-length ratio (**g**) at approximately 3.5 minute intervals over the time course are shown for both treated and untreated wells. The measured area and width-to-length ratio of individual bacteria are plotted for untreated (**e**, **h**) and treated (**f**, **i**) growth conditions. All images were acquired on an Opera Phenix using a 63x water immersion lens.

In this experiment, ciprofloxacin treated wells and untreated controls were imaged approximately every 3.5 minutes for 80 minutes. Harmony analysis pipelines, based on those developed for HCI above but with additional object tracking tools and with a limited number of calculated morphological properties due to the absence of differentiating stains, were utilised across the time course (**Figure 4.12b,c**). In three fields in one untreated well, 16 bacteria could be consistently tracked and measured at 22 time points between 0 and 80 minutes; 206 bacteria could be tracked in a ciprofloxacin treated well. **Figure 4.12d** shows the mean and standard deviation of the bacterial cell area and how this changes over time, increasing for ciprofloxacin treated cells (as above) and falling for untreated cells (possibly due to increased bacterial crowding over the plate surface). However, trends in measurements for individual cells shown in **Figures 4.12e** and **f** illustrate some of the variation in the morphology of bacteria over time. While some differences in bacterial measurements between time points are likely to be artefactual (for example due to bacterial division, imaging artefacts, etc.), there are still clear differences in bacterial morphology within the same population. For example, in the ciprofloxacin treated well there are some cells that retain a bacterial area comparable to their untreated controls. The same phenomenon is shown with bacterial cell width-to-length ratio in **Figures 4.12g-i**.

Further work to optimise live imaging experimental protocols and analysis pipelines, and their utility in investigating AMR in *K. pneumoniae*, was cut short by the COVID-19 pandemic. Work that was planned but not possible to contribute to this thesis is discussed below.

4.7 Discussion

In this chapter I have detailed the optimisation of an experimental method for high content imaging of bacteria in liquid culture. I have demonstrated the refinement of initial growth conditions, the development of bespoke imaging pipelines, morphological characterisation of bacteria, quality control and data analyses. I have shown how pipelines can be used for Gram-negative rods and Gram-positive cocci, and how these methods require adaptations for non-adherent bacteria. These pipeline have then been used to demonstrate a wide range of distinctive phenotypic changes that occur in the presence of different antimicrobials, which can be related to their mechanism of action. I have demonstrated the variation in phenotypic appearances evident in bacterial populations under antimicrobial selection. Provisional experiments have also shown the utility of live bacterial imaging to study morphological changes over time, including the identification of subpopulations of bacteria with distinct phenotypes under antimicrobial selection.

There are a number of potential uses for HCI in bacterial research. The Opera Phenix platform is extremely rapid at capturing high resolution images across a 96 well plate. This enables great flexibility in the construction of experiments, with the ability to customise and compare conditions in each well. However, it also permits relatively standardised growth conditions across the plate, enabling comparison of a large number of bacterial isolates simultaneously in the same experimental conditions. The associated image analysis software enables fast, detailed profiling of individual bacteria as well as quality control checks. The analysis pipelines demonstrated here can reliably segment bacteria and determine detailed morphological characterisation that can distinguish a range of antimicrobials. In practice, this means that hundreds of thousands of individual bacteria can be imaged growing in a variety of conditions on a single plate in under 1 hour, with a comparable time for associated image analysis. Whereas traditional phenotyping methods

rely on aggregate measurements across a population, this method can provide over 60 measurements on each individual bacterium.

The phenotypic changes induced by antimicrobials identified in this work are similar to those demonstrated in previous imaging studies in *Enterobacteriaceae*, including filamentous elongation in the presence of fluoroquinolones[333], bacterial enlargement observed with carbapenems and cephalosporins[112,334] and compaction of the nucleoid observed with antimicrobials targeting the bacterial ribosome[335]. In comparison to previous HCI studies, I have identified very similar morphological changes in this *K. pneumoniae* isolate to those previously reported by Nonejuie *et al* in *E. coli* in response to a range of antimicrobial classes[103]. However, I have simplified the methodology used by Nonejuie *et al* by removing centrifugation steps and imaging bacteria directly in wells rather than on agarose pads, facilitating high sample throughput.

Although HCI has largely been used in compound screening and drug development in bacterial research to date[97], this platform presents exciting opportunities to study a range of phenotypes beyond antimicrobial susceptibility, including assays of biofilm production, therapeutic antibodies and phagocytosis.

Of course, further work is required using a wider range of stains, bacterial isolates and growth conditions to determine additional possibilities of this method. There are multiple challenges to using HCI routinely for bacterial research. As demonstrated here, some species and isolates did not adhere well, or at all, to the plate, leading to a reduction in image quality and the total number of bacteria successfully captured for further analysis. Of note, isolates known for their adhesive properties (such as clinical isolates of *K. pneumoniae*) are more likely to adhere than laboratory-adapted or facultative intracellular bacteria (including *S.* Typhimurium). *K. pneumoniae* possess multiple proteins that facilitate

adhesion to plastic and other surfaces, contributing to their ability to thrive as a hospital associated pathogen[336,337]. While plate coatings can improve bacterial adhesion, further work is required in a wider range of isolates. Also, the effect that certain coating matrices may have on bacterial morphology, such as the aggregative phenotype on certain coatings in this study, needs to be considered and further evaluated.

The combination of stains used in this study, previously employed by Nonejuie et al[103], were successfully applied to all isolates tested. Using separate stains that target nucleic acids (DAPI) and the cell wall (FM4-64), the direct or indirect targets of many antimicrobials, enabled the reliable distinction of morphological changes in the presence of each compound. While individual parameters could distinguish between treated and untreated bacteria at an aggregate level for certain antimicrobials, for others the distinction relied on small variations in a wide range of variables that needed to be aggregated using multidimensional principal component analyses. This emphasises the utility of this approach, taking large numbers of measurements to make distinctions that are not obvious using traditional phenotyping techniques.

There are additional technical challenges specifically associated with image analysis. Segmentation of individual cells is challenging in fields of dense bacterial growth; this overcrowding may also limit the ability of bacteria to develop morphological changes. Optimising initial bacterial inocula is therefore necessary but, due to intra-species variation in doubling time, requires additional preparatory work which may impact high throughput studies on uncharacterised organisms. Additionally, because this method works at the limit of resolution of the Opera Phenix, it can be challenging to distinguish artefacts from unusual bacterial phenotypes that may be of interest, such as persister cells. Despite these challenges, and using only tools that were developed for eukaryotic cell analysis, I have developed pipelines that are capable of segmenting both rod- and coccus-shaped bacteria.

Though the analyses in this study were performed using proprietary software from PerkinElmer, including the image analysis tools Harmony and QC package in TIBCO Spotfire, there are open access software options available, such as CellProfiler[338] and Cellpose[339], which have similar analysis capabilities. In the same way that advances in bioinformatics have made genomic analysis of pathogens faster, easier and more widely available, equivalent advances in the field of imaging are likely to produce similar results in time.

Live bacterial imaging presents exciting opportunities for the study of bacterial response to antimicrobials, including the identification of individual cells that persist in the face of antimicrobial pressure. The experiments presented in this study are only provisional, but highlight how GFP-expressing bacteria can be used to identify cellular replication and morphological changes over time after exposure to ciprofloxacin. Further work is required to optimise bacterial adhesion (with consideration of any impact on morphology), staining (to improve the morphological characterisation of isolates with minimal impact on growth) and image analysis (complicated by factors such as cell division and changes in adhesion over time). Even simple variables require further investigation, such as the optimal interval between image captures: too frequent and the camera will not be able to image further fields elsewhere in the same plate, limiting the number of growth conditions that can be studied in parallel; too infrequent and the software is likely to have difficulty in tracking individual cells over the time course as they change morphology and/or replicate. Such investigations were planned as part of my original PhD fellowship, but have not been possible due to the COVID-19 pandemic.

There are other considerations, common to all HCI studies, that need to be addressed prior to implementation of this approach[99]. The first relates to the large quantity of data generated and analysed, raising the possibility of statistical errors and spurious correlations which

complicate analyses. While data are generated on an individual bacterial level, the majority of the analyses have been conducted on aggregated well-based means and measures of variance. This is due to the computational power and storage required to process data on hundreds of thousands of individual cells. Also, while bacterial genomic data has long been made publicly available for validation and re-analysis, there are challenges in replicating this approach with imaging studies due to the associated file sizes[340]. While desirable, the practical aspects of publishing studies with image and, ideally, linked genomic data in pathogen research remain unclear. Finally, while there is excellent within-plate correlation between wells under the same growth conditions (as illustrated by the tight clustering on the PCA plots in this chapter), the same is not necessarily true for separate plates in the same experiment. This is likely due to small perturbations in experimental techniques, including growth conditions, bacterial inocula and imaging procedures, which can be minimised but not eliminated. A variety of factors are known to influence the cellular morphology of *Enterobacteriaceae*, including growth conditions[341], cell age[342], interaction with host cells[343] and a range of antimicrobials[334,335,344,345]. Care must therefore be taken to set up appropriate controls within each plate to account for plate effects and other systematic artefacts[96,99].

Nevertheless, these methods have demonstrated the possibility of expanding microbial phenotyping from population-level to single-cell analysis. This has particular relevance, as demonstrated here, in the study of AMR. Other HCI techniques have been used in compound screening to identify both novel small molecules and their mechanism of action[97]. This could be expanded to large compounds and/or collections of bacterial isolates under standardised growth conditions. HCI may also be adapted for high throughput antimicrobial susceptibility testing of clinical isolates[346]. The technique has already been adapted by colleagues in the same laboratory to use HCI to screen a therapeutic antibody candidate against a collection of clinical *E. coli* isolates[105]. Because of the flexibility and scalability of the Opera Phenix system, this could be expanded to assess synergistic effects between

libraries of antimicrobials, monoclonal antibodies and other novel therapies against multidrug resistant bacteria, an approach that would be highly challenging using other platforms. The potential for combining this with live bacterial imaging to characterise the variation in effects on bacteria, and identify persister cells and therefore potential resistance mechanisms, is a promising asset in the fight against AMR.

5. Genotypic and phenotypic responses to antimicrobials in *K. pneumoniae*

5.1 Introduction

In **Chapter 3** I demonstrated the substantial differential expression of genes in a single clinical isolate of *K. pneumoniae* in response to ciprofloxacin, identifying a number of upregulated genes that may contribute to antimicrobial resistance. However, it remains unclear whether this is a fluoroquinolone-specific effect, or whether it can be reproduced by other antimicrobials or environmental stressors. Similarly, all of the provisional RNAseq experiments were performed using a single clinical isolate; it is unclear whether these findings can be replicated in other *K. pneumoniae* strains. In particular, it is uncertain whether the substantial differential expression of multiple resistance genes identified on an IncF plasmid in strain KC006 in response to ciprofloxacin is unique to this isolate or replicon type.

HCI, as described in **Chapter 4**, presents an excellent opportunity to add detailed phenotyping approaches to the toolkit of investigations for AMR. However, given the diversity of genotypes and resistance mechanisms harboured by an increasing variety of bacterial species and lineages, it is important to optimise the technology for a range of pathogens. *K. pneumoniae* presents an ideal species to investigate this challenge: in addition to the genetic diversity of the species, all provisional experiments to date have

demonstrated reliable adherence to the plates used in HCI, enabling imaging with greater ease and higher throughput. Additionally, it is unclear how the phenotypic changes induced by antimicrobials relate to transcriptomic changes, such as those demonstrated in **Chapter 3**.

In this chapter, I first investigated the impact of antimicrobials on clinical isolates of *K. pneumoniae*, using transcriptomic approaches in combination with high content imaging. Using RNAseq I demonstrated how the differential expression induced by ciprofloxacin is distinct from other antimicrobial classes in the clinical isolate KC006. I then explored the impact of ciprofloxacin on a wider range of *K. pneumoniae* isolates, again using RNAseq. Using the results of these experiments, I attempted to correlate the transcriptomic and imaging findings to better understand the association between the phenotypic and genotypic changes associated with antimicrobial exposure.

To better understand the impact of antimicrobials on *K. pneumoniae*, I assembled a physical collection of 175 isolates representative of the diversity of the species, derived from a number of studies that have characterised these isolates using whole genome sequencing and antimicrobial susceptibility testing. I used this collection for high content imaging experiments, exposing all 175 isolates to four different growth conditions: treatment with ciprofloxacin, meropenem and tigecycline, in comparison to an untreated control. These experiments were repeated in duplicate at MIC and 4x MIC, to explore whether the morphological changes identified vary in a dose-dependent response. Finally, I outline an approach using quantitative PCR to demonstrate how transcriptomic changes identified using RNAseq can be replicated at scale using the Fluidigm platform.

5.2 Differential gene expression of isolate KC006 in response to antimicrobials

In **Chapter 4** I demonstrated the phenotypic changes in clinical isolate KC006 in response to seven antimicrobials, shown in **Figure 4.10**. In parallel, I used bacteria harvested from the same experiment in identical growth conditions for further RNAseq analysis, to obtain paired phenotypic and genotypic data following exposure to antimicrobials. In brief, overnight cultures of KC006 were diluted into 9 mL of fresh IsoSensitest broth (in separate tubes for each antimicrobial) and incubated for 90 minutes in a shaking incubator at 37°C to ensure that the bacteria were in early exponential growth phase. At this point, 1 mL of antimicrobial was added to each sample; 100 µL was then taken and added to each of 3 wells on a 96 well Cell Carrier Ultra plate for subsequent imaging. Both the plate and the remaining sample were incubated together in the same incubator for a further 2 hours, when the plate was fixed and 1 mL of the remaining sample was added to RNAprotect for later extraction and sequencing. I have assumed that the growth conditions in the well and the test tube are similar to enable a comparison between phenotype and genotype. Additionally, separate untreated samples were incubated at 45°C to determine the impact of heat, a non-pharmacological stressor, in comparison to antimicrobials. The temperature of 45°C was selected based on previous studies investigating the impact of heat on transcription in *E. coli* and on growth in *K. pneumoniae*[347,348]. Each experiment was performed in triplicate.

Representative images and a PCA plot comparing the phenotypic measurements for each growth condition have already been shown in **Figure 4.10**. This demonstrated that each antimicrobial induces a phenotypic change that clearly distinguished it from a control, and from each other. However, antimicrobials with related mechanisms of action clustered more tightly, demonstrating a shared phenotypic change.

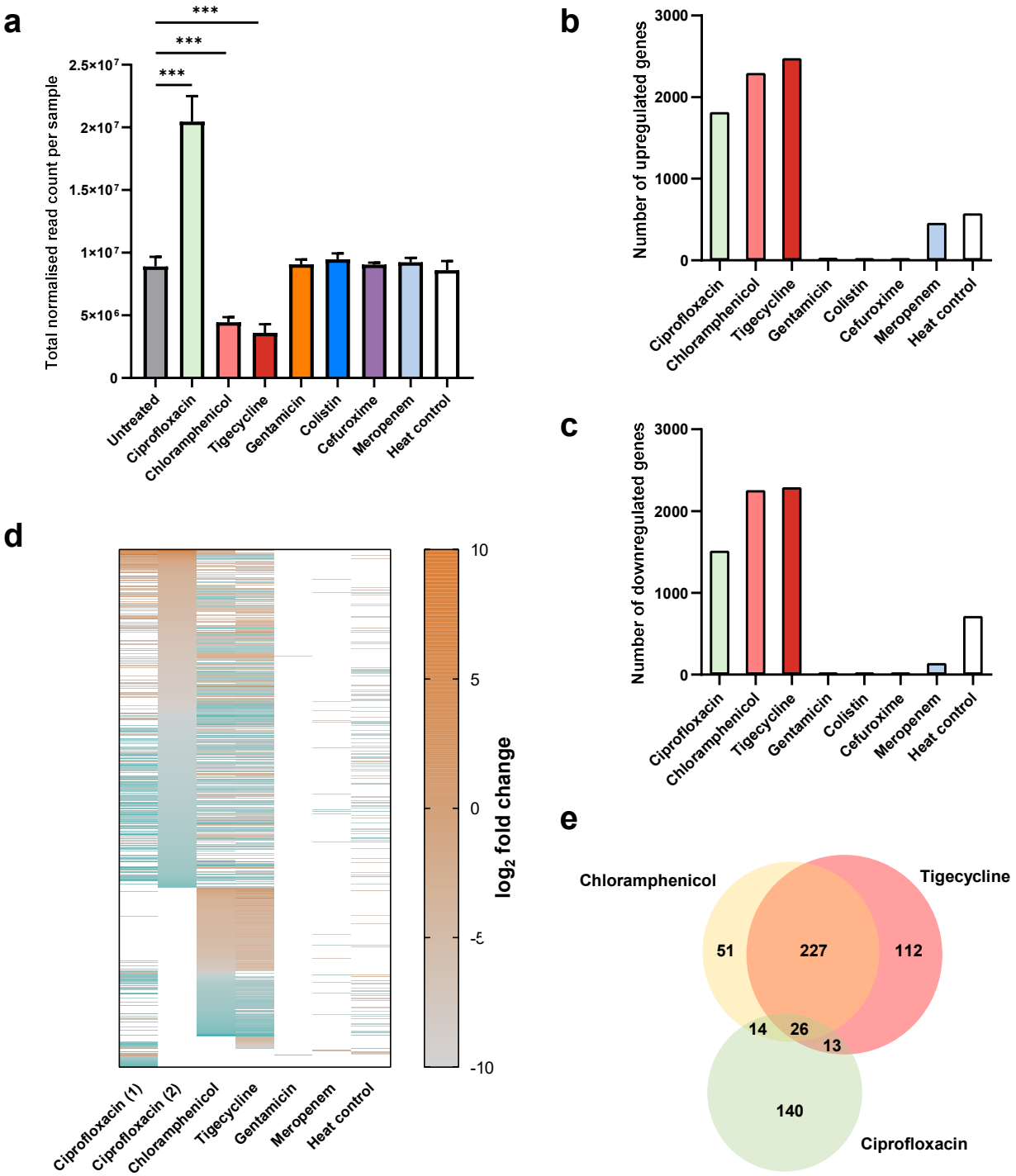


Figure 5.1 RNAseq results for isolate KC006 in the presence of antimicrobials

(**a**) Total normalised mapped read count per sample for each antimicrobial condition. Bar height represents the mean of three biological replicates; error bars represent standard deviation. Significant differences from controls are annotated (P<0.001, two tailed t test). All other comparisons with untreated controls were non-significant. (**b-c**) Total number of significantly upregulated (**b**) and downregulated (**c**) differentially expressed genes in each condition relative to an untreated control. All genes with adjusted P <0.05 at any $\log_2$ fold change in expression are included. (**d**) Heat map demonstrating significantly differentially expressed genes from (**b**) and (**c**). Genes are sorted by $\log_2$ fold change by Ciprofloxacin (2). Previous ciprofloxacin experiment is included for comparison (Ciprofloxacin (1)). Genes that not significantly differentially expressed in each growth condition (adjusted P<0.05) are coloured white. Only genes where at least one growth condition induced a minimum fourfold change in expression are included. Cefuroxime and colistin are omitted due to the limited number of differentially expressed genes and to aid visualisation. (**e**) Quantitative Venn diagram showing the overlapping upregulated genes in bacteria treated with ciprofloxacin (repeat experiment), chloramphenicol and tigecycline, with a minimum $\log_2$ fold change of 2.

Results of differential gene expression analyses for this experiment are shown in **Figure 5.1**. The total normalised mapped read count, which approximates the relative overall level of transcription for each antimicrobial condition in comparison to an untreated control, showed that ciprofloxacin induced a more substantial relative increase in overall transcription in comparison to all other antimicrobials (**Figure 5.1a**). By contrast, tigecycline and chloramphenicol, which share the same cellular target (the bacterial ribosome), induced a significant reduction in overall transcription. There was no significant difference in overall gene expression for any other growth condition in comparison to the untreated control. Similarly, it was ciprofloxacin, tigecycline and chloramphenicol that induced the largest number of differentially expressed genes, both for those that were upregulated (**Figure 5.1b**) and downregulated (**Figure 5.1c**). While there were a large number of differentially expressed genes in the presence of meropenem (n=603) and the heat control (n=1294), there was a much more modest impact of gentamicin (n=38), cefuroxime (n=4) and colistin (n=2), despite a clear phenotypic change evident for all antimicrobials tested (**Figure 4.10**).

Figure 5.1d is a heatmap plotting the individual genes that were differentially expressed for each antimicrobial, sorted by the ciprofloxacin condition in this experiment ("Ciprofloxacin (2)"). The results of the previous ciprofloxacin RNAseq data from **Chapter 3** are shown for comparison ("Ciprofloxacin (1)"). While there were some differences, overall there is a clear trend that upregulated genes in the first experiment were also upregulated in the second, showing reproducibility in results. Of note, many of the most highly upregulated genes were the same in both experiments, including genes associated with DNA damage and repair (*dinI*, *recN*, *yebG*), low fidelity DNA polymerases (*umuC*, *umuD* and *ssb*, as well as their repressor *lexA*) and a number of genes encoding hypothetical proteins of unknown function (**Table 3.3**, **Appendix Table D.1**). Similar findings were also present on the IncF plasmid, including upregulation of a number of genes associated with plasmid partitioning and conjugation (such as *traM*) and antimicrobial resistance (including *qnrB*, *cat* and *bla*).

By contrast, chloramphenicol and tigecycline expressed a distinct pattern of differential expression in comparison to ciprofloxacin, but one that is largely similar to each other. As shown in the heatmap in **Figure 5.1d** and the quantitative Venn diagram in **Figure 5.1e**, there was a considerable overlap in the most upregulated genes for tigecycline and chloramphenicol, which was largely distinct from those for ciprofloxacin. This likely reflects their common cellular target and may also contribute to their similar phenotypic appearances.

The most upregulated genes associated with tigecycline are shown in **Figure 5.2c** and **Appendix Table D.2**. The most substantially differentially expressed genes were *tetA*, which encodes the tetracycline efflux pump, and its neighbouring gene *yedA*, which encodes another drug/metabolite transporter. Other highly expressed genes included *smvA*, encoding an efflux pump that confers resistance to chlorhexidine, other multidrug efflux pumps (*mdtE*, *mdtK*) and the porin *ompK35*, implicated in resistance to beta lactams. As may be expected from a drug targeting the bacterial ribosome, a number of upregulated genes were related to ribonucleotide processing and ribosomal function (such as *nrdF* and *rlmA*). As with ciprofloxacin, some of the most highly upregulated genes encode hypothetical proteins of unknown function.

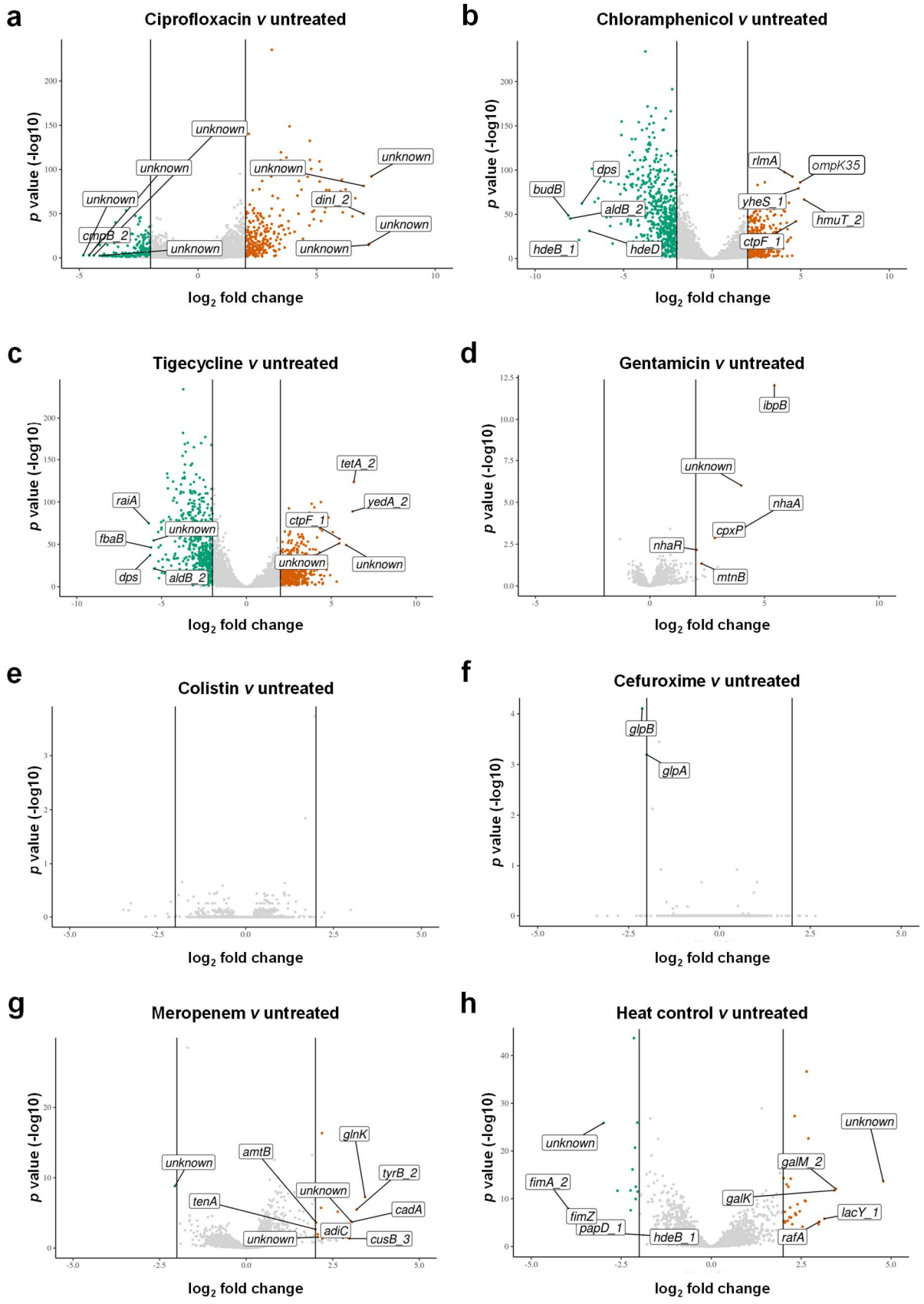


Figure 5.2 Differential expression of isolate KC006 in the presence of antimicrobials

(**a-h**) Volcano plots demonstrating differential expression of genes following exposure to antimicrobial treatment at MIC, in comparison to an untreated control at 2 hours incubation. Growth conditions included are exposure to ciprofloxacin (**a**, using the values from the second experiment), chloramphenicol (**b**), tigecycline (**c**), gentamicin (**d**), colistin (**e**), cefuroxime (**f**), meropenem (**g**) and a heat control (**h**, at a temperature of 45°C). Points are coloured where the adjusted *P* <0.05 and $\log_2$ fold change in expression is ≥2 or ≤-2. Up-regulated genes are coloured in orange; down-regulated genes in green. The genes with the greatest $\log_2$ fold change in expression are annotated. Where genes are labelled as "unknown", they encode hypothetical proteins of unknown function.

Specifically addressing the IncF plasmid in KC006, tigecycline induced an even greater degree of differential expression than ciprofloxacin. There was significant upregulation of the *tra* group of conjugal transfer genes, including a number of neighbouring genes encoding proteins of unknown function (**Figure 5.3**). Additionally, while both ciprofloxacin and tigecycline led to upregulation of *tetA* and its repressor *tetR*, tigecycline additionally led to upregulation of two neighbouring uncharacterised genes, as well as downregulation of the neighbouring resistance genes *qnr*, *cat*, *bla* and *aacA-aphD*. This demonstrates that different antimicrobial classes can yield unique effects on known antimicrobial resistance mechanisms on the same plasmid.

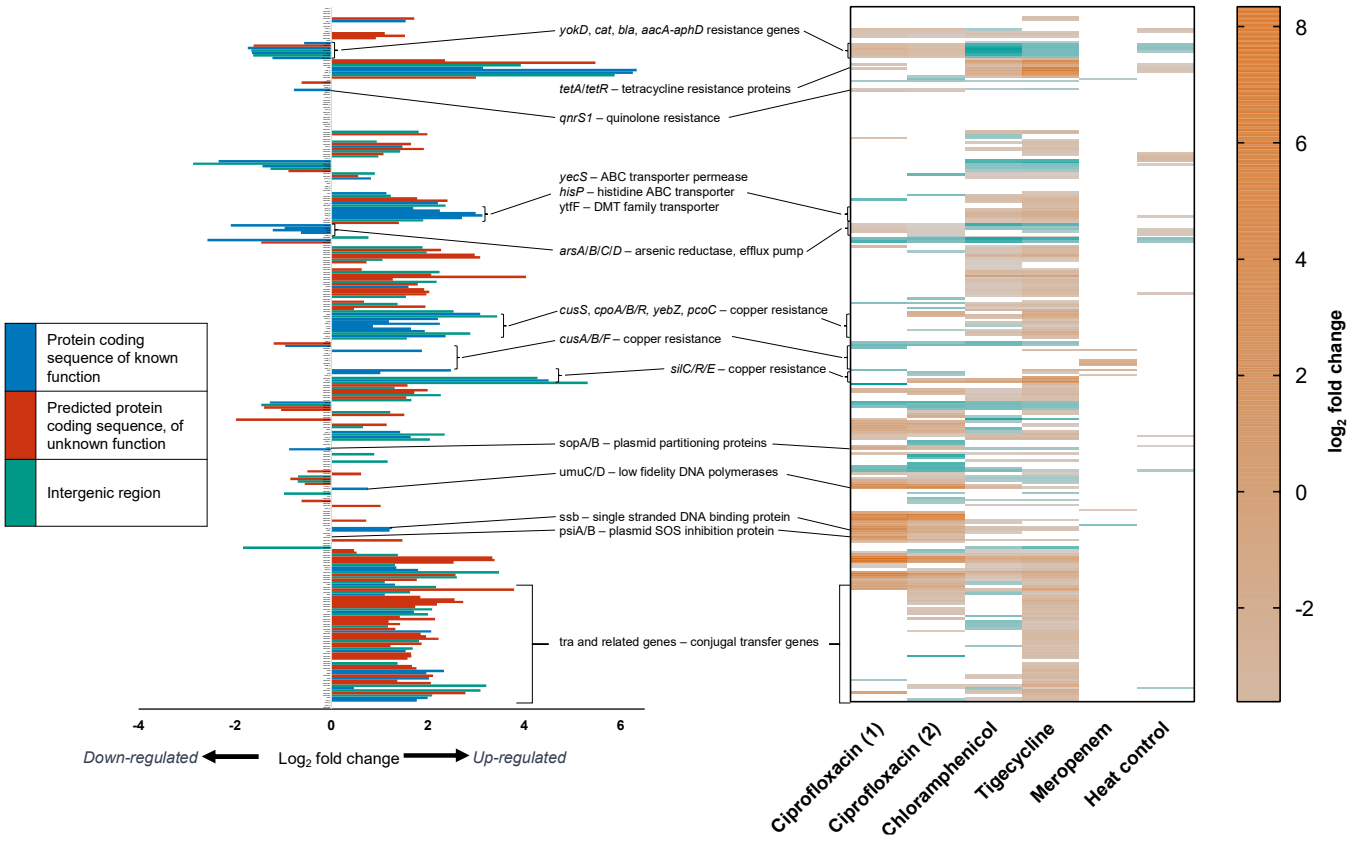


Figure 5.3 Linear map of the 198 kbp plasmid in isolate KC006, with associated differential gene expression results in the presence of antimicrobials

Differential gene expression results compare antimicrobial treatment at MIC with an untreated control after 2 hours of incubation. The length of the coloured bars on annotated on the plasmid map represent the $\log_2$ fold change in expression in the presence of tigecycline, with no limits on significance. Note that the *y* axis represents position relative to other genes and is not to scale, but does reflect the distribution of genes in the associated heat map. Genes of interest are annotated.

As noted above, there was considerable overlap in the differentially expressed genes for chloramphenicol and tigecycline (**Figure 5.2b** and **Appendix Table D.3**), including upregulation of genes encoding porins (*ompK35*), efflux pumps (*mdtE*, *mdtK* and *smvA*), and a range of genes involved in ribonucleotide processing. On the plasmid there is a similar effect of upregulation of *tetA* and *tetR* and downregulation of *qnrB*, *bla* and *aacA-aphD*. Among these resistance genes was the chloramphenicol resistance gene *cat*, encoding the chloramphenicol acetyltransferase, which is also downregulated (log_2 fold change -3.1, adjusted $P = 2.7x10^{-55}$), showing that the transcriptomic changes in resistance genes induced by an antimicrobial are not specific to that drug. In comparison to tigecycline there were much more modest increases in genes encoding the plasmid conjugation machinery, but there remained significant upregulation of *traH*, *traI* and *traD*. As with ciprofloxacin, both drugs show upregulation of both chromosome and plasmid-encoded *umuC/D* and *lexA*.

By contrast, very few differentially expressed genes in the presence of meropenem were associated with these mechanisms of resistance (**Figure 5.2g**, **Appendix Table D.4**). Many of the differentially expressed genes were associated with nitrogen regulation, such as *gln*. Among the most upregulated genes were the plasmid encoded *cusB* and *cusA*, which contribute to the efflux pump CusCFBA. These copper and silver resistance genes are strongly associated with beta lactam resistance as well as an emerging role in tigecycline resistance[349]. While there was also upregulation of *smvA* and the multidrug resistance gene *mdtH*, there was no significant change in expression for other genes implicated in resistance with other antimicrobials, including *umuC/D*, the plasmid conjugation *tra* genes or known antimicrobial resistance mechanisms, including both the chromosomal and plasmid encoded beta lactamases.

Although gentamicin shares the same cellular target as tigecycline and chloramphenicol, the bacterial ribosome, and caused a similar degree of phenotypic change during HCI, there was

a much reduced impact on transcription in comparison to these other antimicrobials (**Figure 5.2d**, **Appendix Table D.5**). The most substantially upregulated gene was *ibpB*, encoding the 16 kDa heat shock protein B, while a number of other genes were involved in ribonucleotide pathways and ribosomal function. None of the known or putative resistance mechanisms discussed above were significantly differentially expressed, nor were any genes on the KC006 IncF plasmid.

For the cefuroxime condition, no genes were significantly upregulated and only 3 genes with known function were significantly downregulated (**Figure 5.2f**). These were all involved in the anaerobic catabolism of glycerol (*glpA*, *B* and *C*). A similar phenomenon has previously been described in RNAseq approaches studying the response of *Klebsiella* to polymixin[350]. This is a clear example of prominent phenotypic change occurring in the absence of a strong transcriptomic impact on bacteria. Similarly, only one gene was significantly differentially upregulated in the presence of colistin, encoding the toxic peptide TisB. This gene was also upregulated in the presence of ciprofloxacin and chloramphenicol, and is under transcriptional control of the protein LexA, which also regulates genes implicated in resistance such as *umuC/D*.

In addition to antimicrobial pressure, I also studied the impact of heat on isolate KC006, to distinguish the impact of antimicrobials from those of a non-pharmacological stress response. There was a distinct pattern of differential expression for heat in comparison to any antimicrobial tested (**Figure 5.2h**, **Appendix Table D.6**). The majority of upregulated genes were not related to known or putative resistance mechanisms and were dominated by heat shock proteins and sugar transporters. On the plasmid, there was upregulation of the *lac* operon (*lacY* and *lacZ*, encoding enzymes in galactoside metabolism), genes that are

associated with carbapenem resistance, although their precise role is unclear[351]. Other genes known to be involved in resistance, including *qnrB*, *umuC* and *umuD*, were not differentially expressed; there was a modest downregulation of *cat* and *bla*, an increase in *tetA*, and a small downregulation in *traT*, but none of the other plasmid conjugation genes (**Figure 5.3**).

5.3 Differential gene expression using other *K. pneumoniae* isolates

To further address the impact of fluoroquinolones on *K. pneumoniae*, and to determine whether the findings using RNAseq in clinical isolate KC006 could be replicated in other lineages, further work was conducted on five clinical isolates and the reference isolate ATCC 43816 (**Table 5.1**). All five clinical isolates were obtained from a prospective surveillance study undertaken at CUH in 2016[31]. The isolates were selected based on: i) ciprofloxacin MIC (this had to be ≤32 μg/mL, due to concerns for highly resistant isolates not mounting any transcriptomic/phenotypic response); ii) possessing at least one Inc plasmid (as determined by PlasmidFinder); and iii) possessing a number of plasmid encoded resistance genes. The final selection included isolate KC036, characterised in **Chapter 3**, as an ST45 isolate that possessed a similar IncF plasmid but fewer resistance genes and had a thousand-fold lower ciprofloxacin MIC. Two further isolates possessed IncF plasmids, but were derived from different sequence types and therefore different genetic backgrounds. The small collection also possessed isolates with a wider range of plasmid types, including IncN, IncHI2 and IncI1. ATCC 43816 was included as it possesses no detectable plasmids, but has produced clear phenotypic changes in the presence of ciprofloxacin, as described in **Chapter 4**.

All of the common fluoroquinolone resistance genes and mutations were determined using Kleborate. **Table 5.1** shows the wide variation in MIC (ranging from 1 to 64 μg/mL) in the presence of only a single *qnrB1* gene, suggesting that there other mechanisms that are not properly understood that may contribute to resistance.

Table 5.1 – Characteristics of *K. pneumoniae* isolates used in RNAseq

MLST and ciprofloxacin resistance genes were determined using Kleborate. Plasmid types were predicted using PlasmidFinder. Ciprofloxacin MICs were determined using Etests, with the exception of KC006 which was determined using agar and broth dilution.

Isolate	Source	MLST	Plasmid type	Ciprofloxacin MIC	Ciprofloxacin resistance genes
KC006	Human, blood	45	IncFIB(K - Kpn3) IncFII(K)	64	*qnrB1*
KC036	Human, blood	45	IncFIB(K - Kpn3) IncFII(K_CP000966) IncFII(pKP91) IncN3	0.032	-
KC153	Human stool/rectal swab	251	ColRNAI IncFIB(K - Kpn3) IncFII(K)	1	*qnrB1*
KC154	Human, blood	240	IncFIA(HI1) IncFIB(K - Kpn3) IncFII(K) IncFII(pMET) IncHI2 IncN	8	*qnrB1*
KC155	Human, blood	4023	Col156 IncI1 IncN	0.38	*qnrS1*
ATCC 43816	Unknown	493	-	0.032	-

MLST = multilocus sequence type; MIC=minimum inhibitory concentration

In keeping with previous RNAseq experiments using isolate KC006, each isolate was grown overnight, inoculated into fresh IsoSensitest broth for 90 minutes of incubation prior to the addition of ciprofloxacin at MIC, and incubated for a further 2 hours. Samples were taken for RNAseq and HCI from the same source culture. The experiment was performed in biological triplicate. To ensure accurate mapping of reads produced during RNAseq, especially to genes encoded on plasmids, hybrid assemblies were produced for the four additional clinical isolates in this experiment using a combination of short reads from Illumina sequencing and long reads from Oxford Nanopore sequencing, assembled using Unicycler[279]. The same PacBio assembly was used for isolate KC006, as previously. A published reference genome from GenBank was used for ATCC 43816[352]. Reads were mapped back to the appropriate reference assembly and analysed using DEAGO[300].

A summary of differential gene expression for each isolate in the presence of ciprofloxacin at MIC is shown in **Figure 5.4**. Two sets of results are shown for KC006: one from the experiment performed in **Chapter 3** (at 2 hours incubation and at MIC, in keeping with the other experiments here) and a second from the repeat experiment outlined above, in which KC006 was exposed to multiple antimicrobials in parallel in otherwise identically controlled conditions. This demonstrates a major and reproducible effect of ciprofloxacin on transcription in isolate KC006. However, there was a more modest transcriptional response in isolates KC036 (a total of 58 differentially expressed genes), KC153 (50 genes) and ATCC 43816 (258 genes), with a comparably minor response in isolates KC154 (17 genes) and KC155 (16 genes). The reason for this difference in expression was not immediately clear; potential explanations are explored below.

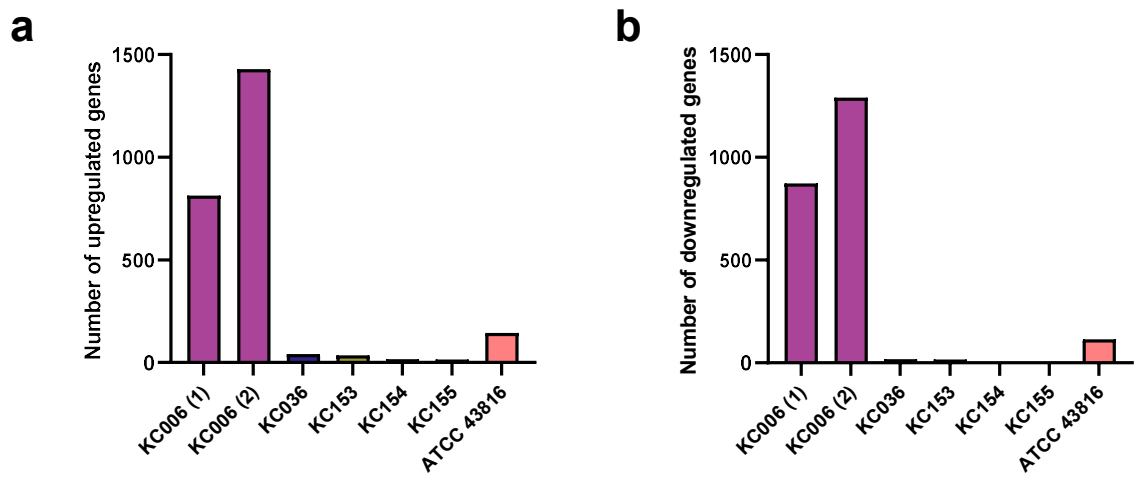


Figure 5.4 Differential gene expression in *K. pneumoniae* isolates following ciprofloxacin exposure
(**a-b**) Bar plots showing the total number of significantly upregulated (**a**) and downregulated (**b**) differentially expressed genes in each isolate in the presence of ciprofloxacin at MIC, relative to an untreated control. All genes with an adjusted *P* <0.05, at any $\log_2$ fold change in expression are included. Bars are coloured by isolate. The inclusion of two bars for isolate KC006 reflects the inclusion of (1) the first RNAseq experiment detailed in **Chapter 3**, and (2) the repeat experiment performed here for KC006 in the presence of different antimicrobials.

Despite the overall difference in expression between isolates, there was considerable overlap in those genes that were significantly upregulated for both chromosome- and plasmid-encoded genes when they were identified in that isolate's genome (**Figure 5.5**, **Appendix Tables D7-D11**). The gene *recN*, encoding a DNA repair protein, was the only chromosomally encoded gene that was significantly differentially expressed in all 6 isolates. However, other significantly upregulated genes in the majority of isolates, including those most highly upregulated in KC006, were *lexA* (the regulator of *umuC/D*), *dinF* (SOS response protein), *tisB* (toxic peptide) and *sbmC* (DNA gyrase inhibitory protein). Additionally, the gene *sulA* was upregulated in 4 isolates. This gene forms part of the SOS response in Gram-negative bacteria and inhibits cell division, blocking FtsZ ring formation (*ftsZ* itself is significantly downregulated in KC006). This may contribute to the filamentous elongation phenotype observed when imaging *Klebsiella*.

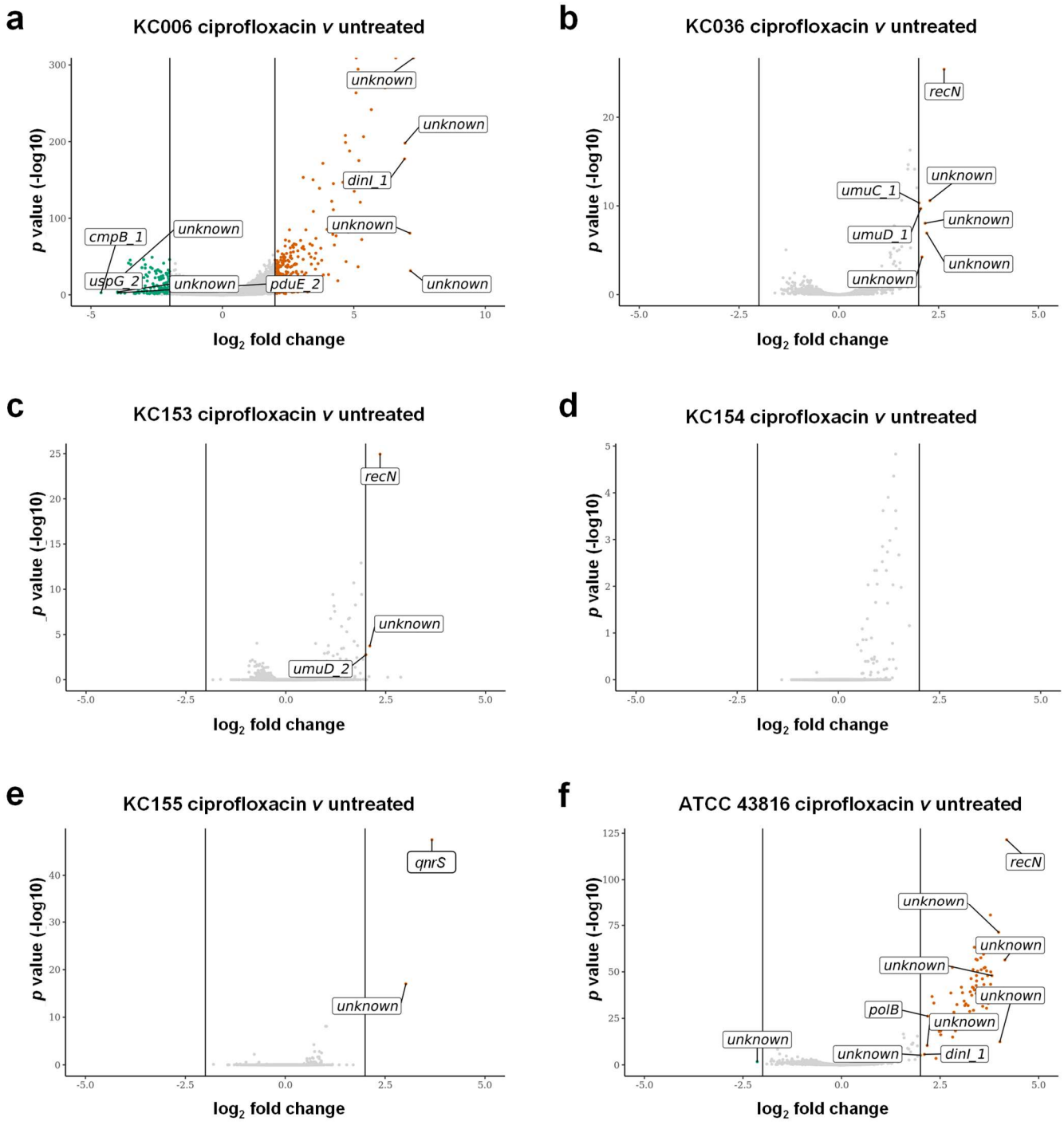


Figure 5.5 Differential gene expression of *K. pneumoniae* isolates in the presence of ciprofloxacin

(**a-f**) Volcano plots demonstrating differential expression of genes following exposure to ciprofloxacin treatment at MIC, in comparison to an untreated control at 2 hours incubation. Isolates shown are KC006 (**a**, using the values from the second experiment), KC036 (**b**), KC153 (**c**), KC154 (**d**), KC156 (**e**), and ATCC 43816 (**f**). Points are coloured where the adjusted *P* <0.05 and $\log_2$ fold change in expression is ≥2 or ≤-2. Up-regulated genes are coloured in orange; down-regulated genes in green. The genes with the greatest log2 fold change in expression are annotated. Where genes are labelled as “unknown”, they encode hypothetical proteins of unknown function.

At least one copy of the genes *umuC* and *umuD*, either on the chromosome or on the plasmid, was also upregulated in all isolates. This included genes present in multiple plasmid types, including IncF (KC006, KC153), IncN (KC154, KC155) and IncI1-I(Alpha) (KC155). Of the plasmids encoding known resistance genes, *qnrS* is strikingly upregulated (log_2 fold change 3.67) in the IncF plasmid of KC155. However, no other genes on this plasmid, including the plasmid conjugation genes and a number of other resistance genes, such as the ESBL CTX-M-15, were differentially expressed.

In isolates that possessed an IncF plasmid, none were associated with upregulation of the plasmid conjugation *tra* gene family apart from KC006. However, ciprofloxacin led to substantial upregulation of a type IV secretion system (associated with DNA transfer between bacteria) and the plasmid mobilisation genes *mobB* and *mobC* on the IncN plasmid of isolate KC154. Additionally, ciprofloxacin induced substantial upregulation of the colicin-encoding gene *ceaC* and its associated immunity protein in the Col plasmid in isolate K153. There were no known resistance genes or pathways substantially upregulated in any of the other isolates that had not already been identified in isolate KC006, with the possible exception of a number of prophage genes in ATCC 43816. A number of genes encoding hypothetical proteins of unknown function, some of which correspond to homologous genes in KC006, were upregulated in other isolates.

5.4 Comparison of differential gene expression and phenotypic changes using HCI

A PCA plot demonstrating the phenotypic changes associated with ciprofloxacin exposure in this small collection of isolates is shown in **Figure 5.6**. All technical and biological replicates

are shown in the same plot, including ciprofloxacin treated wells (lighter shades) and untreated controls (darker shades) on the same axes. The untreated wells cluster together by isolate. However, in comparison to plots shown in **Chapter 4** which included only a single isolate, there is a wider distribution of the morphological features between isolates, demonstrating that untreated bacteria had a slightly different phenotype. While ciprofloxacin treated wells also cluster together, again separately by isolate, there is a general trend for ciprofloxacin treated wells to diverge from their untreated controls in the same direction (i.e. towards the origin of the *x* axis, PC1). The extent of phenotypic change, illustrated by wider divergence between untreated and ciprofloxacin treated wells in the same isolate, appeared to mirror the transcriptomic changes identified in these isolates above: it is greatest for isolate KC006, modest for isolates KC036, KC153 and ATCC 43816, and smallest for isolates KC154 and KC155.

To further investigate the relationship between transcriptomic and phenotypic changes, correlations were made between commonly upregulated genes and morphological changes associated with ciprofloxacin exposure. First, I examined the correlation between *recN*, the only gene significantly upregulated in all isolates, and a range of simple morphological properties with significant Z' scores for differentiating ciprofloxacin-treated bacteria from their untreated controls. The scatter plots in **Figure 5.7a-c** show a significant correlation between differential *recN* expression and bacterial length, area and roundness. There is a weaker correlation between bacterial length and other commonly upregulated genes in the face of ciprofloxacin exposure, namely *umuC*, *sulA* and *tisB* (**Figure 5.7d-f**). This supports the finding above that transcriptomic and phenotypic changes under ciprofloxacin exposure are related. However, the sample size here is small, limiting interpretation.

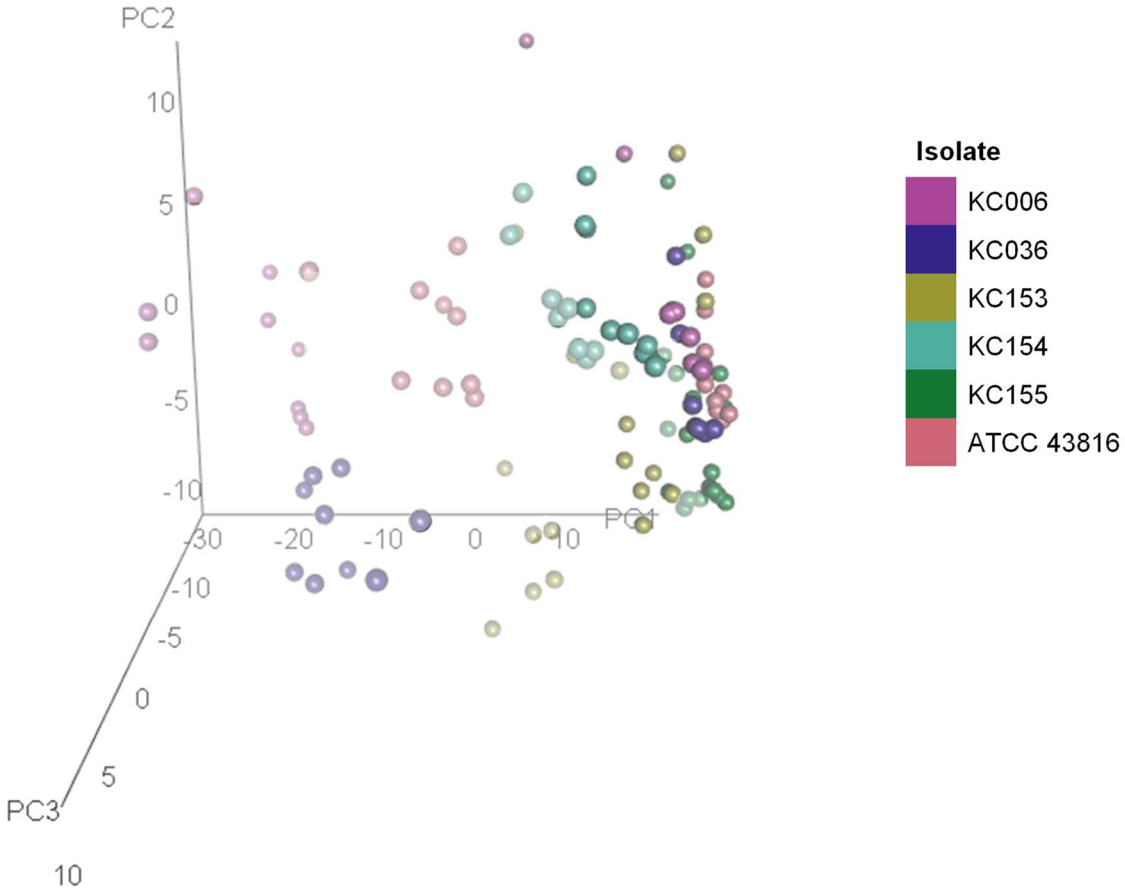


Figure 5.6 Phenotypic effects of ciprofloxacin on isolates of *K. pneumoniae*

Three-dimensional principal component analysis of the mean and standard deviation values of morphological properties measured for single bacterial cells in each well. Technical and biological triplicate repeats are shown. Each point represents a well; lighter shades indicate ciprofloxacin treated isolates and darker shades indicate untreated controls.

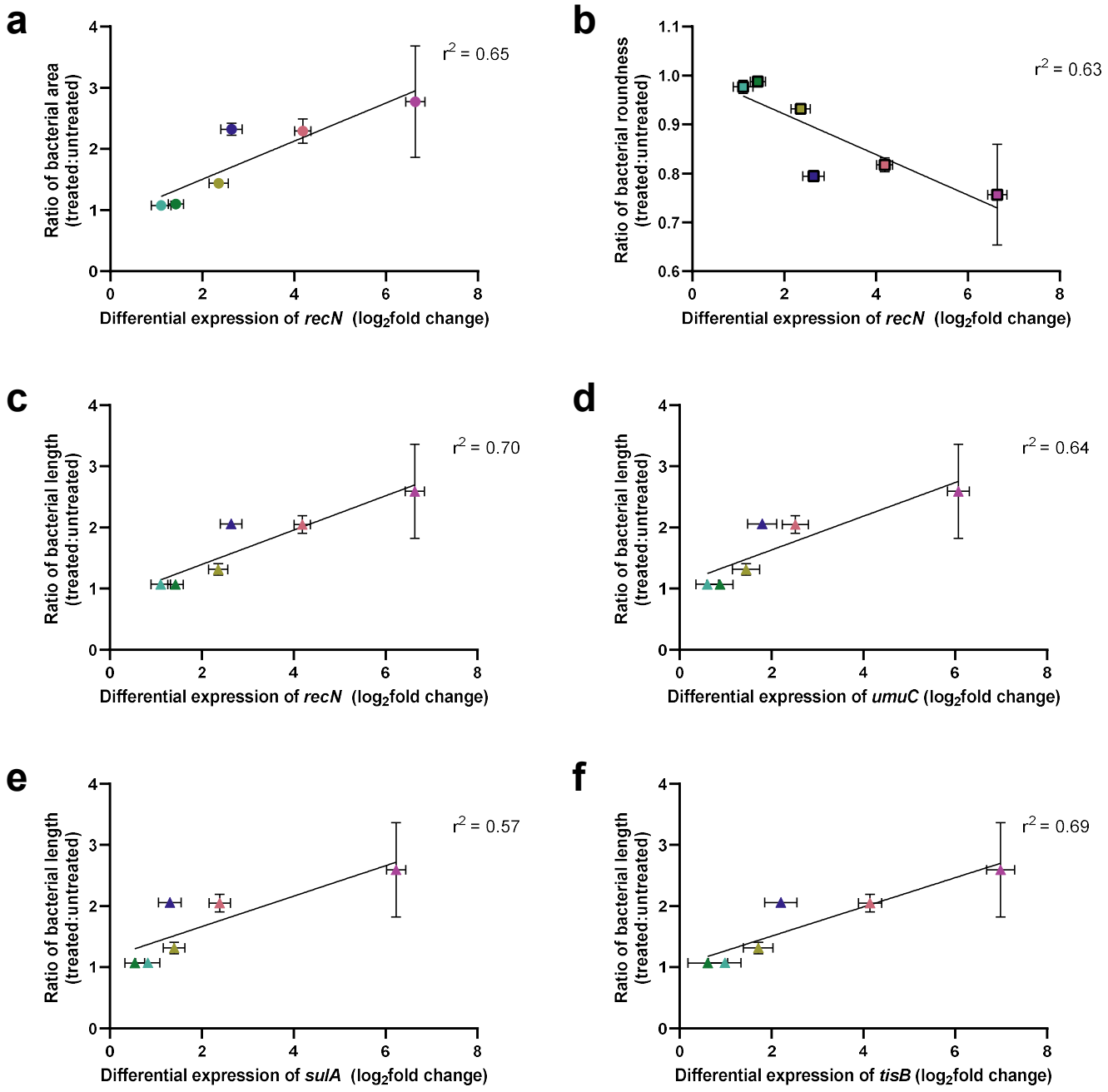


Figure 5.7 Comparing differential gene expression and phenotypic changes for multiple isolates following exposure to ciprofloxacin

(**a-c**) Scatter plots comparing the differential expression of *recN* with morphological properties associated with high Z' scores in high content imaging. These features are expressed as the ratio of the measured phenotype in treated cells compared to its untreated control. Plots are shown for bacterial cell area (**a**), roundness (**b**) and length (**c**).

(**d-f**) Scatter plots comparing the differential expression of other genes significantly upregulated in the majority of isolates, with the ratio of bacteria length in treated versus untreated conditions. Horizontal error bars represent the $\log_2$ fold standard error in differential expression and vertical error bars represent the standard deviation in the measurement over three biological replicates. Points are coloured by isolate using the same legend as **Figure 5.6**.

5.5 Assembling a *K. pneumoniae* collection representative of the species

To further investigate the relationship between antimicrobial exposure, phenotypic change and differential gene expression, I assembled a larger collection of *K. pneumoniae* isolates to more broadly reflect the diversity of the species. In recent years, a number of *K. pneumoniae* isolate collections have been stored on the Cambridge Biomedical Campus and sequenced at the Wellcome Sanger Institute. Both the physical isolates and genome sequences, as well as underlying metadata and antimicrobial susceptibility results, were made available for this study. These collections are summarised in **Table 5.2**.

The isolates in these collections originate from a range of human, animal and environmental samples, predominantly from the 2010s. They represent deep sampling of single clinical sites (in the UK and Vietnam) and two national clinical reference laboratory collections (from the UK and Ireland), with a wide genomic diversity of *K. pneumoniae* isolates. These were supplemented with commonly studied and well characterised reference isolates: ATCC 43816[352]; B5055 (and its non-mucoid variant)[353]; AJ218 (and its non-mucoid variant)[354,355]; ecl8[356]; NTUH-K2044[357]; MKP103[358]; RH201207[92]; and NCTC 13438[359].

To produce a working collection of approximately 150 isolates for further investigation, I collated available WGS data from all isolates stored on the Cambridge Biomedical Campus from the above studies. To establish the species, sequence type and common resistance genes, all available sequences were analysed using Kleborate. These were combined with the associated sample metadata and antimicrobial susceptibility results.

Table 5.2 – Summary of isolate collections used as sources for working collection

Collection	Study site	Isolate source	Year of isolate	No. isolates	AMS testing	Reference
CUH prospective surveillance study	CUH, UK	Clinical samples, screening samples, hospital environment	2016	127	Vitek Disc diffusion	Brodrick *et al.*[31]
One Health study in the East of England	CUH, UK; Various sites in East Anglia	Human blood, faeces, hospital environment, wastewater, sewer, animals	2014-15	249	Vitek	Ludden *et al.*[360]
BSAC blood culture collection	28 hospitals, UK/Ireland	Human blood	2001-11	250	Broth microdilution	Moradigaravand *et al.*[39]
Global phylogeny collection	USA, Vietnam, Australia, Laos, Indonesia, Singapore	Human blood/carriage; animals	Various	247	Vitek	Holt *et al.*[34]
Vietnam ICU prospective surveillance study	2 ICUs, Hanoi, Vietnam	Clinical samples; hospital environment	2017-18	543	Vitek	Roberts *et al.*[361]
Ireland AMR reference laboratory CPE study	Hospitals, Ireland	Clinical samples	2012-17	278	Etests	Ludden *et al.* (unpublished)
Hypervirulent *K. pneumoniae* case report	Medway Hospital, UK	Human blood	2018	1	Vitek	Macleod *et al.*[362]
Reference isolates	Various	Various	Various	10	Vitek	Various
Total				**1,705**		

AMS = Antimicrobial susceptibility. CUH = Cambridge University Hospitals. BSAC = British Society for Antimicrobial Chemotherapy. ICU = Intensive Care Unit

I then selected isolates from each study based on the following criteria to ensure diversity of the species, with a focus on clinically relevant lineages:

i. Proportionate representation of all four subspecies of *K. pneumoniae*;
ii. A range of susceptibilities to a number of antimicrobial classes, including ciprofloxacin and carbapenems, with a diversity of resistance mechanisms (such as mutational and plasmid-encoded quinolone resistance mechanisms and a variety of carbapenemases);
iii. Isolates with high levels of ciprofloxacin resistance in the absence of common fluoroquinolone resistance mechanisms;
iv. Inclusion of hypervirulent strains;
v. A range of isolate sources, in terms of study, country of origin, year of isolation and sample site;
vi. A range of sequence types, broadly representative of the species but with over-representation of strains associated with clinical outbreaks. The decision of which sequence types to include was based on: (1) the global phylogeny of *K. pneumoniae* published by Holt *et al.* in 2015[34]; (2) a preliminary dataset, made available by the authors prior to publication, of over 10,000 publicly available genomes analysed using Kleborate and subsequently published by Lam *et al* in 2021[42]; (3) frequently represented sequence types in the studies above, including known outbreak strains in CUH. The deliberately over-represented sequence types, and the rationale for their selection, were:
    a. CG258 (ST258, ST11, ST512) – global spread of the KPC carbapenemase, with a range of other associated plasmids and antimicrobial resistance genes and clinical outbreaks[29,34,363-365]

b. ST14 – widely globally distributed, associated with a range of plasmid-encoded antimicrobial resistance genes and clinical outbreaks (including the single locus variant ST78, which has caused a clonal outbreak in CUH)[31,35]
c. ST15 – widely globally distributed, associated with a range of plasmid-encoded antimicrobial resistance genes and clinical outbreaks[60,365-368]
d. ST17 – emerging clade, associated with clinical outbreaks[35]
e. ST23 – a clade with clonal expansion, responsible for severe community acquired infection due to associated virulence genes[38,369-371]
f. ST45 – emerging clade, associated with clinical outbreaks[372]; additionally, to provide contextual ST45 isolates for the RNAseq work conducted in this thesis using the clinical isolate KC006
g. ST101 – emerging clade in the UK and Europe[29], a high proportion of cases historically in CUH[313] and a number of resistance and virulence genes
h. ST147 – emerging clade in the UK with associated clinical outbreaks[35,39,365]
i. ST307 – emerging clade in the UK, causing a high proportion of cases in CUH and possessing a number of resistance and virulence genes[318-323]
j. ST874 – emerging clade in the UK, causing a high proportion of cases in CUH and possessing a number of resistance and virulence genes[39,313]

Where multiple isolates meeting the above criteria were included in a single collection, I selected the first isolate listed. Once identified, the species of the stored isolate was confirmed by MALDI-TOF. To ensure consistent AMR phenotyping across the study, antimicrobial susceptibility testing using Vitek was performed on all isolates that had not previously been tested using this platform. A complete list of the 175 isolates selected for inclusion in this study is provided in **Appendix E**.

5.6 Genomic characterisation of *K. pneumoniae* collection

Once the collection had been physically collated, the whole genome sequence assemblies and annotation files produced by the Wellcome Sanger Institute as part of their respective studies (or files uploaded to GenBank for reference isolates) were analysed further. Species, MLST and antimicrobial resistance genes/mutations were determined using Kleborate[42], and capsular (K locus) and O locus antigen types were determined using Kaptive[44]. Plasmid replicon types were determined using SRST2, cross-referencing to the PlasmidFinder database[293,294]. Pangenome analysis was performed using Roary[290]. This identified 3,299 core genes (identified in >99% of isolates), 604 “soft core” genes (95%-99% isolates), 1,694 “shell genes” (15-95% isolates) and 17,115 “cloud” genes (<15% isolates).

A maximum likelihood phylogenetic tree based on single nucleotide polymorphisms from the core genome alignment produced by Roary is shown in **Figure 5.8**. This demonstrates the long branches separating sequence types, typical of a phylogeny of *K. pneumoniae*. Additionally, there was representation from all four subspecies of *K. pneumoniae* and all of the clinically relevant sequence types listed above, from multiple collections and specimen types. Further genotypic analysis has demonstrated a range of K and O antigen loci, plasmid replicon types and genes/mutations associated with antimicrobial resistance. Finally, there was a diversity of phenotypic antimicrobial susceptibility, ranging from pan-susceptible to pan-resistant isolates for the antimicrobials tested.

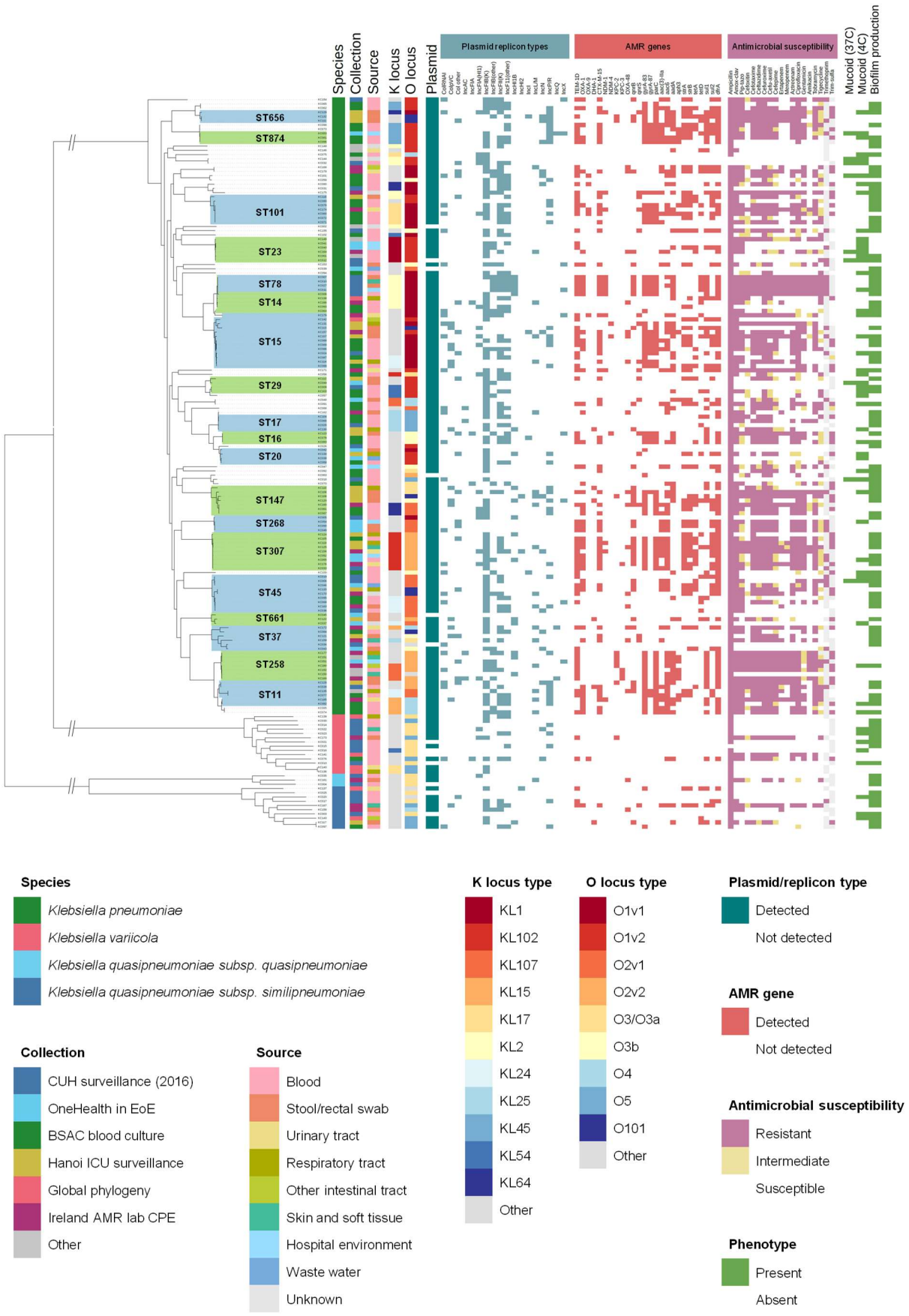


Figure 5.8 Phylogenetic tree of *K. pneumoniae* collection.

Maximum likelihood phylogenetic tree produced by IQTREE, based on SNPs in the core genome determined by Roary. The tree is midpoint rooted. The branches to *K. pneumoniae*, *K. variicola* and *K. quasipneumoniae* are not to scale to facilitate visualisation. Species, MLST and AMR gene profile were determined using Kleborate. K and O locus types were determined using Kaptive. Plasmid replicon types were identified using SRST2 against the PlasmidFinder database. Phenotypic susceptibilities were determined using Vitek. Common and/or clinically important sequence types represented by 3 or more isolates are highlighted in alternating colours and annotated.

Figure 5.9 places this collection in the context of a global phylogeny of *K. pneumoniae* isolates published by Holt *et al.* in 2015[34]. Only isolates sequenced at the Wellcome Sanger Institute from the global collection are included in this tree. This shows that the study collection broadly reflected the diversity of the species as a whole, including the relative proportion and diversity of each of the four major subspecies. There was representation of most major sequence types, including lineages associated with hypervirulence (ST23) and antimicrobial resistance (ST15, ST11, etc.). Additionally, there were some lineages that were absent from the Holt *et al.* collection, but have come to prominence as important causes of clinical outbreaks since its publication in 2015 and were well represented here (e.g. ST307, ST101).

5.7 Further optimisation of *K. pneumoniae* HCl

Having produced a collection of isolates that represented a genomically diverse selection of *K. pneumoniae* isolates, I next sought to investigate the phenotypic variation across the species in both standard growth conditions and following exposure to antimicrobials. Following provisional RNAseq results in isolate KC006 that ciprofloxacin induced a distinct transcriptomic profile to other antimicrobials, as well as the characteristic morphological appearances shown in **Chapter 4**, I sought two additional comparator antimicrobials with different cellular targets and mechanisms of action for further experimental work.

Figure 5.9 Phylogenetic tree of isolates from this study, in the context of a global *K. pneumoniae* collection.

Maximum likelihood phylogenetic tree produced by IQTREE, based on SNPs in the core genome determined by Roary (n=2,015 genes present in >99% isolates). The tree is midpoint rooted. Species and MLST were determined using Kleborate. Common and/or clinically important sequence types represented by 4 or more isolates are highlighted in alternating colours and annotated.

First, I selected meropenem; given the clinical importance of beta lactams (and their associated resistance), I wanted to include one drug from this broad class. However, the majority of *K. pneumoniae* isolates are intrinsically resistant to penicillin and amoxicillin, with a high number of ESBLs in the collection. The disappointing provisional RNAseq results for cefuroxime led me to favour meropenem, with its distinct transcriptomic profile and associated morphological changes using HCI. Second, I wanted to include an antimicrobial targeting the bacterial ribosome. Given the marked changes in differential expression induced by tigecycline and its increasing clinical relevance in the treatment of invasive *K. pneumoniae* infection in comparison to chloramphenicol, this was selected for further work. To determine the MIC of each isolate to the three selected antimicrobials, Etests were performed on all 175 isolates (**Appendix E**).

Provisional imaging experiments were conducted on all isolates to determine the adherence of bacteria to the plate surface, following findings in **Chapter 4** that some species (and some reference isolates of *K. pneumoniae*) require additional plate coatings to obtain satisfactory images. Additionally, I had observed that the final number of untreated bacteria imaged per field was different depending on the isolate: some isolates have comparably high numbers, crowding the image and making bacterial segmentation challenging during image analysis. Different starting dilutions of bacteria from the overnight culture (1:100 and 1:1000) were therefore trialled to determine which one provided the optimal final number of imaged bacteria.

The total number of single bacterial cells (imaged across 10 fields per well) in each of four growth conditions (1:100 and 1:1000 starting inocula from overnight culture, in the presence of absence of ciprofloxacin at MIC) are shown in **Figure 5.10a**. There was a significantly

different distribution of the number of bacteria imaged in each condition. The minimum number of bacteria imaged per well in the untreated condition was defined as 400 (i.e. 40 per field); below this, there were insufficient bacteria in the antimicrobial treated conditions to yield reliable results. In untreated wells at the highest starting concentration of bacteria (1:100) there were 13/175 (7.4%) isolates identified as falling below this threshold. All images were manually reviewed to confirm that these bacteria did not adhere to the surface of the plate. Some of these isolates were from lineages that are associated with a hypermucoid phenotype, including 4 isolates from the hypermucoid ST23 lineage. The mucoid reference isolate B5055 adhered poorly to the plate, whereas its non-mucoid variant bound to a much greater extent. To explore this further, the string test, a simple phenotypic test of hypermucoidy, was performed on all isolates: 7/13 non-adherent isolates were hypermucoid after incubation at 37°C (compared to 2/163 adherent isolates, $P<0.0001$, chi-square test). Three further non-adherent isolates developed a hypermucoid phenotype after incubation at 4°C. The remaining three isolates were not phenotypically hypermucoid and did not belong to a lineage that was known to be associated with hypermucoidy.

Figure 5.10b shows the impact of the density of imaged cells in each field on the morphological characteristics of each bacterium. Using the example of bacterial area, there was a significant difference in the distribution of mean bacterial area per well depending on the starting inoculum of overnight culture, in both treated and untreated conditions. **Figure 5.10c** shows how this relationship may be explained by the number of single cells imaged per well: there appears to be a minimal effect on bacterial area until 10,000 cells were imaged per well, after which the distribution of bacterial cell area falls. At this density of bacteria there was an approximately even monolayer of cells across the entire field, with cells closely abutting one another; higher bacterial densities may impede further bacterial growth and therefore the morphological properties of the imaged bacteria.

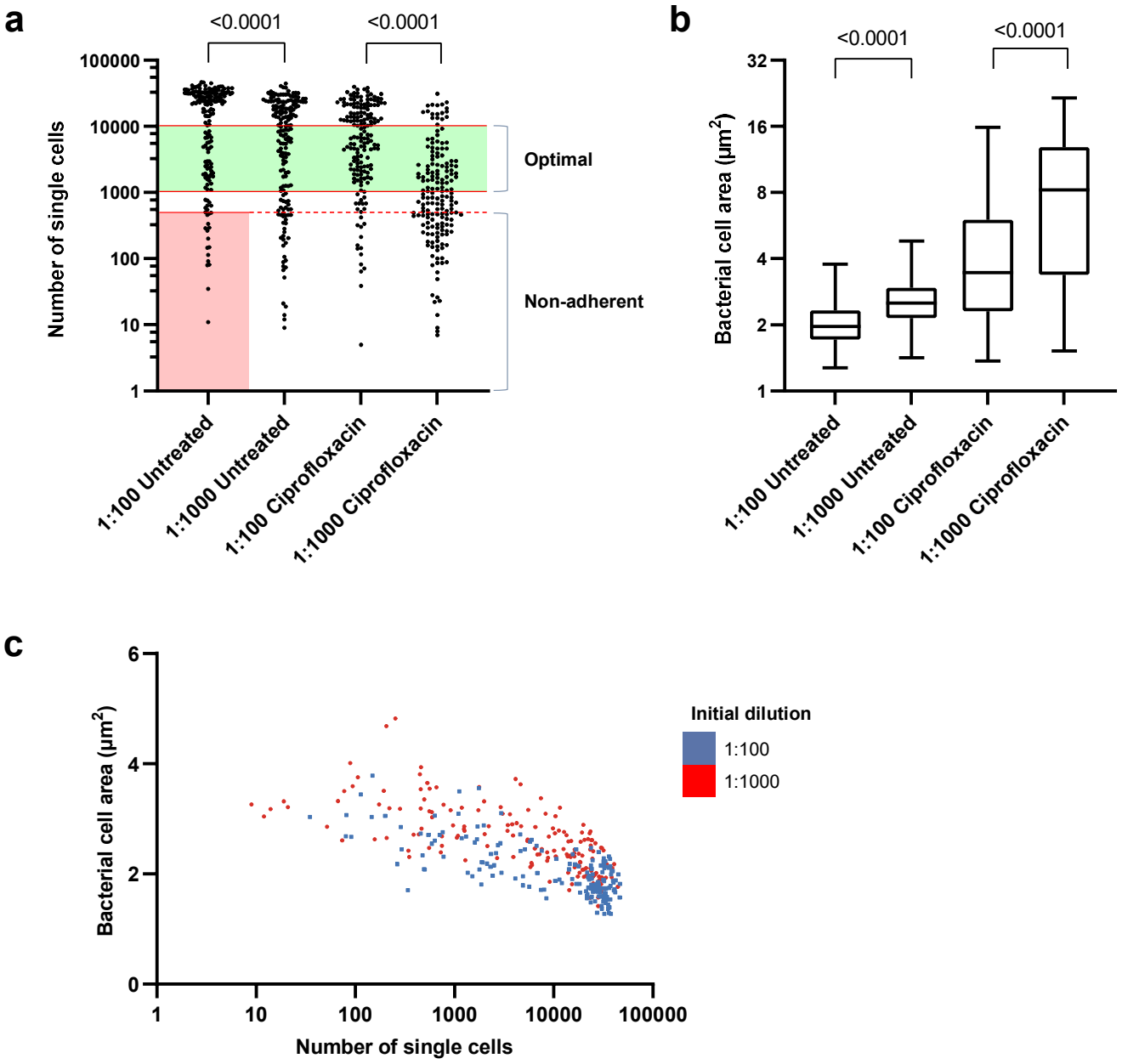


Figure 5.10 Provisional imaging experiments using *K. pneumoniae* collection

(**a**) Scatter plot showing the distribution of imaged bacterial cells in each well (representing 0.4 mm^2 coverage). Each point represents a single well. The red shaded area shows bacterial isolates that do not adhere in the untreated condition, resulting in <400 imaged bacteria per well. The green shaded area represents optimal bacterial number for image analysis. (**b**) Box and whisker plot demonstrating the distribution of bacterial cell area in the four growth conditions. The whiskers represent the range of values. The *y* axis is plotted in a log_2 scale to aid visualisation. Comparisons were made using a two tailed paired *t* test, with the level of statistical significance annotated above each plot. (**c**) Scatter plot showing the relationship between the number of imaged bacterial cells and bacterial cell area. With increased numbers of bacteria, in particular >10,000 imaged cells per well, there is a reduction in mean bacterial cell area.

Combining the maximum and minimum threshold of bacteria imaged per well, each isolate was assigned an optimal starting dilution of overnight culture. For non-adherent bacteria, all further experiments were conducted using plates coated in vitronectin (based on the optimisation experiments performed in **Chapter 4** and further provisional experiments on these isolates to confirm improved adherence).

5.8 Characterising *K. pneumoniae* in response to antimicrobials using HCI

All 175 isolates were imaged in biological duplicate at both MIC and 4x MIC for each antimicrobial. To ensure a reliable comparison between growth conditions, each isolate was imaged in four growth conditions on the same plate: untreated and in the presence of ciprofloxacin, tigecycline and meropenem. Each 96 well plate also included two positive controls (the clinical isolate KC006 and the reference isolate ATCC 43816, grown in the same conditions as the other isolates) and negative controls for LB broth, stain and PBS. All 175 isolates were successfully imaged with three exceptions, all at high concentrations of antimicrobials leading to insufficient bacteria for further phenotyping: KC012 (missing one replicate of ciprofloxacin and meropenem at MICx4); KC108 (missing one replicate of ciprofloxacin at MICx4); KC144 (missing one replicate of tigecycline at MICx4). PCA plots demonstrating the variation in calculated morphological properties for all 175 isolates are shown in **Figure 5.11**.

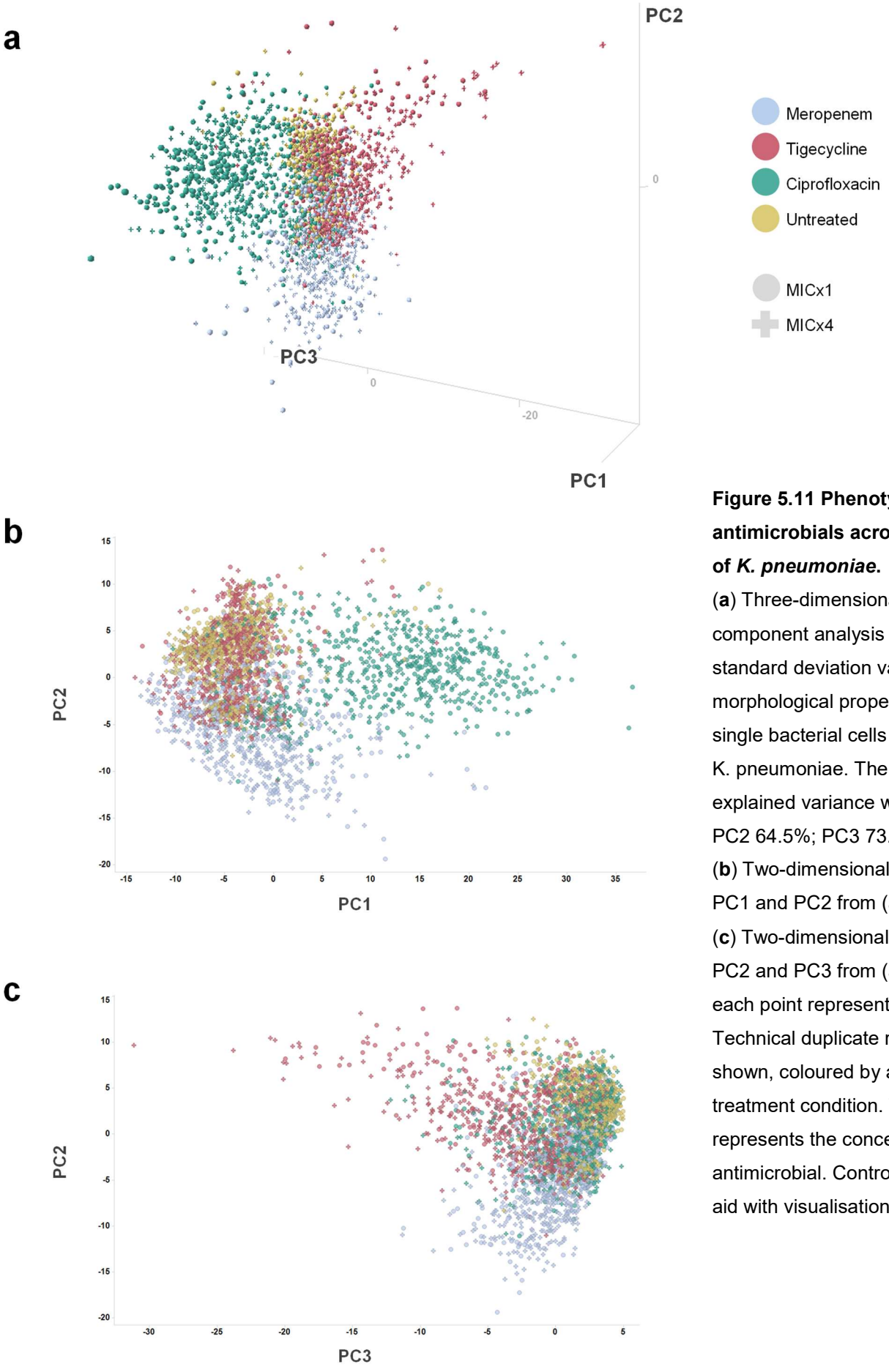


Figure 5.11 Phenotypic effects of antimicrobials across the species of *K. pneumoniae*.

(**a**) Three-dimensional principal component analysis of the mean and standard deviation values of morphological properties measured for single bacterial cells in 175 isolates of K. pneumoniae. The cumulative explained variance was: PC1 47.9%; PC2 64.5%; PC3 73.5%.

(**b**) Two-dimensional PCA comparing PC1 and PC2 from (**a**).

(**c**) Two-dimensional PCA comparing PC2 and PC3 from (a). In all figures, each point represents a well. Technical duplicate repeats are shown, coloured by antimicrobial treatment condition. The shape represents the concentration of antimicrobial. Controls are omitted to aid with visualisation.

In keeping with previous experiments, the untreated wells cluster most tightly together, while data points representing ciprofloxacin, meropenem and tigecycline treated wells extend down each of the three PC axes. To better visualise the separation and clustering of the antimicrobial treatment conditions, two dimensional PC plots are shown in **Figures 5.11b** and **5.11c**. While there was some overlap in clusters, there remained a clear and characteristic phenotypic change associated with each of these three antimicrobials.

In comparison to previous experiments, there was greater variation in the observed phenotype for each antimicrobial condition. There are a number of explanations for this observation. First, there were some isolates that had a distinct morphology from other *K. pneumoniae* isolates in the untreated growth condition. **Figure 5.12a** is a two dimensional PCA plot showing untreated wells across all experiments. There are some clear outliers, highlighted in the orange oval which represents two isolates, KC168 and KC180. These isolates had a reproducible morphological appearance that was consistent across all replicates. On review of these images, both isolates had a phenotypic appearance similar to bacteria treated with ciprofloxacin, with filamentous elongation and higher uptake of SYTOX green. These isolates are readily identifiable in the PCA plots in **Figure 5.11**, shown as outlying untreated wells clustering with ciprofloxacin treated wells at the top of the plots.

Second, the untreated wells in **Figure 5.12a** appeared to form two distinct clusters, with the smaller of these highlighted in the green oval. On review of the morphological properties that differed between these two clusters, there was greater heterogeneity in a variety of features, shown by a wider standard deviation of the measured morphological property.

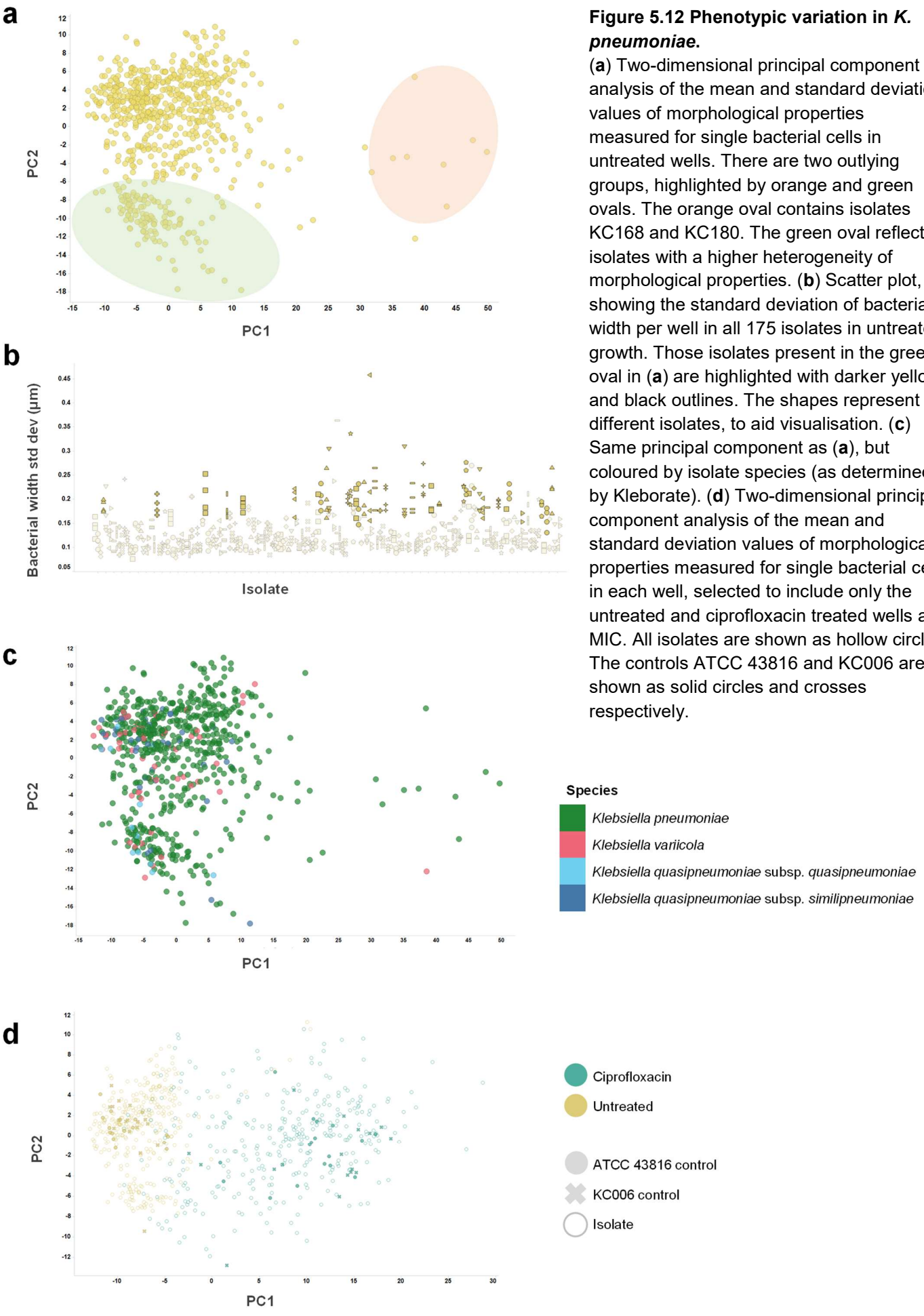


Figure 5.12 Phenotypic variation in *K. pneumoniae*.
(**a**) Two-dimensional principal component analysis of the mean and standard deviation values of morphological properties measured for single bacterial cells in untreated wells. There are two outlying groups, highlighted by orange and green ovals. The orange oval contains isolates KC168 and KC180. The green oval reflects isolates with a higher heterogeneity of morphological properties. (**b**) Scatter plot, showing the standard deviation of bacterial width per well in all 175 isolates in untreated growth. Those isolates present in the green oval in (**a**) are highlighted with darker yellow and black outlines. The shapes represent different isolates, to aid visualisation. (**c**) Same principal component as (**a**), but coloured by isolate species (as determined by Kleborate). (**d**) Two-dimensional principal component analysis of the mean and standard deviation values of morphological properties measured for single bacterial cells in each well, selected to include only the untreated and ciprofloxacin treated wells at MIC. All isolates are shown as hollow circles. The controls ATCC 43816 and KC006 are shown as solid circles and crosses respectively.

An example of this is shown in **Figure 5.12b**, which plots the standard deviation of the bacterial width in each well, grouped by isolate on the *x* axis. Cells in the green oval are highlighted and demonstrate a wider standard deviation in comparison to the other untreated isolates, a finding that was consistent between experimental replicates. While this represents greater heterogeneity of morphology, the underlying cause is unclear. There was no evident association between bacterial morphology and sequence type, antimicrobial susceptibility or sub-species (as shown in **Figure 5.12c**, where the four *K. pneumoniae* sub-species included in this collection show no obvious clustering using PCA).

Additionally, there was more plate-to-plate variation in this experiment in comparison to previous HCI studies in this thesis. To illustrate this, **Figure 5.12d** shows the KC006 and ATCC 43816 controls placed in the context of untreated and ciprofloxacin treated wells at MIC. While the majority of wells cluster together in each condition, there is more heterogeneity than in previous experiments for both isolates. The reasons for this are unclear but may reflect both genuine morphological differences in wells (for example due to small changes in bacterial inoculum or environmental growth conditions, despite attempts to standardise them), or artefactual changes from the imaging process itself. For example, in comparison to previous experiments, a larger number of plates were fixed and imaged on the same day to facilitate high throughput. One of the caveats with this approach is there can be a longer delay from staining to imaging, which leads to a reduction in stain intensity over time, e.g. the mean bacterial DAPI stain relative intensity on the first imaged plate for the MICx4 growth condition was 2,290; by the third plate this had fallen to 1,243. The impact of this reduction in intensity is unclear, as the majority of morphological features can be assessed independent of stain intensity. Further work is required to understand and correct for plate-to-plate variation in this context, which is beyond the scope of this thesis.

Using this large collection, it was possible to examine the diversity of morphological appearances across the species in response to antimicrobials, including the effect of drug concentration. **Figure 5.13** shows the variation in four morphological properties that have previously been identified as key variables that distinguish antimicrobial treated bacteria from untreated controls. Using meropenem as an example, there is a clear and statistically significant dose-dependent increase in the related properties of bacterial cell width, area and width-to-length ratio, as well as a higher uptake of the live/dead differentiation stain SYTOX green. This demonstrates that at a population level, higher doses of meropenem are associated with larger phenotypic changes in bacteria. However, while there were significant, dose-dependent changes associated with tigecycline treatment, the changes associated with ciprofloxacin were not as consistent. There was an increase in bacterial cell width at higher concentrations, as would be anticipated, but a minimally significant difference in cell area, no difference in SYTOX green intensity, and a reversal in the trend of bacterial width-to-length ratio. This suggests that some antimicrobials may reach a “ceiling effect” where there is a maximal change in morphology associated with increasing concentrations of drug. Also, at such high concentrations of bactericidal antimicrobials there is a substantial reduction in the number of cells imaged (as shown in **Figure 5.10a**), which may lead to bacterial cell death before a greater change in phenotype can manifest.

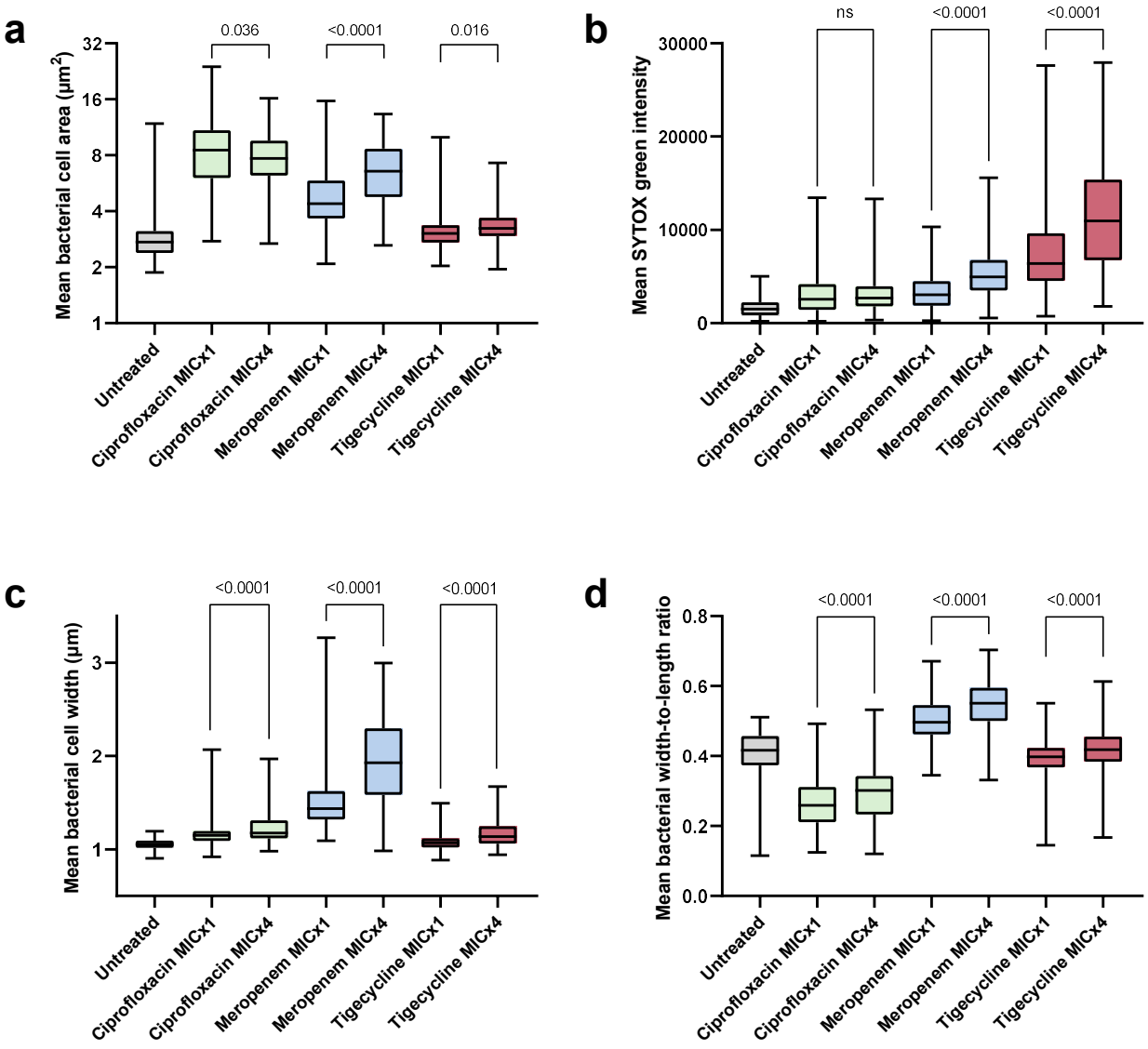


Figure 5.13 Dose dependent variation in *K. pneumoniae* morphology in the presence of antimicrobials.

(**a-d**) Box and whisker plots showing the variation present in a selection of morphological features imaged in all isolates in the collection in the presence of antimicrobials. Whiskers represent the range of values measured. Comparisons between MIC and MICx4 growth conditions are shown for each antimicrobial, calculated using an unpaired two tailed *t* test. (**a**) Mean bacterial cell area. All growth conditions at MIC were significantly different from the untreated control (P<0.0001), apart from tigecycline at MICx1 (P=0.013). (**b**) Mean SYTOX green intensity. All growth conditions at MIC were significantly different from the untreated control (P<0.0001). (**c**) Mean bacterial cell width. All growth conditions at MIC were significantly different from the untreated control (P <0.0001), apart from tigecycline (P=0.0023). (**d**) Mean bacterial width-to-length ratio. All growth conditions at MIC were significantly different from the untreated control (p <0.0001), apart from tigecycline (P=0.0077). All distributions were compared using with a two tailed, unpaired *t* test.

5.9 Characterising the genotypic response of *K. pneumoniae* using qPCR

Further work on this *K. pneumoniae* collection has been impacted by the COVID-19 pandemic. In addition to further analyses of the imaging study above, I had originally intended to compare HCI with differential gene expression to test the hypothesis that changes in bacterial morphology are associated with transcriptomic changes in each isolate. I had also intended to use this wider collection to further investigate genes identified in RNAseq experiments, which show potential mechanisms of antimicrobial resistance that are up-regulated in the presence of antimicrobials.

Undertaking this work with RNAseq would have been prohibitively expensive, so I intended to use a qPCR-based approach using the Fluidigm platform to address a limited number of approximately 90 genes identified from previous experiments. The Fluidigm assay enables analysis of a comparably large number of targets (e.g. 48/96 depending on the chip) in parallel using a small volume of sample. The approach has been used to quantify differential expression in eukaryotic cells and recent studies have demonstrated its application to bacterial samples[269,373-376]. Further work would have been required to develop and validate primer sets for each gene, in addition to the extraction of RNA from hundreds of stored samples and QC and analysis of the resultant qPCR data, which made this work unmanageable in the time of my PhD fellowship.

However, provisional experiments were performed to determine the feasibility of this approach. Samples of the clinical isolate KC006 in two conditions (untreated and with ciprofloxacin at MIC after 2 hours of incubation) were stabilised in RNAprotect and extracted

using the Qiagen RNEasy kit, replicating previous RNAseq experiments. The experiment was performed in biological triplicate. Following a reverse transcription step, RNA concentration was quantified and normalised before being analysed on the Fluidigm platform using a 96x96 chip being validated in the laboratory for the detection of a range of antimicrobial resistance genes in wastewater samples.

Five resistance genes were identified in each sample: the fluoroquinolone resistance gene *qnrB*, the aminoglycoside resistance genes *yokD* and *aacA-aphD*, a chromosomally encoded beta-lactamase, bla_{SHV}, and *bepE* (which is orthologous to *oqxB* in *E. coli* – both are members of the resistance nodulation division family of multidrug efflux pumps[377]). To quantify differential gene expression, a constitutively expressed gene would ordinarily be used to normalise the CT values. Results were therefore normalised to the CT values of the *bla* gene which, of the five genes detected, was the least differentially expressed in previous RNAseq experiments.

Figure 5.14 shows the differential gene expression estimated from the Fluidigm assay in comparison to that measured during previous RNAseq experiments. Despite using a single gene for normalisation, this experiment has again identified that the resistance genes *qnrB* and *aacA-aphD* are significantly upregulated in the presence of antimicrobials. Additionally, the aminoglycoside resistance gene *yokD*, which is present in the antimicrobial resistance locus on the IncF plasmid in KC006 in the same region as *qnrB* and *aacA-aphD*, may also be upregulated (although this was not significantly differentially expressed in previous RNAseq studies). This replication of findings from RNAseq experiments demonstrates that the Fluidigm platform is of potential utility in determining differential gene expression in response to antimicrobials at scale.

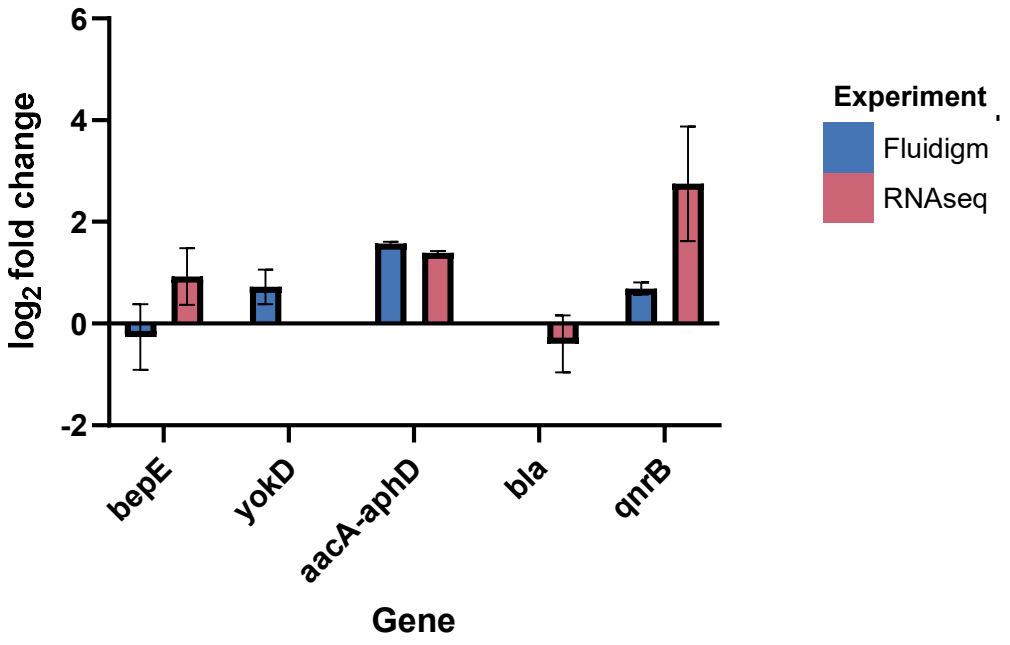


Figure 5.14 Provisional differential expression results for antimicrobial resistance genes in the clinical isolate KC006 in the presence of ciprofloxacin using the Fluidigm platform. Differential gene expression for 5 genes associated with antimicrobial resistance, comparing untreated controls with ciprofloxacin treated cells after 2 hours of incubation, in the clinical isolate KC006. The two platforms presented are: Fluidigm PCR (3 biological replicates) and RNAseq (composite of 2 separate experiments, each with 3 biological replicates). All Fluidigm results are normalised to the gene *bla*. The whiskers represent the range of each measurement.

5.10 Discussion

In this chapter I have demonstrated the morphological and transcriptomic changes that occur in *K. pneumoniae* in response to multiple antimicrobials. I have found that both genotypic and phenotypic responses to antimicrobials are specific to their mechanism of action, with drugs that impact the same cellular pathway eliciting similar responses. I have shown that genes that are upregulated in response to ciprofloxacin at MIC are largely shared between unrelated isolates, although there was marked variation in the extent of differential

expression between isolates. By combining data from RNAseq and HCI experiments I have demonstrated a correlation between differential gene expression and key morphological features that distinguish treated bacteria from untreated controls. I have collated a physical collection of *K. pneumoniae* representative of the diversity of the species, which has been phenotypically and genotypically characterised. Using this collection I have utilised HCI to study the impact of the antimicrobials ciprofloxacin, meropenem and tigecycline on the morphology of 175 isolates. Finally, I have provided preliminary data which shows that differential gene expression can be performed using the Fluidigm platform, a potential avenue for high throughput differential gene analysis that can be coupled with high content imaging.

In **Chapter 3** I demonstrated the transcriptional changes provoked by ciprofloxacin in a single clinical isolate, including upregulation of a number of known and putative resistance mechanisms. Expanding this study to include other antimicrobials has demonstrated that genes associated with resistance are differentially expressed depending on the class of antimicrobial. For example, in addition to many of the resistance mechanisms induced by ciprofloxacin, tigecycline exposure led to significant up-regulation of the efflux pump *smvA*, which confers resistance to chlorhexidine[378]. Interestingly there was also upregulation of the porin *ompK35*, whose loss is associated with ESBL and carbapenem resistance[379]. In the presence of meropenem there was upregulation of *cusA* and *cusB*, genes that are strongly associated with beta lactam resistance as well as playing a potential role in tigecycline resistance[349]. This pattern of gene expression differs between each antimicrobial condition, with some overlap between drugs with a common mechanism of action or cellular target such as tigecycline and chloramphenicol. Other studies have demonstrated a characteristic transcriptomic response of bacteria to different antimicrobials. For example, Molina-Santiago *et al* compared the transcriptomic profiles of *Pseudomonas putida* in the presence of a range of antimicrobial classes and found a distinct profile for each, with some overlap[329].

The finding that plasmid conjugation proteins are upregulated in the presence of both ciprofloxacin (in multiple plasmid replicon types), tigecycline and chloramphenicol demonstrates the potential collateral impact that antimicrobials may play in spreading resistance genes between isolates. Combined with other findings, such as the upregulation of the low fidelity polymerases *umuC/D* (potentially inducing mutational resistance in these isolates), this supports the hypothesis that antimicrobials actively drive the development of resistance, rather than simply passively selecting of isolates that are already resistant. Studies in other Enterobacteriaceae have previously demonstrated that ciprofloxacin increases plasmid conjugation *in vitro* between *E. coli* and *Shigella*, and causes increased rates of mutation in *S.* Typhimurium that disproportionately affect *gyrA*, one of the key genes that are major determinants of fluoroquinolone resistance[331,380]. Further work is required to determine whether antimicrobials induce equivalent changes in *Klebsiella*, and whether that in turn can alter the course of resistance in large bacterial populations, such as in wastewater or the gastrointestinal tract.

In this study I have also identified a number of conditions in which differential gene expression was not substantially altered by antimicrobials, despite inducing a clear phenotypic change on imaging and sharing similar cellular targets to the drugs above (such as cefuroxime and gentamicin). The reason for this is unclear. It may reflect suboptimal growth conditions or incubation times to detect a transcriptomic response, as all time points and drug concentrations in this study were based on observations made during time course experiments using ciprofloxacin. Other antimicrobials may well require different experimental conditions. Colistin, for example, acts on bacterial cell membranes causing a direct lytic effect which, at MIC, may cause cell death before the bacterium has the opportunity to detect any damage and mount a transcriptomic response. Previous studies using RNAseq to

investigate the effect of colistin on *Klebsiella* have used subinhibitory concentrations because of this direct bactericidal effect[87], or noted that the maximal impact on the transcriptome occurs within the first hour after exposure[381].

A similar phenomenon may partly explain the variation in differential expression observed with other *Klebsiella* isolates exposed to ciprofloxacin. Growth rates in different isolates will vary; previous transcriptomic studies have demonstrated major changes in the transcriptome of *K. pneumoniae* depending on the stage of bacterial growth[382], so it may not be appropriate to sample all isolates at the same time point. Additionally, the observation that the extent of transcriptional change correlates with phenotypic changes measured using high content imaging suggests that the bacteria were insufficiently treated with ciprofloxacin to induce a phenotypic and genotypic response. I noted in **Chapter 3** that a sub-inhibitory concentration of ciprofloxacin led to very modest differential expression of genes in comparison to treatments at MIC and 4xMIC in the isolate KC006. It is therefore potentially important to treat bacteria as close to MIC as possible. While it is possible that the MIC was inaccurately recorded for these isolates, the standard methods used tended, if anything, to overestimate the MIC (by rounding up the Etest result to the nearest gradation on the strip, as well as the concentration of ciprofloxacin used in imaging studies). This may reflect small differences in the MIC measured by Etest in comparison to other methods. Comparison with other methods, such as broth microdilution techniques, may be beneficial to investigate this further, but are impractical in the investigation of bacterial isolate collections at scale.

It is important to consider that the variation in differential gene expression may reflect genuine differences in transcription between isolates from different lineages. Additional planned experiments – using qPCR to quantify differential gene expression in a range of isolates at both MIC and 4xMIC – would have addressed this issue. Samples from 60 isolates in all 7 growth conditions used in HCI, derived from the original dilutions of bacteria

used to prepare the images in this chapter and incubated in the same conditions, have been stored with a plan to complete this experiment in the near future.

Combining phenotypic approaches and transcriptomics in the same experiment can yield a greater insight into the mechanisms of resistance. Previous AMR studies have combined RNAseq data with imaging to validate their findings. For example Cain *et al.* have recently demonstrated associations between differential gene expression and transmission electron microscopy findings in *K. pneumoniae* isolates exposed to colistin[87]. Using a different approach, Low *et al*. exposed three carbapenem resistant *K. pneumoniae* isolates to sub-inhibitory concentrations of imipenem and made associations between the changes in the expression of genes involved in metabolite transport and phenotypic measurements of bacterial metabolites using NMR spectroscopy[383].

In this study I have demonstrated that the extent of morphological change induced by ciprofloxacin correlates with the differential expression of *recN*, one of the most upregulated genes across all studied isolates. This is in keeping with the major cellular function of RecN in the early detection and repair of DNA breaks[384], which accumulate after fluoroquinolone inhibition of the bacterial topoisomerase. Similarly, there is a correlation between phenotype and the differential expression of *sulA*, whose expression inhibits bacterial cell division and therefore likely contributes to the filamentous phenotype observed with ciprofloxacin[385]. There is also a weaker correlation between phenotype and the putative resistance gene *umuC*; further work is required to determine how strong this association is across the species and, therefore, whether bacterial phenotyping using HCI can be used as a proxy for the development of antimicrobial resistance that can be used to study bacteria in different growth conditions at scale.

This combined phenotype-genotype approach is likely to provide further insights into bacterial physiology, with greater opportunity for studying a wide range of isolates and experimental conditions using high throughput imaging in comparison to conventional techniques[386]. However, there are unresolved challenges in the statistical analyses required to compare the large datasets provided by RNAseq and high content imaging, especially if they involve large isolate collections. Additionally, the use of RNAseq approaches in this context may provide useful preliminary data, such as in this study, but may be prohibitively expensive to use at scale. Using qPCR strategies to validate provisional findings is an approach that has previously been used by Van Laar and colleagues, who exposed sublethal concentrations of carbapenems to a clinical isolate of *K. pneumoniae*[112]. Following leads from RNAseq data at 2 and 24 hrs of exposure to imipenem, they then used qPCR to investigate a small number of genes of interest following exposure to a wider range of carbapenems. In addition, the authors used scanning electron microscopy in a related experiment to correlate the phenotypic and genotypic effects of these antimicrobials. Focusing on a select number of genes in this way not only reduces the cost but also the complexity of analyses.

Utilising a similar approach, I have presented preliminary data that shows that the Fluidigm platform may be used to quantify transcriptomic changes in response to antimicrobials. There are a number of strengths of the Fluidigm platform, including reduced expense, and parallel testing of multiple samples in the same experiment. Additionally, the assay only requires a small amount of sample in comparison to RNAseq, with the potential to analyse small volumes of bacterial culture such as the 100 µl wells of CellCarrier plates used for the Opera Phenix. This would enable improved direct comparison between the phenotype and genotype of the same bacterial population.

However, there are additional considerations when performing differential expression analyses using qPCR. Of key importance is the identification of genes that can be used to normalise the CT values of other genes. These are commonly housekeeping genes that are ubiquitous across the species. Obvious candidates would therefore be the seven loci that account for the *K. pneumoniae* multilocus sequence typing scheme. However, it is important that these genes are not themselves differentially expressed in the growth conditions under investigation. From the RNAseq data in **Chapter 3**, the genes *pgi*, *rpoB*, *tonB*, *mdh* and *phoE* are all differentially expressed in isolate KC006 after 2 hours of incubation with ciprofloxacin at MIC. The genes *infB* and *gapA* are upregulated at the 4 hour time point in the presence of ciprofloxacin; additionally, *gapA* is significantly downregulated in the presence of tigecycline in RNAseq studies from this chapter. Further work is therefore required to identify the most appropriate genes to normalise any differential gene expression experiments using the Fluidigm platform.

Many previous studies using RNAseq are limited by the investigation of a small number of isolates, leading to challenges in assessing the generalisability of results to the species as a whole. RNAseq experiments in larger collections have demonstrated variation in gene expression in untreated bacteria that reflects the sequence type (and therefore genetic lineage) of *K. pneumoniae*[387]. It is therefore reasonable to assume that there will be equivalent variation in the transcriptome in response to antimicrobials. Supporting this hypothesis, I have demonstrated a wide diversity of phenotypes across treated and untreated conditions, despite overall trends in morphological changes associated with different classes of antimicrobials. Future work in genomically diverse species such as *K. pneumoniae* needs to consider this problem, utilising approaches that investigate a wide range of isolates and growth conditions in parallel.

The availability of large isolate collections that have been well characterised, both phenotypically and genomically, may help to address this problem. One of the advantages of this study is the use of a wide range of isolates, deliberately selected to reflect the diversity of the species. Maintaining such collections, and updating them with emerging lineages of clinical importance (such as the recently described threat of hypervirulent and highly resistant ST11 *K. pneumoniae* causing severe community acquired infections in South East Asia[388]), will be key in the investigation of novel antimicrobials and the inevitable development of resistance to them. An example of the utility of HCI in conjunction with large bacterial libraries has been demonstrated by Maes *et al*., who screened a novel monoclonal antibody targeting the O25b O-antigen against a collection of 86 clinical ST131 lineage *E. coli* isolates[105]. Even within this single sequence type they identified a range of different phenotypes using HCI that were associated with altered binding properties and efficacy of serum killing and phagocytosis, that were subsequently linked back to mutations and insertion sequences in the O antigen using genomic analyses.

Building on work from the previous chapter, I have demonstrated the utility of high content imaging in studying morphological changes associated with antimicrobials across a wide variety of *K. pneumoniae* isolates. I have shown that the antimicrobials ciprofloxacin, meropenem and tigecycline induce characteristic phenotypic changes across the entire species, but with some variation in morphology between isolates that requires further investigation. As noted above, there was clear phenotypic variation in untreated bacteria with some overlap with bacteria treated with antimicrobials. I have also identified a small proportion of this collection that do not adhere to uncoated plates, the majority of which were associated with a hypermucoid phenotype.

HCI approaches have several advantages over other imaging approaches such as electron microscopy in the study of large bacterial populations. They are able to capture a greater

number of cells in a wider variety of controlled growth conditions in the same experiment, which can be performed relatively quickly. For example, once preliminary experiments had been conducted to optimise imaging conditions, all 175 isolates in this study were imaged in 7 different growth conditions in duplicate experiments that took place over three non-consecutive days. In total, over 12,000,000 bacteria across more than 3,000 wells were individually characterised to provide data for this study.

There are some limitations to this approach. As outlined above and in **Chapter 4**, such large volumes of data present challenges in terms of quality control, analysis and storage. Some of these issues could be improved by dimensionality reduction, i.e. limiting the number of measured variables to those that can reliably be used to distinguish between phenotypes of interest. These may include variables that are less susceptible to batch effects (such as those that rely on stain intensity for their measurement), reducing the unresolved issue of plate-to-plate variation in results, and with the additional advantages of reducing image analysis time and computational expense[96]. Other limitations include the number of phenotypes, particularly at a sub-cellular level, that can be captured using confocal imaging. However, a wider range of stains and fluorescent probes, such as labelled monoclonal antibodies targeting specific subcellular structures, are becoming available which can quantify the expression of certain targets or phenotypes. Further work is required to establish the full capabilities of this platform.

In the preceding three chapters I have shown the wide genomic diversity of invasive *Klebsiella* in a large hospital in the East of England. I have demonstrated clear and characteristic morphological and transcriptomic changes in a number of *K. pneumoniae* isolates exposed to multiple antimicrobial classes. In doing so I have developed methods for novel high throughput bacterial phenotyping using high content imaging. I have applied this technique to a well characterised collection of *K. pneumoniae* that I have produced from multiple sources, including clinical isolates from local, national and international collections, and which reflects the diversity of the species. Of importance, I have identified a number of genes that encode known or potential resistance mechanisms that are significantly upregulated in the presence of certain antimicrobials, including the fluoroquinolones. Although the overall use of ciprofloxacin is falling globally, it remains widely available and contributed an estimated 14% of all orally administered antimicrobials in the African region in 2016-18[389]. Due to the increasing prevalence of CPEs, there is increasing reliance on repurposed antimicrobials such as tigecycline, which I have also demonstrated is associated with the upregulation of a range of antimicrobial resistance mechanisms. Further work is urgently needed to better understand and model the impact of these commonly utilised antimicrobials on the development and spread of resistance.

6. Influenza transmission, mortality and management in secondary care

6.1 Introduction

In January 2017, Addenbrooke's Hospital received national press attention with the closure of 5 wards due to suspected outbreaks of influenza A[390]. Initial investigations by the CUH infection control team identified that approximately one third of all inpatient influenza A cases at that time were acquired in hospital. Although the potential for influenza to be spread nosocomially had previously been described[133,391,392], the extent of transmission and the associated high mortality for patients with hospital acquired infections (HAI) were considered unusual by the infection teams at CUH. A serious incident was declared by the Trust, and a joint investigation was conducted between CUH infection specialties, the PHE Cambridge regional microbiology laboratory, the PHE East of England Field Epidemiology Service and the PHE Respiratory Virus Unit (RVU) at Colindale.

In parallel with this investigation, I had noticed in my clinical role that the routine management of influenza patients during the 2016/17 season was poor, with few clinicians following local and national guidelines on isolation and antiviral therapy[135]. By the time an outbreak was recognised in the Trust I had already started auditing patient care using data extracts from the CUH electronic medical record, Epic. Confirming my initial concerns, I noted considerable delays in patient testing, isolation and antiviral therapy that may have

contributed to nosocomial transmission and mortality. I was therefore invited to contribute to further investigations into influenza at the Trust, and have subsequently led a three year programme of staff education and service evaluations of a range of initiatives in the hospital with the aim of improving clinical outcomes.

In this chapter, I will detail the clinical, epidemiological and genomic analyses conducted into influenza A in CUH during the 2016/17 winter, and how they demonstrated complex transmission networks extending through multiple inpatient and outpatients areas. Using data collected from over 1,000 influenza A and B patients from 2016-18, I will demonstrate risk factors for inpatient mortality, and the protective effect associated with the antiviral drug oseltamivir. Finally, I will show how a range of interventions introduced into the hospital over three winters (including a staff education programme and the introduction of point of care molecular testing into acute services) have led to reductions in hospital associated transmission and mortality.

6.2 Clinical characteristics of influenza at CUH, 2016-17

From August 1, 2016 to March 19, 2017, a total of 2,580 patients presenting to CUH were tested for respiratory virus infection. Of these, 348 individuals (13.5%) tested positive for influenza, with 334/348 (96.0%) positive for influenza A(H3N2). The majority of these cases (314/334, 94.0%) were admitted to an inpatient bed. The epidemic curve of influenza A inpatient cases over the 2016/17 winter is shown in **Figure 6.1**. Cases were initially categorised as HAI if the first positive influenza sample was collected ≥48 hours following presentation to hospital, a definition conventionally used for mandatory reporting purposes.

Using this definition, 118/314 cases (37.6%) were attributed as HAI. However, due to concerns for delays in patient testing for respiratory viruses at this time, I manually reviewed the clinical records of all cases to confirm the likely location of viral acquisition, based on the patient's presenting symptoms on admission. The detailed reasons for recategorisation are included in **Appendix F.1**. In brief, 25 cases originally designated as HAI were recategorised as CAI, due to delays in: i) recognition of influenza as a possible cause of symptoms; ii) requesting swabs once influenza was suspected; iii) collecting swabs once they had been requested; iv) collected samples being transported to, or processed in, the laboratory. Six additional cases were re-classified from HAI to CAI, due to discharge and re-admission to hospital within 48 hours, inpatient transfer from another hospital or, in one case, admission without influenza symptoms into a bay with confirmed influenza cases, and symptom onset after nearly two days in hospital. This recategorisation resulted in 216/314 (68.8%) cases attributed to CAI and 98/314 (31.2%) to HAI.

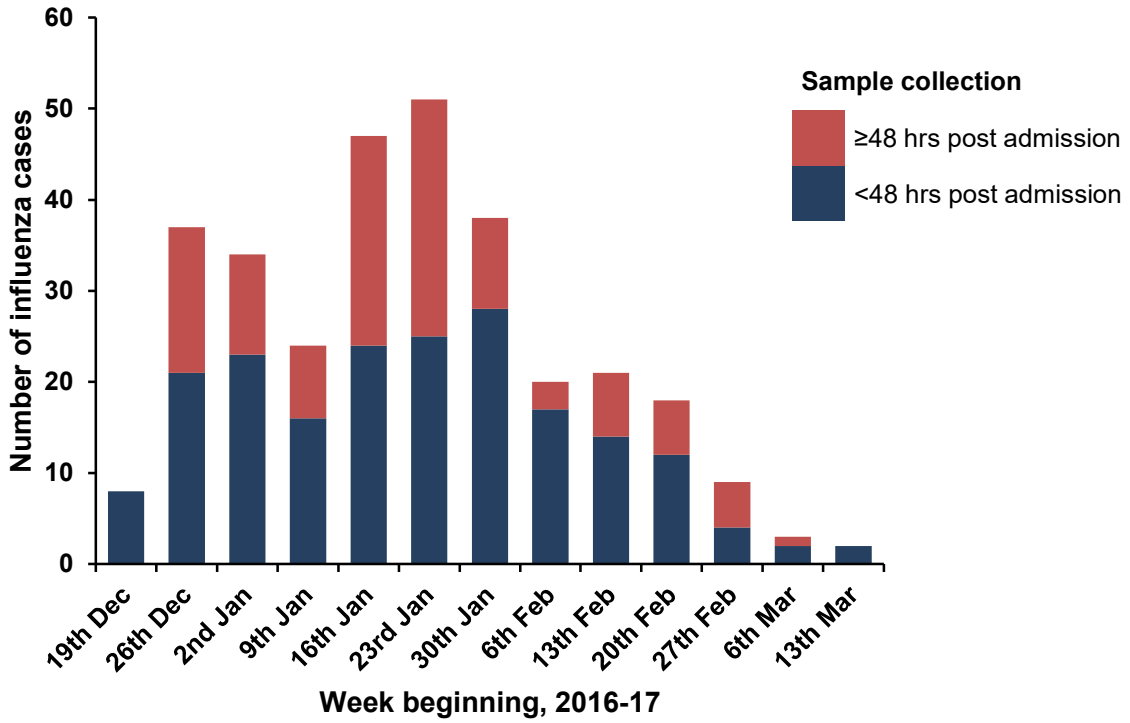


Figure 6.1 Epidemic curve of influenza A(H3N2) cases identified in CUH, 2016-17

Table 6.1 – Characteristics of influenza A(H3N2) positive inpatients in CUH 2016/17

Characteristic	Value	Number recorded	Result	
Sex	Female, n (%)	314	167	53.2
	Male, n (%)	314	147	46.8
Age (years)	Median (range)	314	77	0-102
	Mean (IQR)	314	69.4	59-86
Pregnant	n (%)	167	3	1.8
Acquisition according to 48-hour rule	Community, n (%)	314	196	62.4
	Hospital, n (%)	314	118	37.6
Acquisition according to clinical review	Community, n (%)	314	216	68.8
	Hospital, n (%)	314	98	31.2
From where admitted	Own home, n (%)	312*	262	83.4
	Residential care, n (%)	312*	32	10.2
	Another hospital, n (%)	312*	16	5.1
	Other, n (%)	312*	2	0.6
Current smoker	n (%)	243*	29	11.9
Underlying respiratory disease	Asthma, n (%)	308*	45	14.6
	COPD, n (%)	308*	64	20.8
	Bronchiectasis, n (%)	307*	11	3.6
	Interstitial Lung Disease, n (%)	307*	3	1.0
	Respiratory failure, n (%)	306*	7	2.3
	Long term oxygen therapy, n (%)	305*	3	1.0
	No underlying respiratory disease, n (%)	309*	212	67.5
Immune suppression	Any immune suppression, n (%)	314	57	18.2
	Steroid use, n (%)	57	2	3.5
	Chemotherapy within 6 months, n (%)	57	15	26.3
	Generalised radiotherapy within 6 months, n (%)	57	3	5.3
	Organ transplant and on immune suppressants, n (%)	57	12	21.1
	Bone marrow transplant within 6 months, n (%)	57	4	7.0
	Chronic kidney disease, n (%)	57	10	17.5
	Other immune suppression, n (%)	57	21	36.8
Charlson comorbidity index	Median (IQR)	314	2	1-3
Antiviral prescribed	n (%)	314	259	82.4
Influenza vaccine received**	n (%)	314	117	37.3
Admission to critical care	n (%)	314	29	9.2
Death during admission	n (%)	314	32	10.2

*N=314. A number different from 314 is due to missing data points for some patients

** Documented history of vaccination in hospital medical record or primary care record for 2016/17 winter influenza vaccine. Alternative sources of vaccination (e.g. local pharmacies, occupational health) may not be identified.

The demographic and clinical features of inpatient cases, manually abstracted from the patient record by a team of clinicians, are shown in **Table 6.1**. The median duration of all admissions was 7 days (range 0-353 days), during which 32 patients died. Inpatient deaths occurred in 15/216 (6.9%) CAIs and 17/98 (17.3%) HAIs (*P*=0.0048, two-sided Chi-square test). All patients with inpatient mortality had underlying co-morbidities and a median age 85 years (range 34-102). There were high proportions of patients with underlying lung disease (97/309, 31.4%) and immune suppression (57/314, 18.2%). A minority of patients had a documented influenza vaccine (117/314, 37.3%). Further demographic and clinical features of patients from 2016/17 and 2017/18 winters are included in **Appendix F.2** and discussed below.

6.3 Spatio-temporal analysis of inpatient influenza cases

Ward movement data was extracted from the electronic patient record for all 314 inpatient cases of laboratory-confirmed influenza A(H3N2). Algorithms were developed by Iain Roddick at the PHE Field Service to analyse these patient movements in conjunction with the time of the first positive sample collection, to detect co-location of a potentially infectious patient with someone who was susceptible. A ward-time cluster was defined as a sequence of two or more patients on the same ward where there was an overlap of the intervals of infectiousness and of susceptibility, based on published incubation periods and duration of viral shedding/transmission for influenza[393]. This approach was agnostic to the date of symptom onset and attribution of hospital or community acquired infection described above. Using this method, 190/314 (60.5%) patients were identified in 33 independent spatio-temporal networks of plausible transmission events, ranging in size from 2 cases (10 networks) to the largest cluster of 39 cases (**Figure 6.2** and **6.3**).

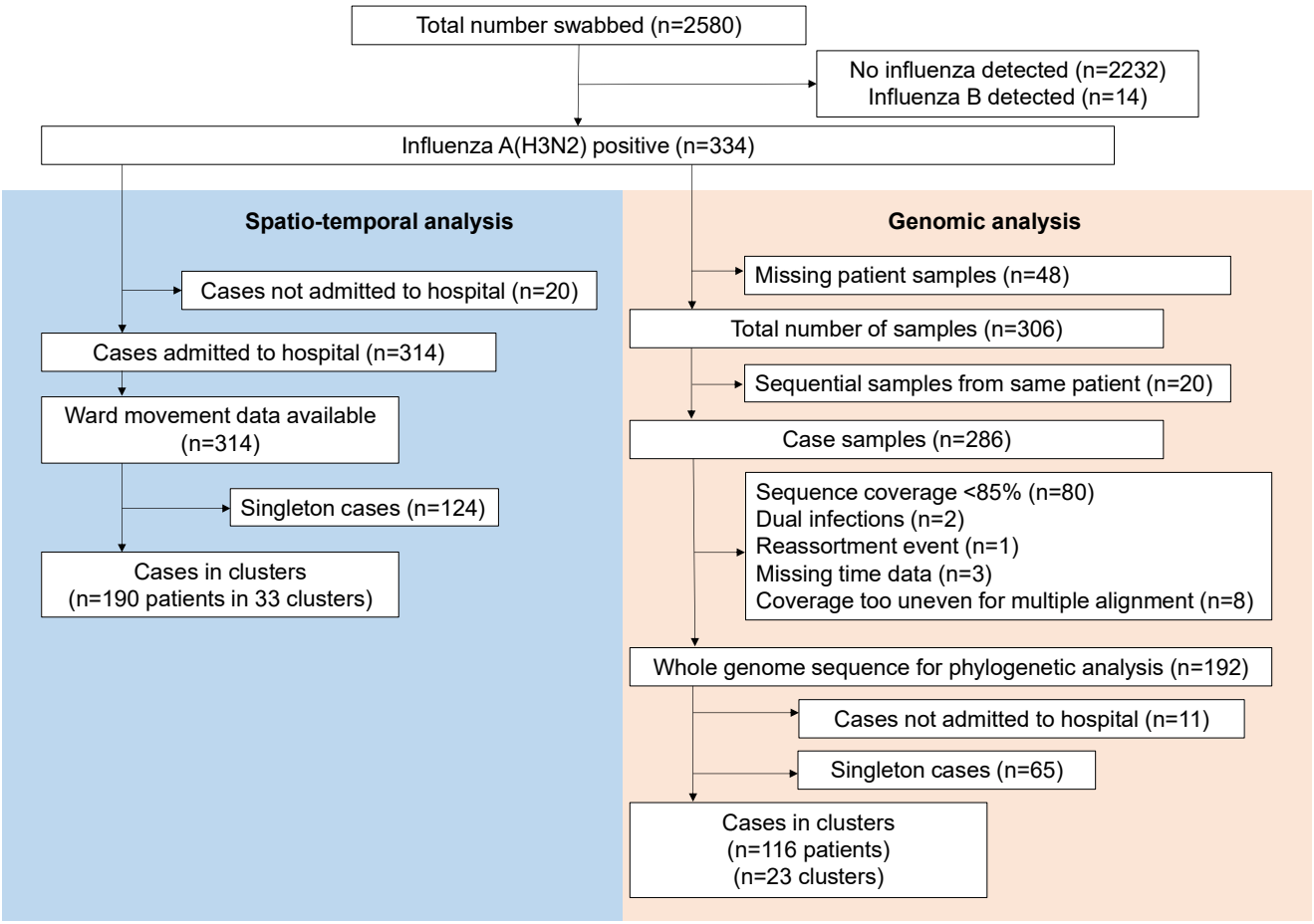


Figure 6.2 Case ascertainment of clusters of confirmed influenza A(H3N2) infection in CUH, 2016-17

Flow chart of case ascertainment and clusters using spatio-temporal and genomic methods. Initial case ascertainment was determined using a comprehensive list of all microbiological samples available from the hospital laboratory information management system.

A total of 260 plausible transmission events were detected, with exposure durations ranging from 7 minutes to 3 days. Six of the spatio-temporal networks included potential between-ward transmission events, in which the index case in the network had been transferred between wards during their infectious period. This analysis identified that 49/98 (50%) HAIs could plausibly be traced back to inpatients who acquired their infections in the community.

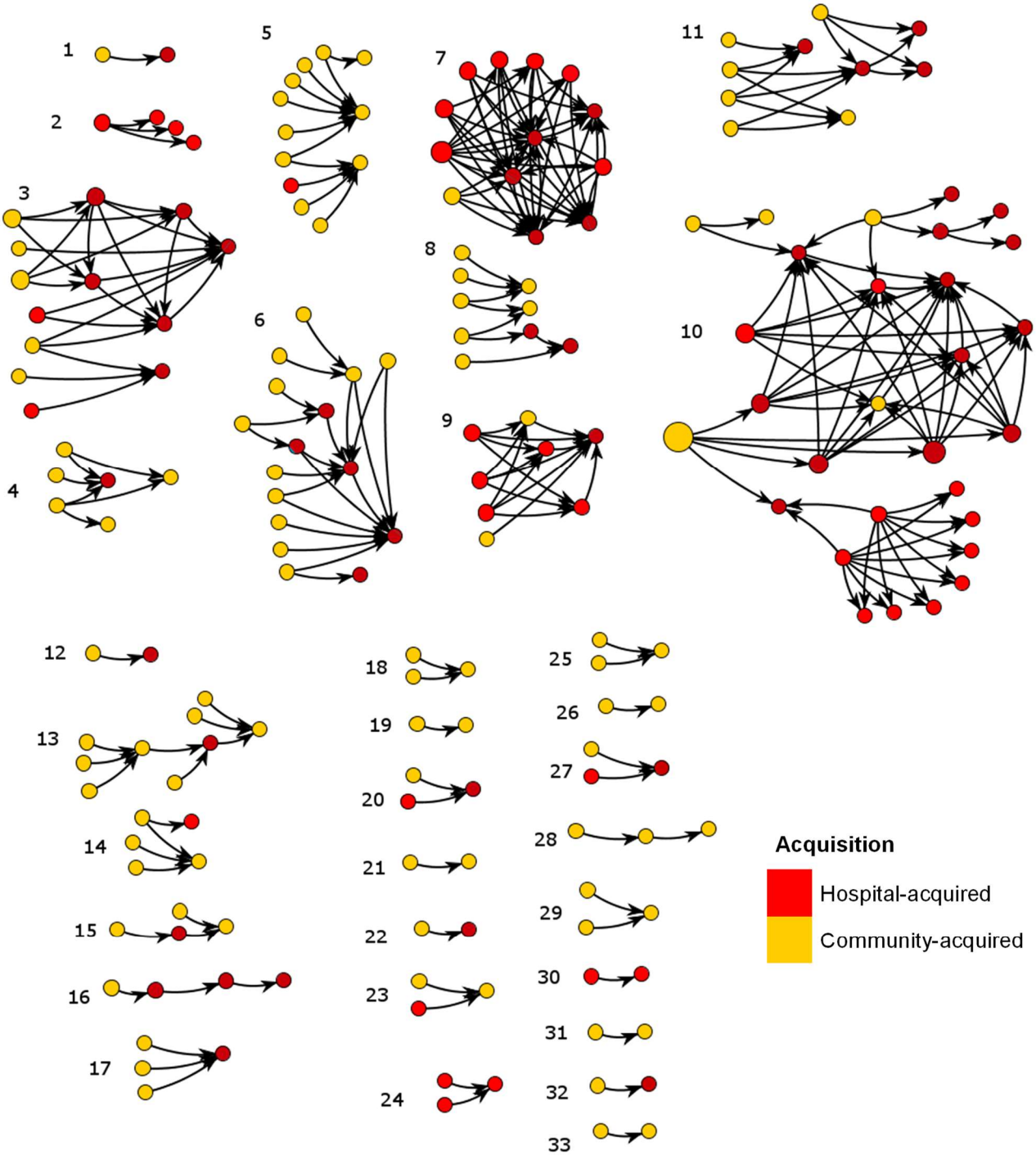


Figure 6.3 Spatio-temporal networks of potential influenza A(H3N2) transmission in CUH, 2016-17

Each circle represents a single patient; each intervening arrow demonstrates co-location in the same ward with an overlap between one patient's infectious period and another patient's susceptibility to infection. Patients are categorised as either community acquired infection (CAI, yellow circles), or hospital acquired infection (HAI, red circles). Figure adapted from network diagrams produced by Iain Roddick, East of England PHE Field Service, and reproduced here with permission.

6.4 Phylogenetic analysis of influenza A(H3N2) sequences

Genome sequencing was attempted using RNA extracted directly from clinical samples, which had been prospectively stored in line with laboratory protocols. Of all patients with confirmed influenza, sufficient sample was available to attempt sequencing in 286/334 cases (85.6%). Further phylogenetic analyses were conducted by David Williams at the PHE RVU using sequences derived from 181 inpatients, where suitable whole genome sequences were available (**Figure 6.2**).

Of note, two isolates were excluded due to co-infection with multiple H3N2 lineages, characterized by groups of minority variants each at two distinct frequencies. In one additional isolate, the SNPs identified in genome segments HA, MP and PA were inconsistent with the range of SNPs in other segments, indicating a genomic reassortment event. This sequence was sampled late in the study period and excluded from further analysis, but did not contribute to any spatio-temporal networks.

In total, genomes from 54/98 (55.1%) HAIs and 137/216 (63.4%) CAIs were available for analysis. A phylogenetic tree of these sequences is shown in **Figure 6.4**. This demonstrated a number of low diversity clades consistent with recent transmission events.

Nosocomial transmission between two sequences was considered plausible when fewer SNPs than expected were observed since their most recent common ancestor, based on the known rate of molecular evolution of influenza. A substitution rate of 0.00607 nucleotide substitutions/site/year was selected, as this had been previously inferred as the upper limit of the fastest evolving influenza A genome segment[307]. Given the influenza genome size is

11,070 nucleotides, the expected number of nucleotide substitutions per genome per year (μ) is 67.19. As in the spatio-temporal analysis above, 7 days was considered the maximum interval between dates of symptom onset or positive tests for transmission pairs (t) [305]. The precise limit was set where 95% of the expected evolutionary trajectories would have accumulated fewer substitutions, given the stochasticity of molecular evolution. The equation below, which incorporates the gamma distribution and e constant, models the stochastic accumulation of nucleotide substitutions over time. This provides an upper limit of plausible transmission within the same transmission cluster (n) of 3.64 nucleotide substitutions across the 11,070 position genome alignment.

$$\frac{(e - \mu(\frac{t}{365.25}))\left(\mu\left(\frac{t}{365.25}\right)\right)n}{gamma(n+1)}$$

Grouping samples with 3.64 substitutions in comparison to a common ancestor yielded 24 clusters, involving 116/181 (64.1%) sequenced isolates and with cluster sizes ranging from 2 to 31 sequences. The median cluster size was 2, excluding singletons. Fatal cases were distributed throughout the entire genetic diversity noted in the hospital and were not associated with any individual genomic or spatio-temporal clusters. With the exception of the emergency department, individual wards showed low genomic diversity.

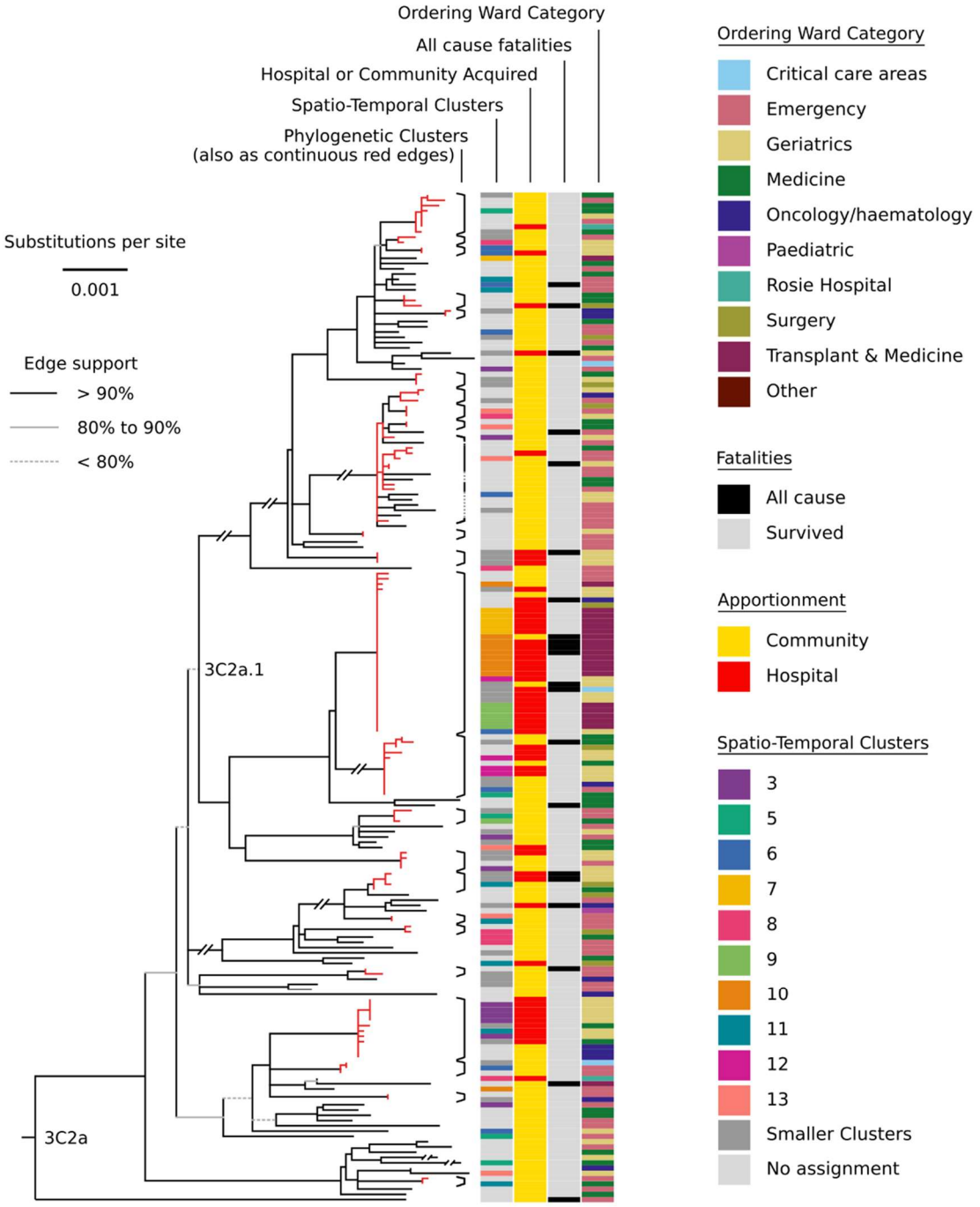


Figure 6.4 Phylogenetic tree of influenza A(H3N2) isolates from CUH, 2016-17

Rooted phylogenetic reconstruction for 181 influenza A(H3N2) samples from CUH. The 2014 strain "A/Hong Kong/4801/2014" was used as an outgroup for rooting. Genome sequences included in the multiple alignment provided a minimum 85% coverage of the reference sequence. Longer branch lengths are truncated to facilitate visualisation of possible nosocomial clusters. The phylogenetically inferred putative nosocomial transmission clusters are indicated with red edges and by brackets. Other sequence classifications, including spatio-temporal clusters, are indicated by coloured labels opposite each tip. Phylogeny produced by David Williams, PHE RVU, and reproduced here with permission.

6.5 Combined spatio-temporal and phylogenetic analyses

Spatio-temporal and genomic analyses were conducted in parallel by separate groups, to develop independent evidence of transmission networks. Once completed, I combined the results to compare their networks and, where there were discrepancies, manually reviewed the patient records to identify any epidemiological associations that had not previously been identified.

Of the 24 genomic clusters, 6 were partly or wholly congruent with one or more of the 33 spatio-temporal clusters. Of the 181 inpatients in whom both genomic and spatio-temporal information were available, 108 (59.7%) were found in spatio-temporal clusters and 116 (64.1%) in genomic clusters (**Figure 6.5**); 82 patients (45.3%) were located in both a genomic and a spatio-temporal cluster, of which 42 (51.2%) were HAIs. These 42 patients fell into 7 genomic clusters (range 2 to 26 cases) and were distributed across 11 spatio-temporal clusters (range 2 to 29 cases). The largest cluster confirmed with both genomic sequencing and spatio-temporal analysis involved at least 26 cases with transmission events taking place over 5 weeks and involving 5 wards. The index case was transferred from a rehabilitation hospital as an acute admission via the Emergency Department.

Combined genomic and spatio-temporal analysis demonstrated that 42/54 (78%) HAIs with full genome sequence were part of both spatio-temporal and genomic clusters; 37/42 had congruent genomic and spatio-temporal clusters indicating 6 likely transmission chains across 8 wards, including transplant, care of the elderly, rehabilitation and hepatology specialist wards. The remaining 5 cases were linked genomically to other cases, but fell into 5 five different ward-time clusters so no transmission network could be identified.

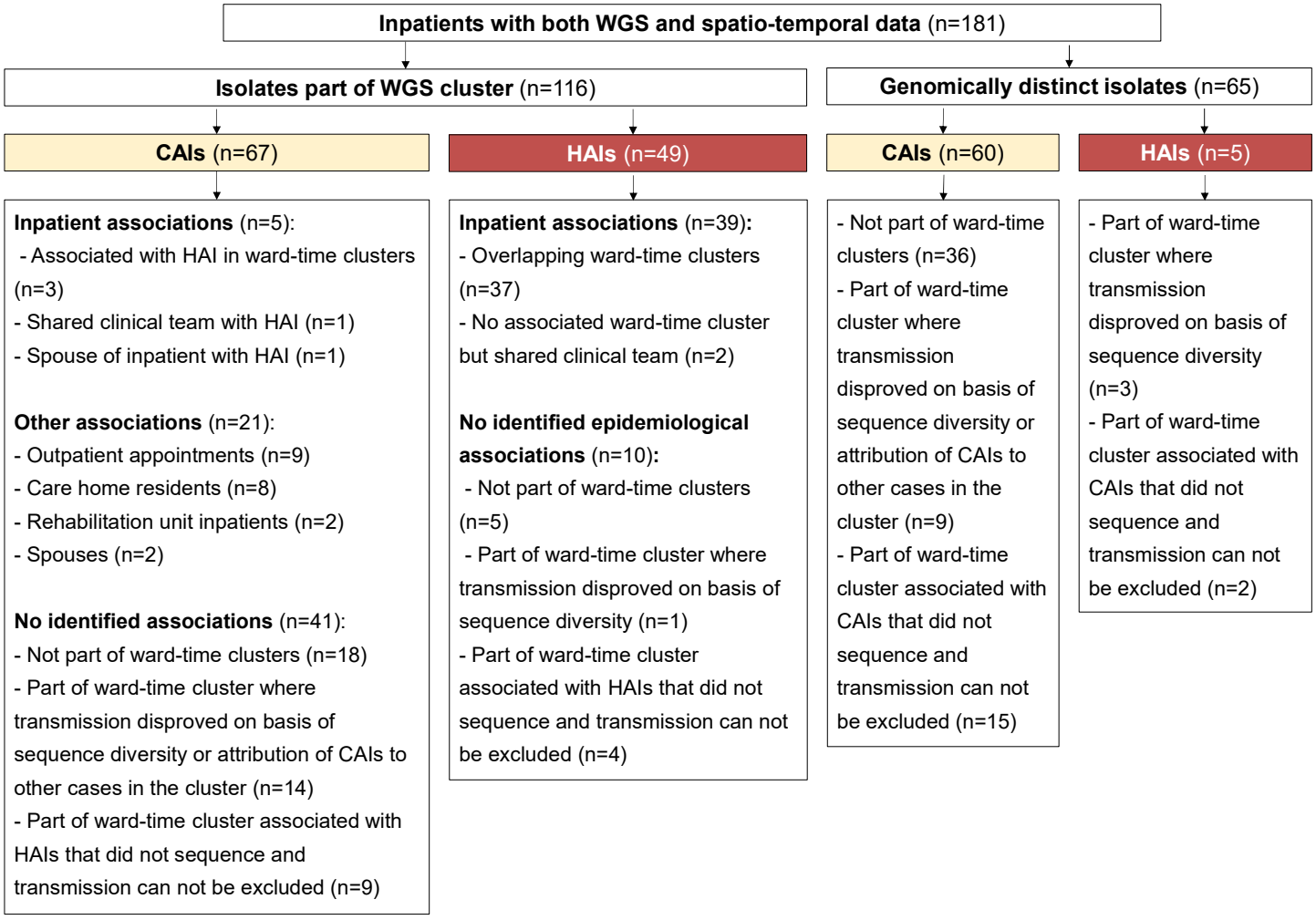


Figure 6.5 Combined spatio-temporal and genomic analysis of influenza A(H3N2) transmission networks in CUH 2016-17

CAI = community acquired infection; HAI = hospital acquired infection

Of the remaining 12/52 HAIs, 7 were linked genomically but had no spatio-temporal connections. Further review of the medical records of these cases found that two of the seven patients could be linked to other genomically-related cases through other epidemiological associations in hospital (one was on a neighbouring ward with shared staff; the other shared the same primary medical team but was located on a separate ward),

raising the possibility of nosocomial transmission from healthcare workers. The final five HAIs shared ward-time events with other influenza patients, but were genomically distinct from the other isolates in the study. However, four of these five cases were part of spatio-temporal clusters with isolates for which no genomic information is available, and transmission events can therefore not be excluded.

Genomic analysis identified highly related cases with community acquired infection that were not identified through spatio-temporal analysis. On detailed review of the medical record, epidemiological associations between some of these cases were identified prior to their admission to hospital, arising from co-habitation or links to health and care facilities. These included transmission networks between spouses (2 cases), nursing home residents (8 cases in 3 clusters), rehabilitation hospital inpatients (2 cases), patients sharing the same dialysis unit (3 cases) and haem-oncology outpatients sharing appointment dates (5 cases in 2 clusters). Genomic analysis also identified the large majority of cases that were admitted from the community that did not lead to transmission events within the hospital. A summary of all cases and possible transmission events is shown in **Figure 6.5**.

6.6 Risk factors for inpatient mortality

Having identified a high proportion of deaths in influenza inpatients during the 2016/17 winter, a retrospective cohort study was undertaken to identify risk factors that were associated with inpatient mortality. Data were abstracted from the patient medical record by myself and a group of clinicians using a structured case report form (**Appendix F.2**) for all 334 cases. Database compiling and cleaning was undertaken by myself and the PHE Field Service. Multivariable analysis using methods described by Hosmer and Lemeshow[260] was

then performed by Neville Verlander, statistician at PHE Colindale, demonstrating the odds of mortality associated with a range of demographic and clinical variables. This work was subsequently published[394].

However, during the following 2017/18 winter there were more than twice as many cases of influenza diagnosed at CUH compared to 2016/17. This was due, at least in part, to three circulating influenza viruses during the 2017/18 winter[116]: influenza A lineages H1N1pdm09 and H3N2 (comprising 373/763 cases, 48.9%), and influenza B (390/763, 51.1%). High rates of hospital admission and mortality were observed, with 52/730 (7.1%) inpatients dying during their hospital stay. To better understand the impact of seasonality and influenza species on mortality, the observational study outlined above was repeated for the 2017/18 winter. Results were combined for both years and yielded a wider range of results with greater degrees of confidence than the 2016/17 winter alone. I have therefore opted to include this latter analysis of both winters in this thesis.

The combined cohort from 2016-18 comprised 1,048 cases with mean age 66 years (range 0 to 102 years) of which 53.3% were female and 83/1,048 (7.9%) patients died during their admission. The median time from admission to death, discharge or transfer to other hospitals was 7 days (interquartile range (IQR) 3-23 days).

A complete single variable analysis of the factors associated with mortality is provided in **Appendix F.3**. Significantly raised odds for inpatient death were associated with increasing age (OR 1.04 per year, 95% CI 1.01-1.06, P<0.001), and Charlson co-morbidity index (OR 2.03 per unit, 95% CI 1.43-4.35, P=0.001). Hospital acquired infection was associated with increased odds of inpatient mortality compared to community acquired infection (OR 2.48,

95% CI 1.55-3.98, *P*=0.003). In comparison to admission from a patient's own home, transfer from another care setting was also associated with mortality (OR 3.78, 95% CI 2.11-6.77). Odds of inpatient death among patients infected with influenza A(H3N2) compared to influenza B were similar (OR 0.96, 95% CI 0.60-1.54), with no deaths in patients infected with influenza A(H1N1pdm09). There was no significant difference in the odds of inpatient death between the 2016-17 and 2017-18 influenza seasons (OR 0.71, 95% CI 0.45-1.13, *P*=0.16).

Clinical features associated with increased risk of mortality included admission to critical care (OR 3.4, 95% CI 1.89-6.12, *P*<0.001) and radiological evidence of pneumonia (OR 3.56, 95% CI 2.25-5.62, *P*<0.001). For blood parameters, complex non-linear associations were observed for serum haemoglobin, total white cell count, serum creatinine and serum urea (**Appendix F.3**). Of note, no association with mortality was identified for some established risk factors for severe disease, including patient pregnancy, recent chemotherapy, radiotherapy, organ transplantation, bone marrow transplantation, and body mass index ≥40 kg/m^2, although absolute numbers in each of these categories were small. Similarly, documented receipt of a seasonal influenza vaccine was not protective for inpatient mortality, although interpretation is limited as comprehensive vaccination history was not available in all patients. Variables in **Appendix F.3** with *P* values in bold met our criteria for being carried forward to the multivariable analysis.

Two multivariable analyses were performed: the first, using categorisation of oseltamivir treatment (not given, standard course, non-standard course) is shown in **Table 6.2**.
The second analysis used the delay from symptom onset to first dose of oseltamivir (to test previous observations that such a delay may impact mortality[131,132]), and is shown in **Appendix F.4**.

Table 6.2 Multivariable analysis of independent factors associated with mortality in patients with confirmed influenza, 2016-18

Original analysis and table produced by Neville Verlander. Total number of patients with complete data = 954.

Variable	Category	OR	95% CI	*P*-value
Age		1.05	1.02-1.07	<0.001
Acquisition	Community	1.00		0.02
	Hospital	2.08	1.12-3.87	
Admitted from	2016/17			0.01*
	Own home	1.00		
	Residential care	0.44	0.10-1.85	
	Another hospital	0.42	0.07-2.67	
	Other	0.00	n.e.	
	2017/18			
	Own home	1.00		
	Residential care	4.65	1.85-11.7	
	Another hospital	2.87	0.59-14.0	
	Other	4.79	0.49-46.5	
Non-invasive ventilation	Yes	6.83	1.91-24.5	0.004
	No	1.00		
Admitted to critical care	Yes	3.16	1.23-8.13	0.02
	No	1.00		
Radiological evidence of pneumonia	Yes	2.08	1.13-3.83	0.02
	No	1.00		
Excessive alcohol use	Yes	22.7	4.67-110	<0.001
	No	1.00		
Oseltamivir course completed	Not given	1.00		0.003
	Non-standard course	0.39	0.14-1.14	
	Standard course	0.18	0.07-0.51	
Haemoglobin g/L		QF[a]		0.009*
Total white cell count		1.08	1.03-1.13	0.002
Urea		QF[b]		<0.001*
Cycle Threshold (CT) value		CF[c]		0.03

**P*-value for interaction n.e.= not estimable; QF Quadratic function, CF Cubic function

QF[a] Haemoglobin g/L 2016/17:- Linear: OR 1.03 (0.82, 1.30) Quadratic: 1.00 (1.00, 1.00)

2017/18:- Linear: OR 0.85 (0.75, 0.96) OR Quadratic: 1.00 (1.00, 1.00)

QF[b] Urea mmol/L 2016/17:-Linear: OR 2.44 (1.22, 4.91), Quadratic: OR 0.96 (0.93, 0.99)

2017/18:- Linear: OR 1.34 (1.15, 1.56) Quadratic: OR 0.99 (0.99, 1.00)

CF[c] CT value Linear: OR 22.7 (0.97, 525) Quadratic: OR 0.88 (0.79, 0.99) Cubic: OR 1.00 (1.00, 1.00)

The variables independently associated with increased odds of inpatient death in our final multivariable model included increasing age, hospital acquisition, and admission from a care home, as detailed above. Clinical features associated with disease severity, including non-invasive ventilation, admission to critical care, radiological evidence of pneumonia and blood markers, were also associated with mortality. Including influenza species had no meaningful effect on the odds ratios and *P* values of the final multivariable model. Although non-significant in the univariable model, the use of oseltamivir was identified as having a protective effect in the multivariable model, with a standard course of treatment reducing odds of mortality by 82% (OR 0.18, 95% CI 0.07-0.51, *P*=0.003).

In the second multivariable model, the category of oseltamivir course received was replaced by the delay between onset of symptoms and first dose of oseltamivir (**Appendix F.4**). This had little impact on the odds ratios and confidence intervals for variables identified in the first analysis. The odds ratio for delay in starting oseltamivir following onset of symptoms was 1.05 (95% CI 0.97-1.14) per day of delay, *P*=0.01.

Oseltamivir was not prescribed for 104/1,048 cases (9.9%), a non-standard course was given for 167/1,048 (15.9%) and a standard course for 777/1,047 (74.1%). While inconsistently documented in the medical record, the reasons identified for not prescribing oseltamivir in the 2016/17 winter were: more than 48 hours having elapsed since the onset of symptoms (13.7%); the patient being discharged before the positive PCR test result for influenza was available (24.7%); clinical improvement (11%); patient dying before the test result was received (2.7%); and medications stopped for palliative care (1.4%).

6.7 Evaluation of influenza control measures

In parallel with the investigations conducted above, I performed a service evaluation of clinical management of patients with influenza to determine the standard of care in CUH over the 2016/17 winter. This focused on 3 key areas: 1) identification and testing for influenza in community acquired cases on presentation to hospital; 2) appropriate isolation of cases on admission to the ward; 3) oseltamivir prescribing practices. All influenza testing at the time was conducted by PCR at the PHE regional microbiology laboratory on the Cambridge biomedical campus. The service evaluation covered a 9 week period from December 18 2016 to February 19 2017, over which time the majority of the winter's confirmed cases of influenza A (283/334, 84.7%) were identified.

For community acquired cases, only 48.0% were swabbed in the Emergency Department (ED) prior to admission to hospital. This proportion remained largely constant over the study period, despite recognition of suspected hospital transmission events and internal staff communications on the importance of screening. Similarly, the proportion of patients with influenza who were admitted to an appropriate isolation bed (either a side room or cohort bay) was 48.0%. There were substantial delays in the time taken for samples to be booked into the laboratory once they had been collected (median 10.6 hours, IQR 5.3-16.3 hours) and for the laboratory to process the sample and release a result (median 19.7 hours, IQR 12.3-22.7 hours).

National and hospital guidance at the time of the study was to commence treatment with oseltamivir in admitted patients who have been symptomatic for less than 48 hours and who are at high risk of complications from influenza[135]. These included the recommendation to

start treatment empirically prior to results becoming available. In 64/283 (22.6%) patients, oseltamivir was not given at all; in 152/219 (69.4%) patients with whom the drug was given, there was a delay in administration of the first dose until after the positive test result was available. Additionally, there was a median delay from result verification to first dose administration of 7.8 hours (IQR 3.3-20.6 hours). Combined with the delays in sample collection and result verification, this led to a total median delay from hospital presentation to administration of first dose of definitive treatment of 46 hours.

The results of this service evaluation, combined with preliminary findings from the combined genomic and epidemiological analyses, were presented to clinical teams in infection specialties, emergency and acute medicine and CUH senior management in the summer of 2017. As a result, five key multidisciplinary interventions were introduced prior to the 2017/18 winter:

1. **Improved isolation procedures.** Cohorting of confirmed influenza cases and the isolation of influenza-like illness was sporadically and inconsistently used during the 2016/17 season. Procedures for repurposing the use of ward N2 (a 22-bedded unit composed entirely of side rooms) as an isolation facility for admissions of suspected influenza cases were endorsed by senior management and the unit was opened in November 2017.

2. **Laboratory practices.** The frequency of respiratory virus PCR testing was increased from one to two batches every day during the influenza season. The timing of the batches was moved earlier in the day to ensure that results were available in working hours, to facilitate isolation decisions and maintain patient flow through side rooms, in particular ward N2. In addition, the development of the “Respiratory Virus Tracker”, a web-based data visualisation tool that could be used by all staff, enabled rapid

visualisation of influenza activity in the Trust to guide infection control and operational decisions.

3. **Pharmacy and prescribing practices**. Interventions developed with the Trust's lead antimicrobial pharmacist led to wider dissemination of prescribing guidelines (including an antimicrobial app available for smartphones and widely used by junior doctors) and broader availability of oseltamivir (with a stock maintained in 11 clinical areas in comparison to 3 in 2016/17).

4. **Vaccination of healthcare workers.** The occupational health service led a well-publicised campaign of seasonal influenza vaccination within the Trust. During the 2016/17 season, the national target of vaccinating 75% healthcare workers was not reached until March, in the final weeks of the influenza season. By contrast, in 2017/18 this target was reached in early November.

5. **Education.** The results of the 2016/17 service evaluation and changes to practice listed above were disseminated in a programme of presentations to infection specialities and clinical teams, supported by written summaries sent to all medical consultants and trainees in the Trust at the onset of the 2017/18 influenza season. Daily updates and brief training sessions on influenza management were delivered to the acute medical team in the ED by infection specialists from December 2017 to the end of the influenza season in March 2018. Nursing leads provided personal training with all nurses in the ED and utilised social media platforms to update staff. Trust influenza guidelines were made widely available on the CUH intranet and clinical

teams were directed to them in a note placed in the patient's medical record by the duty virologist at the time that the diagnosis was confirmed in the laboratory.

Progress in influenza management was evaluated in a second service evaluation covering the 2017/18 winter. Although improvements were seen in patient care over this time (shown in **Figure 6.6**), there remained long delays from patient presentation to hospital and a result being made available. This led to long waits to move patients out of the isolation facility on N2 and into cohort bays or standard hospital wards depending on the result. Following further discussions with the various parties involved in the trust's influenza pathway, the main recommendation of this service evaluation was the introduction of a molecular point of care test (POCT) for influenza in the ED. Recent evidence had shown that an influenza POCT substantially reduces result turnaround times, while reducing overall healthcare costs (through admission avoidance and earlier discharges), and rationalising infection control decisions when side rooms are a limited resourced[395]. After the evaluation of two PCR-based assays on-site in the POCT laboratory over the 2017-18 winter, a Cepheid GeneXpert rapid diagnostic platform[254] was acquired and became fully operational in December 2018. All other interventions, including a further education programme using the same format as that delivered in 2017/18, remained in place for the 2018/19 winter.

The performance of the Cepheid GeneXpert in comparison to the standard of care PHE laboratory PCR test (based on the Rotor-Gene Q platform (Qiagen) using primers and probes developed in the laboratory[256]) is shown in **Table 6.3**. The results between the two platforms were largely concordant, with the Cepheid providing 95.2% sensitivity and 99.9% specificity in comparison to the PHE laboratory test. The majority of discordant results had low levels of virus (Ct values 32-35). The mean delay from sample collection to result verification for the hospital's main laboratory was 28 hours, in comparison to 50 minutes for the Cepheid GeneXpert. Overall, 79% of confirmed influenza diagnoses in CAIs were made using the Cepheid platform during the 2018/19 winter.

Table 6.3 Comparison of the performance of the Cepheid GeneXpert Flu Assay against the PHE laboratory PCR assay, 2018/19

		PHE laboratory		Total
		Positive	Negative	
Cepheid GeneXpert	Positive	546	2	**548**
	Negative	27	1,460	**1,487**
Total		**573**	**1,462**	**2,035**

Improvements in adult patient management on admission over the course of three winters from 2016-2019 are shown in **Figure 6.6a**. Isolation of patients on admission to the ward rose from 48% of influenza cases in 2016-17 to 93% in 2018-19. Over the same period, the median time from admission to swabbing fell from 9 hours to 2 hours. The use of a POCT in the ED enabled more prompt cohorting of influenza patients in bays in the 2018/19 winter, increasing the availability of side rooms for other indications and improving flow through the hospital.

The median delay from admission to first antiviral dose fell from 46.0 hours in 2016/17 to 7.1 hours in 2018/19 (**Figure 6.6b**). Rapid diagnosis of influenza also led to a reduction in empiric oseltamivir prescribing: 1527 patients without influenza were treated with oseltamivir in 2017-18, compared to 473 patients in 2018-19.

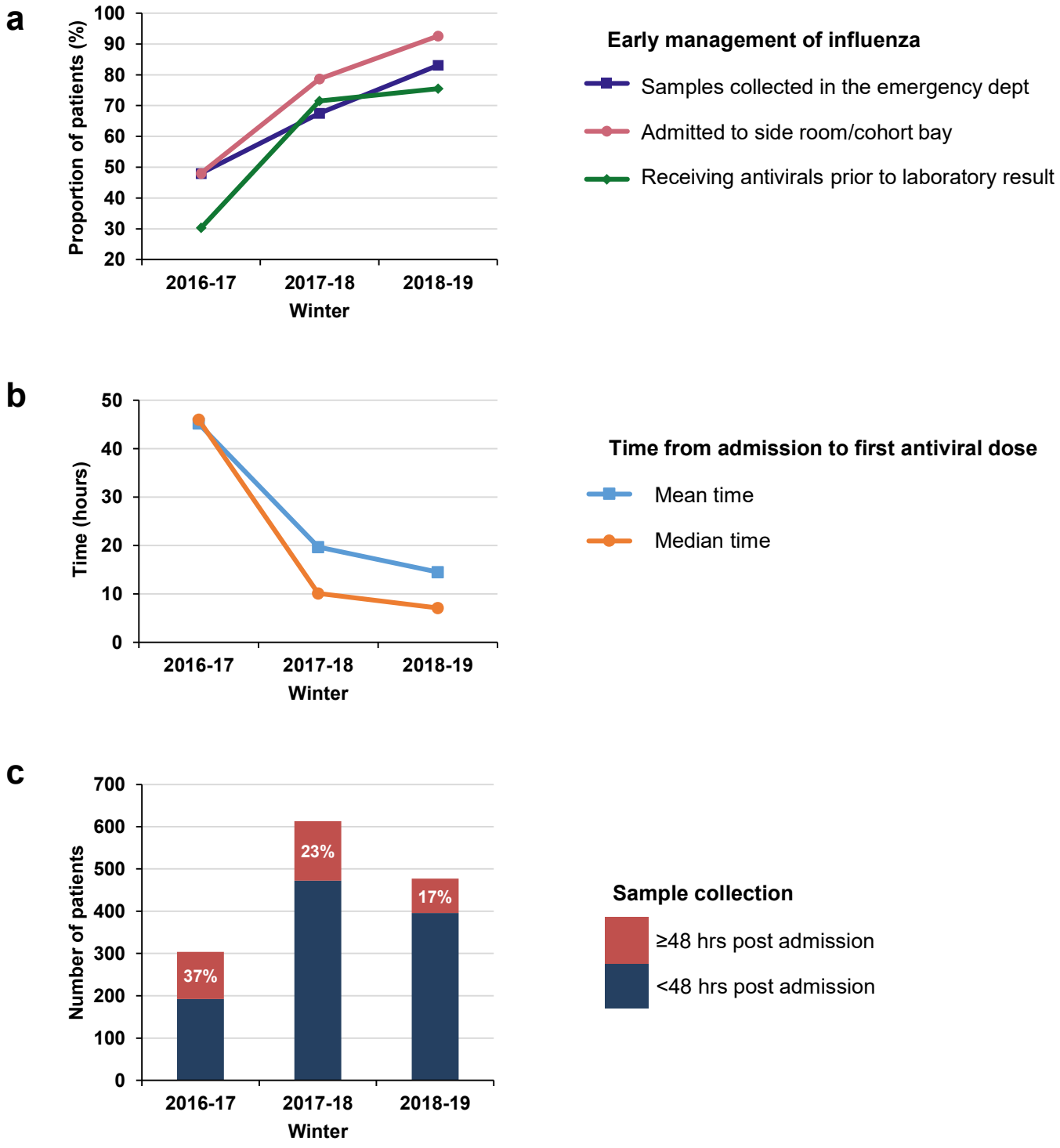


Figure 6.6 Changes in influenza management at CUH over 3 winters, 2016-19

(**a**) Early management of patients with influenza. Cases shown are limited to confirmed, community acquired influenza infections in adult inpatients admitted via the Emergency Department. (**b**) Delay from first presentation to hospital to first dose administration of antiviral, using the same criteria as (**a**). (**c**) Total number of adult admissions to CUH with confirmed influenza, with the proportion of hospital acquired cases (defined as samples collected ≥48 hours from initial presentation to hospital) shown in red. Inpatient all-cause mortality for each influenza season is annotated in white at the top of each bar.

The characteristics of patients with influenza in CUH over these three winters is shown in **Table 6.4**. There has been a trend of increased numbers of patients being identified as influenza positive since 2016. This is, in part, due to increased burden of influenza in the community (especially in 2017-18), but another major factor relates to change in clinical practice. The introduction of a POCT in the ED led to a larger number of swabs being performed on young adults who were subsequently discharged directly from ED, who may not have been swabbed in previous years (contributing to the lower proportion of adult admissions and fall in median age of all cases). The largest change in practice was in paediatrics, who had a 12-fold increase in the number of cases of influenza diagnosed over this period. Of the 290 patients with a positive influenza result who were not admitted, 92% were discharged from ED and 35% were paediatric patients.

Despite the limitations in the use of the 48 hour cut-off definition of hospital acquired infection described above, a manual review of cases could not be performed for the 2018/19 winter. Using the 48 hour cut-off for comparison of cases across winters, the proportion of adult influenza infections acquired in hospital fell from 36.8% in 2016-17 to 17.0% in 2018-19 (**Figure 6.6c**). Overall adult inpatient mortality fell over the same period from 9.7% to 3.6%.

Table 6.4 Characteristics of confirmed influenza cases in CUH by season, 2016-19

	2016-17	2017-18	2018-19
Total number of patients with confirmed influenza	348	763	823
Influenza A	334 (96%)	373 (49%)	818 (99%)
Influenza B	14 (4%)	390 (51%)	5 (1%)
Paediatric cases	12 (3%)	40 (5%)	157 (19*)
Adults admitted	315 (94%)	613 (85%)	477 (72%)
Female adults	54%	52%	53%
Median adult age	77 yrs	74 yrs	58 yrs
Community acquired cases (samples taken <48 hours from admission, adult inpatients)	192 (63%)	472 (77%)	396 (83%)
Hospital acquired cases (samples taken >=48 hours from admission, adult inpatients)	112 (37%)	141 (23%)	81 (17%)
Overall inpatient mortality	32 (9.7%)	52 (7.1%)	19 (3.6%)
Mortality in hospital acquired cases	18 (56%)	18 (35%)	5 (26%)

6.8 Discussion

In this chapter I have demonstrated the importance of influenza in causing mortality and hospital-associated transmission in a large teaching hospital over three winters. Combined genomic and epidemiological analyses from a large outbreak investigation of the 2016/17 winter have identified complex transmission networks in multiple inpatient and outpatient clinical areas, as well as transmission events in the community. Using detailed clinical data

from over 1,000 influenza patients over two years, I have demonstrated the risk factors associated with mortality from an admission with influenza and the protective effect of oseltamivir. Finally, I have shown how a combination of low cost, multidisciplinary interventions and the introduction of point of care testing have improved the clinical management of influenza, associated with a 54% reduction in the proportion of hospital acquired cases and a 63% reduction in inpatient mortality over three years.

The 2016/17 winter was characterised by high numbers of community and hospital cases of influenza A(H3N2) across Europe, in a season in which the seasonal influenza vaccine was estimated to have only moderate efficacy of 30-40%, with disproportionately poor efficacy in people aged over 75[396,397]. This is reflected in our predominantly elderly cohort, in whom the seasonal vaccine was not associated with protection against mortality. Over one third of these cases were categorised as being HAIs, based on time from presentation to sample collection. This is higher than previous studies of influenza in secondary care in high income settings, which have demonstrated a range of 6% to 24%, although direct comparison is challenging due to methodological differences and lack of a standard definition of HAI[133,391,392,398]. Indeed, in this study I have shown that there are multiple delays in the recognition of influenza and associated laboratory testing that can lead to false categorisation of HAI. While detailed reviews of all cases may not be feasible or timely in all cases, there is increasing recognition of the importance of root cause analysis of HAIs[399] and duty of candour to affected patients[400], in which understanding of such categorisation is imperative.

Where epidemiological and genomic data were available, 37/54 (68.5%) HAIs fell within 6 transmission networks across 8 wards. This shows that, despite incomplete availability of

sequences, the majority of HAIs could be linked to a transmission network. This highlights the utility of an approach combining phylogenetics with routinely collected patient movement data to assess influenza transmission, a critical step in guiding infection control practices. This study, which has attempted to review all influenza cases in a large tertiary referral centre over a whole influenza season, is larger than previously published studies using genomic epidemiology, which have either focused on small outbreaks in a limited number of clinical areas[150,151], or have a lower ascertainment of viral sequences[401]. However, these approaches also highlight the complexity of transmission: their individual use can over-estimate patient clustering, while combined they reveal gaps in networks which may represent asymptomatic or untested patients, healthcare workers (HCWs) and hospital visitors. While genomic investigation is highly specific, it suffers from under-representation of isolates through sampling and technical biases which, when combined with the high level of sequence identity in many clusters, limits interpretation of the directionality of transmission between individual patients. A particular limitation of this study is the lack of data from HCWs and hospital visitors. Given the high prevalence of community infection and incomplete vaccine efficacy, it is entirely possible that HCWs and hospital visitors may contribute to transmission networks[402]. This is suggested in our study by a number of transmission networks for which no index CAI can be identified.

A further limitation was the retrospective nature of the study. Because of the short incubation period of influenza in comparison to other pathogens, real-time sequencing and analysis is required to maximise infection control interventions to impact ongoing transmission events. However, genome sequencing is currently comparably expensive and challenging to implement for influenza. By contrast, ward movement data is routinely collected in real-time in electronic medical record systems and could be utilised by infection control teams to detect outbreaks and target interventions. The spatio-temporal analysis shown here used simple algorithms based on patient co-location in time and ward, combined with a basic

understanding of the incubation and infectivity periods of the virus. Although they lack specificity, such approaches could be used to study a range of pathogens implicated in nosocomial infection. As genome sequencing becomes more readily available in the clinical setting, the combination of the two are likely to be highly effective in understanding and intervening in nosocomial transmission.

This investigation has demonstrated the extent and severity of nosocomial influenza. The clinical areas in which transmission occurred provided cared for patients in high risk groups for developing severe influenza, such as geriatric patients and solid organ transplant recipients. Nosocomial transmission was associated with high rates of inpatient mortality across the study period. Previous studies have identified a higher risk of mortality with HAI, again identifying patients of older age, or with chronic kidney disease and immunodeficiency, as being at particular risk of mortality[403]. Additionally, I identified several clusters imported into hospital, including transmission in three outpatient units and three care homes. These again represent high risk groups, including outbreaks in a dialysis unit and haematology outpatients as well as long term care institutions, for which there is considerable existing evidence of institution-based transmission[404,405].

To understand the risk factors for mortality in our influenza cohort, a retrospective cohort study including data from over 1,000 inpatients was completed using single variable and stepwise multivariable logistic regression. This is one of the largest studies to comprise largely of the epidemic H3N2 subtype and contained thorough measurement and adjustment for demographic, social and clinical associations with poor outcome. Many of the risk factors identified here had previously been reported in cohort studies during the H1N1pdm09 pandemic, including age, underlying comorbidities, elevated inflammatory markers,

radiological evidence of pneumonia, and the need for supplementary oxygen and ventilation[134].

Of note, we have identified the considerable benefit of a standard course of oseltamivir therapy, which reduced the adjusted odds of mortality by 82% in comparison to those who did not receive the drug in a multivariable model that accounted for many confounders of mortality. The effect of oseltamivir was highest when administered on the day of admission, with increased odds of mortality for each day's delay in treatment. While the indiscriminate use of oseltamivir in treating influenza remains controversial, there is a lack of randomised controlled trial data in high risk populations in hospital settings. Most published trials have been conducted in populations of young adults with few co-morbidities, in outpatient settings, or using an inconsistent approach to diagnosis (sometimes relying on symptomatology rather than laboratory confirmed disease), limiting their application to a predominantly elderly, co-morbid hospital population[141,142,406]. This study supports previous observations made following the H1N1pdm09 pandemic that oseltamivir carries a protective effect on mortality[131,132], and that this may extend beyond the currently recommended treatment window of two days from symptom onset, to at least five days. The additional benefit found in our cohort may reflect the higher proportions of elderly and immunocompromised patients who have longer times from symptom onset to cessation of viral replication, supporting the use of oseltamivir for longer periods[144,407].

Although recommended by CUH and national guidelines[135], 10% of the cohort from 2016-18 did not receive oseltamivir during their admission. Incomplete adherence to guidelines may be explained by prescribing inhibition due to incomplete evidence to support their implementation[139], and the widespread coverage of concerns for oseltamivir stockpiling for pandemic influenza preparedness[140,408,409].

Influenza B has previously been considered a cause of milder illness than influenza A subtypes. However, the proportion of inpatient deaths we observed were similar for influenza A(H3N2) (51/610 (8.4%)) and influenza B (30/371 (8.1%). This supports the emerging view that influenza B can cause as much severe disease and death as influenza A(H3N2)[410]. The role of oseltamivir in the treatment of influenza B has previously been poorly studied. In order to examine if influenza species altered the risk estimates for the variables in the final multivariable model, influenza species was added in a supplementary analysis. This made no meaningful difference to the odds ratios and *P* values of the variables. This observation is consistent with the protective association of oseltamivir treatment and odds of inpatient mortality being similar for influenza B and influenza A(H3N2).

Observational studies are recognised to be prone to multiple sources of error including selection and measurement biases and incomplete adjustment for confounding. However, by reviewing all patients admitted with confirmed seasonal influenza without any restriction, selection bias was avoided. The use of a clear standardised case record form completed by a number of practising physicians using electronic medical records mitigated against selection and measurement biases as far as possible.

Through these studies the hospital obtained a far greater understanding of the transmission and mortality impact of influenza and, consequently, how management needed to improve. Initial interventions (such as staff education, improved isolation procedures and adjustments to laboratory procedures) were largely cost-neutral and could therefore readily be translated to other settings. However, they were also complex and relied on communication between a number of disparate hospital departments (laboratory, clinical, operations, etc.) for which dedicated organisational structures were required. Additionally, CUH successfully introduced

an accurate influenza POCT that reduced delays from admission to diagnosis by approximately 24 hours. Previous studies have demonstrated the additional benefits of POCT for influenza, including shorter lengths of stay, reduced delays from admission to antiviral therapy and isolation, leading to an overall cost-saving despite higher testing costs[395,411,412].

Education programmes were likely key to the success of the interventions in CUH. Providing updates to clinical teams was challenging due to staff turnover and large numbers of contributing clinicians, but the utilisation of multiple approaches (such as opportunistic face-to-face teaching, social media platforms, formal training days, dissemination of electronic and paper versions of the new guidelines) ensured that any new policies were being used by teams caring for influenza patients. Additionally, the use of local data obtained from the above studies was instrumental in staff engagement and education.

Following these interventions there was a substantial change in clinical practice, including a 73% increase in the proportion of inpatients screened in the ED, a 93% increase in the proportion admitted to isolation facilities, and an 85% reduction in the median delay from admission to first dose of antiviral. This has been associated with a substantial fall in the rates of HAI and inpatient mortality. However, the rate of HAI remained relatively high at 17% in 2018/19. There are multiple possible explanations for this, including: ongoing inconsistent management of influenza cases in the ED, allowing spread of undiagnosed influenza to wards; sources of influenza other than patients (including healthcare workers and visitors) that have not been accounted for; and the delay in recognition and therefore management of influenza outbreaks on hospital wards.

Although this work has focused on influenza, in **Chapter 3.1** I showed that other respiratory viruses have contributed a substantial burden of disease in CUH. For example, similar combined genomic epidemiology approaches demonstrated an outbreak of parainfluenza on

a paediatric ward in CUH in 2017[413]. Further discussions between CUH, PHE and academic collaborators during the 2019/20 winter centred on expansion of the influenza studies presented here to other pathogens. However, this work was interrupted by the emergence of a novel respiratory virus at the end of 2019: SARS-CoV-2. In **Chapter 7** I demonstrate the impact that COVID-19 has had on CUH, how combined genomic-epidemiology studies were used to evaluate nosocomial transmission of SARS-CoV-2 and how lessons learned from seasonal influenza could be translated into the hospital's pandemic response.

7. Hospital acquired COVID-19 in CUH

7.1 Introduction

On December 31 2019 I received an email from the internet-based reporting system ProMED (the Program for Monitoring Emerging Diseases), describing a cluster of unexplained pneumonia cases in Wuhan, China, resembling the Severe Acute Respiratory Syndrome[414,415]. This was the first global report of COVID-19, a disease that, by October 2021, has caused over 240 million confirmed cases and 4.9 million deaths[9].

Reports from Wuhan in early 2020 provided alarming statistics about rates of hospitalisation, critical care admission, nosocomial transmission and excess mortality that were placing unprecedented demand on hospital services[167,416,417]. At the time, however, CUH was in the midst of a winter characterised by endemic respiratory viruses such as influenza and RSV (**Figure 3.1**), causing similar numbers of admissions and pressure on hospital beds as the preceding two winters[418]. Although rising cases in Northern Italy indicated that SARS-CoV-2 transmission within the UK was increasingly likely, guidance from central government and national public health agencies in preparation for the pandemic was initially lacking. Internal discussions in winter planning sessions turned from influenza to the emerging threat of the novel coronavirus. While previous work on influenza had provided some infrastructure and interdisciplinary communication between groups that would become key in the pandemic response (including laboratory specialties, infection control, medical teams, operational leads and senior management) it was readily apparent that many of the interventions

successfully introduced to control influenza were unavailable for the control of COVID-19. At the time of our first case on March 10 2020 there were no rapid diagnostics, proven pharmacological therapies or vaccines, and only limited data on the clinical features and pathogenicity associated with this novel virus on which to base clinical and infection control decisions.

Nevertheless, teams from across the Cambridge biomedical campus were able to rapidly mobilise clinical and academic expertise to both study and manage the local effects of the pandemic. Like many of my clinical academic colleagues, I placed my existing research on hold to return to CUH at the start of March 2020. Drawing on my experience with influenza and AMR, I played a multifaceted role in the hospital's pandemic response during the first wave. In this chapter I will demonstrate how large amounts of clinical data from our COVID-19 cohort were automatically extracted from the patient medical record system, analysed and utilised for a variety of applications in CUH. I will illustrate the considerable burden of SARS-CoV-2 as a nosocomial pathogen and how combined genomic and epidemiological approaches were used to describe detailed networks of hospital transmission. In comparison to the influenza study presented in **Chapter 6**, these analyses were performed in real time, and could therefore be integrated into a novel HAI assessment process that could make rapid patient safety and infection control policy decisions to reduce transmission. Finally, I will describe how these data have been used for other research into SARS-CoV-2, and how CUH became a sentinel site for the study of hospital-associated COVID-19 during the first wave of the pandemic, informing national policy.

7.2 Burden of SARS-CoV-2 at Cambridge University Hospitals

One of the earliest interventions made during my involvement in pandemic preparedness was to establish a system of prospective detailed extraction of clinical and demographic data from the medical record to better understand our COVID-19 patient cohort, and the demand they would place on local services. Previous work on influenza had demonstrated the utility of collating data on the clinical features, management and outcome of patients for a variety of purposes, including research, service evaluation and implementation of new control interventions. While some of the data presented in **Chapter 6** could be extracted electronically from the patient record system, Epic, the process was time-consuming and limited to a small number of variables at any one time. A considerable amount of the data therefore relied on manual abstraction that was both time consuming and prone to human error. In late 2019 I began work with the data management team at CUH to develop automated extracts of demographic, clinical, laboratory, ward movement and outcome data from all patients identified as having a confirmed influenza result within the laboratory information management system. At the onset of the pandemic work to develop this data extract was prioritised by the hospital, with influenza cases being exchanged for confirmed SARS-CoV-2 diagnoses, and the build was completed during March 2020. Updated daily, this extract provided over 100 data points per patient. In addition to providing variables in a tabular format that could be imported into statistical analysis software, additional dashboards with interactive graphical interfaces were produced by the data management team that could provide rapid overviews of trends in variables of operational interest, including testing requirements, confirmed case numbers and bed occupancy. An overview of the COVID-19 pandemic at CUH from the first case on March 10 2020 to April 8 2021 is shown in **Figure 7.1**.

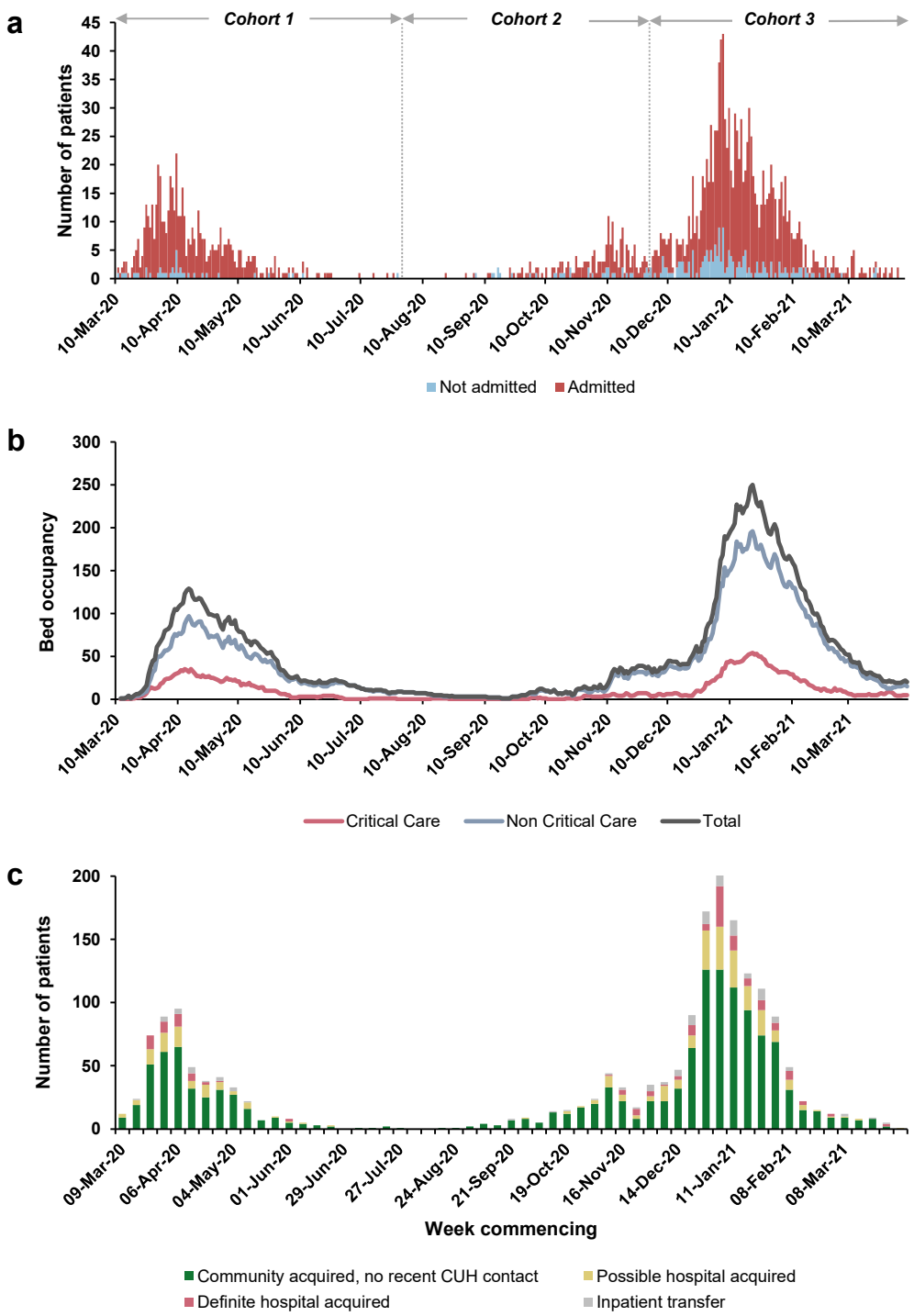


Figure 7.1 Burden of SARS-CoV-2 infection in CUH, March 10 2020 – April 8 2021

(**a**) Daily number of newly confirmed SARS-CoV-2 cases at CUH, by date of swab collection. Bar stacks are coloured by the number of patients admitted to hospital (red), or diagnosed in CUH without formal admission (blue). The cohorts described in **Table 7.1** are shown by the grey arrows. (**b**) Bed occupancy of confirmed SARS-CoV-2 cases, by critical care (red line), general medical (blue line) and total occupied beds. (**c**) Epidemic curve of cases by acquisition of virus, using weekly totals. A simplified categorisation is used, where possible hospital acquired infection refers to any patient swabbed between 2 and 14 days of admission, or following a recent inpatient encounter at CUH within the 2 weeks prior to swab collection. Definite hospital acquired infections were swabbed more than 14 days into their admission. All cases are limited to patients with confirmed SARS-CoV-2 infection using molecular tests performed in CUH.

For the purposes of this thesis, patients have been presented as three “cohorts” over the course of the first 13 months of the pandemic, reflecting the national picture of COVID-19 in the UK. Cohort 1 represents the period from our first confirmed case in CUH on March 10 2020 to July 31 2020, the end of the first wave of cases in the UK and easing of national lockdown restrictions. Cohort 2 spans the period from August 1 to November 30 2020, representing a period of slowly increasing patient numbers until the second national lockdown in November 2020, as well as the first evidence-based changes to patient management (such as the use of dexamethasone following the initial results of the RECOVERY trial[419]). Cohort 3 is taken from December 1 2020 to April 8 2021, reflecting a rise in admissions and the increasing national prevalence of the SARS-CoV-2 alpha variant (lineage B.1.1.7)[420], associated with a third national lockdown.

As shown in **Figure 7.1a** and **Table 7.1**, 1,932 patients were diagnosed with SARS-CoV-2 infection in CUH between March 10 2020 and April 8 2021. There was an overall associated mortality (either during the admission or within 30 days of the first positive swab) of 17%. The majority of cases were admitted to hospital, although the higher proportion in cohort 1 may reflect changes to diagnostic policy in the first wave when testing was initially restricted to patients admitted to hospital. During summer 2020, when elective procedures such as endoscopy were recommenced, a programme of asymptomatic screening of patients prior to their hospital attendance was adopted leading to an increased number of outpatient diagnoses of COVID-19 shown in cohorts 2 and 3.

Table 7.1 – Summary of confirmed COVID-19 cases in CUH patients, 2020-21

Patient Outcome (as of April 8 2021)	Cohort 1	Cohort 2	Cohort 3	Total
Current inpatients on medical wards	0	1 (1%)	11 (1%)	12 (1%)
Current inpatients on critical care	0	0	5 (<1%)	5 (<1%)
Discharged alive from inpatient stay	345 (65%)	129 (65%)	738 (61%)	1212 (63%)
Swabbed in ED/N2/outpatients only*	43 (8%)	38 (19%)	185 (15%)	266 (14%)
Died as inpatients or within 30 days of first positive swab	118 (22%)	22 (11%)	196 (16%)	336 (17%)
Inpatient transfer from CUH to another hospital	26 (5%)	10 (5%)	65 (5%)	101 (5%)
Total	**532**	**200**	**1200**	**1932**

*Note that an increasing number of outpatients have been swabbed as part of routine asymptomatic screening during Cohorts 2 and 3.

Bed occupancy data shown in **Figure 7.1b** illustrates the high demand in both general medicine and critical care. There was a lag of 1-2 weeks between peaks of admissions and bed occupancy in each cohort; at the maximum occupancy in the third week of January 2021 there were 54 critical care beds and 194 general beds occupied by COVID-19 patients.

Figure 7.1c shows the burden of hospital associated infection (HAI) during the first and second waves. Given the variable incubation period of SARS-CoV-2 (median 5 days, range 2 to 14 days[167]), distinguishing CAI and HAI can be challenging in patients who develop symptoms during their stay. Excluding inpatient transfers from other hospitals, 147/1,564 (9.4%) inpatient infections were definitely acquired in hospital (defined as a delay from admission to sample collection of first positive result >14 days). An additional 294/1,564 cases (18.8%) may have been acquired in hospital (defined as swab collection between 2 and 14 days from admission, or any previous admission in the two weeks prior to swab collection – see **Table 7.3**). These issues are discussed in detail below.

7.3 Clinical features of SARS-CoV-2 infection

The characteristics of 1,666 inpatients with COVID-19 in CUH are shown in **Table 7.2**. While the range of ages seen in all cohorts was wide, there was a disposition towards older adults (median age 68, IQR 51-80). This distribution has changed over the course of the pandemic, with older patients in cohort 1 (median age 70, IQR 55-82) in comparison to cohort 2 (between the first and second waves of the pandemic, median age 60, IQR 44-78). A minority of patients were paediatric (30/1666, 2%).

COVID-19 patients placed a considerable burden on beds within the hospital, with an overall mean length of stay of 9.6 days. This has changed over the pandemic, being 14.4 days in cohort 1 but 8.2 days in cohort 3. The reasons for this are likely multifactorial, including the younger age of patients, lower thresholds for admission, improved diagnostics and treatments available in the second wave. A high proportion of patients (302/1,666, 18%) were admitted to critical care, where the mean length of stay was 14.0 days over all three cohorts, but again with a trend of reduced duration as the pandemic progressed (**Appendix H.1**). Reflecting CUH's role as a tertiary referral hospital and its assignation as a surge centre for COVID-19, 18% of all confirmed cases were patients who were resident outside the hospital's normal emergency catchment area.

These data were used by the hospital to estimate key operational requirements during the course of the pandemic, in particular the modelling of bed requirements to plan for additional general medical COVID-19 cohort wards and the expansion of critical care beds. Additional modelling was conducted to estimate requirements for staffing, oxygen, ventilators and personal protective equipment (PPE). Similarly, these data were used in recovery phases between pandemic waves, to plan for re-opening of elective services.

Table 7.2 – Patient characteristics of COVID-19 cases admitted to CUH, 2020-21

	Cohort 1 (n=489)	**Cohort 2 (n=162)**	**Cohort 3 (n=1,015)**	**Total (n=1,666)**
Age, mean (range) median (IQR)	67 (0-98) 70 (55-82)	60 (3-93) 60 (44-78)	64 (0-102) 66 (49-80)	65 (0-102) 68 (51-80)
Paediatric (age <16 yrs)	10 (2%)	2 (1%)	18 (2%)	30 (2%)
Male gender (%)	288 (59%)	78 (48%)	539 (53%)	905 (54%)
Female gender (%)	201 (41%)	84 (52%)	476 (47%)	761 (46%)
Ethnicity – White	374 (76%)	100 (62%)	675 (67%)	1,149 (69%)
Ethnicity – Black, Asian and minority ethnic	46 (9%)	30 (19%)	93 (9%)	169 (10%)
Ethnicity – missing/not stated	69 (14%)	32 (20%)	247 (24%)	348 (21%)
Charlson co-morbidity index, median (IQR)	1 (0-3)	1 (0-2)	1 (0-2)	1 (0-2)
Critical care admissions	95 (19%)	21 (13%)	186 (18%)	302 (18%)
Admissions from outside CUH emergency catchment area	89/489 (18%)	31/161 (19%)	187/1,013 (18%)	307/1,663 (18%)
Length of stay	**Cohort 1 (n=238)**	**Cohort 2 (n=104)**	**Cohort 3 (n=528)**	**Total (n=870)**
LOS* (days), mean (range) median (IQR)	14.4 (0.3-164.6) 8.8 (2.8-17.3)	6.7 (0.3-46.0) 2.6 (1.0-8.7)	8.0 (0.3-79.8) 4.7 (1.3-10.3)	9.6 (0.3-164.6) 5.2 (1.5-11.9)
Length of stay (min 30 days post first positive result)**	**Cohort 1 (n=238)**	**Cohort 2 (n=104)**	**Cohort 3 (n=507)**	**Total (n=849)**
LOS**(days), mean (range) median (IQR)	14.4 (0.3-164.6) 8.8 (2.8-17.3)	6.7 (0.3-46.0) 2.6 (1.0-8.7)	8.2 (0.3-79.8) 4.9 (1.3-10.8)	9.7 (0.3-164.6) 5.3 (1.3-10.8)

*discharged patients, not including readmissions and inpatient transfers, with community acquired infections;

**patients diagnosed >30 days prior to report; discharged patients, not including readmissions, with community acquired infections); LOS = Length of stay, IQR = interquartile range

Table 7.3 summarises the patient cohorts that died during their admission with COVID-19, or within 30 days of their first positive SARS-CoV-2 swab. These patients were older than the general inpatient population, with a higher proportion of male cases. In our cohort, there was no major difference in mortality between ethnic groups, although interpretation is limited by small numbers of patients from ethnic minorities and incomplete data documented in the patient record. The mean length of stay for these patients was 11.6 days, comparable to 9.6 days for all inpatients. The majority of deaths occurred on general medical wards, although there was 32% mortality in all patients requiring critical care admission (**Appendix H.1**).

Table 7.3 Characteristics of COVID-19 fatalities in CUH, 2020-21

Deaths occurred during the patient's hospital admission or within 30 days of their first positive COVID swab collection

	Cohort 1 (n=118)	**Cohort 2 (n=22)**	**Cohort 3 (n=196)**	**Total (n=336)**
Age, mean (range) median (IQR)	78 (0-97) 80 (73-87)	73 (43-93) 76 (63-83)	77 (25-101) 78 (67-86)	77 (0-101) 79 (71-86)
Male gender (%)	75 (64%)	14 (64%)	122 (62%)	211 (63%)
Female gender (%)	43 (36%)	8 (36%)	74 (38%)	125 (37%)
Ethnicity – White	90 (76%)	14 (64%)	135 (69%)	239 (71%)
Ethnicity – Black, Asian and minority ethnic	4 (3%)	4 (18%)	18 (9%)	26 (8%)
Ethnicity – missing/not stated	24 (20%)	4 (18%)	43 (22%)	71 (21%)
Critical care admission	28 (24%)	9 (41%)	56 (29%)	93 (28%)
Length of stay (days)*	**Cohort 1 (n=77)**	**Cohort 2 (n=8)**	**Cohort 3 (n=95)**	**Total (n=180)**
mean (range) median (IQR)	10.2 (0.5-65.9) 8.4 (4.3-13.5)	11.2 (3.3-23.5) 8.7 (4.1-18.7)	12.8 (0.1-61.6) 8.7 (4.4-18.3)	11.6 (0.1-65.9) 8.7 (4.4-16.0)

*community acquired cases only

Used in conjunction with the above data on COVID-19 admissions and bed occupancy, understanding mortality rates in CUH was important in planning for end of life care during the pandemic and additional modelling was conducted to estimate mortuary requirements. Guidance around escalation and resuscitation decisions for admissions with COVID-19 was reviewed in light of this data, as were provision of palliative care services. Adaptations were required in all departments providing specialist care during the COVID-19 pandemic, in the face of challenges such as staff shortages (through acute infection, contact isolation or shielding), additional PPE requirements and limitations on patient movements to specialist care areas. **Table 7.4** shows common co-morbidities among COVID-19 inpatients during the first wave of the pandemic. Among them are high numbers of patients with diabetes mellitus, asthma, chronic obstructive pulmonary disease (COPD), chronic kidney disease (including dialysis patients) and dementia. These patients would ordinarily receive face-to-face reviews from specialist teams, or be moved to specialist care areas during long admissions. Given tighter infection control procedures restricting staff and patient movement, adaptations and new guidelines were made, bespoke to each specialty.

Table 7.4 Co-morbidities of COVID-19 inpatients

Data from patients admitted with confirmed SARS-CoV-2, March 10-June 21 2020, (n=489).

Condition	n (%)
Hypertension	162 (33%)
Ischaemic heart disease	73 (15%)
Cardiac failure	40 (8%)
Asthma	61 (12%)
Chronic obstructive pulmonary disease	43 (9%)
Diabetes mellitus	102 (21%)
Chronic kidney disease	49 (10%)
Cirrhosis	21 (4%)
Dementia	55 (11%)

7.4 Hospital acquired SARS-CoV-2 infection at CUH

As outlined above, the definition of HAI in the context of SARS-CoV-2 is challenging due to the comparably long incubation period in comparison to other respiratory viruses. **Table 7.5** shows definitions of HAI, produced by the SAGE Hospital-Onset COVID-19 Working Group and subsequently adopted by PHE and NHSE, which reflects the increasing likelihood of HAI with longer inpatient stays from admission to sample collection, up to the maximum incubation period of 14 days. These definitions were used as a rapid tool to assess likelihood of HAI and target infection control investigations. An additional hospital dashboard identifying HAIs was created by the CUH data management team to facilitate this work.

Table 7.5 Definitions of COVID-19 hospital acquired infection categories

Category	Definition
Hospital onset, healthcare associated	Positive specimen date >14 days from admission
Hospital onset, suspected healthcare associated	Positive specimen date 8-14 days after admission or specimen date 3-14 days after admission, with prior admission in previous 14 days
Hospital onset, indeterminate healthcare associated	Positive specimen date 3-7 days after admission, with no prior hospital admission in previous 14 days
Community onset, suspected healthcare associated	Positive specimen date ≤2 days after admission, with prior hospital admission in previous 14 days
Community onset, community associated	Positive specimen date ≤2 days after admission, with no healthcare contact during the previous 14 days

Figure 7.2b shows the epidemic curve of COVID-19 cases at CUH during the first wave of the pandemic. The different categories of HAI are shown, demonstrating the higher proportions of suspected or definite HAI in the 8 weeks following the first confirmed case in CUH on March 10 2020. In addition, numbers of CAIs are broken down into patients admitted from their own home (in light green) or from long term care facilities (dark green). A timeline of key events that may impact HAI, reflecting changes in hospital policy or management, are shown above the epidemic curve (**Figure 7.2a**).

Of note, there was a steep rise in cases between our first case and the peak number of admissions four weeks later. Following the introduction of a national lockdown, there was a plateauing of cases admitted from their own home, followed by a similarly steep fall in cases two weeks later, demonstrating the efficacy of the lockdown in reducing hospitalisations among the local population. However, there was a continued increase in the number and proportion of cases admitted from care facilities until the end of April. The number and proportion of HAIs also peaked later than CAIs, demonstrating ongoing transmission in the hospital. Following the introduction of HCW testing and asymptomatic screening in the first week of April, there was a large increase in the number of confirmed SARS-CoV-2 infections over the following 4 weeks. While this needs to be interpreted with caution, as there was no community-based screening before this time and HCW testing capacity was limited, there was a clear reduction in positive HCW tests through May and June.

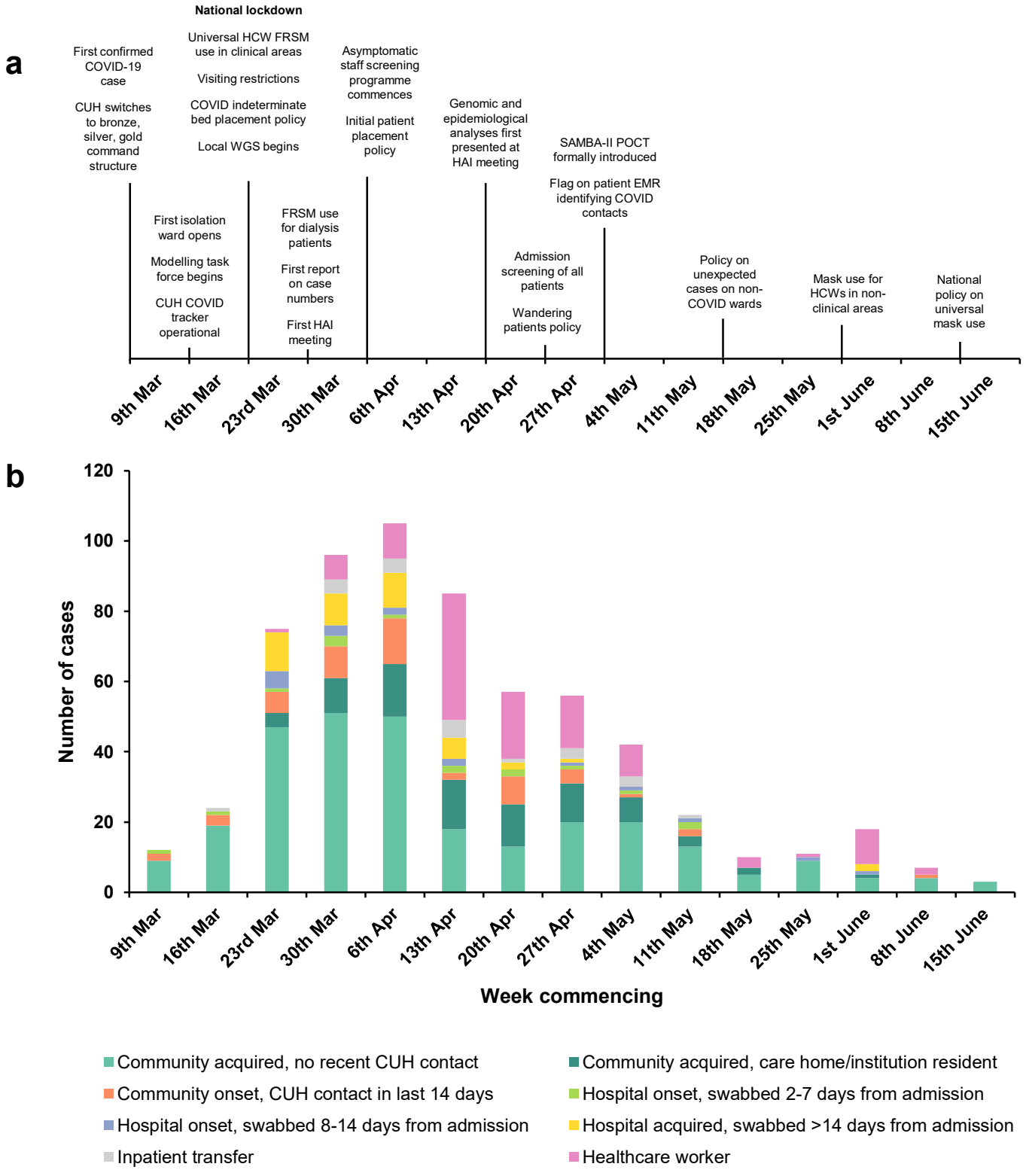


Figure 7.2 Epidemic curve of hospital-associated infections and timeline of infection control interventions in the first wave of the COVID-19 pandemic at CUH.

(**a**) Timeline of key interventions in the management of COVID-19 at CUH during the first wave of the pandemic, starting from the first case on March 10, 2020. Events are grouped by week. (**b**) Epidemic curve of cases over the same time period as (**a**). Bars are coloured by the likelihood of hospital associated infection, using the same definitions as **Table 7.5**. Additionally, dark green bars represent CAIs admitted from care homes an other long term care facilities. Pink bars represent HCWs, the majority of whom were diagnosed via staff testing or asymptomatic screening. EMR = electronic medical record. FRSM = fluid resistant surgical mask. HAI = hospital associated infection. HCW = healthcare worker. POCT = point of care test.

7.5 Epidemiologic and phylogenetic analysis of SARS-CoV-2 transmission at CUH

The high numbers of HAIs in patients and HCWs, especially at a time of reducing transmission in the community, raised concerns for transmission events in the hospital involving staff and patients. To investigate nosocomial transmission, epidemiological and genomic analyses on prospectively stored clinical samples, analogous to those detailed in **Chapter 6**, were performed. However, unlike influenza, which was conducted retrospectively some months after the final case had been identified, the mean turnaround time from sample collection to sequencing was approximately 48 hours. Samples were sequenced using the MinION platform using Oxford Nanopore Technology. The PHE Field Service again provided specialist epidemiological input using ward movement data extracted from the patient's electronic medical record. The initial investigations shown here reflect cases identified between March 10 and April 24 2020.

Figure 7.3 shows the selection of samples for combined sequencing and epidemiological analyses. Of the 374 COVID-19 patients diagnosed during that time, 262 (70.1%) had associated sequences (**Table 7.6**). 57/374 (15.2%) of admissions were suspected or highly likely hospital acquired, of which 49 (86.0%) had associated sequences. Additionally, 46/73 (63.0%) sequences were available from HCWs, 9 of whom were diagnosed as patients and 37 through CUH's own HCW testing programme[220].

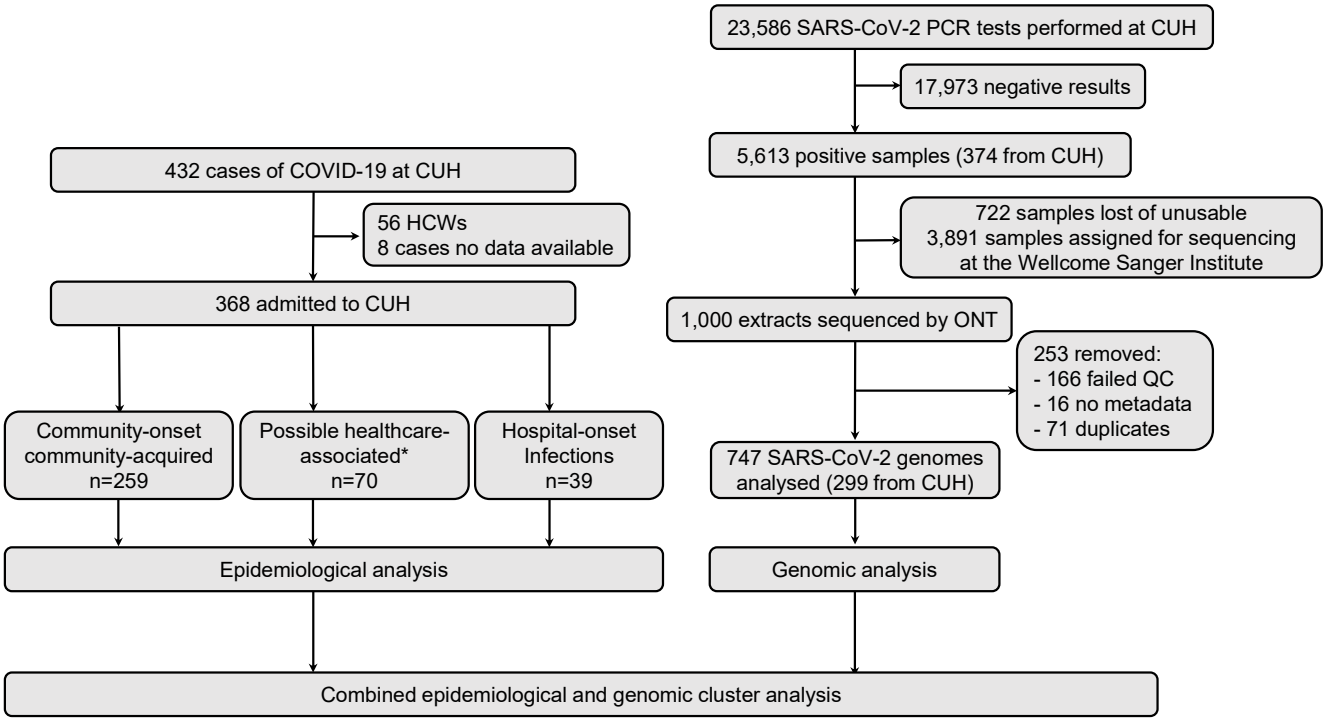


Figure 7.3 Case ascertainment of genomic epidemiology analysis of COVID-19 cases at CUH
Of the 299 genomic sequences derived from CUH, 37 were from HCWs identified in the screening programme and 262 were CUH patients. 9/262 patients were also HCWs, giving a total of 46 HCWs.

Table 7.6 COVID-19 infections in CUH and sequence data availability

Classification of infection	No.	No. with available sequence (%)
Community onset, community associated	261	174 (66.7%)
Community onset, suspected healthcare-associated	40	27 (67.5%)
Hospital onset, indeterminate healthcare-associated	11	7 (63.6%)
Hospital onset, suspected healthcare-associated	13	11 (84.6%)
Hospital onset, healthcare-associated	39	36 (92.3%)
Healthcare worker	9	6 (66.7%)
Unknown	1	1 (100%)
Total	**374**	**262 (70.1%)**

To facilitate rapid review of transmission networks, a sample set was locked each week in order to perform QC and downstream analysis. Lineage assignation was performed using Pangolin[312]. The majority of sequences belonged to lineage B.1. Phylogenetic trees were created in IQTREE[291] to determine genetic clustering by ward location and category of HAI (**Figure 7.4**). Samples collected early in the patient admission, predominantly from the emergency department (ED), were phylogenetically dispersed, which likely reflected unconnected transmission events in the community. In contrast, sequences derived from samples collected on the same wards in CUH were closely related (0-1 SNPs). These clusters contained a number of cases categorised as suspected or definite HAI, strongly suggestive of linked, nosocomial transmission in hospital.

At the time of this study, SARS-CoV-2 had only recently been identified in the human population. Genetic diversity of sequenced isolates was therefore low. The median genetic difference between any two samples in this study was 8 SNPs (range 0-24 SNPs). Interpreting putative clusters was therefore challenging, as identical sequences may have occurred “by chance” rather than through genuine transmission. To maximise confidence, a strict cluster definition was used of 0 SNP differences between isolates. Epidemiological data, using similar spatio-temporal approaches to those employed for influenza in **Chapter 6**, provided additional confidence in any genetic clusters. Where genomically identical isolates were identified without associated spatio-temporal links, I reviewed the records of all patients in that cluster to identify epidemiological associations (including residential address, social setting, clinical details and outpatient appointments) that could not be provided from ward-movement data.

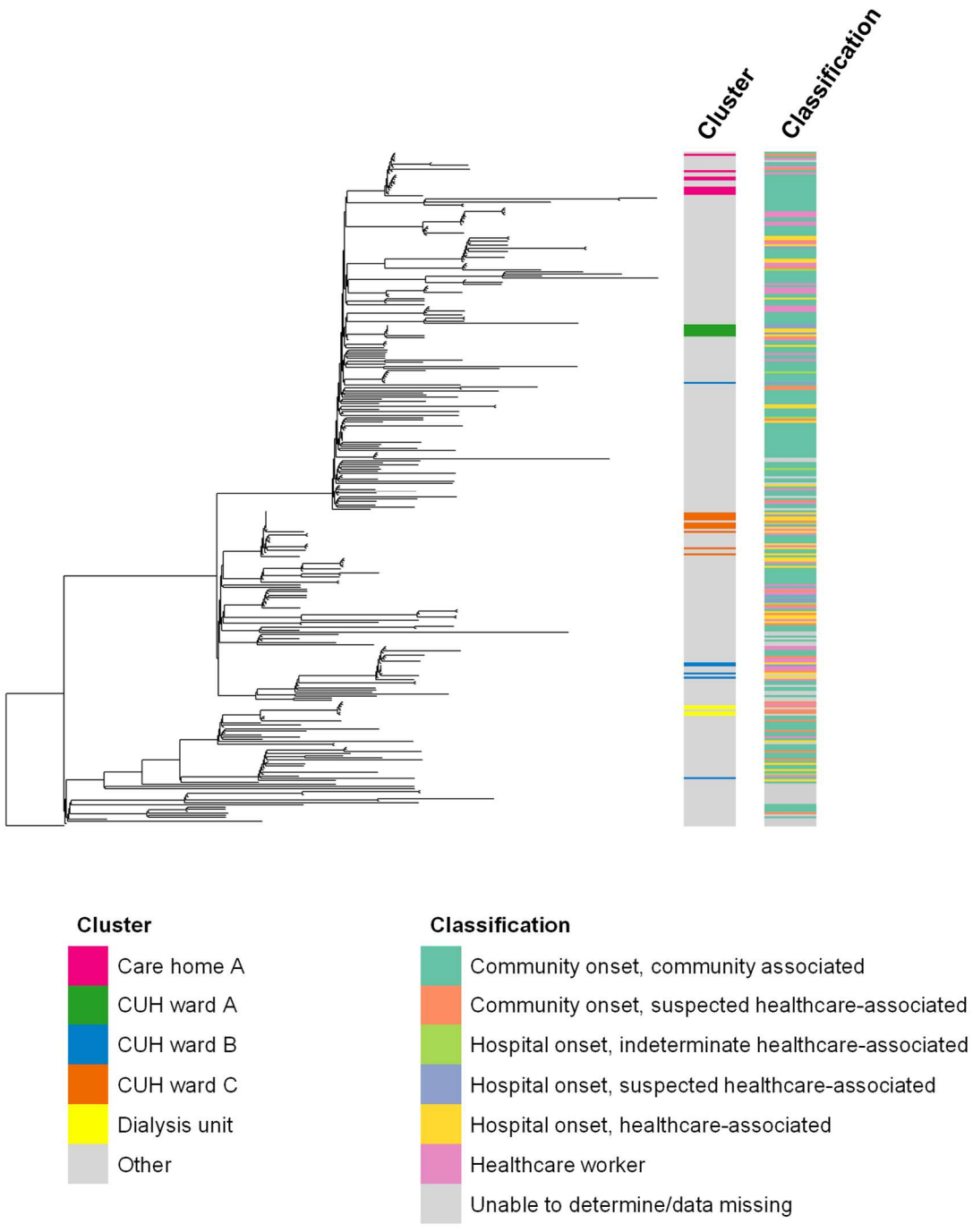


Figure 7.4 Phylogenetic tree of SARS-CoV-2 genomes from CUH, first pandemic wave 2020.
Maximum likelihood phylogenetic tree of 299 SARS-CoV-2 genomes from CUH. The tree is rooted on a December 2019 genome from Wuhan, China. The left hand column highlights epidemiologically associated clusters of cases linked to four hospital clinical areas and one community care home. Wards A, B and C all had genomic clusters (defined as viruses ≤1 SNP different) of epidemiologically linked HAI cases. Eight Ward C cases were contained within one of the largest genomic clusters of identical viruses (n=15). Genomic clusters containing the cases for Wards B and C, the dialysis unit and Care home A all include HCWs. The right hand column shows the classification of infection, consistent with the definitions in **Table 7.5**. Phylogeny produced by William Hamilton, and reproduced here with permission.

In total, 114/262 genomes (43.5%) were identified that had an identical sequence to at least one other isolate, across 26 clusters (range 2-15 genomes per cluster). In 63/114 cases (55.3%) there were strong epidemiological associations between cases to support recent transmission, while 27/114 (23.7%) had intermediate evidence and in 24/262 no associations could be identified. Suspected transmission events were identified on 12 inpatient wards, 3 of which were COVID-19 cohort “red” wards and 9 were non-COVID-19 “green” wards. Additional transmission events were identified in a dialysis unit, in two local care homes and between paramedics. Some genomic clusters involved transmission events between patients, patients and HCWs, or HCWs alone. The details of all genomic clusters are included in **Appendix G.1**, with illustrative examples discussed below.

7.6 Genomic epidemiology in the implementation of infection control measures

While nosocomial transmission events were identified early in the first wave of the pandemic, the optimal method of reviewing these cases by clinicians and infection control teams was unclear. There was concern that the standard route of investigation through root cause analysis would be too slow and time consuming. Additionally, there were data available from genomics and epidemiological analyses that provided insight into transmission and which the clinical teams were keen to utilise to better understand and prevent nosocomial spread.

To address this issue, I worked with other infection clinicians at CUH and the patient safety team to develop a process for rapid review of suspected HAIs using a modified gap analysis. Using a structured proforma, infection clinicians reviewed the medical records of all possible

HAIs identified on the CUH COVID-19 dashboard. Medical records were reviewed to establish the likely mechanism of transmission, and standards of care were established to identify lapses in infection control practices that may have contributed to SARS-CoV-2 transmission. Gap analyses were presented alongside data from the genomic and epidemiological analyses in a weekly meeting attended by infection control, other infection specialists, patient safety representatives, the PHE Field Service and academics involved in the genomic sequencing and analysis. By combining clinical, genomic and spatio-temporal data, decisions were made regarding the likelihood of a patient acquiring their infection in hospital and any factors that contributed to transmission. Where gaps in care were identified, recommendations were made for the rapid deployment of measures to enhance infection control and prevent further transmission. This process is outlined in **Figure 7.5**.

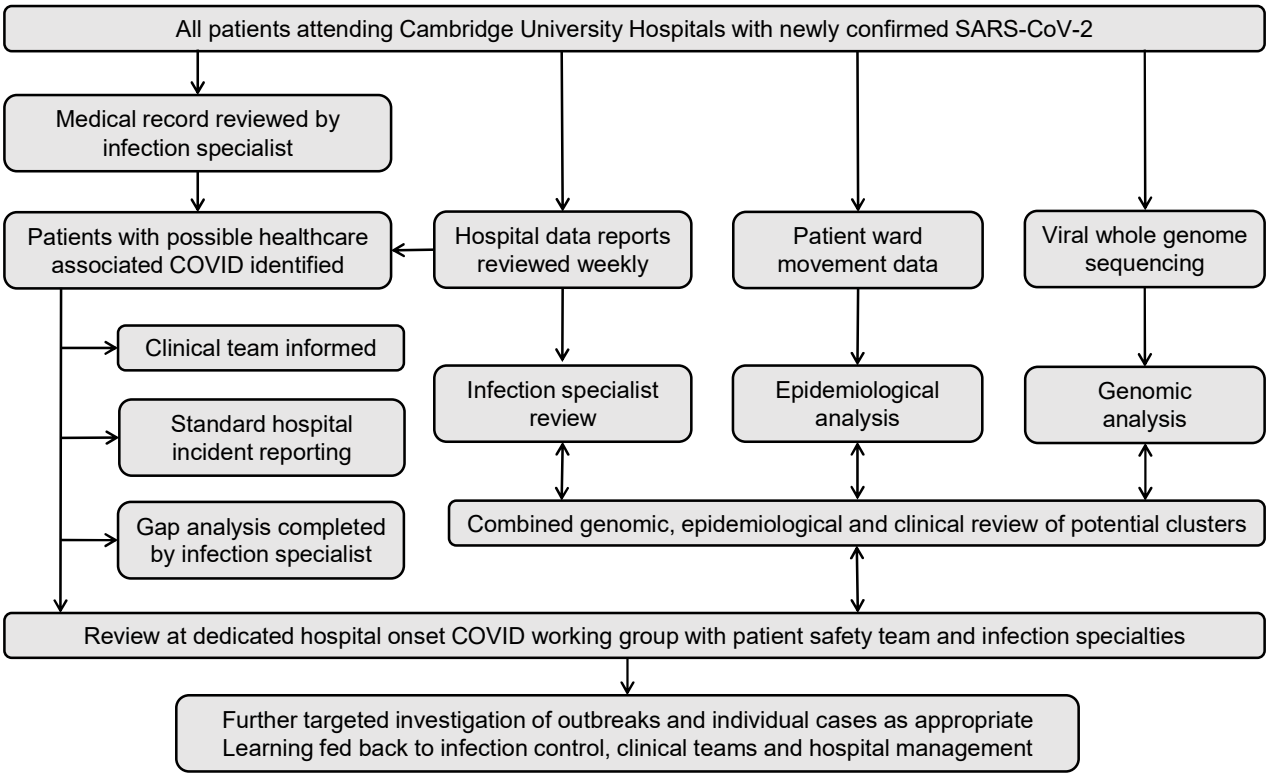


Figure 7.5 Flow diagram summarising the review process for healthcare associated COVID at CUH

Examples of genomic and epidemiological clusters are shown in **Figure 7.4** and described in further detail below. In ward A (a non-COVID ward), six surgical patients were diagnosed with COVID-19 infection more than seven days after admission to hospital (**Appendix G.1**, cluster 5). All patients had been co-located in the same bay at different times over a two week period. Sequences were available for all patients, and revealed that five were genomically identical and the sixth differed by a single SNP, consistent with recent transmission. However, the index patient could not be identified. In response to this and similar clusters, universal mask use was introduced and a policy of screening of all patients admitted to green wards was adopted (when testing capacity was available).

In a second cluster, four solid organ transplant patients on ward B (a non-COVID area) were diagnosed with COVID-19 infection between April 3 and April 20 2020 (**Appendix G.1**, cluster 3). A fifth patient subsequently presented to the ED with COVID-19, having recently been discharged from ward B. Sequences from these patients were genomically identical. Additionally, two staff members on ward B had identical genomes in the same cluster, as did a third HCW who had professional contact with one of the ward B staff. These findings led to a review of infection control and PPE procedures for staff and patients in the transplant service. To illustrate the complexity and limitations of this approach, one HCW and one patient with genomically identical sequences could not be epidemiologically associated with the other members of this cluster.

Renal dialysis patients are among the most vulnerable to COVID-19. Infection control practices in dialysis settings are challenging, as most units comprise large open rooms with no barriers between patients. Six patients attending the same dialysis unit were diagnosed with COVID-19 between April 1 and April 20 2020, with identical viral genomes (**Appendix**

G.1, cluster 10). Epidemiological investigation revealed they dialysed on the same days of the week, though not all at the same time and not in the same part of the unit. The common association linking three of the patients was shared transport to the unit. These findings led to a review of infection control procedures in dialysis patients, including the introduction of surgical masks for patients on the unit and in shared transport. Of note, in this case the genomic analysis was able to exclude transmission between the dialysis unit and the renal ward (which had a cluster of COVID-19 patients and shares some patients admitted to CUH from the dialysis unit). However, the dialysis unit genomes belonged to lineage B.2 (relatively rare in the community at this time), whereas the renal ward genomes were the common B.1 lineage, making it highly unlikely that infections between the two patient groups were related.

A more complicated transmission network is shown at the top of **Figure 7.4** (**Appendix G.1**, cluster 2). Fifteen patients were diagnosed in CUH between April 7 and 21 2020 with genetically identical viruses, of which eight were resident in the same home (care home A). Another patient worked in care home A, while a further case worked in an unspecified care home. Two cases were paramedics and two were nurses (who worked in different CUH ward but lived with paramedics). The final case did not have any discernible epidemiological association with any other individual in this cluster. This evidence of transmission arising in the same care home, with plausible associations between HCWs and the care home, had not been detected by clinicians or infection control using routine practices.

Over the course of the first wave, data from these and other clusters led to changes in policy regarding staff social distancing in communal areas, identifying COVID-19 cases in non-COVID areas, the management of confused patients with COVID-19, identification and isolation of contacts of confirmed cases, mask usage for patients and staff in non-clinical

areas. Many of these interventions were introduced before similar recommendations from national guidelines.

Additionally, 58 cases reviewed using this process during the first wave were attributed to HAI, leading to further patient safety reviews and duty of candour reporting to patients and relatives. Of the 58 patients, 14 died during their hospital stay; in 9 cases the deaths were directly attributed to COVID-19 acquired during their admission, while 3 more patients came to severe harm necessitating admission to intensive care. Further details of the themes identified during these patient safety reviews are beyond the scope of this thesis, but included identification of vulnerable patients, inconsistent communication, variable implementation of infection prevention and control measures, incomplete guidelines and organisational response[421].

7.7 Discussion

In this chapter I have described the burden of COVID-19 on a large teaching hospital in the UK over the first year of the pandemic, in terms of admissions, bed occupancy, mortality and hospital acquired infection. I have described the clinical and demographic features of these patients and how these have changed over three phases of the pandemic. Using a combination of genomic, epidemiological and clinical methods I have identified multiple transmission networks occurring within the hospital, involving patients and staff in both COVID and non-COVID areas, as well as complex transmission chains involving care home residents, staff, paramedics and CUH HCWs. I have shown how insights from genomic and

epidemiological analyses have been combined in real time with clinical reviews to inform improvements in infection control practices ahead of national policy changes.

The clinical features and outcomes in the first pandemic wave in CUH are comparable to a large contemporaneous dataset of 20,133 inpatients included in the ISARIC study from February to April 2020[172]. There were similar distributions of age (median 70 in CUH versus 73 in ISARIC), gender (59% *v* 60% male) and the relative frequency of specific co-morbidities, as well as outcomes of critical care admission (19% *v* 17%) and death (24% *v* 26%). Data from the same study between February and August 2020 also demonstrated the high burden of HAI; using the most conservative definition (sample collection more than 14 days after hospital admission), they estimate that 9.7% of all COVID-19 diagnoses made in acute trusts were HAIs, comparable to the 7.9% of cases in CUH using the same criteria over the same time period[189].

The combined genomic epidemiology study at CUH was perhaps the first to investigate SARS-CoV-2 hospital transmission networks across a large acute hospital[257]. While the utility of viral genome sequencing has been demonstrated in previous epidemics[194-196,422], this was, to the best of my knowledge, the first to prospectively apply sequencing of a respiratory virus to the hospital setting in real-time to influence hospital policy. In comparison to previous retrospective genomic studies at CUH[21,83,423], and the work presented on influenza in **Chapter 6**, the timeframe from sample collection to actionable outcome from genomic analysis was in the order of 10 days, as opposed to many months.

One of the great strengths of this study was the combination of data from multiple sources. Previous work on influenza had shown the utility of combining genomic and epidemiological data, but the role of spatio-temporal analyses in this case was critical to support inferences on transmission in a viral population with low genetic diversity. Additionally, this study

benefited from high case ascertainment in both patients and HCWs. CUH was one of the first hospitals in the UK to undertake both symptomatic testing and asymptomatic screening of its staff[220,223], providing access to sequences that confirmed that HCWs play a role in nosocomial transmission of respiratory viruses. This study benefited from the availability of complex demographic, clinical and ward movement data on all confirmed COVID-19 cases through daily automated extracts from the electronic medical record and associated dashboards of key COVID-19 statistics at CUH.

In response to a rapidly evolving situation and changing national guidance, the HAI review process enabled prompt identification and correction of gaps in hospital policies to improve patient safety. Change was rapidly implemented, typically within a week of the review and often prior to the announcement of similar policies at a national level. Engagement of a multidisciplinary team, including patient safety and infection control, were central to this process, as they included participants with the skills and influence within hospital management to ensure swift policy implementation. By contrast, existing practices, such as the standard RCA process which can take many weeks or even months, would have been too slow to have an impact[399].

Of note, some clinical areas were identified as the likely location of transmission in both influenza and SARS-CoV-2 transmission studies, including wards of high risk patients (including geriatrics and solid organ transplantation) and the same dialysis unit. As with the influenza study, imported outbreaks from care homes and other cryptic transmission events were identified using sequencing that were not initially apparent. Genomic data was also used to refute suspected transmission events, which was important in focusing interventions at a time of unprecedented demand on infection control teams.

There are several limitations to this study. In keeping with most transmission studies using genomics, incomplete sequence availability may have led to missed or incomplete transmission networks. Similarly, limited resolution of spatio-temporal data (largely to ward level) for patients and HCWs may overlook other epidemiological links, both in hospital and the community. Limited genomic diversity meant that there was insufficient evidence to infer transmission in the absence of confirmatory epidemiological associations. Although some clusters had strong epidemiological and genomic associations, the mechanism and direction of transmission remained unclear. Additionally there were limitations in the HAI review process, in particular the primary focus on suspected HAIs rather than the role of HCWs or CAIs admitted to the Trust and seeding infection into the hospital. Despite insights available through genomic epidemiology, the exact reason for an individual infection may be difficult to discern through any process, especially given the long incubation period, incomplete case ascertainment and asymptomatic transmission of SARS-CoV-2.

Despite its limitations, the combination of genomic, epidemiological and clinical data in both patients and HCWs at CUH has proven valuable, and the HAI review process has continued into second and subsequent waves of the pandemic[424]. Additionally, I have contributed data from this work to a variety of academic and clinical projects in the study and management of the pandemic, outlined below.

Addressing the limitations on ascribing direction of transmission to clusters identified in this study, a new bioinformatic tool has been developed for the rapid analysis of SARS-CoV-2 transmission using both genomic and epidemiological data. A2B-COVID combines knowledge about infection dynamics, ward movement, and novel approaches to genome sequence data to assess whether or not cases are consistent or inconsistent with linkage via transmission[425]. This method has been applied to clusters from the first wave of the pandemic at CUH and demonstrated an uneven pattern of transmission between individuals,

with patients being much more likely to be infected by other patients than by HCWs. Additionally, A2BCOVID has demonstrated a pattern of superspreading, with 21% of individuals causing 80% of transmission events in the largest hospital associated transmission networks in this study[426]. Further work has also been conducted on the high numbers of admissions from care homes. Combining data from CUH with samples collected from care homes and local hospitals in the East of England identified 409 viral clusters affecting 292 homes. Approximately 70% of care home residents in this analysis were admitted to hospitals, with considerable evidence of transmission in both directions between care homes and secondary care[427].

The CUH data described here have been used to evaluate other interventions made over the course of the pandemic, including the introduction of a novel point of care testing platform, the SAMBA-II, into acute admissions[255], and the impact of HCW screening and vaccination on COVID-19 infection in CUH staff[220,223,428,429]. The medical record extracts described above were used to identify and recruit newly diagnosed patients to the ISARIC cohort study[172] and the RECOVERY trial[419]. In collaborations between the hospital and the University of Cambridge Institute for Manufacturing, data were used to model COVID-19 admission numbers, bed occupancy, PPE usage and oxygen supplies, to help with pandemic preparedness. Additionally, data were used for internal and external communications with staff, through regular emails, intranet posts, formal teaching, public engagement and liaison with the local media. Reports summarising COVID-19 numbers and patient characteristics were circulated to all staff weekly during the first wave. An example is given in **Appendix H.2**.

The quantity of data available, and the rapidity with which it could be obtained, led to a collaboration between myself and members of the Scientific Pandemic Influenza Group on Modelling (SPI-M). Using CUH as one of a small number of sentinel secondary care sites, data from the hospital was used to model the burden of nosocomial transmission on case numbers and the impact of interventions to reduce transmission[430]. Through my position on the Scientific Advisory Group for Emergencies (SAGE) Hospital-Onset COVID-19 Working Group from April to July 2020, studies and data from CUH were regularly presented to senior government advisors and COVID-19 teams working at Public Health England and NHS England. In summer 2020, data from CUH were used to validate HAI dashboards produced by NHSE and PHE based on mandatory notifications and hospital admissions data. These dashboards now provide real-time updates on nosocomial infection at hospital-level across the UK.

In this chapter I have shown how multidisciplinary collaboratives from academic and clinical groups have produced extensive clinical, genomic and epidemiological datasets on COVID-19 in CUH. These have been used to respond in real time to the many and varied challenges that the hospital has faced during the course of the pandemic, as well as informing national strategy on HAIs. In **Chapter 8** I will demonstrate how similar approaches were applied in a community setting, to study the burden of SARS-CoV-2 transmission within the University of Cambridge and to evaluate the impact of a novel programme of asymptomatic student screening using pooled sampling.

8. Transmission of SARS-CoV-2 at the University of Cambridge

8.1 Introduction

In **Chapter 7** I characterised the burden of SARS-CoV-2 in a large UK teaching hospital during the first wave of the pandemic. For the majority of this period, the country was under a national lockdown with strict public health measures to limit viral transmission in the community. When these restrictions were gradually lifted in summer 2020, attention turned to interventions to control SARS-CoV-2 in populations at high risk of transmission. One of the key groups of concern was university students, due to their high levels of shared accommodation, frequent social contacts and potential for minimally symptomatic infection that may not have been be identified through existing community testing pathways. The prevalence of COVID-19 among university-aged individuals in the UK rose steeply in autumn 2020, reaching approximately 1.5% by October 1, 2020[431]. However, transmission dynamics of SARS-CoV-2 in higher education institutions (HEIs) were not well understood, including the risk of spread from HEIs to surrounding communities. The optimal method of controlling viral transmission in this setting was therefore also unclear, including the role of mass testing.

Following the work that I had conducted in CUH at the start of the pandemic, I was invited to take up the position of clinical lead for the University of Cambridge asymptomatic COVID-19 screening programme in August 2020. Through a collaboration between the University, AstraZeneca and GlaxoSmithKline (and later Charles River Laboratories), Cambridge had established a laboratory for high throughput processing of SARS-CoV-2 PCR tests, part of the UK's lighthouse laboratory network that provided the majority of the UK's testing capacity in the second half of 2020. As part of that collaboration, the University was provided access to a fraction of that capacity, equating to approximately 2,500 PCR tests per week. While insufficient to screen the University's student population individually, a screening programme was established based on the use of pooled swab sampling of students in groups (based on household structure), to maximise the utility of that testing capacity.

In this chapter, I detail how this screening programme was implemented for students living in College accommodation during Michaelmas Term 2020. I will present detailed descriptions of the student populations that participated and risk factors for subsequently testing positive, linking secondary attack rates within households to the characteristics of the index case. With high case ascertainment from both asymptomatic screening and symptomatic testing, a combination of whole genome sequencing and routinely collected epidemiological data from the University and national contact tracing provided detailed insights into the spread of the virus in this cohort. Using this approach, I describe viral introductions into the University, amplification of transmission associated with key student activities and, through phylogenetic comparison with contemporaneous sequences from across the local area, identify the extent of transmission between the University and community. Through a combination of data sources, I demonstrate the efficacy of asymptomatic screening and other local and national interventions in controlling the spread of SARS-CoV-2 in a higher education setting.

8.2 Implementation of asymptomatic screening and characteristics of participants

A detailed description of the asymptomatic screening programme is provided in **Chapter 2**. In brief, students were grouped into pools of up to ten individuals, based on household structures within College accommodation. They were screened using self-administered nose and throat swabs that were pooled into the same tube of viral transport medium at the time of sample collection. Swabs were collected, tested and resulted within 24 hours. Students in a pool that resulted positive were immediately isolated and invited for same-day individual confirmatory testing. Any student with a positive individual result was managed in line with national guidance on symptomatic individuals with confirmed SARS-CoV-2 infection, including isolation of the student and their household, and associated contact tracing. If all students in a positive pool tested negative, they were released from isolation following 1-2 rounds of individual tests. An independent, parallel pathway of individual PCR testing for students and staff at the University who were experiencing the cardinal symptoms of COVID-19 (fever, cough, anosmia and/or ageusia) was available throughout the study period.

During the 9 weeks of Michaelmas Term 2020 (October 5 to December 6, 2020), 12,781/15,561 (82.1%) eligible students living in College accommodation participated in the screening programme (**Figure 8.1**). Additionally, 1,031/12,781 (8.1%) study participants took at least one University symptomatic test during this period. College accommodation was divided into 3,094 households (mean 5.0 household members, range 1-20) based on shared communal facilities such as bathrooms and kitchens (**Figure 8.2a**). For the purposes of asymptomatic screening, students were organised into 2,275 pools (mean 6.8 students, range 1-10) (**Figure 8.2b**). To maximise the efficiency of the programme, 1,476 small households were combined to form merged pools.

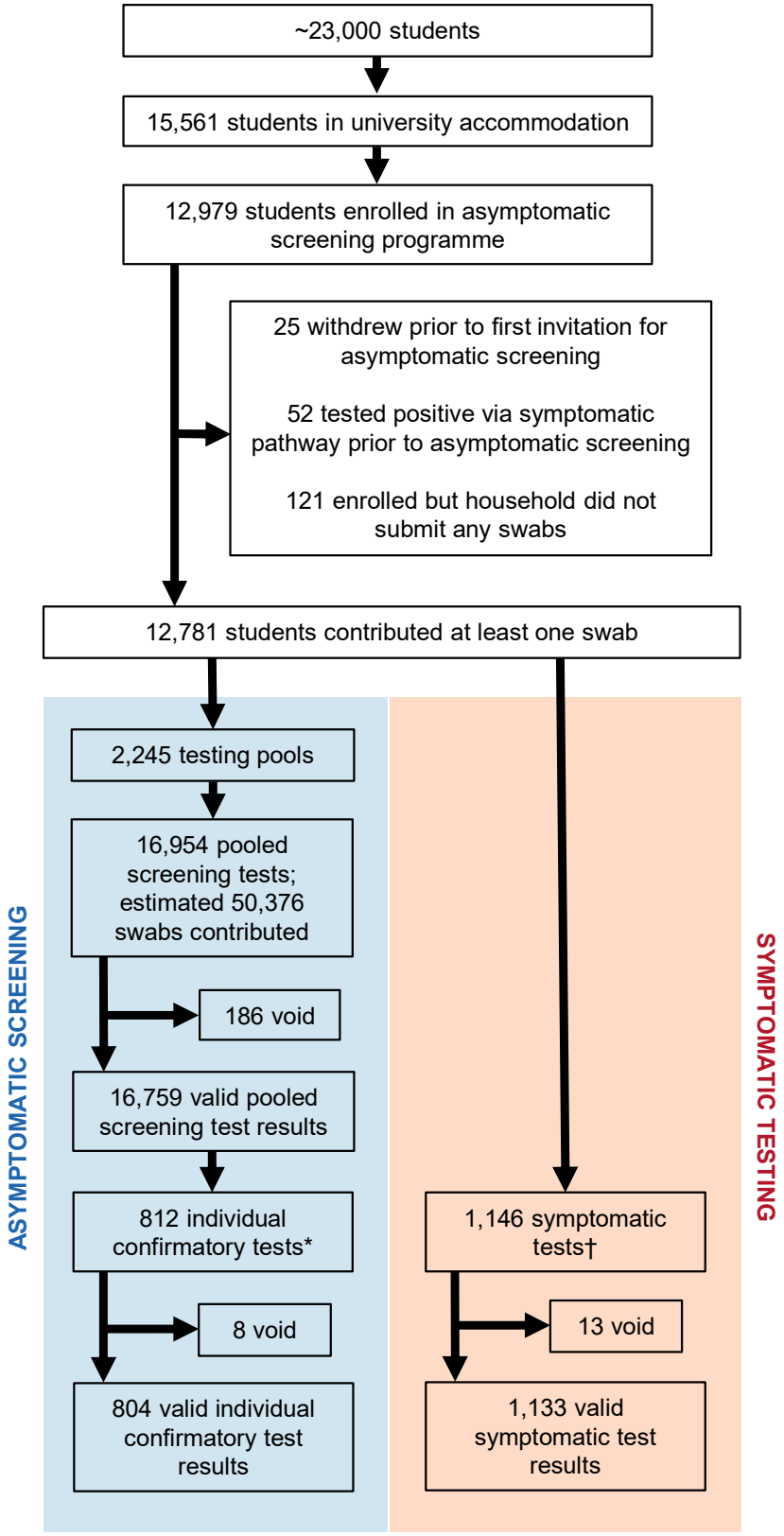


Figure 8.1 University of Cambridge asymptomatic COVID-19 screening programme study cohort
Participation in University of Cambridge asymptomatic COVID-19 screening programme and symptomatic testing by students living in College accommodation from October 5, 2020 to December 5, 2020. *Includes 19 repeat individual confirmatory tests. †Excludes symptomatic tests conducted by NHS pathways.

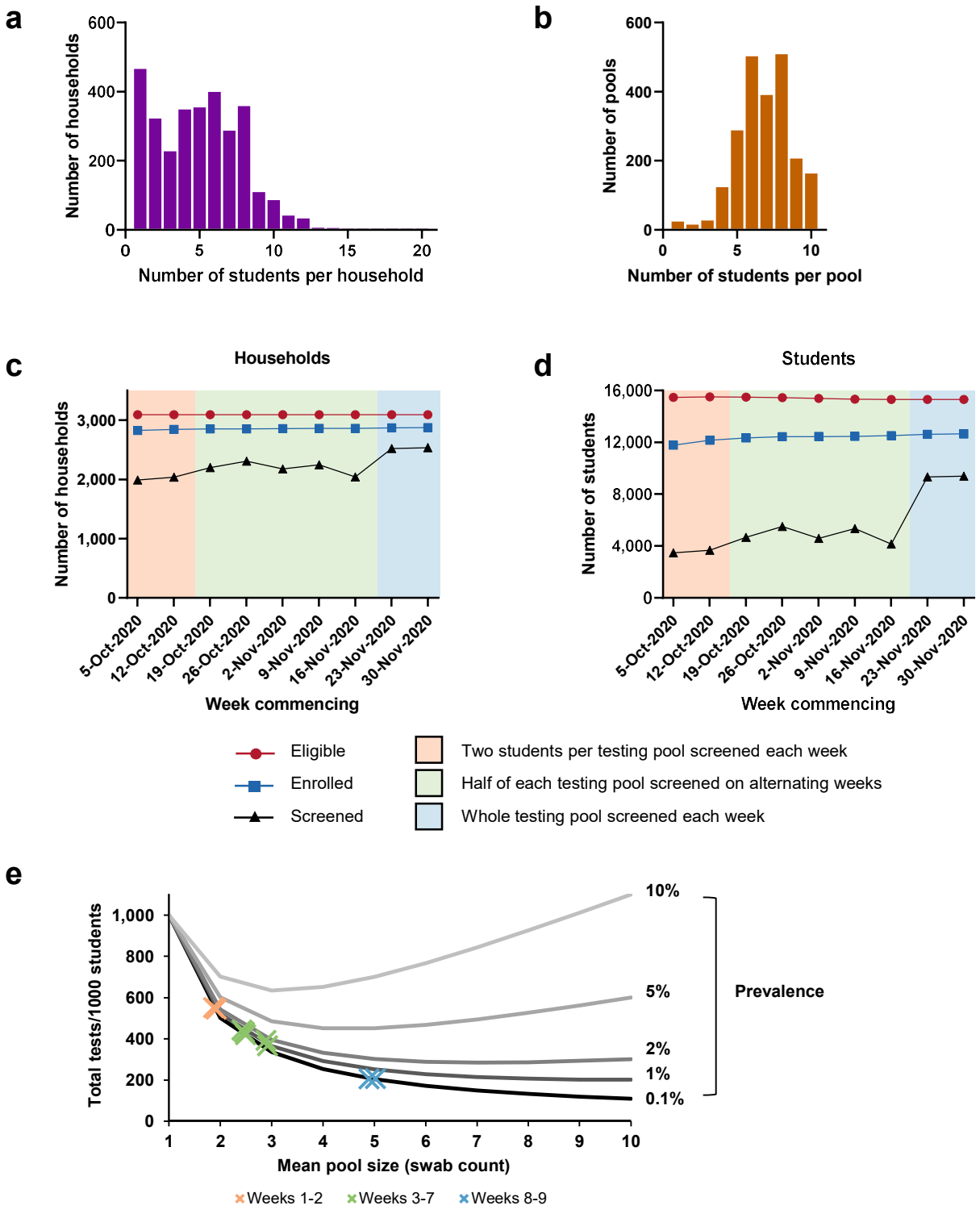


Figure 8.2 Student participation and implementation of swab pooling

(**a-b**), Distribution of (**a**) household size and (**b**) testing pool size for eligible students in College accommodation. (**c-d**) Weekly participation rates by household (**c**) and individual student (**d**). The number of eligible (red circles), enrolled (blue squares) and screened (black triangles) students and households is shown during the phased implementation of the screening programme (indicated by the shaded areas in each plot). Some eligible students were excluded from screening each week because of a recent positive test for SARS-CoV-2, symptoms of possible COVID-19, or a requirement to self-isolate. (**e**) Predicted total number of pooled screening and individual confirmatory tests required to screen 1,000 students in the presence of different testing pool sizes and SARS-CoV-2 prevalences (grey lines, 0.1-10%). Crosses indicate the actual mean pool sizes and tests utilised at the indicated time points. In the absence of swab pooling (equivalent to a mean pool size of 1), 1,000 tests would be required.

Because the physical capacity of sample tubes was limited to a maximum of ten swabs, 151 large households were divided into multiple pools. Overall, 65% of pools corresponded exactly with households, 21% of pools included multiple households, 10% of pools included a proportion of one household, and 4% of pools included students from more than one household and part of one household. Participation was higher per household than per individual student, with over 90% of households contributing consistently across the term (**Figure 8.2c**).

Students were first made aware of the programme in an email on September 9, 2020, four weeks before the start of term. Over 3,500 students consented for the programme on the first day of recruitment, September 26, 2020, when an email providing access to the consent form and study documents was sent to all eligible students. Further spikes in recruitment were observed after subsequent email correspondence about the programme (**Figure 8.3**). The proportion of students consenting to participate was therefore high at the outset of the programme, but continued to rise through the term from 75.2% of eligible students in week 1, to 81.9% in week 9 (**Figure 8.2d**, **Appendix I.1**). A comparison of participating and non-participating students is shown in **Table 8.1**. Although the two groups were similar, undergraduates were better represented (especially those studying science and technology subjects and first year students) and there were small but significant differences in sex, age, ethnicity and UK versus international home residency.

Table 8.1 – Characteristics of asymptomatic COVID-19 screening study population.

		Eligible	Participating	Not-participating	*P* value
N		15,561	12,781	2,780	
Sex	Male	8,060 (51.8%)	6,504 (50.9%)	1,556 (56.0%)	<0.001
	Female	7,213 (46.4%)	6,125 (47.9%)	1,088 (39.1%)	
	Other/unknown	288 (1.9%)	152 (1.2%)	136 (4.9%)	
Age, years, median (IQR, range)		20.6 (19.4-22.6, 16.4-71.7)	20.5 (19.3-22.3, 16.4-58.1)	21.3 (19.8-24.3, 17.4-71.7)	<0.001
Ethnicity	White	9,453 (60.7%)	8,168 (63.9%)	1,285 (46.2%)	<0.001
	Black, Asian and minority ethnic	5,302 (34.1%)	4,097 (32.1%)	1,205 (43.3%)	
	Refused/unknown	806 (5.2%)	516 (4.0%)	290 (10.4%)	
Residency	UK	9,870 (63.4%)	8,504 (66.5%)	1,366 (49.1%)	<0.001
	International	5,496 (35.3%)	4,205 (32.9%)	1,291 (46.4%)	
	Unknown	195 (1.3%)	72 (0.6%)	123 (4.4%)	
Year of study	1st	3,788 (24.3%)	3,336 (26.1%)	452 (5.5%)	<0.001
	2nd	3,239 (20.8%)	2,673 (20.9%)	566 (20.4%)	
	3rd	2,832 (18.2%)	2,434 (19.0%)	398 (14.3%)	
	4th (or above)	997 (6.4%)	844 (6.6%)	153 (5.5%)	
	Postgraduate	4,369 (28.1%)	3,366 (26.3%)	1,003 (36.0%)	
	Missing	336 (2.2%)	126 (1.0%)	210 (7.5%)	
Stage	Undergraduate	10,857 (69.8%)	9,288 (72.7%)	1,569 (56.4%)	<0.001
	Postgraduate	4,369 (28.1%)	3,368 (26.3%)	1,001 (36.0%)	
	Missing	335 (2.2%)	125 (1.0%)	210 (7.5%)	
Course*	Undergraduate arts and humanities	5,490 (35.3%)	4,553 (35.6%)	937 (33.7%)	<0.001
	Undergraduate science and technology	5,454 (35.1%)	4,806 (37.6%)	648 (23.3%)	
	Postgraduate vocational	466 (3.0%)	343 (2.7%)	123 (4.4%)	
	Postgraduate (other)	3,815 (24.5%)	2,953 (23.1%)	862 (31.0%)	
	Unknown	336 (2.2%)	126 (1.0%)	210 (7.6%)	
Household size	1 person	472 (3.0%)	299 (2.3%)	173 (6.2%)	<0.001
	2 to 5 people	4,535 (29.1%)	3,703 (29.0%)	832 (29.9%)	
	6 to 10 people	9,228 (59.3%)	7,688 (60.2%)	1,540 (55.4%)	
	>10 people	1,326 (8.5%)	1,091 (8.5%)	235 (8.5%)	

Student characteristics associated with programme participation were assessed by comparing participating and non-participating students using unpaired 2 sample 2 tailed *t*-tests (continuous variables) or chi-square tests (categorical variables).

*Courses are grouped as: undergraduate arts and humanities (undergraduate students in the School of Arts and Humanities and the School of Humanities and Social Sciences); undergraduate science and technology (undergraduate students in the School of Biological Sciences, the School of Physical Sciences and the School of Technology); postgraduate vocational courses (students in clinical medicine, clinical veterinary medicine and postgraduate certificates in education); and other postgraduate courses (all other postgraduate students, including those in doctoral and masters programmes).

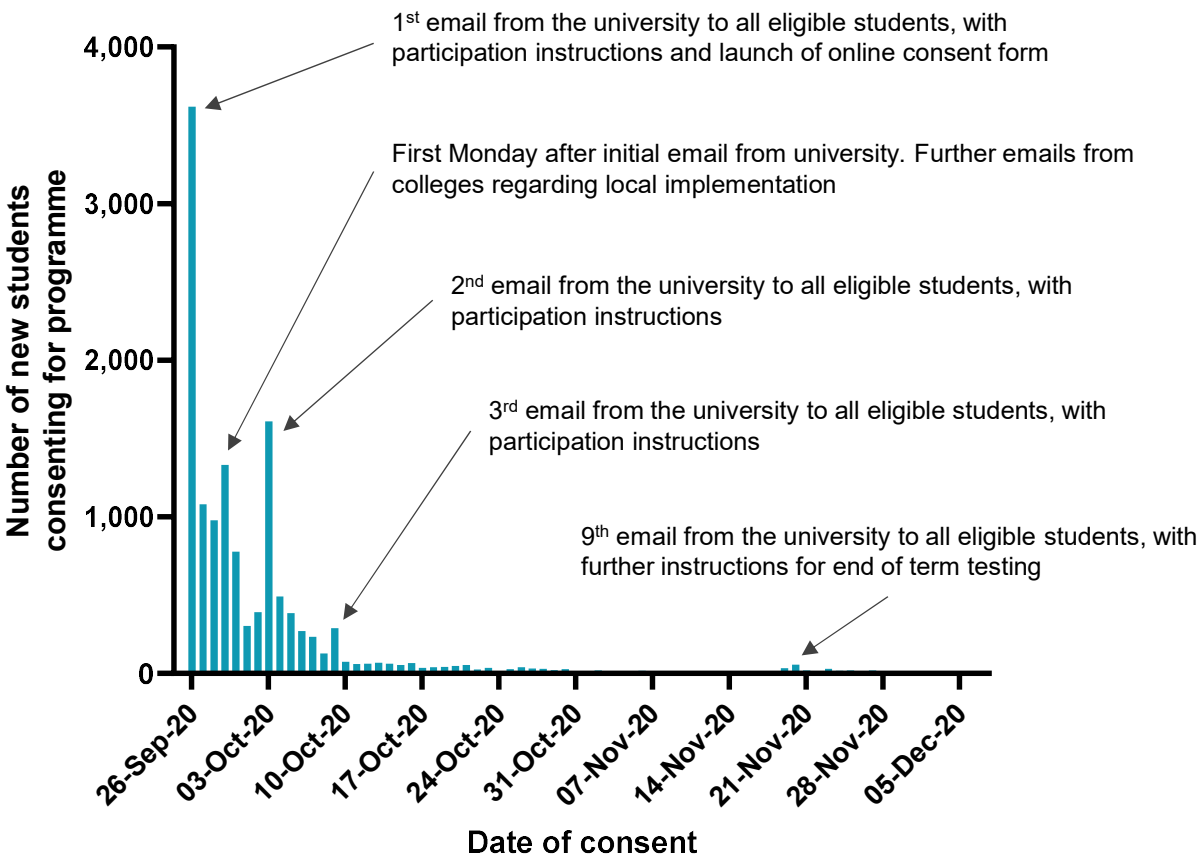


Figure 8.3 Impact of communications on asymptomatic screening participation.
Daily number of students consenting to participate in the asymptomatic screening programme. Key communications are highlighted. Where students consented more than once, only their first consent is included.

In principle, the total testing capacity required to run the programme each week (including both pooled and individual confirmatory tests) varied based on the prevalence of infection among the screened population and the mean size of the testing pool (**Figure 8.2e**). Due to an unclear prevalence of infection in students arriving at the start of term, initial constraints on testing capacity and a need to ensure that the various, untested aspects of the programme were running effectively, the programme began at a reduced scale that was expanded over the course of the term. The number of students invited to participate each week increased in stages: in weeks 1-2, 2 students from each testing pool were invited; in

weeks 3-7, half the students from each testing pool were invited, alternating week by week; in weeks 8-9, all the students from each testing pool were invited every week (**Figure 8.2c,d**). When running at full scale (weeks 8-9), utilisation of testing capacity was approximately five times more efficient than screening with individual tests (**Figure 8.2e**).

8.3 Performance of the pooled sampling strategy

Over the course of the term, 252/16,945 pooled screening tests were positive (**Appendix I.2**). In 205/252 positive pools (81.3%), at least one individual student was identified as being positive using confirmatory testing (**Figure 8.4a** and **Appendix I.2**). This equates to an operational “false positive” rate (i.e. an initial positive pooled sample followed by negative individual confirmatory tests from all members of the pool) of 1/360 samples. The majority of these operational false positives were in pools with an initial cycle threshold (CT) value of >36. By using a CT value of ≤36, the proportion of confirmed positive pools rose to 183/189 (96.8%). The cause of operational false positive results was likely multifactorial; because pooled and individual tests were conducted on separate samples approximately 24 hours apart, this may have led to true negative individual tests from students with falling viral loads, false negative individual tests from students with borderline viral loads, as well as technical false positive test results for CT values at the limit of detection of the assay. The proportion of operational false positives was lowest when the prevalence of SARS-CoV-2 infection was highest (i.e. in weeks 2 (1/28, 3.6%), 3 (2/38, 5.3%) and 6 (8/67, 11.9%)); conversely, the proportion of operational false positives was highest in weeks 8-9 (14/17, 82.4%), when the prevalence of infection was lowest (**Appendix I.2**).

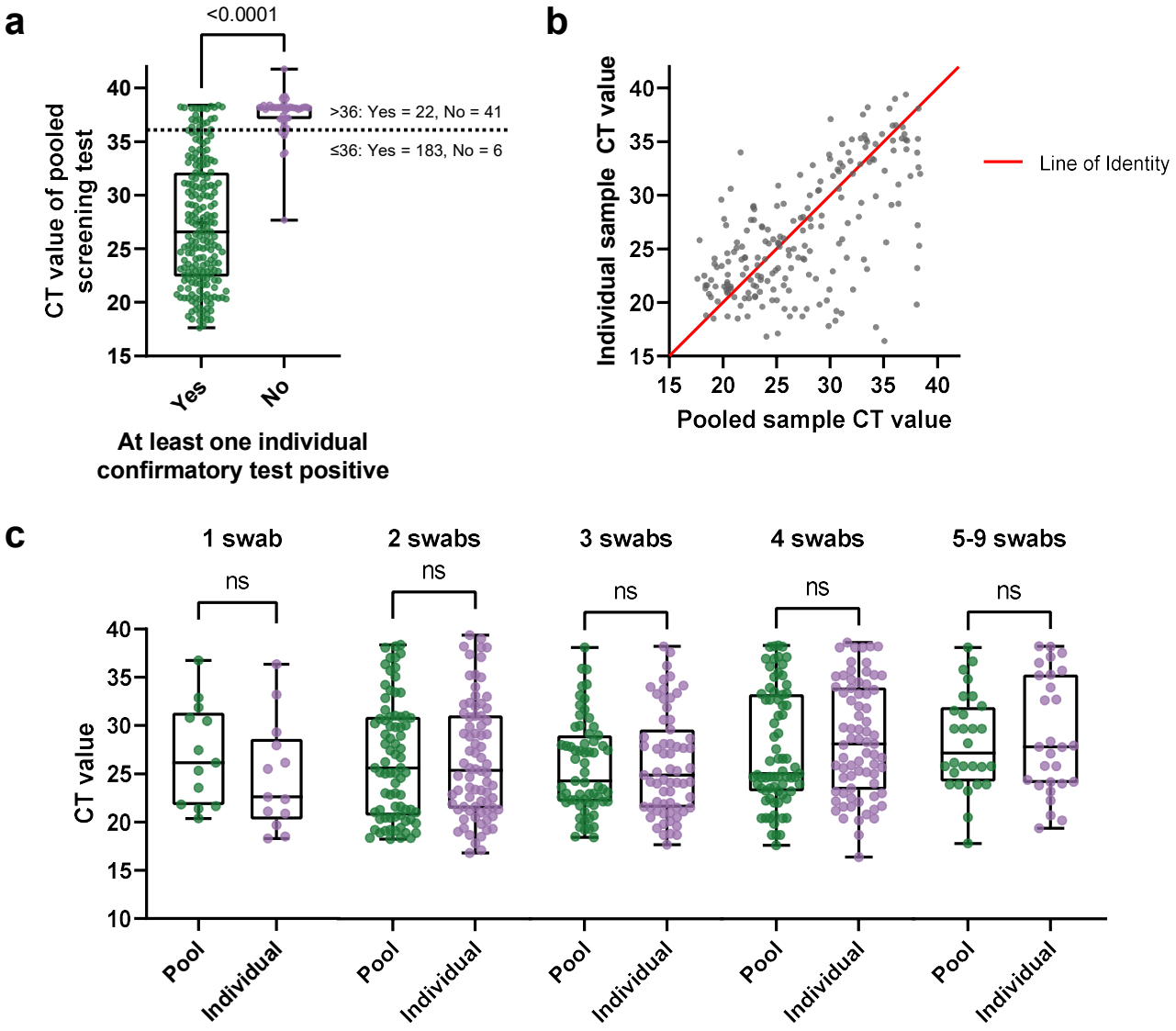


Figure 8.4 Clinical validation of swab pooling

(**a**) SARS-CoV-2 cycle threshold (CT) values for pooled screening tests, comparing 203 pooled samples resulting in positive individual confirmatory tests with 47 operational false positives (no positive individual confirmatory tests). Centre line indicates median, boxes indicate interquartile range, whiskers indicate range. Distributions were compared using an unpaired 2-tailed *t*-test. (**b**) Comparison of CT values between positive paired pooled screening and individual confirmatory tests. Where a positive pooled screening test resulted in >1 positive individual confirmatory test, the lowest individual CT value is plotted.

(**c**) Comparison of CT values between positive paired pooled screening and individual confirmatory tests, stratified by number of swabs in the pooled sample. Distributions were compared using paired 2-tailed *t*-tests. ns = not significant, $P>0.05$.

It was not possible to determine an operational false negative rate for this pooled screening approach, as students contributing swabs to negative pooled tests did not undertake confirmatory individual tests. However, there was a reasonable correlation between CT values from pooled tests and their paired individual confirmatory tests (**Figure 8.4b**). Additionally, no differences were observed in the distributions of CT values between paired tests when stratified by the number of swabs included per sample (**Figure 8.4c**). Supported by findings in previous studies[246-248], these data indicate that swab pooling does not result in any substantial loss of PCR test sensitivity.

8.4 Characteristics of students with confirmed SARS-CoV-2

Across asymptomatic screening and symptomatic testing via University and other pathways, 671/12,781 (5.2%) of participating students were diagnosed with SARS-CoV-2 infection over the term (**Appendix I.2**). Cases rose steeply in the second week of term, before gradually falling during weeks 3-5 and a subsequent increase in week 6 (**Figure 8.5a**). Following the introduction of a second national lockdown on November 5, there was a dramatic decline in cases until the end of term. This is described in detail below. Single variable and multivariable models demonstrating associations with testing positive for SARS-CoV-2 are shown in **Tables 8.2** and **8.3**. Testing positive was associated with being male, of white ethnicity, having permanent residency in the UK, being in the student's first year at the University, studying arts and humanities at undergraduate level and living in a larger household.

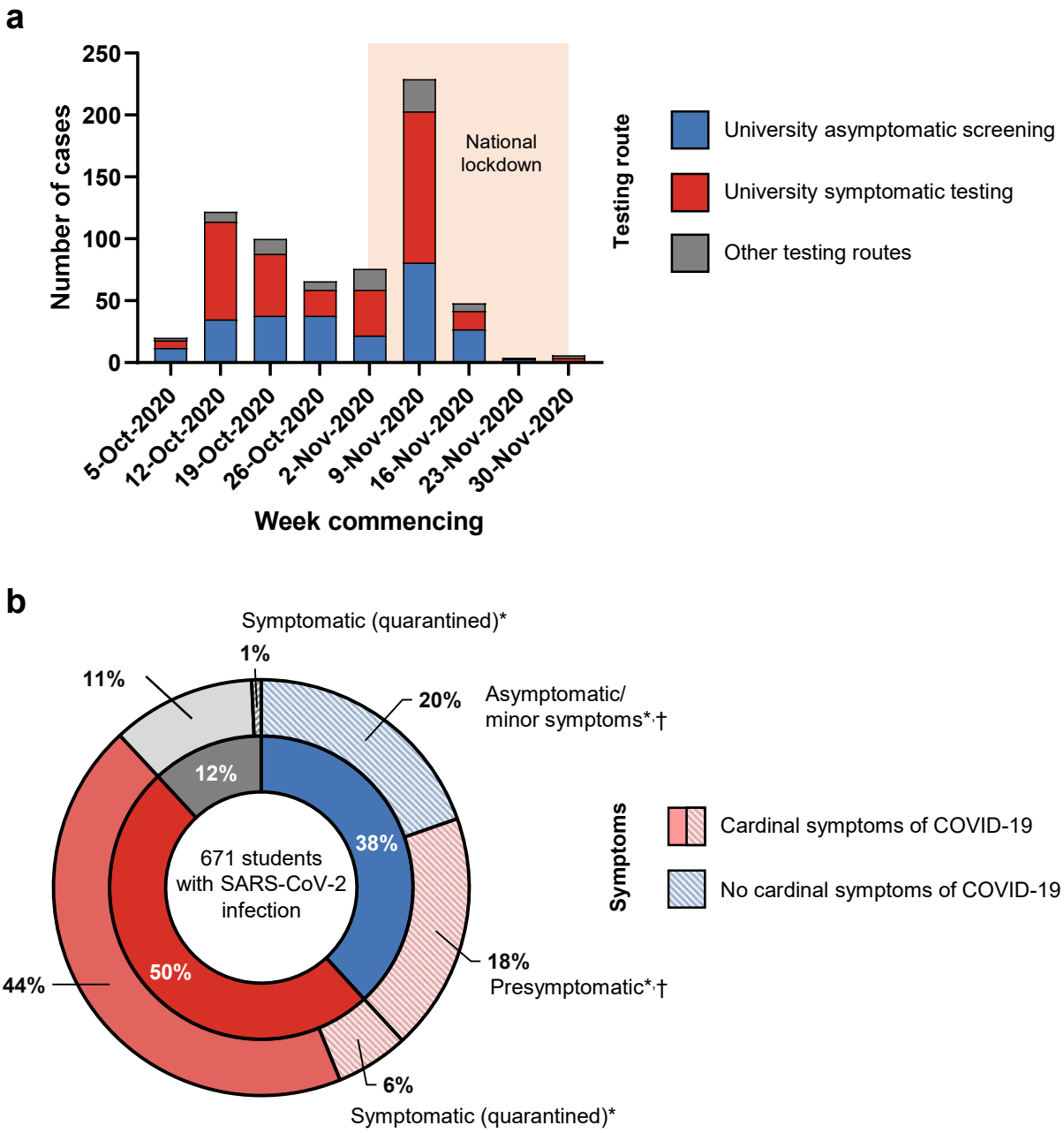


Figure 8.5 Case ascertainment by asymptomatic screening and symptomatic testing.
(**a**) Epidemic curve of weekly SARS-CoV-2 cases amongst participating students identified by asymptomatic screening (blue), symptomatic testing (red) or other testing routes (grey). The steep decline in cases coinciding with the second UK national lockdown is indicated by the shaded area. (**b**) Classification of SARS-CoV-2 cases amongst participating students according to testing pathway (inner doughnut) and symptoms (outer doughnut). Inner doughnut: same colours as (**a**). Outer doughnut: red, cardinal symptoms of SARS-CoV-2 (fever, cough and/or anosmia/ageusia) experienced at some time during infection; blue, no symptoms or minor symptoms only (no cardinal symptoms). *Cases where potential for onward transmission was interrupted by asymptomatic screening are indicated by patterned areas. †Estimated numbers based on retrospective telephone survey (sampling 140/256 students).

Table 8.2 – Single variable logistic regression analysis of student characteristics associated with a positive SARS-CoV-2 result.

n=12,781

Variable	Category	Positive	Not positive	OR	(95% CI)	*P* value
Sex	Female	325	5,800			
	Male	346	6,158	1.00	0.86-1.17	0.97
	Other/unknown	5	147	0.61	0.21-1.34	0.28
Ethnicity	White	509	7,659			
	Other	153	3,944	0.58	0.48-0.70	<0.001
	Unknown	14	502	0.42	0.23-0.69	0.002
Residency	UK	570	7,934			
	International	103	4,102	0.35	0.28-0.43	<0.001
	Unknown	3	69	0.61	0.15-1.63	0.40
Year of study	1st	367	5,195			
	2nd	166	2,999	0.78	0.65-0.95	0.01
	3rd or higher	142	3,786	0.53	0.43-0.65	<0.001
	Unknown	1	125	0.11	0.006-0.51	0.03
Course	Undergraduate arts and humanities	348	4,205			
	Undergraduate science and technology	246	4,560	0.65	0.55-0.77	<0.001
	Postgraduate vocational	22	321	0.83	0.52-1.26	0.41
	Postgraduate (other)	59	2,894	0.25	0.18-0.32	<0.001
	Unknown	1	125	0.10	0.0055-0.43	0.02
Household size	Minimum	1	1	1.08	1.05-1.11	<0.001
	25th centile	6	5			
	Median	7	7			
	75th centile	8	8			
	Maximum	20	20			

Table 8.3 – Multivariable logistic regression analysis of student characteristics associated with a positive SARS-CoV-2 result.

Variable	Category	aOR	95% CI	*P* value
Sex	Female			
	Male	1.21	1.03-1.42	0.02
	Other/unknown	0.44	0.07-1.42	0.26
Ethnicity	White			
	Other	0.69	0.57-0.84	<0.001
	Unknown	0.46	0.23-0.81	0.01
Residency	UK			
	International	0.53	0.42-0.67	<0.01
	Unknown	12.4	1.63-113	0.01
Year of study	1st			
	2nd	0.63	0.52-0.76	<0.001
	3rd or higher	0.44	0.36-0.54	<0.001
	Unknown	0.04	0.002-0.24	0.004
Course	Undergraduate arts and humanities			
	Undergraduate science and technology	0.74	0.62-0.88	0.001
	Postgraduate vocational	0.73	0.45-1.11	0.16
	Postgraduate (other)	0.28	0.21-0.38	<0.001
	Unknown	NA	-	-
Household size		1.06	1.03-1.09	<0.001

Of all confirmed SARS-CoV-2 cases in students participating in the programme across the term, 256/671 (39.2%) were identified by asymptomatic screening (**Figure 8.5b**, **Appendix I.2**). To determine the proportion of students who remained truly asymptomatic during the course of their infection, a retrospective telephone survey was conducted of all students testing positive through asymptomatic screening, with 140/256 (54.7%) responses. Of these,

29/140 (20.7%) remained entirely asymptomatic, 43/140 (30.7%) reported only minor symptoms and 68/140 (48.6%) reported developing cardinal symptoms of COVID-19 (fever, cough and/or anosmia/ageusia) (**Figure 8.6a**,**b**). The distribution of time intervals between pooled screening sample collection and symptom onset in these students suggests a variable presymptomatic infectious period, with a mean of 3.6 days (**Figure 8.6c**). This period is consistent with, though slightly longer than, estimates from other studies (typically 1-4 days)[432]. This disparity may reflect differences in methodology, or the specific characteristics of this study population.

The distribution of CT values was higher for students identified via asymptomatic screening in comparison to those tested via the University's symptomatic testing pathway (mean CT values 24.8 vs 27.2, P <0.0001; 59% vs 42% CT values <25, P <0.0001), suggesting lower viral loads (**Figure 8.7a**). However, stratifying the results of students who responded to our telephone survey showed that CT values were lower for presymptomatic students (mean CT value 25.5; 53% CT values <25), similar to students identified by symptomatic testing (**Figure 8.7b**). This is consistent with previous studies that have demonstrated that RNA shedding peaks early in SARS-CoV-2 infection, most commonly at, or immediately prior to, symptom onset, and therefore approximately coinciding with testing of symptomatic students or screening of presymptomatic students[433]. By contrast, within the limits of the screening interval (most commonly 14 days for the majority of cases identified during the study period), truly asymptomatic cases may be sampled at any point during their infection.

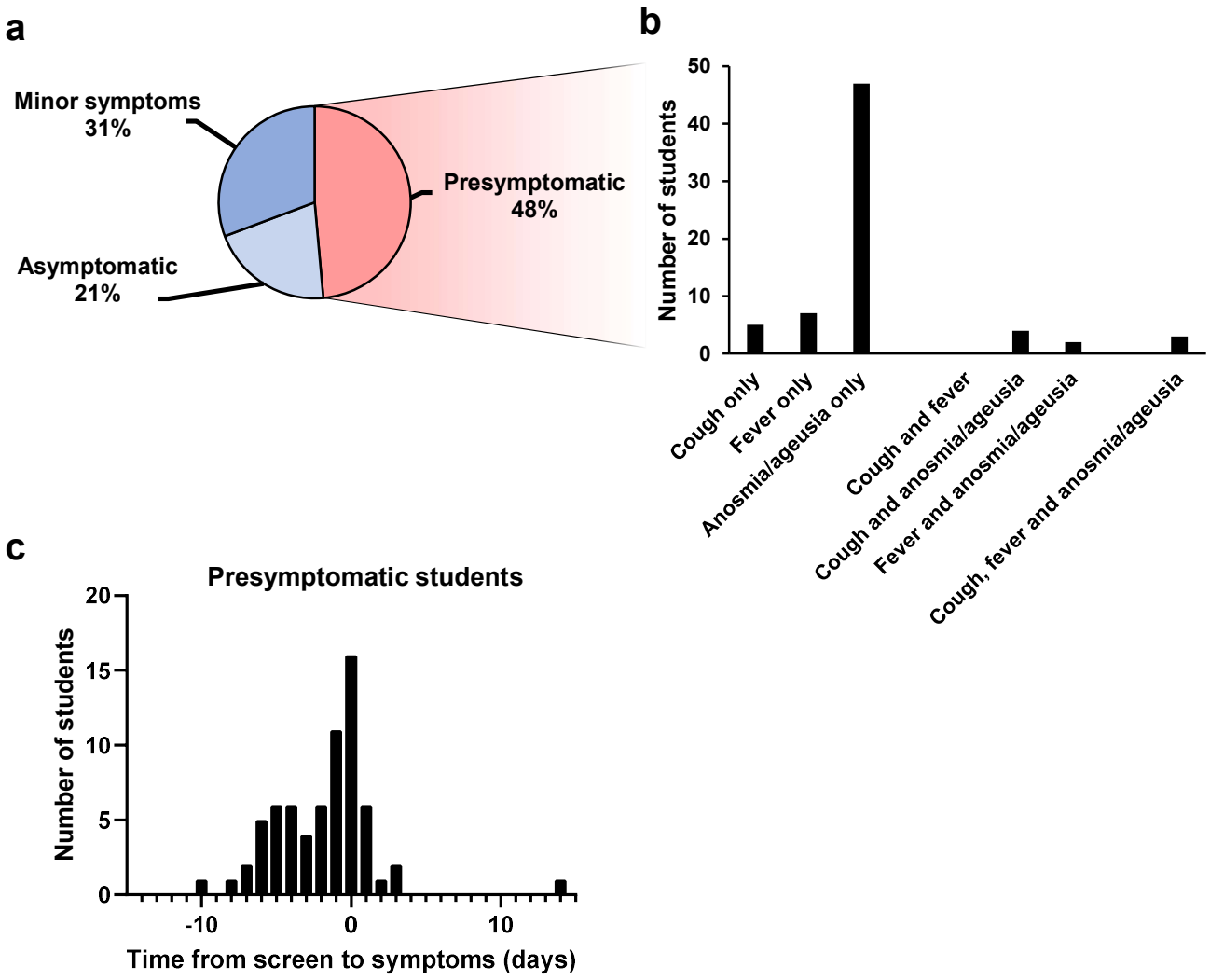


Figure 8.6 Symptoms identified by telephone survey.

(**a**) Proportion of students identified by asymptomatic screening who went on to develop cardinal symptoms of COVID-19 (fever, cough, and/or anosmia/ageusia) at some time during infection (presymptomatic, red), those who experienced no symptoms at all (asymptomatic, pale blue), and those who reported only minor symptoms (minor symptoms, dark blue). Data are shown for 140 students who responded to the retrospective telephone survey. (**b**) Frequency of individual symptoms reported by 68 presymptomatic students from (**a**). (**c**) Time interval between positive screening test and onset of cardinal symptoms of COVID-19 for 68 presymptomatic students from (**a**). Since presymptomatic students may be sampled at any time during their presymptomatic infectious period, and (on average) halfway through, the mean presymptomatic infectious period is expected to be twice the mean observed time interval.

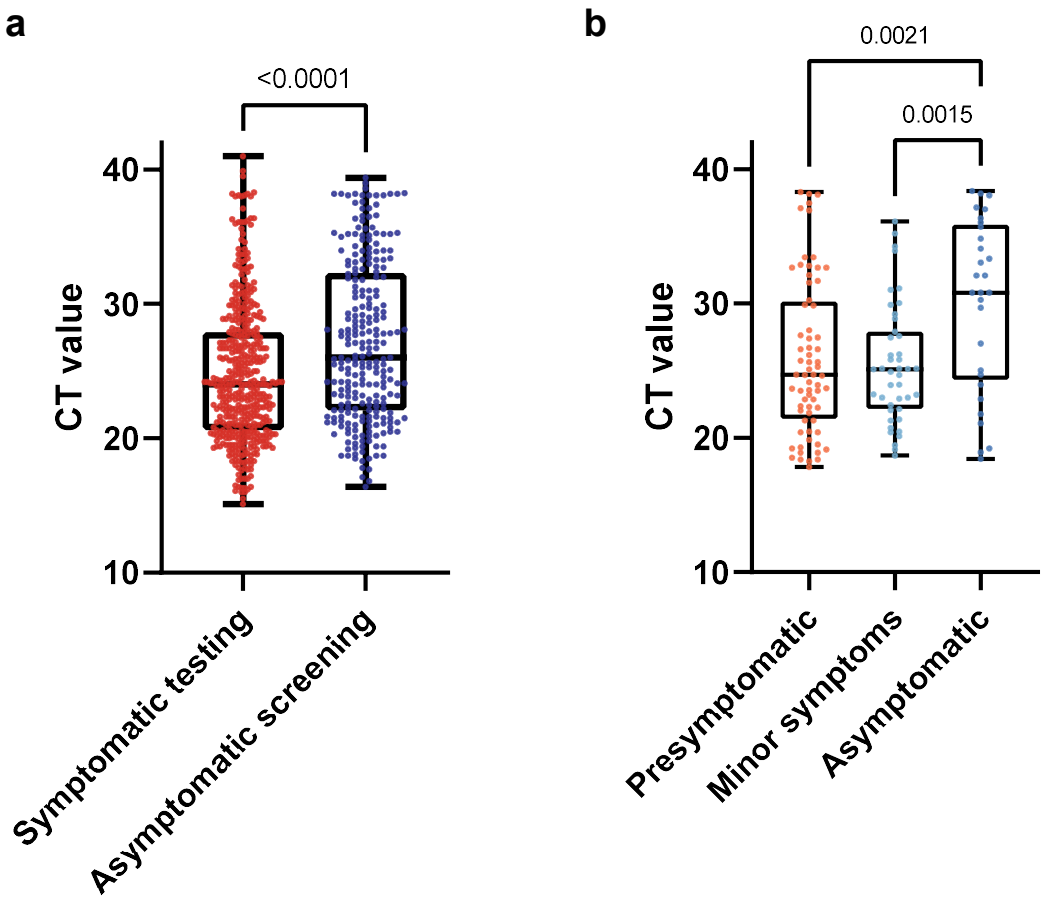


Figure 8.7 Comparison of CT values by testing pathway and symptoms.
(**a**) SARS-CoV-2 cycle threshold (CT) values for positive University symptomatic tests, compared with positive individual confirmatory tests from the asymptomatic screening programme. (**b**) Comparison of CT values for students identified by asymptomatic screening who went on to develop cardinal symptoms of COVID-19 (fever, cough, and/or anosmia/ageusia) at some time during infection (presymptomatic, red), those who experienced no symptoms at all (asymptomatic, dark blue), and those who reported only minor symptoms (minor symptoms, pale blue). Data are shown for 140 students who responded to the retrospective telephone survey. Centre line indicates median, boxes indicate interquartile range, whiskers indicate range. *P* values were calculated for pairwise comparisons using unpaired two-tailed *t*-tests (**a**) and ordinary one-way ANOVA (**b**).

8.5 Efficacy of asymptomatic screening in controlling transmission

Applying data collected from the telephone survey to all cases identified during the term, we estimate that 20% (95% CI 16.4-22.9%) of all students with confirmed COVID-19 were asymptomatic or only experienced minor symptoms (**Figure 8.5b**). Of the students who developed disease with one of the cardinal symptoms of COVID-19, 23% (95% CI 19.0-27.1%), were detected by the asymptomatic screening programme while in the pre-symptomatic phase of their illness. An additional 8% of these students were detected by symptomatic testing, but had been quarantined at the time of symptom onset they were household contacts of a case identified through asymptomatic screening. Overall, the asymptomatic screening programme may therefore have contributed to the interruption of onward transmission in 45% of students with confirmed SARS-CoV-2 infection.

To further assess the impact of asymptomatic screening on detecting pre-symptomatic infection, I reviewed cases from weeks 3-7 of term. Over this time, students were screened on alternate weeks with a screening interval of 14 days. As shown in **Figure 8.8**, the risk of a positive symptomatic test in students participating in the programme was lowest immediately following a negative pooled screen. The likelihood of a positive symptomatic test was approximately 51% lower in the first week after a negative screening test compared with the second week (52 *v* 106 students). This reduction likely corresponds to the successful identification of presymptomatic cases through screening. Additionally, it suggests that asymptomatic screening will detect approximately half of all students who subsequently develop cardinal symptoms of COVID-19 while they are still presymptomatic, which supports our observation that the mean presymptomatic infectious period in this study is 3.6 days.

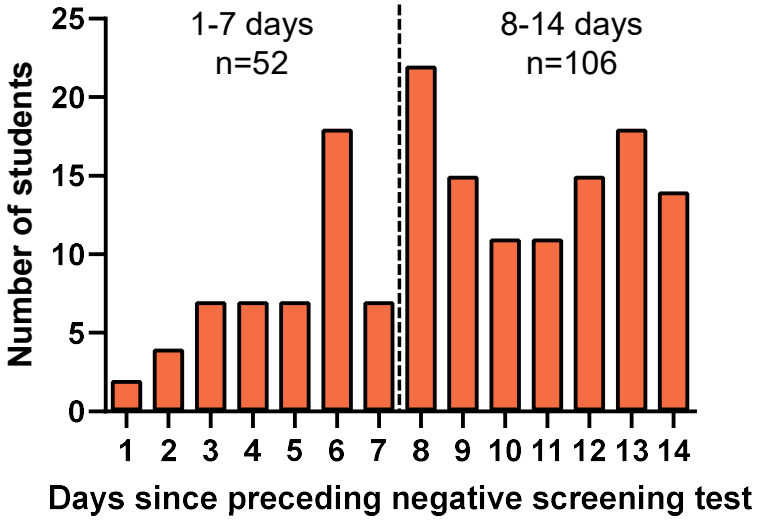


Figure 8.8 Efficacy of asymptomatic screening.
Time between positive symptomatic test and most recent negative screening test for 158 students with symptomatic SARS-CoV-2 infection identified during weeks 3 to 7. Half of each testing pool was screened on alternating weeks over this period, corresponding to a screening interval of 14 days for individual students.

8.6 Genomic analysis of SARS-CoV-2 cases

To further investigate transmission dynamics of SARS-CoV-2 in the University of Cambridge, and the impact of interventions such as the asymptomatic screening programme, a combined genomic and epidemiological analysis was attempted on all confirmed cases, including students and staff members who did not participate or were not eligible for asymptomatic screening. In addition to 448 sequences derived from University testing routes, 24 sequences were obtained from community testing in the COG-UK dataset and 12 sequences from the CUH healthcare worker (HCW) testing programme[220]. In total, 484 high

quality genome sequences were available for University members (**Figure 8.9**). Additionally, 972 contextual sequences were available from the local community, including 747 from the general public identified through community testing, 121 from HCWs and 104 from patients identified by hospital testing (71 from CUH and 33 from other medical facilities in Cambridgeshire). Of all positive SARS-CoV-2 test results from Cambridgeshire during the study period (including the community and University), an estimated 8% were successfully sequenced and included in this analysis.

A maximum likelihood phylogenetic tree of all sequences in this study is shown in **Figure 8.10**. Over the study period, 62 Pango lineages were identified across all sequences (**Figures 8.10** and **8.11**). In the University, 23 lineages were identified, although the majority (438/482, 90.9%) were from just 4 lineages (B.1.60.7, B.1.177, B.1.36, B.1.177.16), all of which were detected by the second week of term. Genomes with spike protein mutations associated with increased transmission or reduced sensitivity to antibody-mediated immunity were observed in the university population: 3 sequences from the B.1.258 lineage with the N439K mutation and ΔH69/ΔV70; 2 cases of B.1.1.7/alpha variant and its associated mutations[434]; and 88 cases of B.1.177 with the A222V mutation[435]. Of these, lineage B.1.1.7 is most reliably associated with increased transmission[436]; both cases of B.1.1.7 were amongst postgraduate students with no epidemiological associations, diagnosed during national lockdown, and with no evidence of further transmission within the University.

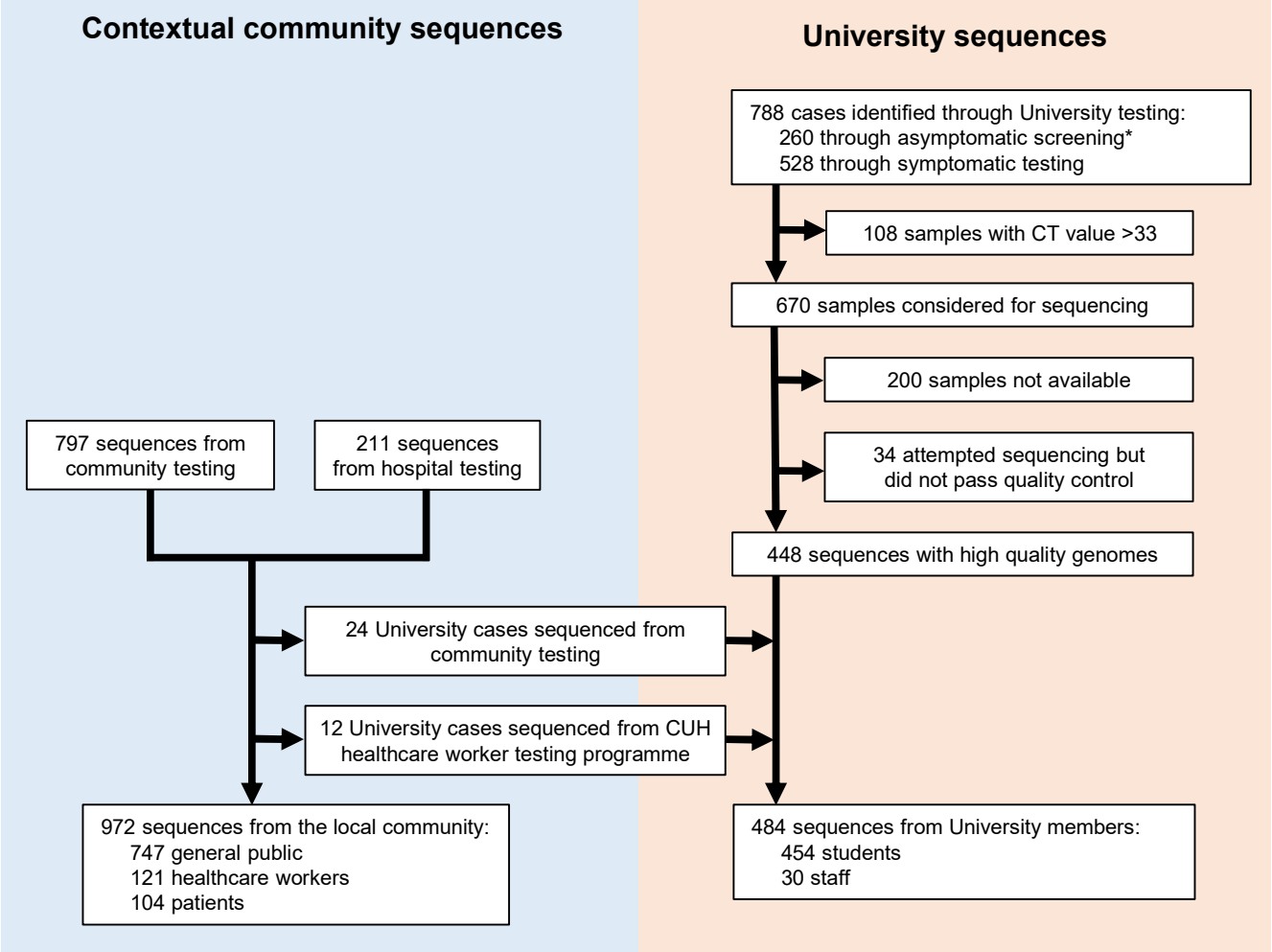


Figure 8.9 Study cohort and available genome sequences.

*Includes 14 students identified through ad hoc asymptomatic screening conducted as part of an outbreak investigation by the University of Cambridge in conjunction with local public health authorities, responding to increased rates of infection in a block of student accommodation (described in further detail in cluster 2 below). **includes 2 students associated with a single sequenced pooled sample. CUH = Cambridge University Hospitals

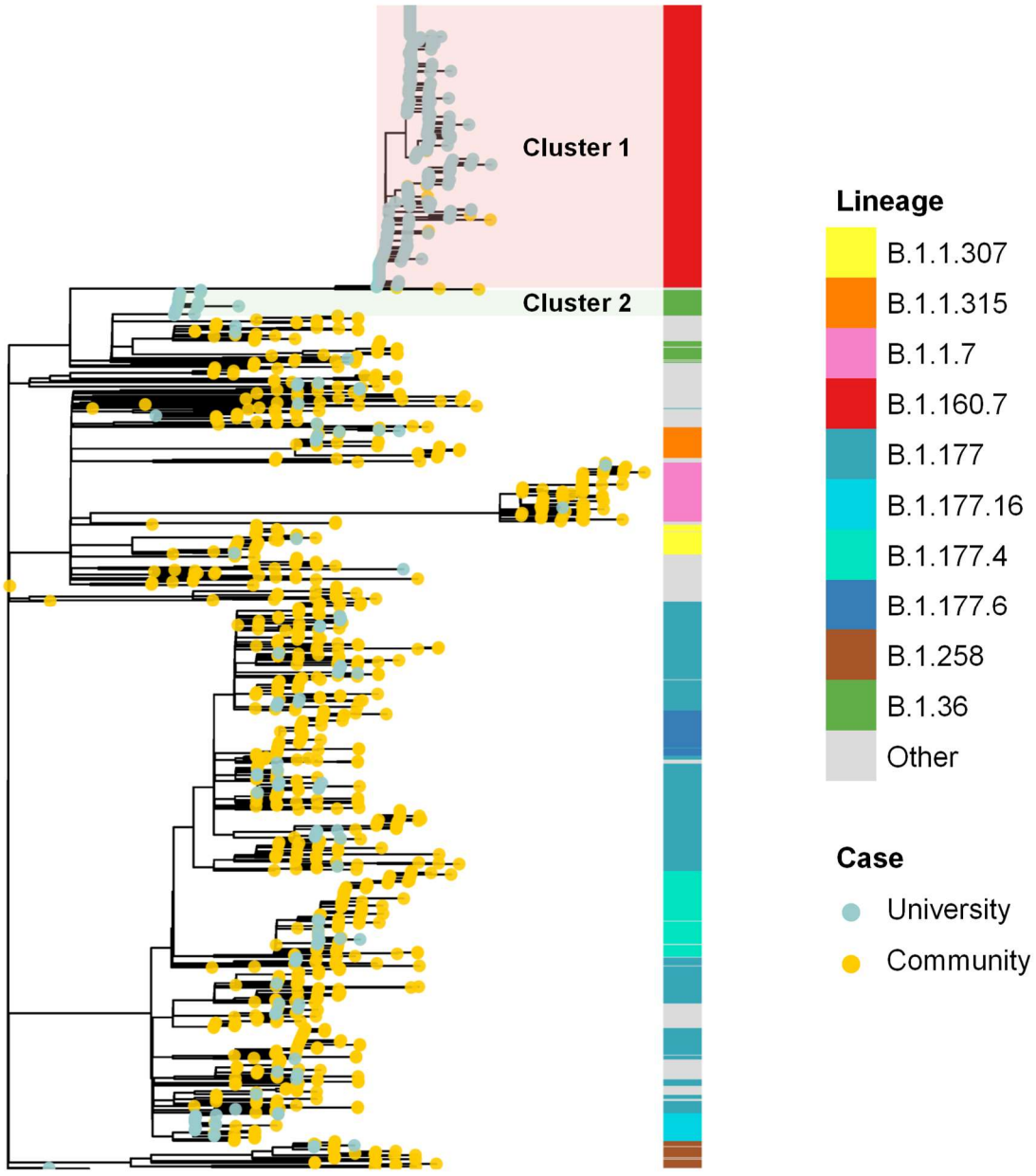


Figure 8.10 Phylogenetic tree of SARS-CoV-2 genomes from University of Cambridge and local community

Maximum likelihood phylogenetic tree of 484 SARS-CoV-2 genomes from the University of Cambridge and 972 genomes from the local community, produced using IQTREE. The tree is rooted using Wuhan Hu-1 (MN908947.3) as an outgroup. The branch tips are coloured by the origin of the sample (University in aqua, local community in yellow), demonstrating that the majority of sequences derived from University samples are from distinct lineages to those derived from the community. The global PANGO lineage is illustrated in the vertical bar. Two large University clusters are annotated. Phylogeny produced by Dinesh Aggarwal, and reproduced here with permission.

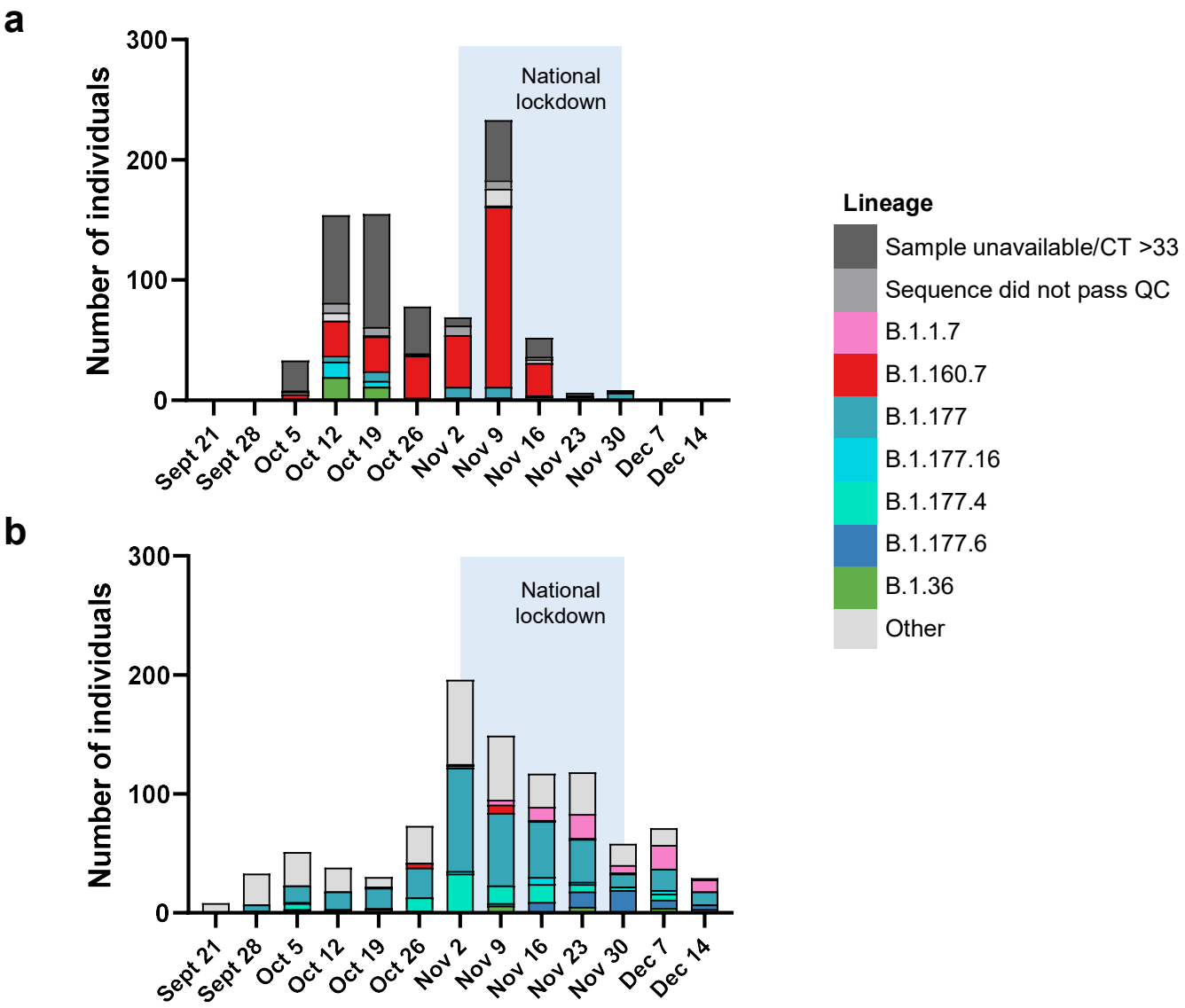


Figure 8.11 Epidemic curves of SARS-CoV-2 and associated lineage in the University compared to the local community

(**a**) Epidemic curve of all confirmed SARS-CoV-2 cases in University members from October 5 to December 6. (**b**) Epidemic curve of community cases, extending beyond the end of the University term. Only community cases with available genomes are included. The light blue shaded area reflects a 4-week national lockdown in the UK, which was associated with a large fall in COVID-19 cases in University students. Specific lineages highlighted are the 4 largest lineages within the University (minimum 20 cases over the study period) and the community (minimum 50 cases over the study period).

A total of 198 putative transmission clusters were identified by CIVET (https://github.com/artic-network/civet), with 36 clusters including one or more members of the University. Of these, 8/36 contained 5 or more University members (range 6-337 individuals), which together represent 91.3% of all University cases. This indicates that the majority of introductions into the University did not cause ongoing transmission. To further investigate the largest of these clusters (cluster 1, described below), I identified subgroups of identical samples (0 SNP differences) which produced 19 additional clusters (a grand total of 34 clusters with 2 or more University cases) for further analysis. A detailed summary of all 34 genomic clusters and any epidemiological associations between cases is provided in **Appendix G.2**; specific illustrative clusters are discussed below.

8.7 Introductions of SARS-CoV-2 into the University

To determine transmission dynamics following the introduction of viral lineages into the University, we performed further detailed investigation of the largest genomic cluster (cluster 1), which accounted for 337/484 (69.6%) sequences from the University. This was first detected in week 1 of term and had spread widely four weeks later, affecting students from 29/31 Colleges, 28 undergraduate courses and 208 households in College accommodation (**Figure 8.11a**). Isolates from cluster 1 were from lineage B.1.160.7. No mutations known to be associated with increased transmissibility were identified in this lineage compared to other genomes in the study. Interrogation of the COG-UK dataset showed that this lineage was first identified in the UK on October 4, 2020. It was not identified in the local Cambridgeshire community until term week 3 (October 19-25, 2020) and is phylogenetically

distinct from other lineages in the dataset (**Figure 8.10**). These data suggest that the B.1.160.7 lineage was introduced into the University from outside Cambridgeshire. Subsequent analysis with A2B-COVID[425], which uses genomic data paired with date of symptom onset or sample collection to evaluate the likelihood of transmission between individuals, has shown that these sequences were consistent with a single introduction into the University[437].

Contact tracing data from the University and NHS test and trace were interrogated to identify potential sources of this isolate and how it propagated through the University. Ten students testing positive in the first two weeks of term reported visiting the same nightclub (venue A); nine of these students either had a sequence from cluster 1, or (in the event that their sample did not yield a high quality sequence) shared a household of an individual with an associated sequence from cluster 1. No data were available for the tenth student.

Ongoing transmission of cluster 1 was evident from the first week of term until lockdown was enforced on November 5, 2020 (**Figure 8.11a**). There was a spike in cases in week 6, following the announcement of the impending lockdown made on October 31, coinciding with Halloween festivities. Students testing positive in the two week period around the introduction of lockdown reported that the majority of common exposure events occurring outside University grounds (25/59, 42.4% events reported by 48 students) were linked to nightclubs. Venue A was again identified as the most common venue visited by students where contact tracing information was available. Of the 16 students reporting visiting venue A, 9 had sequences in cluster 1 and a further 5 (where no genome sequences were available) were household contacts of individuals in cluster 1. All sequences from students who reported attending venue A, or who were a household contact of an individual who attended venue A, were present in cluster 1 (**Figure 8.12**).

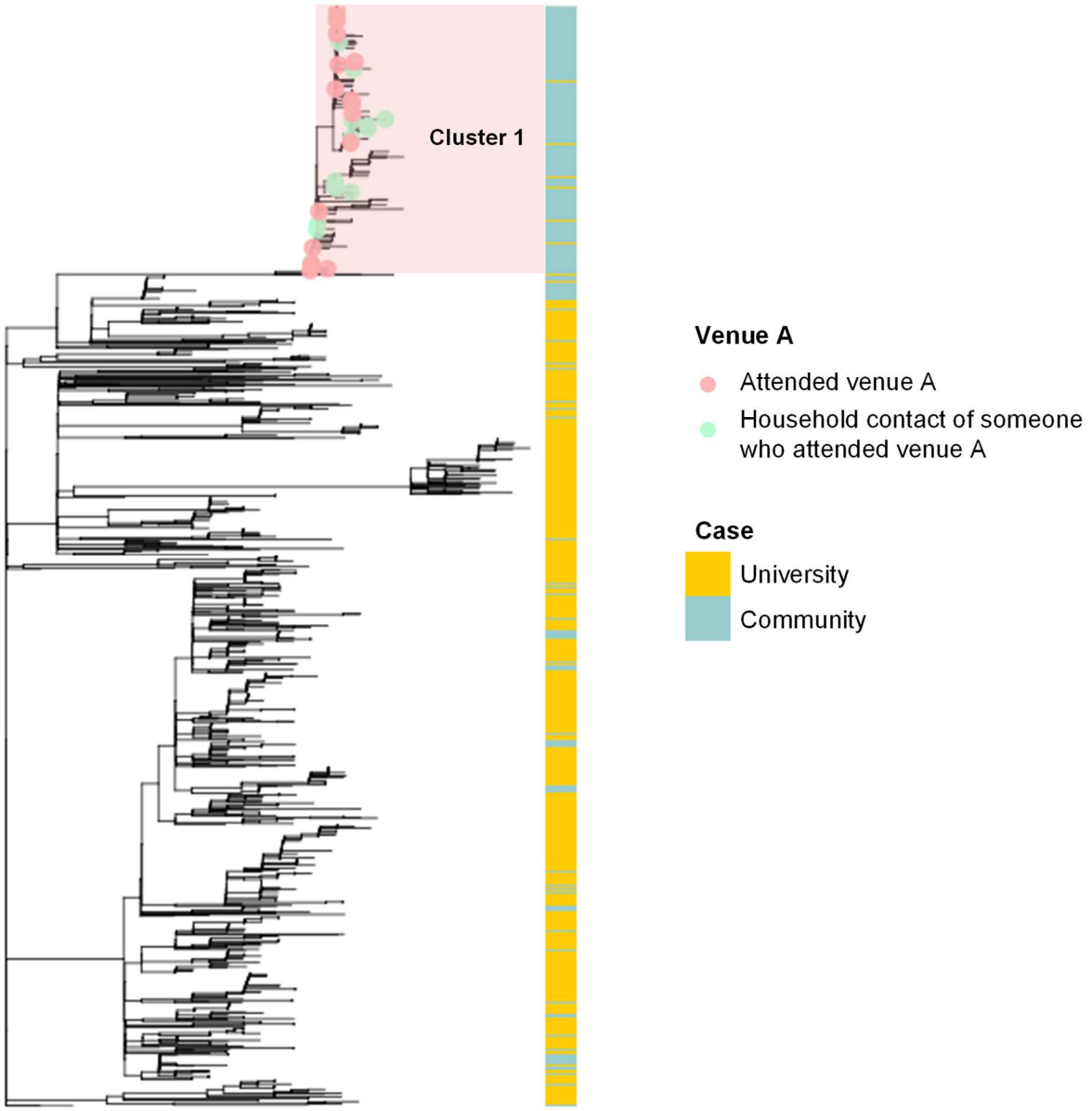


Figure 8.12 Phylogenetic tree of SARS-CoV-2 genomes demonstrating cases associated with venue A

The same maximum likelihood phylogenetic tree as **Figure 8.10** is presented again here, but annotated with associations with the nightclub venue A. Individuals who attended venue A (pink tree tips) and household contacts of individuals who visited Venue A but were not sequenced (green tree tips) are highlighted on the tree, all of which are located in cluster 1. This venue was implicated as a possible source for the dispersion of SARS-CoV-2 across the University at the start of term, and increased transmission in the weeks around the introduction of national lockdown. The vertical panel represents cases by location (general community and University members). Phylogeny produced by Dinesh Aggarwal, and reproduced with permission.

8.8 Transmission within student households

In 18/34 (52.9%) University genomic clusters there was evidence of transmission of SARS-CoV-2 in student accommodation. In cluster 1, 169/337 (50.1%) students had an associated viral sequence identical to at least one other student living in the same or neighbouring household (with clusters ranging from 2-11 students).

The largest single network associated with transmission in student accommodation was cluster 2 (lineage B.1.36, **Figure 8.10**). By term week 3, this cluster involved 30 students, of whom 24/30 (80%) lived in the same accommodation block in College A, while 4/30 students lived in 2 separate households in the same College. The first individuals to be identified were screened via the asymptomatic screening programme, with subsequent students identified through the University's symptomatic testing pathway later that week. All suspected and confirmed cases, and their households, were immediately isolated and contact tracing initiated. Within four days of the index case testing positive an extraordinary meeting was held between members of the College, University and the local public health authority who agreed an immediate lockdown of the affected accommodation block in its entirety. Students were supported with deliveries of food and drink, their educational needs were discussed individually between students and their tutors, and additional psychological support was provided as necessary. However, students were not allowed to leave the accommodation block unless they were attending an appointment for SARS-CoV-2 testing or another valid medical reason. In addition to the existing availability of symptomatic testing from the University, individual asymptomatic screening was offered to all students living in the accommodation block over the following four days. These measures were successful at reducing the number of cases within both the accommodation block and the wider College. After the end of term week 3, no further genomically-related isolates were identified in the

University or surrounding community, indicating that these interventions were successful at halting transmission.

All other genomic clusters where likely transmission events were identified between household members are detailed in **Appendix G.2**. They follow similar patterns to cluster 2, with groups of cases confined to a single College not leading to sustained transmission, through a range of interventions similar to those described above.

To better understand transmission within households, I used data from the asymptomatic screening programme to assess the risk of students with SARS-CoV-2 infection (index cases) causing secondary infections amongst household contacts. Compared with index cases identified by symptomatic testing, the secondary attack rate identified by asymptomatic screening was lower (6.9% vs 12.0% of household contacts, P=0.0003) (**Figure 8.13a**). Stratifying index cases based on their symptomatology (based on samples from the telephone survey), the secondary attack rate was much higher for presymptomatic students (15.2%), similar to index cases identified by symptomatic testing (**Figure 8.13b**). Conversely, the secondary attack rate for persistently asymptomatic students was much lower (2.3%), consistent with the differences observed in CT values shown in **Figure 8.7b** and suggesting a lower infectiousness of students who never develop symptoms (relative risk 0.17 compared with symptomatic students, 95% C.I. 0.08-0.38). This is consistent with studies in other populations[169].

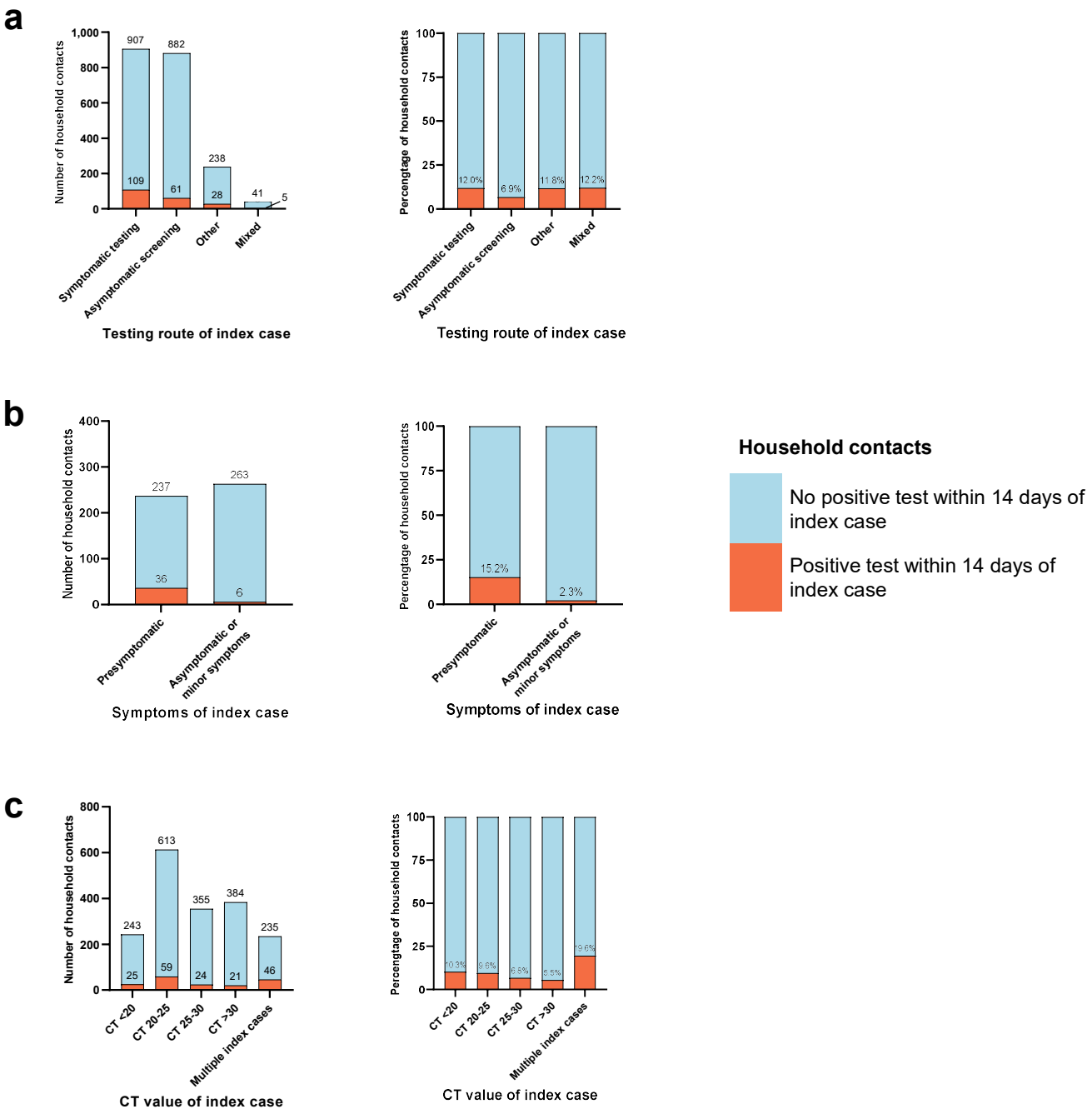


Figure 8.13 Secondary household attack rates.

(**a**) Total number (left panel) and proportion (right panel) of household contacts of students with confirmed SARS-CoV-2 infection who tested positive for SARS-CoV-2 within 2 weeks of the index case (any testing pathway). Total household contacts (orange plus blue bars) and household contacts testing positive for SARS-CoV-2 (orange bars) are shown. Data are stratified by testing pathway of the index cases. Other, students identified from notifications to Public Health England. Mixed, multiple students from same household identified on the same index day, using different testing pathways. The proportion of household contacts who tested positive was significantly different for asymptomatic (6.9%) versus symptomatic (12.0%) index cases (P=0.0003, Fisher's exact test). (**b**) As for (**a**), but results are shown for 140 index cases identified by asymptomatic screening who responded to the retrospective telephone survey. Data are stratified according to whether index cases went on to develop cardinal symptoms of COVID-19 (fever, cough, and/or anosmia/ageusia) at some time during infection (presymptomatic), or remained asymptomatic/reported only minor symptoms. The proportion of household contacts who tested positive was significantly different for contacts of presymptomatic index cases (15.2%) versus contacts of index cases who remained asymptomatic/reported only minor symptoms (2.3%) (P<0.0001, Fisher's exact test). (**c**), As for (**a**), but data are stratified according to CT value of the index case. Where multiple students from the same household were identified on the same index day, data are shown separately (multiple index cases).

Infectious virus has been found to be more readily recovered in laboratory settings from patient samples with low CT values (reflecting higher viral loads); individuals with low CT values are therefore presumed to be more infectious[438]. To test this hypothesis in our own cohort, I examined the relationship between the CT values of samples from index cases with the secondary attack rate amongst household contacts. There was an inverse relationship between the CT value and the secondary attack rate, supporting this hypothesis, although the magnitude of the effect was relatively modest, ranging from 10.3% (CT<20) to 5.5% (CT>30) (**Figure 8.13c**).

8.9 Other transmission routes among University members

In addition to transmission between household contacts, I identified networks of genomically related sequences in students that share the same course and year of study in 14/34 (41.2%) genomic clusters. In cluster 1, sequences from 203/337 (60.2%) students were identical to at least one other student sharing the same course and year (cluster size range 2-14 sequences), disproportionately affecting students in their first year of study. Additionally, at postgraduate level I identified two small clusters of students working in the same University department. However, in the majority of clusters I was not able to determine the likely location of transmission: there was considerable overlap between course and household clusters, and complex social and study networks are known to exist between students (as described in **Appendix G.1**, for example in clusters 3, 4 and 10).

In comparison to students, the number of SARS-CoV-2 sequences from University staff members was limited (n=30). In four genomic clusters I identified evidence of transmission between staff members working in the same University department, College or ancillary role. In two separate clusters I identified staff members who shared the same household. There were 8 genomic clusters involving both University staff and students. However, corresponding epidemiological associations between staff and students could only be identified in one genomic cluster: a shared household between a student and staff member working in separate University departments.

While the majority of students could be associated with at least one other case by household or course, 23/34 genomic clusters contained at least one University member that could not be epidemiologically linked with any other case in their cluster. This demonstrates potential cryptic transmission events that could not be identified in this study.

8.10 Transmission between the University and the surrounding community

Having identified transmission networks within the University, I next sought to investigate the extent of transmission between the University and the local community. The phylogenetic tree in **Figure 8.10** shows that the majority of sequences derived from the University are genetically distinct from those derived from the community, demonstrating a lack of substantial transmission between the two settings. In total, 29/198 (14.6%) clusters contained both University and community cases; of those, only 6 contained 5 or more University members and 3 or more individuals from the community (**Appendix G.2**).

Further focused analyses were performed to identify transmission events involving the University and local hospitals. 17 clusters were identified that involved both University members and hospital patients or staff (**Appendix G.2**). Cluster 1, representing the majority of student cases, contained only one patient and one HCW, with no identified epidemiological links to students. The second largest cluster of University members (n=21) included 2 patients, 5 HCWs and 9 clinical medical students at the University. Phylogenetically, the medical students and one of the HCWs were closely linked. All 9 medical students were on clinical rotations at the time that the index case was diagnosed; 7/9 students lived in neighbouring households in the same College, while the remaining 2 were identified as contacts of the index student.

To further investigate epidemiological associations in genomic clusters involving University members and the community, 1,243/1,455 (85.4%) individuals sequenced over the study period (excluding hospital cases) were linked to national contact tracing data. Of these, 219 (17.6%) cases reported 127 common exposure events. In cluster 1, which represented 69.6% of cases within the University, but only 18/976 (1.8%) cases from the community, only one individual from the community had a common exposure with a University student: dining at the same public restaurant on the same day. No other epidemiological links were identified in all other genomic clusters. Transmission events between University and community cases suspected in 19 epidemiologically linked clusters (defined by common exposures identified through national contact tracing) were refuted by genomic variation (defined by separate transmission clusters as determined by CIVET).

8.11 Discussion

In this chapter I have presented a comprehensive description of the introduction and evaluation of a regular, large-scale asymptomatic screening programme for COVID-19 in the University of Cambridge. I have demonstrated how participation of students and households was high throughout the first term of its implementation. The approach of screening using swabs pooled at the time of sample collection is a valid strategy for mass screening, using a fraction of the tests required in individual screening. I have demonstrated the efficacy of this screening programme in identifying and isolating asymptomatic and presymptomatic students. With these high levels of case ascertainment, I have then provided a large combined epidemiological and genomic analysis of SARS-CoV-2 transmission in a university setting. Following a limited number of viral introductions into the University, the majority of cases were linked to a single cluster that was likely to have rapidly spread across the University following multiple gatherings at a nightclub. Further transmission was associated with student accommodation and courses, with minimal evidence of transmission within departments, or between students and staff. There was little evidence of substantial transmission between the University and local community. Finally, I have shown that the combination of asymptomatic screening, other measures employed by the University to control COVID-19 and national lockdown were effective at reducing SARS-CoV-2 transmission.

Across the cohort of students participating in the screening programme, the proportion of truly asymptomatic SARS-CoV-2 infections (20%) was comparable with other studies[168,169], but possibly lower than that predicted in the age range of university students. Direct comparisons are challenging due to methodological differences, and this study is limited by the retrospective ascertainment of symptoms, including subjective changes in smell or taste

which have been reported to be more frequent in young adults and those with mild disease[439,440]. The proportion of presymptomatic individuals identified by screening who go onto develop symptoms of COVID-19 (25%) is consistent with the observed reduction in likelihood of symptomatic disease in the week after a negative screening result (51%), as well as the observed presymptomatic infectious period (3.6 days). CT values in presymptomatic students were low, comparable to those in symptomatic students, which is consistent with the known peak in SARS-CoV-2 viral load at or shortly before symptom onset[433]. This is in keeping with the observation that presymptomatic and symptomatic index cases were associated with higher secondary attack rates among household contacts, in comparison to index cases who remained asymptomatic. This suggests that screening programmes should be designed to target individuals with presymptomatic infection, for example by minimising the testing interval.

Evaluating the impact of the programme in comparison to other universities is not possible, due to a variety of differences in setting, case ascertainment and local infection prevention and control measures. However, in this chapter I have provided direct evidence for the efficacy of asymptomatic screening at reducing transmission and case numbers within the University. Asymptomatic screening led to improved case ascertainment and isolation of asymptomatic and, critically, presymptomatic individuals (totalling 39% of all cases in participating students). Additionally, a number of symptomatic students were pre-emptively isolated at the time of their diagnosis, as contacts of SARS-CoV-2 positive students identified through asymptomatic screening (7% of all cases). I have shown a reduction in the risk of symptomatic COVID-19 in the week following a negative screening result (likely corresponding to the successful detection of presymptomatic cases). Finally, combined genomic and epidemiological evidence has shown that asymptomatic screening could be

used for the prompt identification of outbreaks. In combination with other measures, it likely contributed to the cessation of transmission within Colleges once outbreaks had been identified, through improved case ascertainment and isolation of cases and contacts.

Additional evidence for the benefit of frequent asymptomatic screening has been provided by modelling analyses, parameterised using real-world data from this study[441,442]. These have estimated that weekly screening of all students using PCR resulted in a 31% reduction in the basic reproduction number (R0) from a median of 1.78 (95% CIs 1.37-2.23) to a median of 1.22 (95% CIs 0.82-1.62). By screening half the students from each testing pool each week (as implemented during weeks 3-7 of the programme), there was an intermediate reduction (median R0 1.45, 95% CIs – 1.02-1.88). A frequently voiced concern among students during the course of this study was that increased testing would lead to a rise in the requirement for isolation. However, these modelling studies have demonstrated that the number of students quarantined was reduced, rather than increased, by more frequent screening.

The pooled sample screening approach utilised in this study could be applied to other settings. This programme was deployed in 8 weeks from inception to screening, without any pre-existing clinical, technical or logistics infrastructure. Swab pooling in groups of up to 10 individuals was possible using standard flocked nasopharyngeal swabs and viral transport medium tubes. In addition to facilitating procurement of consumables, this also enables laboratories to process standardised samples without changes to their workflows. By pooling swabs and performing secondary individual confirmatory testing on samples, this approach has increased efficiency in testing three to five fold, up to SARS-CoV-2 prevalence of 2%. Because the testing pool in this study correlated with the unit of isolation, there were limited numbers of people asked to quarantine as a result of a positive pooled result (in a setting where household contacts of positive cases were required to isolate). This means that swab

pooling is well suited to mass testing in residential settings, such as universities, long-term care facilities, prisons, military bases or private households. Where household contacts are not required to self-isolate, this approach may be applied in additional setting where individuals are grouped together, such as schools or workplaces[443].

This study has demonstrated that, despite voluntary involvement and the use of nose and throat swabs rather than less invasive sampling approaches (such as saliva), rates of student participation were high across the collegiate University throughout the term. There were a variety of factors that may have enhanced participation, including the convenience of testing in student accommodation (as opposed to centralised testing locations), regular and open communication strategies, multidisciplinary involvement in the programme oversight committee (including representatives from University departments with expertise in legal, logistic, clinical and information technology issues, as well as student representation) and the collegiate structure of the University, which was utilised to support various aspects of programme delivery and the management of positive cases/contacts. During its first term of operation, the programme was appraised as part of an independent ethical review into asymptomatic COVID-19 screening programmes in HEIs, which was broadly supportive of many of the approaches taken in the development, operation and evaluation of the programme[444].

Community-based mass testing programmes for COVID-19 have been a subject of contention[445,446]. Some of the major criticisms have centred on the rates of false positive and false negative results (leading either to periods of unnecessary isolation or missed cases, respectively)[228,229]. Using the two-step process in this programme (initial pooled screen followed by a confirmatory individual test) substantially reduced the risk of false positive

results. The high sensitivity of PCR, relative to other screening tests such as antigen-based lateral flow tests (LFTs), reduces the risk of false negatives. Additionally, centralised laboratory testing through PCR facilitates reliable and prompt reporting to public health bodies. Residual sample can also be used for genome sequencing to support surveillance and transmission studies, as demonstrated here. By comparison, LFTs are relatively cheap, can return results in under an hour and have been deployed at scale in different settings, without the need for additional infrastructure[227,447]. They have comparable rates of false positive results to our study, which may be further mitigated by confirmatory PCR testing[237]. However, their main limitation is in reduced sensitivity: even when performed by trained individuals, they detect approximately 50% SARS-CoV-2 infections in comparison to PCR tests[448-450]. While LFTs are more accurate at higher viral loads, in this study I have shown a modest difference in secondary attack rates in household contacts of index cases with high or low CT values, suggesting either that infectiousness is not directly related to CT, or that one-off sampling may miss the peak viral load and/or infectiousness of an individual.

One of the strengths of this study was high case ascertainment. SARS-CoV-2 is typically shed from the upper respiratory tract for 2-3 weeks[451,452]. Therefore the combination of regular asymptomatic screening and accessible symptomatic testing was likely to have detected the majority of SARS-CoV-2 infections in students participating in the screening programme. A high proportion of these samples were successfully sequenced, with genomic data that could then be mapped to known University structures, including households and courses, to effectively determine transmission networks.

Over two thirds of all sequenced University cases were associated with the same genetic cluster. The introduction of this lineage into the University, likely from elsewhere in the UK, was coincident with the arrival of students. Viral spread was amplified at multiple points

during the term by students from multiple Colleges, courses and year groups mixing at the same nightclub. Previous studies have also identified similar venues as being high risk for SARS-CoV-2 transmission[453,454], demonstrating that public health interventions may need to target these venues during respiratory virus pandemics, particularly in the context of young susceptible student populations.

The epidemic curve of case numbers changed considerably over the course of the term, reflecting both changes in student behaviour and interventions that successfully reduced viral transmission. There were a limited number of viral introductions into the University, as evidenced by low lineage diversity compared to the surrounding community. This may reflect successful infection prevention and control measures, including social distancing, mask wearing and quarantine of international students (who make up approximately 35% of the student population in College accommodation) at the beginning of term.

Following the first week of term there was an initial increase in SARS-CoV-2 incidence across the University. This may be attributed to increased social mixing during Fresher's week, in venues inside and outside University and College premises. Following the identification and management of College-based transmission networks in term weeks 2 and 3, alongside a likely reduction in the number of social interactions, there was a gradual reduction in cases until term week 5. The approach to identification and management of clusters of cases in Colleges was complex and bespoke to each situation, but relied on prompt case ascertainment (through asymptomatic screening and readily available symptomatic testing in the centre of town), central oversight of case numbers by the University's COVID Incident Management Team in conjunction with public health authorities,

University and national contact tracing and multifaceted support provided by Colleges to help affected students and their contacts to isolate.

Such measures were less successful in controlling transmission in settings outside University premises. There was a sharp rise in cases coinciding with the announcement of a national lockdown on October 31 that was not implemented for a further five days. This delay between a public statement and the introduction of a major socially restrictive public health measure, alongside pre-planned events around Halloween, may have led to higher rates of behaviour associated with increased risk of transmission. This may have been mitigated by reducing the time from announcement to implementation of the lockdown, or a targeted public health campaign to limit high risk activities.

The national lockdown remained in place for the majority of the remainder of term, during which there was a rapid fall in University case numbers, at a faster rate than the surrounding community, to almost zero. This demonstrates that lockdown was effective and likely associated with high levels of compliance in our study population. Contemporaneous studies conducted elsewhere in the UK have shown that levels of adherence to measures to limit SARS-CoV-2 transmission are mixed[455]. While young age is a risk factor for poor adherence, other recognised associations (such as financial hardship, working in a key sector and having a dependent child in the household) were much lower in our University population. Although no direct incentives were provided to students in this cohort, the expectation of individuals to adhere to rules was communicated widely in both national and University media. Additional support provided by the collegiate University was also likely to have ensured high rates of adherence, ranging from practical provision of food and drink through to pastoral and community support provided by established networks of staff, tutors and student representatives.

Limited transmission was observed between the University and the local community. Lineages associated with students and the community were largely phylogenetically distinct. Additionally, there was limited epidemiological evidence from national contact tracing data of shared exposures. The only student group disproportionately represented in community groups was clinical medical students. These students represent an important epidemiological link between secondary care and the University, being at higher risk of acquisition and transmission of SARS-CoV-2, as evidenced by the transmission networks involving HCWs in **chapter 7**. Clinical medical students should therefore be prioritised for public health interventions to minimise transmission, including vaccination.

In keeping with previous analyses of pathogen transmission presented in this thesis, a combination of genomic data (with relatively high case ascertainment) and routinely collected epidemiological data (from the University and national contact tracing data) was key in understanding viral spread within the University and in refuting transmission within epidemiologically linked, but genomically distinct clusters. This combined genomic epidemiology approach is of great utility in informing outbreak investigations[456].

However, there remain a number of limitations to this approach. While this study has benefited from contextual community isolates sequenced through COG-UK[249,288], incomplete sampling (especially in tests performed via community pathways) limits interpretation of transmission networks. In comparison to other studies, active case ascertainment of asymptomatic cases should have mitigated this limitation in the University setting. The available epidemiological data also lacks fine resolution. For example, shared student courses were recognised as a risk factor for transmission, but it has not been possible to identify the setting of this transmission, such as educational or social activities. Additionally,

contact tracing data are dependent on self-reporting and therefore some epidemiological associations may have been missed. Finally, the University is unusual in its collegiate structure and potential limited integration with the community; this study setting should be considered when making any generalisation of conclusions about transmission.

Nevertheless, the combination of detailed genomic and epidemiological analyses of SARS-CoV-2 transmission and the evaluation of a large, successfully implemented mass testing programme involving asymptomatic screening, has provided enormous insight into COVID-19 in higher education settings. These experiences have been shared with SAGE, Universities UK and the government's mass testing programme[457], and informed national understanding and debate in this area[458]. Insights obtained during these studies are likely to be informative in the event of subsequent waves of the COVID-19 pandemic and for future pandemic preparedness.

Data presented in this chapter were derived in the final quarter of 2020. Since then, the UK has seen at least two successive waves of SARS-CoV-2 variants with mutations enhancing transmissibility and enabling at least partial escape from immunity[171,459]. Despite the development and rollout of SARS-CoV-2 vaccines, the evidence that they reduce viral transmission is limited[460], uptake among university-aged individuals is lower than the general population[461] and there is concern for waning immunity within months of a second vaccine dose[462]. The need for additional measures to reduce transmission is therefore likely to continue and university students remain a population at high risk of viral exposure and transmission. Regular asymptomatic mass testing has fewer social, economic and educational impacts than measures such as lockdowns and social distancing and therefore presents an effective and acceptable means to control SARS-CoV-2 transmission in universities, with potential for application to a range of other settings.

9. Conclusions and future directions

In this thesis I have used a variety of methods to investigate three major infectious disease challenges of great public health importance: the rising tide of AMR, exemplified by the bacterium *K. pneumoniae*; seasonal epidemics of influenza and their burden in secondary care; and global pandemics, recently and overwhelmingly demonstrated by the emergence and spread of SARS-CoV-2. By combining large isolate collections, phylogenetically ordered whole genome sequences, detailed epidemiological data and thorough clinical information, I have obtained considerable insights into pathogen transmission, control and response to antimicrobial therapy. Observations made during this PhD fellowship have inevitably led to additional research questions and opportunities for further work.

The high content imaging methods described here have provided detailed phenotypic studies of bacteria at scale. I have demonstrated the strengths and challenges associated with this platform, but its full potential has yet to be explored. There are opportunities to address a wider range of pathogens, staining techniques and live imaging methodologies. One of the strengths of the work presented in this thesis is the harmonisation of imaging and transcriptomic data, developing a greater understanding of the relationship between bacterial genotype and phenotype. Further work is required to make such genotyping scalable, building on promising preliminary experiments using the Fluidigm platform for high throughput differential gene expression using qPCR. In addition to the planned work in 60 isolates imaged in **Chapter 5**, there is the opportunity to study a broader range of species,

antimicrobial growth conditions and even single bacterial cell transcriptomic studies to obtain greater resolution of phenotype-genotype associations.

A key outstanding question is the optimal method of comparing large morphological and transcriptomic datasets in a statistically robust manner. While I have provided individual examples in this thesis, more complex analyses of differential gene expression of cellular and regulatory pathways (in addition to individual genes) are desirable to further understand the impact of antimicrobials on bacterial transcription.

Work to date has shown that commonly used antimicrobials cause major phenotypic and transcriptomic changes that may accelerate the development of AMR through increased rates of mutagenesis, plasmid conjugation and efflux pump expression. This raises fundamental questions about our understanding of the association between antimicrobial therapy and resistance. Further work is needed to determine whether these transcriptional changes are associated with altered bacterial behaviour, and whether they could have a significant impact on AMR in real world settings.

One of the great strengths of this thesis has been its interdisciplinary nature. Through collaborations between academic groups, and clinical and laboratory teams at CUH, I have developed pipelines for isolate identification, storage, sequencing and phenotyping, as well as large extracts of clinical metadata from the patient record. I have used this pipeline to demonstrate the burden of common infectious diseases in the hospital, complex underlying transmission networks of hospital acquired infection and interventions that may improve patient outcome. In particular, I have identified that influenza A and SARS-CoV-2 are major contributors to nosocomial infection, with high associated mortality and involvement of

hospital staff in transmission networks; however, I have also shown how multidisciplinary interventions can reduce transmission and improve clinical outcomes. During the first wave of the COVID-19 pandemic, near real-time integration of genomics, epidemiology and clinical investigations helped to develop rapid and responsive infection control measures. The research described here has impacted policy decisions at CUH and contributed to discussions regarding nosocomial transmission at a national level. This pipeline provides a framework for the detailed investigation of infectious diseases at CUH, which could be adapted for a variety of pathogens. With the ongoing burden of both acute and hospital-associated infections, these pipelines will likely have a range of clinical and academic applications in the future.

From a clinical perspective, the optimal method of combining large and complex data for routine infection control purposes requires further consideration. Advances in genomics and health data informatics make this prospect increasingly plausible. However, the relative costs and benefits of these approaches need to be evaluated, as do the challenges of ensuring data quality, robust statistical methods in analysing these datasets, and the most appropriate infrastructure and governance to ensure the greatest clinical benefit.

Access to large libraries of clinical isolates, that have been well characterised in terms of their phylogeny, genotype, phenotype and their associated clinical course of infection, are of great value in research studies of genomically diverse species such as *K. pneumoniae*. Capturing this diversity is important for *in vitro* research into the effect of existing antimicrobials, as demonstrated here, as well as the evaluation of diagnostics, novel therapeutics and other strategies to tackle AMR[12,80].

Finally, I have applied these approaches to study SARS-CoV-2 transmission in a university, demonstrating that methods developed for hospital settings can also be applied to large community populations. This study has, again, benefited from access to routinely collected epidemiological data and close to real-time genomic sequencing. Utilising this approach demonstrates the possibilities for studying infectious disease transmission in high risk or high impact community settings (such as universities, schools, prisons and care homes) that can be used to understand pathogen spread in these cohorts and develop measures by which to interrupt it. One such approach is regular mass testing using pooled samples, which I have demonstrated is a viable screening approach at scale, and effectively identifies asymptomatic and presymptomatic infections thereby reducing transmission from affected individuals.

Despite the many advances made in modern medicine, major human pathogens have shown time and again the burden they place on global health. There is a persistent threat of novel species or lineages with epidemic or pandemic potential that will necessitate innovative, interdisciplinary approaches to their management and control. I sincerely hope that this thesis will provide some foundation for future work to address challenges posed by these pathogens, in the ongoing fight against infectious diseases.

References

1 Gerberding, J. L. & Haynes, B. F. Vaccine Innovations - Past and Future. *N Engl J Med* **384**, 393-396, doi:10.1056/NEJMp2029466 (2021).

2 Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* **51**, 72-80, doi:10.1016/j.mib.2019.10.008 (2019).

3 Gardy, J. L. & Loman, N. J. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* **19**, 9-20, doi:10.1038/nrg.2017.88 (2018).

4 Simonsen, L., Gog, J. R., Olson, D. & Viboud, C. Infectious Disease Surveillance in the Big Data Era: Towards Faster and Locally Relevant Systems. *J Infect Dis* **214**, S380-S385, doi:10.1093/infdis/jiw376 (2016).

5 Fauci, A. S. & Morens, D. M. The perpetual challenge of infectious diseases. *N Engl J Med* **366**, 454-461, doi:10.1056/NEJMra1108296 (2012).

6 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204-1222, doi:10.1016/S0140-6736(20)30925-9 (2020).

7 Baker, R. E. *et al.* Infectious disease in an era of global change. *Nat Rev Microbiol*, doi:10.1038/s41579-021-00639-z (2021).

8 World Health Organization. *Ten threats to global health in 2019*, Available at: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 (2019).

9 World Health Organization. *WHO Coronavirus (COVID-19) Dashboard*, Available at: https://covid19.who.int/ (2021).

10 Iredell, J., Brown, J. & Tagg, K. Antibiotic resistance in Enterobacteriaceae: mechanisms and clinical implications. *BMJ* **352**, h6420, doi:10.1136/bmj.h6420 (2016).

11 UN News. *Antimicrobial resistance a 'global health emergency,' UN, ahead of awareness week*, Available at: https://news.un.org/en/node/1025511/antimicrobial-resistance-a-global-health-emergency-un-ahead-of-awareness-week-2 (2018).

12 The Review on Antimicrobial Resistance, O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations.* HM Government (2016).

13 Dadgostar, P. Antimicrobial Resistance: Implications and Costs. *Infect Drug Resist* **12**, 3903-3910, doi:10.2147/IDR.S234610 (2019).

14 Jonas, O. B., Irwin, A., Berthe, F. C. J., Le Gall, F. G. & Marquez, P. V. *Drug-resistant infections : a threat to our economic future (Vol. 2) : final report (English)*, 2017).

15 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**, 56-66, doi:10.1016/S1473-3099(18)30605-4 (2019).

16 Rosello, A. *et al.* Impact of long-term care facility residence on the antibiotic resistance of urinary tract Escherichia coli and Klebsiella. *J Antimicrob Chemother* **72**, 1184-1192, doi:10.1093/jac/dkw555 (2017).

17 Kizny Gordon, A. E. *et al.* The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature. *Clin Infect Dis* **64**, 1435-1444, doi:10.1093/cid/cix132 (2017).

18 Peleg, A. Y. & Hooper, D. C. Hospital-acquired infections due to gram-negative bacteria. *N Engl J Med* **362**, 1804-1813, doi:10.1056/NEJMra0904124 (2010).

19 Tamma, P. D. *et al.* Comparing the Outcomes of Patients With Carbapenemase-Producing and Non-Carbapenemase-Producing Carbapenem-Resistant Enterobacteriaceae Bacteremia. *Clin Infect Dis* **64**, 257-264, doi:10.1093/cid/ciw741 (2017).

20 Knight, G. M. *et al.* Quantifying where human acquisition of antibiotic resistance occurs: a mathematical modelling study. *BMC Med* **16**, 137, doi:10.1186/s12916-018-1121-8 (2018).

21 Coll, F. *et al.* Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aak9745 (2017).

22 Van Boeckel, T. P. *et al.* Global trends in antimicrobial resistance in animals in low- and middle-income countries. *Science* **365**, doi:10.1126/science.aaw1944 (2019).

23 Woolhouse, M., Ward, M., van Bunnik, B. & Farrar, J. Antimicrobial resistance in humans, livestock and the wider environment. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140083, doi:10.1098/rstb.2014.0083 (2015).

24 Chaves, J. *et al.* SHV-1 beta-lactamase is mainly a chromosomally encoded species-specific enzyme in Klebsiella pneumoniae. *Antimicrob Agents Chemother* **45**, 2856-2861, doi:10.1128/AAC.45.10.2856-2861.2001 (2001).

25 Nordmann, P., Cuzon, G. & Naas, T. The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. *Lancet Infect Dis* **9**, 228-236, doi:10.1016/S1473-3099(09)70054-4 (2009).

26 Nordmann, P., Poirel, L., Walsh, T. R. & Livermore, D. M. The emerging NDM carbapenemases. *Trends Microbiol* **19**, 588-595, doi:10.1016/j.tim.2011.09.005 (2011).

27 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, doi:10.1016/S1473-3099(15)00424-7 (2016).

28 Snitkin, E. S. *et al.* Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* **4**, 148ra116, doi:10.1126/scitranslmed.3004129 (2012).

29 Oteo, J. *et al.* The spread of KPC-producing Enterobacteriaceae in Spain: WGS analysis of the emerging high-risk clones of Klebsiella pneumoniae ST11/KPC-2, ST101/KPC-2 and ST512/KPC-3. *J Antimicrob Chemother* **71**, 3392-3399, doi:10.1093/jac/dkw321 (2016).

30 Stoesser, N. *et al.* Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. *Antimicrob Agents Chemother* **58**, 7347-7357, doi:10.1128/AAC.03900-14 (2014).

31 Brodrick, H. T., E. in *Applied Bioinformatics and Public Health Microbiology* (Hinxton, Cambridgeshire, UK, 2017).

32 Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis* **18**, 318-327, doi:10.1016/S1473-3099(17)30753-3 (2018).

33 Moradigaravand, D., Martin, V., Peacock, S. J. & Parkhill, J. Population structure of multidrug resistant Klebsiella oxytoca within hospitals across the UK and Ireland identifies sharing of virulence and resistance genes with K. pneumoniae. *Genome Biol Evol* **9**, 574-587, doi:10.1093/gbe/evx019 (2017).

34 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**, E3574-3581, doi:10.1073/pnas.1501049112 (2015).

35 Wyres, K. L. & Holt, K. E. Klebsiella pneumoniae Population Genomics and Antimicrobial-Resistant Clones. *Trends Microbiol* **24**, 944-956, doi:10.1016/j.tim.2016.09.007 (2016).

36 Rodriguez-Medina, N., Barrios-Camacho, H., Duran-Bedolla, J. & Garza-Ramos, U. Klebsiella variicola: an emerging pathogen in humans. *Emerg Microbes Infect* **8**, 973-988, doi:10.1080/22221751.2019.1634981 (2019).

37 Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A. & Brisse, S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. *J Clin Microbiol* **43**, 4178-4182, doi:10.1128/JCM.43.8.4178-4182.2005 (2005).

38 Bialek-Davenet, S. *et al.* Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. *Emerg Infect Dis* **20**, 1812-1820, doi:10.3201/eid2011.140206 (2014).

39 Moradigaravand, D., Martin, V., Peacock, S. J. & Parkhill, J. Evolution and Epidemiology of Multidrug-Resistant Klebsiella pneumoniae in the United Kingdom and Ireland. *MBio* **8**, doi:10.1128/mBio.01976-16 (2017).

40 Woodford, N., Turton, J. F. & Livermore, D. M. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. *FEMS Microbiol Rev* **35**, 736-755, doi:10.1111/j.1574-6976.2011.00268.x (2011).

41 Holden, M. T. *et al.* A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Res* **23**, 653-664, doi:10.1101/gr.147710.112 (2013).

42 Lam, M. M. C. *et al.* A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. *Nat Commun* **12**, 4188, doi:10.1038/s41467-021-24448-3 (2021).

43 Follador, R. *et al.* The diversity of Klebsiella pneumoniae surface polysaccharides. *Microb Genom* **2**, e000073, doi:10.1099/mgen.0.000073 (2016).

44 Wyres, K. L. *et al.* Identification of Klebsiella capsule synthesis loci from whole genome data. *Microb Genom* **2**, e000102, doi:10.1099/mgen.0.000102 (2016).

45 Chen, L. *et al.* Carbapenemase-producing Klebsiella pneumoniae: molecular and genetic decoding. *Trends Microbiol* **22**, 686-696, doi:10.1016/j.tim.2014.09.003 (2014).

46 Podschun, R. & Ullmann, U. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev* **11**, 589-603 (1998).

47 Fuxench-Lopez, Z. & Ramirez-Ronda, C. H. Pharyngeal flora in ambulatory alcoholic patients: prevalence of gram-negative bacilli. *Arch Intern Med* **138**, 1815-1816, doi:10.1001/archinte.1978.03630370033017 (1978).

48 Lin, Y. T. *et al.* Seroepidemiology of Klebsiella pneumoniae colonizing the intestinal tract of healthy Chinese and overseas Chinese adults in Asian countries. *BMC Microbiol* **12**, 13, doi:10.1186/1471-2180-12-13 (2012).

49 Feldman, C. *et al.* Klebsiella pneumoniae bacteraemia at an urban general hospital. *J Infect* **20**, 21-31, doi:10.1016/s0163-4453(90)92258-m (1990).

50 Hansen, D. S., Gottschau, A. & Kolmos, H. J. Epidemiology of Klebsiella bacteraemia: a case control study using Escherichia coli bacteraemia as control. *J Hosp Infect* **38**, 119-132, doi:10.1016/s0195-6701(98)90065-2 (1998).

51 Lin, Y. T., Liu, C. J., Yeh, Y. C., Chen, T. J. & Fung, C. P. Ampicillin and amoxicillin use and the risk of Klebsiella pneumoniae liver abscess in Taiwan. *J Infect Dis* **208**, 211-217, doi:10.1093/infdis/jit157 (2013).

52 Russo, T. A. & Marr, C. M. Hypervirulent Klebsiella pneumoniae. *Clin Microbiol Rev* **32**, doi:10.1128/CMR.00001-19 (2019).

53 Lee, I. R. *et al.* Differential host susceptibility and bacterial virulence factors driving Klebsiella liver abscess in an ethnically diverse population. *Sci Rep* **6**, 29316, doi:10.1038/srep29316 (2016).

54 Lin, J. C. *et al.* High prevalence of phagocytic-resistant capsular serotypes of Klebsiella pneumoniae in liver abscess. *Microbes Infect* **6**, 1191-1198, doi:10.1016/j.micinf.2004.06.003 (2004).

55 Fang, C. T., Chuang, Y. P., Shun, C. T., Chang, S. C. & Wang, J. T. A novel virulence gene in Klebsiella pneumoniae strains causing primary liver abscess and septic metastatic complications. *J Exp Med* **199**, 697-705, doi:10.1084/jem.20030857 (2004).

56 Shankar-Sinha, S. *et al.* The Klebsiella pneumoniae O antigen contributes to bacteremia and lethality during murine pneumonia. *Infect Immun* **72**, 1423-1430, doi:10.1128/IAI.72.3.1423-1430.2004 (2004).

57 Holden, V. I., Breen, P., Houle, S., Dozois, C. M. & Bachman, M. A. Klebsiella pneumoniae Siderophores Induce Inflammation, Bacterial Dissemination, and HIF-1alpha Stabilization during Pneumonia. *mBio* **7**, doi:10.1128/mBio.01397-16 (2016).

58 Sebghati, T. A., Korhonen, T. K., Hornick, D. B. & Clegg, S. Characterization of the type 3 fimbrial adhesins of Klebsiella strains. *Infect Immun* **66**, 2887-2894, doi:10.1128/IAI.66.6.2887-2894.1998 (1998).

59 Jagnow, J. & Clegg, S. Klebsiella pneumoniae MrkD-mediated biofilm formation on extracellular matrix- and collagen-coated surfaces. *Microbiology (Reading)* **149**, 2397-2405, doi:10.1099/mic.0.26434-0 (2003).

60 Lam, M. M. C. *et al.* Convergence of virulence and MDR in a single plasmid vector in MDR Klebsiella pneumoniae ST15. *J Antimicrob Chemother*, doi:10.1093/jac/dkz028 (2019).

61 Banerjee, T., Wangkheimayum, J., Sharma, S., Kumar, A. & Bhattacharjee, A. Extensively Drug-Resistant Hypervirulent Klebsiella pneumoniae From a Series of Neonatal Sepsis in a Tertiary Care Hospital, India. *Front Med (Lausanne)* **8**, 645955, doi:10.3389/fmed.2021.645955 (2021).

62 Wyres, K. L. *et al.* Genomic surveillance for hypervirulence and multi-drug resistance in invasive Klebsiella pneumoniae from South and Southeast Asia. *Genome Med* **12**, 11, doi:10.1186/s13073-019-0706-y (2020).

63 Lan, P., Jiang, Y., Zhou, J. & Yu, Y. A global perspective on the convergence of hypervirulence and carbapenem resistance in Klebsiella pneumoniae. *J Glob Antimicrob Resist* **25**, 26-34, doi:10.1016/j.jgar.2021.02.020 (2021).

64 Gomez-Simmonds, A. & Uhlemann, A. C. Clinical Implications of Genomic Adaptation and Evolution of Carbapenem-Resistant Klebsiella pneumoniae. *J Infect Dis* **215**, S18-S27, doi:10.1093/infdis/jiw378 (2017).

65 Public Health England. *Annual epidemiological commentary: Gram-negative bacteraemia, MRSA bacteraemia, MSSA bacteraemia and C. difficile infections, up to and including financial year April 2020 to March 2021* (2021).

66 Public Health England. *Laboratory surveillance of Klebsiella spp. bacteraemia in England, Wales and Northern Ireland: 2018*, (2020).

67 David, S. *et al.* Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. *Nat Microbiol* **4**, 1919-1929, doi:10.1038/s41564-019-0492-8 (2019).

68 Temkin, E. *et al.* Estimating the number of infections caused by antibiotic-resistant Escherichia coli and Klebsiella pneumoniae in 2014: a modelling study. *Lancet Glob Health* **6**, e969-e979, doi:10.1016/S2214-109X(18)30278-X (2018).

69 Huang, W. *et al.* In-hospital Medical Costs of Infections Caused by Carbapenem-resistant Klebsiella pneumoniae. *Clin Infect Dis* **67**, S225-S230, doi:10.1093/cid/ciy642 (2018).

70 Santos, W. M. D. & Secoli, S. R. Economic burden of inpatients infected with Klebsiella pneumoniae carbapenemase. *Einstein (Sao Paulo)* **17**, eGS4444, doi:10.31744/einstein_journal/2019GS4444 (2019).

71 Zarkotou, O. *et al.* Predictors of mortality in patients with bloodstream infections caused by KPC-producing Klebsiella pneumoniae and impact of appropriate antimicrobial treatment. *Clin Microbiol Infect* **17**, 1798-1803, doi:10.1111/j.1469-0691.2011.03514.x (2011).

72 Pham, M. H. *et al.* Evidence of widespread endemic populations of highly multidrug-resistant *Klebsiella pneumoniae* seen concurrently through the lens of two hospital intensive care units in Vietnam. *medRxiv*, 2021.2006.2029.21259521, doi:10.1101/2021.06.29.21259521 (2021).

73 Wyres, K. L. & Holt, K. E. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr Opin Microbiol* **45**, 131-139, doi:10.1016/j.mib.2018.04.004 (2018).

74 Long, S. W. *et al.* Population Genomic Analysis of 1,777 Extended-Spectrum Beta-Lactamase-Producing Klebsiella pneumoniae Isolates, Houston, Texas: Unexpected Abundance of Clonal Group 307. *MBio* **8**, doi:10.1128/mBio.00489-17 (2017).

75 Doi, Y. Treatment Options for Carbapenem-resistant Gram-negative Bacterial Infections. *Clin Infect Dis* **69**, S565-S575, doi:10.1093/cid/ciz830 (2019).

76 Hooper, D. C. & Jacoby, G. A. Mechanisms of drug resistance: quinolone resistance. *Ann N Y Acad Sci* **1354**, 12-31, doi:10.1111/nyas.12830 (2015).

77 Correia, S., Poeta, P., Hebraud, M., Capelo, J. L. & Igrejas, G. Mechanisms of quinolone action and resistance: where do we stand? *J Med Microbiol* **66**, 551-559, doi:10.1099/jmm.0.000475 (2017).

78 Wong, V. K. *et al.* Phylogeographical analysis of the dominant multidrug-resistant H58 clade of Salmonella Typhi identifies inter- and intracontinental transmission events. *Nat Genet* **47**, 632-639, doi:10.1038/ng.3281 (2015).

79 He, M. *et al.* Emergence and global spread of epidemic healthcare-associated Clostridium difficile. *Nat Genet* **45**, 109-113, doi:10.1038/ng.2478 (2013).

80 HM Government. *UK 5-year action plan for antimicrobial resistance 2019 to 2024. 2019. Available at: https://www.gov.uk/government/publications/uk-5-year-action-plan-for-antimicrobial-resistance-2019-to-2024*. (2019)

81 Joint Programming Initiative on Antimicrobial Resistance. *Strategic Research and Innovation Agenda on Antimicrobial Resistance*, Available at: https://www.jpiamr.eu/about/sria/ (2021).

82 Argimón, S. *et al.* Rapid Genomic Characterization and Global Surveillance of *Klebsiella* Using Pathogenwatch. *bioRxiv*, 2021.2006.2022.448967, doi:10.1101/2021.06.22.448967 (2021).

83 Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis* **13**, 130-136, doi:10.1016/S1473-3099(12)70268-2 (2013).

84 Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. *J Antimicrob Chemother* **68**, 2234-2244, doi:10.1093/jac/dkt180 (2013).

85 Weigel, L. M., Steward, C. D. & Tenover, F. C. gyrA mutations associated with fluoroquinolone resistance in eight species of Enterobacteriaceae. *Antimicrob Agents Chemother* **42**, 2661-2667 (1998).

86 Wright, M. S. *et al.* Genomic and transcriptomic analyses of colistin-resistant clinical isolates of Klebsiella pneumoniae reveal multiple pathways of resistance. *Antimicrob Agents Chemother* **59**, 536-543, doi:10.1128/AAC.04037-14 (2015).

87 Cain, A. K. *et al.* Morphological, genomic and transcriptomic responses of Klebsiella pneumoniae to the last-line antibiotic colistin. *Sci Rep* **8**, 9868, doi:10.1038/s41598-018-28199-y (2018).

88 Hall, C. W., Zhang, L. & Mah, T. F. PA3225 Is a Transcriptional Repressor of Antibiotic Resistance Mechanisms in Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **61**, doi:10.1128/AAC.02114-16 (2017).

89 Li, L. *et al.* RNA-seq-based analysis of drug-resistant Salmonella enterica serovar Typhimurium selected in vivo and in vitro. *PLoS One* **12**, e0175234, doi:10.1371/journal.pone.0175234 (2017).

90 Sinel, C. *et al.* Subinhibitory Concentrations of Ciprofloxacin Enhance Antimicrobial Resistance and Pathogenicity of Enterococcus faecium. *Antimicrob Agents Chemother* **61**, doi:10.1128/AAC.02763-16 (2017).

91 Jensen, P. A., Zhu, Z. & van Opijnen, T. Antibiotics Disrupt Coordination between Transcriptional and Phenotypic Stress Responses in Pathogenic Bacteria. *Cell Rep* **20**, 1705-1716, doi:10.1016/j.celrep.2017.07.062 (2017).

92 Jana, B. *et al.* The secondary resistome of multidrug-resistant Klebsiella pneumoniae. *Sci Rep* **7**, 42483, doi:10.1038/srep42483 (2017).

93 Earle, S. G. *et al.* Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol* **1**, 16041, doi:10.1038/nmicrobiol.2016.41 (2016).

94 Nguyen, M. *et al.* Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. *Sci Rep* **8**, 421, doi:10.1038/s41598-017-18972-w (2018).

95 Brauner, A., Fridman, O., Gefen, O. & Balaban, N. Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat Rev Microbiol* **14**, 320-330, doi:10.1038/nrmicro.2016.34 (2016).

96 Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nature methods* **14**, 849 (2017).

97 Ang, M. L. & Pethe, K. Contribution of high-content imaging technologies to the development of anti-infective drugs. *Cytometry A* **89**, 755-760, doi:10.1002/cyto.a.22885 (2016).

98 van Vliet, E. *et al.* Current approaches and future role of high content imaging in safety sciences and drug discovery. *ALTEX* **31**, 479-493, doi:10.14573/altex.1405271 (2014).

99 Bray, M.-A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nature protocols* **11**, 1757 (2016).

100 Clare, R. H. *et al.* Development and validation of a high-throughput anti-Wolbachia whole-cell screen: a route to macrofilaricidal drugs against onchocerciasis and lymphatic filariasis. *J Biomol Screen* **20**, 64-69, doi:10.1177/1087057114551518 (2015).

101 Christophe, T. *et al.* High content screening identifies decaprenyl-phosphoribose 2' epimerase as a target for intracellular antimycobacterial inhibitors. *PLoS Pathog* **5**, e1000645, doi:10.1371/journal.ppat.1000645 (2009).

102 Mellouk, N. *et al.* Shigella subverts the host recycling compartment to rupture its vacuole. *Cell Host Microbe* **16**, 517-530, doi:10.1016/j.chom.2014.09.005 (2014).

103 Nonejuie, P., Burkart, M., Pogliano, K. & Pogliano, J. Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. *Proc Natl Acad Sci U S A* **110**, 16169-16174, doi:10.1073/pnas.1311066110 (2013).

104 Zoffmann, S. *et al.* Machine learning-powered antibiotics phenotypic drug discovery. *Sci Rep* **9**, 5013, doi:10.1038/s41598-019-39387-9 (2019).

105 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**, 12414, doi:10.1038/s41598-020-69300-8 (2020).

106 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, doi:10.1038/ja.2015.116 (2016).

107 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, doi:10.1016/j.ebiom.2016.01.020 (2016).

108 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**, doi:10.1128/AAC.02310-18 (2019).

109 Booth, H. D. E. *et al.* RNA sequencing reveals MMP2 and TGFB1 downregulation in LRRK2 G2019S Parkinson's iPSC-derived astrocytes. *Neurobiol Dis* **129**, 56-66, doi:10.1016/j.nbd.2019.05.006 (2019).

110 Sun, Y. *et al.* Potent anti-tumor efficacy of palbociclib in treatment-naive H3.3K27M-mutant diffuse intrinsic pontine glioma. *EBioMedicine* **43**, 171-179, doi:10.1016/j.ebiom.2019.04.043 (2019).

111 Acharya, D. *et al.* Shape dependent physical mutilation and lethal effects of silver nanoparticles on bacteria. *Sci Rep* **8**, 201, doi:10.1038/s41598-017-18590-6 (2018).

112 Van Laar, T. A., Chen, T., You, T. & Leung, K. P. Sublethal concentrations of carbapenems alter cell morphology and genomic expression of Klebsiella pneumoniae biofilms. *Antimicrob Agents Chemother* **59**, 1707-1717, doi:10.1128/AAC.04581-14 (2015).

113 Nelson, M. I. & Holmes, E. C. The evolution of epidemic influenza. *Nat Rev Genet* **8**, 196-205, doi:10.1038/nrg2053 (2007).

114 Lafond, K. E. *et al.* Global Role and Burden of Influenza in Pediatric Respiratory Hospitalizations, 1982-2012: A Systematic Analysis. *PLoS Med* **13**, e1001977, doi:10.1371/journal.pmed.1001977 (2016).

115 Lam, T. T. & Pybus, O. G. Genomic surveillance of avian-origin influenza A viruses causing human disease. *Genome Med* **10**, 50, doi:10.1186/s13073-018-0560-3 (2018).

116 Public Health England. *Surveillance of influenza and other respiratory viruses in the UK: Winter 2017 to 2018*, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/740606/Surveillance_of_influenza_and_other_respiratory_viruses_in_the_UK_2017_to_2018.pdf (2018).

117 Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol Rev* **56**, 152-179, doi:10.1128/mr.56.1.152-179.1992 (1992).

118 Ghebrehewet, S., MacPherson, P. & Ho, A. Influenza. *BMJ* **355**, i6258, doi:10.1136/bmj.i6258 (2016).

119 Piroth, L. *et al.* Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: a nationwide, population-based retrospective cohort study. *Lancet Respir Med* **9**, 251-259, doi:10.1016/S2213-2600(20)30527-0 (2021).

120 Mertz, D. *et al.* Populations at risk for severe or complicated influenza illness: systematic review and meta-analysis. *BMJ* **347**, f5061, doi:10.1136/bmj.f5061 (2013).

121 Jester, B. J., Uyeki, T. M. & Jernigan, D. B. Fifty Years of Influenza A(H3N2) Following the Pandemic of 1968. *Am J Public Health* **110**, 669-676, doi:10.2105/AJPH.2019.305557 (2020).

122 Simonsen, L., Fukuda, K., Schonberger, L. B. & Cox, N. J. The impact of influenza epidemics on hospitalizations. *J Infect Dis* **181**, 831-837, doi:10.1086/315320 (2000).

123 Taubenberger, J. K., Reid, A. H., Janczewski, T. A. & Fanning, T. G. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philos Trans R Soc Lond B Biol Sci* **356**, 1829-1839, doi:10.1098/rstb.2001.1020 (2001).

124 Ferguson, N. M., Fraser, C., Donnelly, C. A., Ghani, A. C. & Anderson, R. M. Public health risk from the avian H5N1 influenza epidemic. *Science* **304**, 968-969, doi:10.1126/science.1096898 (2004).

125 Lafond, K. E. *et al.* Global burden of influenza-associated lower respiratory tract infections and hospitalizations among adults: A systematic review and meta-analysis. *PLoS Med* **18**, e1003550, doi:10.1371/journal.pmed.1003550 (2021).

126 Paget, J. *et al.* Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *J Glob Health* **9**, 020421, doi:10.7189/jogh.09.020421 (2019).

127 Iuliano, A. D. *et al.* Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* **391**, 1285-1300, doi:10.1016/S0140-6736(17)33293-2 (2018).

128 Woodman, R. What caused the winter crisis in the NHS? *BMJ* **318**, 145, doi:10.1136/bmj.318.7177.145 (1999).

129 Iacobucci, G. Winter pressure: high demand, pension crisis, and flu could be "perfect storm" for NHS, warns BMA. *BMJ* **367**, l6399, doi:10.1136/bmj.l6399 (2019).

130 Mahase, E. Winter pressure: RSV, flu, and covid-19 could push NHS to breaking point, report warns. *BMJ* **374**, n1802, doi:10.1136/bmj.n1802 (2021).

131 Hsu, J. *et al.* Antivirals for treatment of influenza: a systematic review and meta-analysis of observational studies. *Ann Intern Med* **156**, 512-524, doi:10.7326/0003-4819-156-7-201204030-00411 (2012).

132 Muthuri, S. G. *et al.* Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: a meta-analysis of individual participant data. *Lancet Respir Med* **2**, 395-404, doi:10.1016/S2213-2600(14)70041-4 (2014).

133 Mitchell, R. *et al.* Understanding the burden of influenza infection among adults in Canadian hospitals: a comparison of the 2009-2010 pandemic season with the prepandemic and postpandemic seasons. *Am J Infect Control* **41**, 1032-1037, doi:10.1016/j.ajic.2013.06.008 (2013).

134 Myles, P. R. *et al.* Predictors of clinical outcome in a national hospitalised cohort across both waves of the influenza A/H1N1 pandemic 2009-2010 in the UK. *Thorax* **67**, 709-717, doi:10.1136/thoraxjnl-2011-200266 (2012).

135 National Institute for Health and Care Excellence. *Amantadine, oseltamivir and zanamivir for the treatment of influenza. [Technology appraisal guidance TA168],* (2009).

136 Centers for Disease Control and Prevention. *Influenza antiviral medications: summary for clinicians.* , Available at: https://www.cdc.gov/flu/professionals/antivirals/summary-clinicians.htm (Accessed November 1 2021)

137 Uyeki, T. M. *et al.* Clinical Practice Guidelines by the Infectious Diseases Society of America: 2018 Update on Diagnosis, Treatment, Chemoprophylaxis, and Institutional Outbreak Management of Seasonal Influenzaa. *Clin Infect Dis* **68**, 895-902, doi:10.1093/cid/ciy874 (2019).

138 World Health Organization. *Guidelines for Pharmacological Management of Pandemic Influenza A(H1N1) 2009 and other Influenza Viruses* Available at: http://www.who.int/csr/resources/publications/swineflu/h1n1_guidelines_pharmaceutical_mngt.pdf (Accessed November 1 2021)

139 Hurt, A. C. & Kelly, H. Debate Regarding Oseltamivir Use for Seasonal and Pandemic Influenza. *Emerg Infect Dis* **22**, 949-955, doi:10.3201/eid2206.151037 (2016).

140 Dyer, O. What did we learn from Tamiflu? *BMJ* **368**, m626, doi:10.1136/bmj.m626 (2020).

141 Hernan, M. A. & Lipsitch, M. Oseltamivir and risk of lower respiratory tract complications in patients with flu symptoms: a meta-analysis of eleven randomized clinical trials. *Clin Infect Dis* **53**, 277-279, doi:10.1093/cid/cir400 (2011).

142 Dobson, J., Whitley, R. J., Pocock, S. & Monto, A. S. Oseltamivir treatment for influenza in adults: a meta-analysis of randomised controlled trials. *Lancet* **385**, 1729-1737, doi:10.1016/S0140-6736(14)62449-1 (2015).

143 Chaves, S. S. *et al.* Patients hospitalized with laboratory-confirmed influenza during the 2010-2011 influenza season: exploring disease severity by virus type and subtype. *J Infect Dis* **208**, 1305-1314, doi:10.1093/infdis/jit316 (2013).

144 Lee, N. *et al.* Viral loads and duration of viral shedding in adult patients hospitalized with influenza. *J Infect Dis* **200**, 492-500, doi:10.1086/600383 (2009).

145 Viasus, D. *et al.* Timing of oseltamivir administration and outcomes in hospitalized adults with pandemic 2009 influenza A(H1N1) virus infection. *Chest* **140**, 1025-1032, doi:10.1378/chest.10-2792 (2011).

146 Salgado, C. D., Farr, B. M., Hall, K. K. & Hayden, F. G. Influenza in the acute hospital setting. *Lancet Infect Dis* **2**, 145-155, doi:10.1016/s1473-3099(02)00221-9 (2002).

147 Pollara, C. P. *et al.* Nosocomial outbreak of the pandemic Influenza A (H1N1) 2009 in critical hematologic patients during seasonal influenza 2010-2011: detection of oseltamivir resistant variant viruses. *BMC Infect Dis* **13**, 127, doi:10.1186/1471-2334-13-127 (2013).

148 Eibach, D. *et al.* Routes of transmission during a nosocomial influenza A(H3N2) outbreak among geriatric patients and healthcare workers. *J Hosp Infect* **86**, 188-193, doi:10.1016/j.jhin.2013.11.009 (2014).

149 Cunney, R. J., Bialachowski, A., Thornley, D., Smaill, F. M. & Pennie, R. A. An outbreak of influenza A in a neonatal intensive care unit. *Infect Control Hosp Epidemiol* **21**, 449-454, doi:10.1086/501786 (2000).

150 Houlihan, C. F. *et al.* Use of Whole-Genome Sequencing in the Investigation of a Nosocomial Influenza Virus Outbreak. *J Infect Dis* **218**, 1485-1489, doi:10.1093/infdis/jiy335 (2018).

151 Sansone, M. *et al.* Molecular characterization of a nosocomial outbreak of influenza B virus in an acute care hospital setting. *J Hosp Infect* **101**, 30-37, doi:10.1016/j.jhin.2018.06.004 (2019).

152 Oguma, T. *et al.* Molecular characteristics of outbreaks of nosocomial infection with influenza A/H3N2 virus variants. *Infect Control Hosp Epidemiol* **32**, 267-275, doi:10.1086/658671 (2011).

153 Rodriguez-Sanchez, B. *et al.* Genotyping of a nosocomial outbreak of pandemic influenza A/H1N1 2009. *J Clin Virol* **52**, 129-132, doi:10.1016/j.jcv.2011.07.001 (2011).

154 Valley-Omar, Z., Nindo, F., Mudau, M., Hsiao, M. & Martin, D. P. Phylogenetic Exploration of Nosocomial Transmission Chains of 2009 Influenza A/H1N1 among Children Admitted at Red Cross War Memorial Children's Hospital, Cape Town, South Africa in 2011. *PLoS One* **10**, e0141744, doi:10.1371/journal.pone.0141744 (2015).

155 Houghton, R., Ellis, J., Galiano, M., Clark, T. W. & Wyllie, S. Haemagglutinin and neuraminidase sequencing delineate nosocomial influenza outbreaks with accuracy equivalent to whole genome sequencing. *J Infect* **74**, 377-384, doi:10.1016/j.jinf.2016.12.015 (2017).

156 Seong, M. W. *et al.* Genotyping Influenza Virus by Next-Generation Deep Sequencing in Clinical Specimens. *Ann Lab Med* **36**, 255-258, doi:10.3343/alm.2016.36.3.255 (2016).

157 Jonges, M., Rahamat-Langendoen, J., Meijer, A., Niesters, H. G. & Koopmans, M. Sequence-based identification and characterization of nosocomial influenza A(H1N1)pdm09 virus infections. *J Hosp Infect* **82**, 187-193, doi:10.1016/j.jhin.2012.08.004 (2012).

158 Chen, L. F. *et al.* Cluster of oseltamivir-resistant 2009 pandemic influenza A (H1N1) virus infections on a hospital ward among immunocompromised patients--North Carolina, 2009. *J Infect Dis* **203**, 838-846, doi:10.1093/infdis/jiq124 (2011).

159 Vanhems, P., Benet, T. & Munier-Marion, E. Nosocomial influenza: encouraging insights and future challenges. *Curr Opin Infect Dis* **29**, 366-372, doi:10.1097/QCO.0000000000000287 (2016).

160 Peiris, J. S., Yuen, K. Y., Osterhaus, A. D. & Stohr, K. The severe acute respiratory syndrome. *N Engl J Med* **349**, 2431-2441, doi:10.1056/NEJMra032498 (2003).

161 Zumla, A., Hui, D. S. & Perlman, S. Middle East respiratory syndrome. *Lancet* **386**, 995-1007, doi:10.1016/S0140-6736(15)60454-8 (2015).

162 Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat Rev Microbiol* **19**, 685-700, doi:10.1038/s41579-021-00630-8 (2021).

163 Fung, T. S. & Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* **73**, 529-557, doi:10.1146/annurev-micro-020518-115759 (2019).

164 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269, doi:10.1038/s41586-020-2008-3 (2020).

165 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403-1407, doi:10.1038/s41564-020-0770-5 (2020).

166 World Health Organization. *Tracking SARS-CoV-2 variants*, Available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (Accessed November 1 2021)

167 Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* **382**, 1199-1207, doi:10.1056/NEJMoa2001316 (2020).

168 Johansson, M. A. *et al.* SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. *JAMA Network Open* **4**, e2035057-e2035057, doi:10.1001/jamanetworkopen.2020.35057 (2021).

169 Buitrago-Garcia, D. *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med* **17**, e1003346, doi:10.1371/journal.pmed.1003346 (2020).

170 Belouzard, S., Millet, J. K., Licitra, B. N. & Whittaker, G. R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4**, 1011-1033, doi:10.3390/v4061011 (2012).

171 Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* **19**, 409-424, doi:10.1038/s41579-021-00573-0 (2021).

172 Docherty, A. B. *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* **369**, m1985, doi:10.1136/bmj.m1985 (2020).

173 Gupta, A. *et al.* Extrapulmonary manifestations of COVID-19. *Nat Med* **26**, 1017-1032, doi:10.1038/s41591-020-0968-3 (2020).

174 Oran, D. P. & Topol, E. J. The Proportion of SARS-CoV-2 Infections That Are Asymptomatic : A Systematic Review. *Ann Intern Med* **174**, 655-662, doi:10.7326/M20-6976 (2021).

175 Antonelli, M. *et al.* Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *Lancet Infect Dis*, doi:10.1016/S1473-3099(21)00460-6 (2021).

176 Ruan, Q., Yang, K., Wang, W., Jiang, L. & Song, J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* **46**, 846-848, doi:10.1007/s00134-020-05991-x (2020).

177 Levin, A. T. *et al.* Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol* **35**, 1123-1138, doi:10.1007/s10654-020-00698-1 (2020).

178 Williamson, E. J. *et al.* Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430-436, doi:10.1038/s41586-020-2521-4 (2020).

179 Lutchmansingh, D. D. *et al.* A Clinic Blueprint for Post-Coronavirus Disease 2019 RECOVERY: Learning From the Past, Looking to the Future. *Chest* **159**, 949-958, doi:10.1016/j.chest.2020.10.067 (2021).

180 Crook, H., Raza, S., Nowell, J., Young, M. & Edison, P. Long covid-mechanisms, risk factors, and management. *BMJ* **374**, n1648, doi:10.1136/bmj.n1648 (2021).

181 HM Government. *UK Coronavirus Dashboard*, Available at: https://coronavirus.data.gov.uk/ (Accessed October 21 2021).

182 NHS Engand. *Bed Availability and Occupancy Data – Overnight*, Available at: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/ (Accessed October 21 2021)

183 Mateen, B. A. *et al.* Hospital bed capacity and usage across secondary healthcare providers in England during the first wave of the COVID-19 pandemic: a descriptive analysis. *BMJ Open* **11**, e042945, doi:10.1136/bmjopen-2020-042945 (2021).

184 Department of Health and Social Care (DHSC) and Office for National Statistics (ONS). *Direct and Indirect health impacts of COVID-19 in England. 2021.*, Available at: https://www.gov.uk/government/publications/dhsc-direct-and-indirect-health-impacts-of-covid-19-in-england-long-paper-9-september-2021 (2021)

185 Chen, S. *et al.* The early impact of COVID-19 on mental health and community physical health services and their patients' mortality in Cambridgeshire and Peterborough, UK. *J Psychiatr Res* **131**, 244-254, doi:10.1016/j.jpsychires.2020.09.020 (2020).

186 Neal;, K. The Collateral Damage of COVID-19. *J Public Health (Oxf)* **42**, 659, doi:10.1093/pubmed/fdaa208 (2020).

187 Editors. Clinical research: further COVID-19 collateral damage. *Lancet Gastroenterol Hepatol* **5**, 875, doi:10.1016/S2468-1253(20)30274-0 (2020).

188 Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061-1069, doi:10.1001/jama.2020.1585 (2020).

189 Read, J. M. *et al.* Hospital-acquired SARS-CoV-2 infection in the UK's first COVID-19 pandemic wave. *Lancet* **398**, 1037-1038, doi:10.1016/S0140-6736(21)01786-4 (2021).

190 Carter, B. *et al.* Nosocomial COVID-19 infection: examining the risk of mortality. The COPE-Nosocomial Study (COVID in Older PEople). *J Hosp Infect* **106**, 376-384, doi:10.1016/j.jhin.2020.07.013 (2020).

191 Rickman, H. M. *et al.* Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-acquired Cases in a London Teaching Hospital. *Clin Infect Dis* **72**, 690-693, doi:10.1093/cid/ciaa816 (2021).

192 Sikkema, R. S. *et al.* COVID-19 in health-care workers in three hospitals in the south of the Netherlands: a cross-sectional study. *Lancet Infect Dis* **20**, 1273-1280, doi:10.1016/S1473-3099(20)30527-2 (2020).

193 Islam, M. S. *et al.* Current knowledge of COVID-19 and infection prevention and control strategies in healthcare settings: A global analysis. *Infect Control Hosp Epidemiol* **41**, 1196-1206, doi:10.1017/ice.2020.237 (2020).

194 Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261-1276, doi:10.1038/nprot.2017.066 (2017).

195 Arias, A. *et al.* Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evol* **2**, vew016, doi:10.1093/ve/vew016 (2016).

196 Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* **530**, 228-232, doi:10.1038/nature16996 (2016).

197 Duchene, S. *et al.* Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol* **6**, veaa061, doi:10.1093/ve/veaa061 (2020).

198 Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat Med* **26**, 450-452, doi:10.1038/s41591-020-0820-9 (2020).

199 Basta, N. E. *et al.* Immunogenicity of a Meningococcal B Vaccine during a University Outbreak. *N Engl J Med* **375**, 220-228, doi:10.1056/NEJMoa1514866 (2016).

200 Mandal, S. *et al.* Prolonged university outbreak of meningococcal disease associated with a serogroup B strain rarely seen in the United States. *Clin Infect Dis* **57**, 344-348, doi:10.1093/cid/cit243 (2013).

201 Donahue, M. *et al.* Notes from the Field: Complications of Mumps During a University Outbreak Among Students Who Had Received 2 Doses of Measles-Mumps-Rubella Vaccine - Iowa, July 2015-May 2016. *MMWR Morb Mortal Wkly Rep* **66**, 390-391, doi:10.15585/mmwr.mm6614a4 (2017).

202 Smith, J. *et al.* Public health response to a measles outbreak on a university campus in Australia, 2015. *Epidemiol Infect* **146**, 314-318, doi:10.1017/S0950268817003089 (2018).

203 Vusirikala, A. *et al.* Seroprevalence of SARS-CoV-2 antibodies in university students: Cross-sectional study, December 2020, England. *J Infect* **83**, 104-111, doi:10.1016/j.jinf.2021.04.028 (2021).

204 Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine* **26**, 1205-1211, doi:10.1038/s41591-020-0962-9 (2020).

205 Poletti, P. *et al.* Association of Age With Likelihood of Developing Symptoms and Critical Disease Among Close Contacts Exposed to Patients With Confirmed SARS-CoV-2 Infection in Italy. *JAMA Netw Open* **4**, e211085, doi:10.1001/jamanetworkopen.2021.1085 (2021).

206 Wilson, E. *et al.* Multiple COVID-19 Clusters on a University Campus - North Carolina, August 2020. *MMWR Morb Mortal Wkly Rep* **69**, 1416-1418, doi:10.15585/mmwr.mm6939e3 (2020).

207 Fox, M. D., Bailey, D. C., Seamon, M. D. & Miranda, M. L. Response to a COVID-19 Outbreak on a University Campus - Indiana, August 2020. *MMWR Morb Mortal Wkly Rep* **70**, 118-122, doi:10.15585/mmwr.mm7004a3 (2021).

208 Vang, K. E. *et al.* Participation in Fraternity and Sorority Activities and the Spread of COVID-19 Among Residential University Communities - Arkansas, August 21-September 5, 2020. *MMWR Morb Mortal Wkly Rep* **70**, 20-23, doi:10.15585/mmwr.mm7001a5 (2021).

209 Hamer, D. H. *et al.* Control of COVID-19 transmission on an urban university campus during a second wave of the pandemic. *medRxiv*, 2021.2002.2023.21252319, doi:10.1101/2021.02.23.21252319 (2021).

210 Yamey, G. & Walensky, R. P. Covid-19: re-opening universities is high risk. *BMJ* **370**, m3365, doi:10.1136/bmj.m3365 (2020).

211 Sahu, P. Closure of Universities Due to Coronavirus Disease 2019 (COVID-19): Impact on Education and Mental Health of Students and Academic Staff. *Cureus* **12**, e7541, doi:10.7759/cureus.7541 (2020).

212 Task and Finish Group on Higher Education/Further Education. *Principles for managing SARS-CoV-2 transmission associated with higher education, 3 September 2020*, Available at: https://www.gov.uk/government/publications/principles-for-managing-sars-cov-2-transmission-associated-with-higher-education-3-september-2020 (2020)

213 Benneyan, J., Gehrke, C., Ilies, I. & Nehls, N. Community and Campus COVID-19 Risk Uncertainty Under University Reopening Scenarios: Model-Based Analysis. *JMIR Public Health Surveill* **7**, e24292, doi:10.2196/24292 (2021).

214 Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, doi:10.1126/science.abb6936 (2020).

215 Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* **8**, e488-e496, doi:10.1016/S2214-109X(20)30074-7 (2020).

216 Kretzschmar, M. E. *et al.* Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health* **5**, e452-e459 (2020).

217 Paltiel, A. D., Zheng, A. & Walensky, R. P. Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. *JAMA Netw Open* **3**, e2016818, doi:10.1001/jamanetworkopen.2020.16818 (2020).

218 Peto, J. *et al.* Weekly COVID-19 testing with household quarantine and contact tracing is feasible and would probably end the epidemic. *Royal Society Open Science* **7**, 200915 (2020).

219 Grassly, N. C. *et al.* Comparison of molecular testing strategies for COVID-19 control: a mathematical modelling study. *Lancet Infect Dis* **20**, 1381-1389, doi:10.1016/S1473-3099(20)30630-7 (2020).

220 Rivett, L. *et al.* Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *Elife* **9**, doi:10.7554/eLife.58728 (2020).

221 World Health Organization. *Infection prevention and control guidance for long-term care facilities in the context of COVID-19: interim guidance, 21 March 2020*, Available at: https://apps.who.int/iris/handle/10665/331508 (2020).

222 Centers for Disease Control and Prevention. *Interim infection prevention and control recommendations for healthcare personnel during the coronavirus disease 2019 (COVID-19) pandemic*, Available at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control.html (2020).

223 Jones, N. K. *et al.* Effective control of SARS-CoV-2 transmission between healthcare workers during a period of diminished community prevalence of COVID-19. *Elife* **9**, doi:10.7554/eLife.59391 (2020).

224 Hagan, L. M. *et al.* Mass Testing for SARS-CoV-2 in 16 Prisons and Jails - Six Jurisdictions, United States, April-May 2020. *MMWR Morb Mortal Wkly Rep* **69**, 1139-1143, doi:10.15585/mmwr.mm6933a3 (2020).

225 Kissler, S. M. *et al. Viral dynamics of SARS-CoV-2 infection and the predictive value of repeat testing.* , Available at: https://www.medrxiv.org/content/medrxiv/early/2020/10/23/2020.10.21.20217042.full.pdf (2020).

226 Torjesen, I. What do we know about lateral flow tests and mass testing in schools? *BMJ* **372:n706** (2021).

227 Iacobucci, G. Covid-19: Government rolls out twice weekly rapid testing to all in England. *BMJ* **doi: 10.1136/bmj.n902** (2021).

228 Raffle, A. E., Pollock, A. M. & Harding-Edgar, L. Covid-19 mass testing programmes. *BMJ* **370**, m3262, doi:10.1136/bmj.m3262 (2020).

229 Gill, M. & Gray, M. Mass testing for covid-19 in the UK. *BMJ* **371**, m4436, doi:10.1136/bmj.m4436 (2020).

230 Denny, T. N. *et al.* Implementation of a Pooled Surveillance Testing Program for Asymptomatic SARS-CoV-2 Infections on a College Campus - Duke University, Durham, North Carolina, August 2-October 11, 2020. *MMWR Morb Mortal Wkly Rep* **69**, 1743-1747, doi:10.15585/mmwr.mm6946e1 (2020).

231 Gillam, T. B. *et al.* Norwich COVID-19 testing initiative pilot: evaluating the feasibility of asymptomatic testing on a university campus. *J Public Health (Oxf)* **43**, 82-88, doi:10.1093/pubmed/fdaa194 (2021).

232 Blake, H. *et al.* Perceptions and Experiences of the University of Nottingham Pilot SARS-CoV-2 Asymptomatic Testing Service: A Mixed-Methods Study. *Int J Environ Res Public Health* **18**, doi:10.3390/ijerph18010188 (2020).

233 Rennert, L. *et al.* Surveillance-based informative testing for detection and containment of SARS-CoV-2 outbreaks on a public university campus: an

observational and modelling study. *Lancet Child Adolesc Health*, doi:10.1016/S2352-4642(21)00060-2 (2021).

234 Hemani, G. *et al.* Modelling pooling strategies for SARS-CoV-2 testing in a university setting. *Wellcome Open Research* **6**, 70 (2021).

235 Hill, E. M., Atkins, B. D., Keeling, M. J., Tildesley, M. J. & Dyson, L. Modelling SARS-CoV-2 transmission in a UK university setting. *Epidemics* **36**, 100476, doi:10.1016/j.epidem.2021.100476 (2021).

236 Ford, L. *et al.* Epidemiologic characteristics associated with SARS-CoV-2 antigen-based test results, rRT-PCR cycle threshold values, subgenomic RNA, and viral culture results from university testing. *Clin Infect Dis*, doi:10.1093/cid/ciab303 (2021).

237 Pray, I. W. *et al.* Performance of an Antigen-Based Test for Asymptomatic and Symptomatic SARS-CoV-2 Testing at Two University Campuses - Wisconsin, September-October 2020. *MMWR Morb Mortal Wkly Rep* **69**, 1642-1647, doi:10.15585/mmwr.mm695152a3 (2021).

238 Deeks, J. J. & Raffle, A. E. Lateral flow tests cannot rule out SARS-CoV-2 infection. *BMJ* **371**, m4787, doi:10.1136/bmj.m4787 (2020).

239 Dinnes, J. *et al.* Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev* **3**, CD013705, doi:10.1002/14651858.CD013705.pub2 (2021).

240 Barak, N. *et al.* Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci Transl Med* **13**, doi:10.1126/scitranslmed.abf2823 (2021).

241 Ben-Ami, R. *et al.* Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2 detection. *Clinical Microbiology and Infection* **26**, 1248-1253 (2020).

242 Gupta, E. *et al.* Pooled RNA sample reverse transcriptase real time PCR assay for SARS CoV-2 infection: A reliable, faster and economical method. *Plos one* **15**, e0236859 (2020).

243 Hogan, C. A., Sahoo, M. K. & Pinsky, B. A. Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. *JAMA* **323**, 1967-1969, doi:10.1001/jama.2020.5445 (2020).

244 Lohse, S. *et al.* Pooling of samples for testing for SARS-CoV-2 in asymptomatic people. *Lancet Infect Dis* **20**, 1231-1232, doi:10.1016/S1473-3099(20)30362-5 (2020).

245 Perchetti, G. A. *et al.* Pooling of SARS-CoV-2 samples to increase molecular testing throughput. *J Clin Virol* **131**, 104570, doi:10.1016/j.jcv.2020.104570 (2020).

246 Christoff, A. P. *et al.* Swab pooling: A new method for large-scale RT-qPCR screening of SARS-CoV-2 avoiding sample dilution. *PLoS One* **16**, e0246544, doi:10.1371/journal.pone.0246544 (2021).

247 Thanh, T. T. *et al.* The Application of Sample Pooling for Mass Screening of SARS-CoV-2 in an Outbreak of COVID-19 in Vietnam. *Am J Trop Med Hyg*, doi:10.4269/ajtmh.20-1583 (2021).

248 Chen, F. *et al.* Comparing two sample pooling strategies for SARS-CoV-2 RNA detection for efficient screening of COVID-19. *J Med Virol* **93**, 2805-2809, doi:10.1002/jmv.26632 (2021).

249 COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**, e99-e100, doi:10.1016/S2666-5247(20)30054-9 (2020).

250 Care Quality Commission. *Cambridge University Hospitals NHS Foundation Trust inspection report*, Available at: https://www.cqc.org.uk/sites/default/files/20190214%20Cambridge%20University%20Hospitals%20NHSFT%20-%20Evidence%20Appendix%20FINAL%20for%20publication.pdf (2019).

251 Andrews, J. M., Howe, R. A. & Testing, B. W. P. o. S. BSAC standardized disc susceptibility testing method (version 10). *J Antimicrob Chemother* **66**, 2726-2757, doi:10.1093/jac/dkr359 (2011).

252 Brown, D. F., Wootton, M. & Howe, R. A. Antimicrobial susceptibility testing breakpoints and methods from BSAC to EUCAST. *J Antimicrob Chemother* **71**, 3-5, doi:10.1093/jac/dkv287 (2016).

253 The European Committee on Antimicrobial Susceptibility Testing. *Breakpoint tables for interpretation of MICs and zone diameters. Version 11.0*, Available at: https://eucast.org/clinical_breakpoints/ (2021).

254 Novak-Weekley, S. M. *et al.* Evaluation of the Cepheid Xpert Flu Assay for rapid identification and differentiation of influenza A, influenza A 2009 H1N1, and influenza B viruses. *J Clin Microbiol* **50**, 1704-1710, doi:10.1128/JCM.06520-11 (2012).

255 Collier, D. A. *et al.* Point of Care Nucleic Acid Testing for SARS-CoV-2 in Hospitalized Patients: A Clinical Validation Trial and Implementation Study. *Cell Rep Med* **1**, 100062, doi:10.1016/j.xcrm.2020.100062 (2020).

256 Clark, T. W. *et al.* Adults hospitalised with acute respiratory illness rarely have detectable bacteria in the absence of COPD or pneumonia; viral infection

predominates in a large prospective UK sample. *J Infect* **69**, 507-515, doi:10.1016/j.jinf.2014.07.023 (2014).

257 Meredith, L. W. *et al.* Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis* **20**, 1263-1271, doi:10.1016/S1473-3099(20)30562-4 (2020).

258 Sridhar, S. *et al.* A blueprint for the implementation of a validated approach for the detection of SARS-Cov2 in clinical samples in academic facilities. *Wellcome Open Res* **5**, 110, doi:10.12688/wellcomeopenres.15937.2 (2020).

259 Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* **40**, 373-383, doi:10.1016/0021-9681(87)90171-8 (1987).

260 Hosmer, D. W. & Lemeshow, S. Goodness of fit tests for the multiple logistic regression model. *Communs Statist* **9**, 1043-1069 (1980).

261 Hosmer, D. W., Lemeshow, S. & Sturdivant, R. X. *Applied logistic regression*. (Wiley New York, 2000).

262 Jorgensen, J. H. & Ferraro, M. J. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. *Clin Infect Dis* **49**, 1749-1755, doi:10.1086/647952 (2009).

263 CLSI. Methods for Determining Bactericidal Activity of Antimicrobial Agents. Approved Guideline, CLSI document M26-A. (Clinical and Laboratory Standards Institute, 950 West Valley Road, Suite 2500, Wayne, Pennsylvania 19087, USA, 1999).

264 Zhou, X. Y. *et al.* In vitro characterization and inhibition of the interaction between ciprofloxacin and berberine against multidrug-resistant Klebsiella pneumoniae. *J Antibiot (Tokyo)* **69**, 741-746, doi:10.1038/ja.2016.15 (2016).

265 Heery, D. M., Powell, R., Gannon, F. & Dunican, L. K. Curing of a plasmid from E.coli using high-voltage electroporation. *Nucleic Acids Res* **17**, 10131, doi:10.1093/nar/17.23.10131 (1989).

266 Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67-73, doi:10.1177/108705719900400206 (1999).

267 R Core Team. *R: A language and environment for statistical computing*, Available at: https://www.R-project.org/ (2021).

268 Conway-Morris, A. *et al.* The removal of airborne SARS-CoV-2 and other microbial bioaerosols by air filtration on COVID-19 surge units. *medRxiv*, 2021.2009.2016.21263684, doi:10.1101/2021.09.16.21263684 (2021).

269 Sanchez-Freire, V., Ebert, A. D., Kalisky, T., Quake, S. R. & Wu, J. C. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. *Nat Protoc* **7**, 829-838, doi:10.1038/nprot.2012.021 (2012).

270 Page, A. J. *et al.* Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom* **2**, e000083, doi:10.1099/mgen.0.000083 (2016).

271 Gladman, S. & Seemann, T. *Velvet Optimiser: For automatically optimising the primary parameter options for the Velvet de novo sequence assembler* Available at: http://bioinformatics.net.au/software.velvetoptimiser.shtml (2008).

272 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829, doi:10.1101/gr.074492.107 (2008).

273 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579, doi:10.1093/bioinformatics/btq683 (2011).

274 Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol* **13**, R56, doi:10.1186/gb-2012-13-6-r56 (2012).

275 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).

276 Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res* **40**, D130-135, doi:10.1093/nar/gkr1079 (2012).

277 Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**, R46, doi:10.1186/gb-2014-15-3-r46 (2014).

278 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).

279 Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *Plos Comput Biol* **13**, doi:ARTN e100559510.1371/journal.pcbi.1005595 (2017).

280 Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* **20**, 129, doi:10.1186/s13059-019-1727-y (2019).

281 Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* **3**, e000132, doi:10.1099/mgen.0.000132 (2017).

282 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

283 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).

284 Hunt, M. *et al.* Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* **16**, doi:ARTN 29410.1186/s13059-015-0849-0 (2015).

285 Zhou, B. *et al.* Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. *J Virol* **83**, 10309-10313, doi:10.1128/JVI.01109-09 (2009).

286 Quick, J. *ARTIC amplicon sequencing protocol for MinION for nCoV-2019*, Available at: https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w (

287 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).

288 Nicholls, S. M. *et al.* CLIMB-COVID: continuous integration supporting decentralised sequencing for SARS-CoV-2 genomic surveillance. *Genome Biol* **22**, 196, doi:10.1186/s13059-021-02395-y (2021).

289 Connor, T. R. *et al.* CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* **2**, e000086, doi:10.1099/mgen.0.000086 (2016).

290 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).

291 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534, doi:10.1093/molbev/msaa015 (2020).

292 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296, doi:10.1093/nar/gkab301 (2021).

293 Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* **6**, 90, doi:10.1186/s13073-014-0090-6 (2014).

294 Carattoli, A. *et al.* In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* **58**, 3895-3903, doi:10.1128/AAC.02412-14 (2014).

295 Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350-3352, doi:10.1093/bioinformatics/btv383 (2015).

296 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

297 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

298 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).

299 Pathogen Informatics, Wellcome Sanger Institute. Bio-RNASeq: A toolkit for RNA-Seq analysis. Available at: https://github.com/sanger-pathogens/Bio-RNASeq

300 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

301 Pathogen Informatics, Wellcome Sanger Institute. Deago - R package for generating HTML reports from differential expression analyis and GO term enrichment. Available at: https://rdrr.io/github/sanger-pathogens/deago/

302 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).

303 Pathogen Informatics, Wellcome Sanger Institute. VR Codebase. Available at: https://github.com/sanger-pathogens/vr-codebase

304 Pathogen Informatics, Wellcome Sanger Institute. Circlator: A softare toolkit to circularize genome assemblies. Available at: https://sanger-pathogens.github.io/circlator/

305 Centers for Disease Control and Prevention. *How Flu Spreads*, Available at: https://www.cdc.gov/flu/about/disease/spread.htm (Accessed November 1 2021)

306 Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* **26**, 1569-1571, doi:10.1093/bioinformatics/btq228 (2010).

307 Rambaut, A. *et al.* The genomic and epidemiological dynamics of human influenza A virus. *Nature* **453**, 615-619, doi:10.1038/nature06945 (2008).

308 Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90-95 (2007).

309 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

310 Argimon, S. *et al.* Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* **2**, e000093, doi:10.1099/mgen.0.000093 (2016).

311 Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28-36 (2017).

312 O'Toole, A. *et al.* Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol* **7**, veab064, doi:10.1093/ve/veab064 (2021).

313 Ellington, M. J. *et al.* Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. *Genome Biol* **20**, 184, doi:10.1186/s13059-019-1785-1 (2019).

314 Yi, H., Cho, Y. J., Yong, D. & Chun, J. Genome sequence of Escherichia coli J53, a reference strain for genetic studies. *J Bacteriol* **194**, 3742-3743, doi:10.1128/JB.00641-12 (2012).

315 Herridge, W. P., Shibu, P., O'Shea, J., Brook, T. C. & Hoyles, L. Bacteriophages of Klebsiella spp., their diversity and potential therapeutic uses. *J Med Microbiol* **69**, 176-194, doi:10.1099/jmm.0.001141 (2020).

316 Diago-Navarro, E. *et al.* Antibody-Based Immunotherapy To Treat and Prevent Infection with Hypervirulent Klebsiella pneumoniae. *Clin Vaccine Immunol* **24**, doi:10.1128/CVI.00456-16 (2017).

317 Rollenske, T. *et al.* Cross-specificity of protective human antibodies against Klebsiella pneumoniae LPS O-antigen. *Nat Immunol* **19**, 617-624, doi:10.1038/s41590-018-0106-2 (2018).

318 Wyres, K. L. *et al.* Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307. *J Antimicrob Chemother* **74**, 577-581, doi:10.1093/jac/dky492 (2019).

319 Villa, L. *et al.* Diversity, virulence, and antimicrobial resistance of the KPC-producing Klebsiella pneumoniae ST307 clone. *Microb Genom* **3**, e000110, doi:10.1099/mgen.0.000110 (2017).

320 Haller, S. *et al.* Extensively drug-resistant Klebsiella pneumoniae ST307 outbreak, north-eastern Germany, June to October 2019. *Euro Surveill* **24**, doi:10.2807/1560-7917.ES.2019.24.50.1900734 (2019).

321 Strydom, K. A. *et al.* Klebsiella pneumoniae ST307 with OXA-181: threat of a high-risk clone and promiscuous plasmid in a resource-constrained healthcare setting. *J Antimicrob Chemother* **75**, 896-902, doi:10.1093/jac/dkz550 (2020).

322 Baek, E. H. *et al.* Successful control of an extended-spectrum beta-lactamase-producing Klebsiella pneumoniae ST307 outbreak in a neonatal intensive care unit. *BMC Infect Dis* **20**, 166, doi:10.1186/s12879-020-4889-z (2020).

323 Boonstra, M. B. *et al.* An outbreak of ST307 extended-spectrum beta-lactamase (ESBL)-producing Klebsiella pneumoniae in a rehabilitation center: An unusual source and route of transmission. *Infect Control Hosp Epidemiol* **41**, 31-36, doi:10.1017/ice.2019.304 (2020).

324 Weterings, V. *et al.* An outbreak of colistin-resistant Klebsiella pneumoniae carbapenemase-producing Klebsiella pneumoniae in the Netherlands (July to December 2013), with inter-institutional spread. *Eur J Clin Microbiol Infect Dis* **34**, 1647-1655, doi:10.1007/s10096-015-2401-2 (2015).

325 Tran, J. H. & Jacoby, G. A. Mechanism of plasmid-mediated quinolone resistance. *Proc Natl Acad Sci U S A* **99**, 5638-5642, doi:10.1073/pnas.082092899 (2002).

326 Padilla, E. *et al.* Klebsiella pneumoniae AcrAB efflux pump contributes to antimicrobial resistance and virulence. *Antimicrob Agents Chemother* **54**, 177-183, doi:10.1128/AAC.00715-09 (2010).

327 Cirz, R. T. *et al.* Inhibition of mutation and combating the evolution of antibiotic resistance. *PLoS Biol* **3**, e176, doi:10.1371/journal.pbio.0030176 (2005).

328 Long, H. *et al.* Antibiotic treatment enhances the genome-wide mutation rate of target cells. *Proc Natl Acad Sci U S A* **113**, E2498-2505, doi:10.1073/pnas.1601208113 (2016).

329 Molina-Santiago, C. *et al.* Differential transcriptional response to antibiotics by Pseudomonas putida DOT-T1E. *Environ Microbiol* **17**, 3251-3262, doi:10.1111/1462-2920.12775 (2015).

330 Garcia-Sureda, L., Juan, C., Domenech-Sanchez, A. & Alberti, S. Role of Klebsiella pneumoniae LamB Porin in antimicrobial resistance. *Antimicrob Agents Chemother* **55**, 1803-1805, doi:10.1128/AAC.01441-10 (2011).

331 Thanh Duy, P. *et al.* Commensal Escherichia coli are a reservoir for the transfer of XDR plasmids into epidemic fluoroquinolone-resistant Shigella sonnei. *Nat Microbiol* **5**, 256-264, doi:10.1038/s41564-019-0645-9 (2020).

332 Blair, J. M. & Piddock, L. J. How to Measure Export via Bacterial Multidrug Resistance Efflux Pumps. *mBio* **7**, doi:10.1128/mBio.00840-16 (2016).

333 Zahller, J. & Stewart, P. S. Transmission electron microscopic study of antibiotic action on Klebsiella pneumoniae biofilm. *Antimicrob Agents Chemother* **46**, 2679-2683, doi:10.1128/aac.46.8.2679-2683.2002 (2002).

334 Horii, T., Kobayashi, M., Sato, K., Ichiyama, S. & Ohta, M. An in-vitro study of carbapenem-induced morphological changes and endotoxin release in clinical isolates of gram-negative bacilli. *J Antimicrob Chemother* **41**, 435-442, doi:10.1093/jac/41.4.435 (1998).

335 Zimmerman, S. B. Shape and compaction of Escherichia coli nucleoids. *J Struct Biol* **156**, 255-261, doi:10.1016/j.jsb.2006.03.022 (2006).

336 Di Martino, P., Cafferini, N., Joly, B. & Darfeuille-Michaud, A. Klebsiella pneumoniae type 3 pili facilitate adherence and biofilm formation on abiotic surfaces. *Res Microbiol* **154**, 9-16, doi:10.1016/s0923-2508(02)00004-9 (2003).

337 Desai, S., Sanghrajka, K. & Gajjar, D. High Adhesion and Increased Cell Death Contribute to Strong Biofilm Formation in Klebsiella pneumoniae. *Pathogens* **8**, doi:10.3390/pathogens8040277 (2019).

338 Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100, doi:10.1186/gb-2006-7-10-r100 (2006).

339 Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100-106, doi:10.1038/s41592-020-01018-x (2021).

340 Williams, E. *et al.* The Image Data Resource: A Bioimage Data Integration and Publication Platform. *Nat Methods* **14**, 775-781, doi:10.1038/nmeth.4326 (2017).

341 Mattick, K. L. *et al.* Survival and filamentation of Salmonella enterica serovar enteritidis PT4 and Salmonella enterica serovar typhimurium DT104 at low water activity. *Appl Environ Microbiol* **66**, 1274-1279, doi:10.1128/aem.66.4.1274-1279.2000 (2000).

342 Tavaddod, S. & Naderi-Manesh, H. Evidence of Multi-Domain Morphological Structures in Living Escherichia coli. *Sci Rep* **7**, 5660, doi:10.1038/s41598-017-05897-7 (2017).

343 Justice, S. S. *et al.* Differentiation and developmental pathways of uropathogenic Escherichia coli in urinary tract pathogenesis. *Proc Natl Acad Sci U S A* **101**, 1333-1338, doi:10.1073/pnas.0308125100 (2004).

344 Lleo, M. M., Canepari, P. & Satta, G. Bacterial cell shape regulation: testing of additional predictions unique to the two-competing-sites model for peptidoglycan assembly and isolation of conditional rod-shaped mutants from some wild-type cocci. *J Bacteriol* **172**, 3758-3771, doi:10.1128/jb.172.7.3758-3771.1990 (1990).

345 Spratt, B. G. & Pardee, A. B. Penicillin-binding proteins and cell shape in E. coli. *Nature* **254**, 516-517, doi:10.1038/254516a0 (1975).

346 Choi, J. *et al.* A rapid antimicrobial susceptibility test based on single-cell morphological analysis. *Sci Transl Med* **6**, 267ra174, doi:10.1126/scitranslmed.3009650 (2014).

347 Zhang, Y. *et al.* Ribosome Profiling Reveals Genome-wide Cellular Translational Regulation upon Heat Stress in Escherichia coli. *Genomics Proteomics Bioinformatics* **15**, 324-330, doi:10.1016/j.gpb.2017.04.005 (2017).

348 Tsuji, A., Kaneko, Y., Takahashi, K., Ogawa, M. & Goto, S. The effects of temperature and pH on the growth of eight enteric and nine glucose non-fermenting species of gram-negative rods. *Microbiol Immunol* **26**, 15-24, doi:10.1111/j.1348-0421.1982.tb00149.x (1982).

349 Chen, D. *et al.* CusS-CusR Two-Component System Mediates Tigecycline Resistance in Carbapenem-Resistant Klebsiella pneumoniae. *Front Microbiol* **10**, 3159, doi:10.3389/fmicb.2019.03159 (2019).

350 Ramos, P. I. *et al.* The polymyxin B-induced transcriptomic response of a clinical, multidrug-resistant Klebsiella pneumoniae involves multiple regulatory elements and intracellular targets. *BMC Genomics* **17**, 737, doi:10.1186/s12864-016-3070-y (2016).

351 Navon-Venezia, S., Kondratyeva, K. & Carattoli, A. Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiol Rev* **41**, 252-275, doi:10.1093/femsre/fux013 (2017).

352 Broberg, C. A., Wu, W., Cavalcoli, J. D., Miller, V. L. & Bachman, M. A. Complete Genome Sequence of Klebsiella pneumoniae Strain ATCC 43816 KPPR1, a

Rifampin-Resistant Mutant Commonly Used in Animal, Genetic, and Molecular Biology Studies. *Genome Announc* **2**, doi:10.1128/genomeA.00924-14 (2014).

353 Clements, A. *et al.* The major surface-associated saccharides of Klebsiella pneumoniae contribute to host cell association. *PLoS One* **3**, e3817, doi:10.1371/journal.pone.0003817 (2008).

354 Jenney, A. W. *et al.* Seroepidemiology of Klebsiella pneumoniae in an Australian Tertiary Hospital and its implications for vaccine development. *J Clin Microbiol* **44**, 102-107, doi:10.1128/JCM.44.1.102-107.2006 (2006).

355 Wilksch, J. J. *et al.* MrkH, a novel c-di-GMP-dependent transcriptional activator, controls Klebsiella pneumoniae biofilm formation by regulating type 3 fimbriae expression. *PLoS Pathog* **7**, e1002204, doi:10.1371/journal.ppat.1002204 (2011).

356 Forage, R. G. & Lin, E. C. DHA system mediating aerobic and anaerobic dissimilation of glycerol in Klebsiella pneumoniae NCIB 418. *J Bacteriol* **151**, 591-599, doi:10.1128/jb.151.2.591-599.1982 (1982).

357 Wu, K. M. *et al.* Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis. *J Bacteriol* **191**, 4492-4501, doi:10.1128/JB.00315-09 (2009).

358 Ramage, B. *et al.* Comprehensive Arrayed Transposon Mutant Library of Klebsiella pneumoniae Outbreak Strain KPNIH1. *J Bacteriol* **199**, doi:10.1128/JB.00352-17 (2017).

359 Woodford, N. *et al.* Arrival of Klebsiella pneumoniae producing KPC carbapenemase in the United Kingdom. *J Antimicrob Chemother* **62**, 1261-1264, doi:10.1093/jac/dkn396 (2008).

360 Ludden, C. *et al.* A One Health study of the genetic relatedness of Klebsiella pneumoniae and their mobile elements in the East of England. *Clin Infect Dis*, doi:10.1093/cid/ciz174 (2019).

361 Roberts, L. W. *et al.* A genomic epidemiology study of multidrug-resistant Escherichia coli, Klebsiella pneumoniae and Acinetobacter baumannii in two intensive care units in Hanoi, Vietnam. *medRxiv*, 2020.2012.2009.20246397, doi:10.1101/2020.12.09.20246397 (2021).

362 Macleod, C. K. *et al.* Rapid Whole Genome Sequencing of Serotype K1 Hypervirulent Klebsiella pneumoniae from an Undocumented Chinese Migrant. *Case Rep Infect Dis* **2021**, 6638780, doi:10.1155/2021/6638780 (2021).

363 Gaiarsa, S. *et al.* Genomic epidemiology of Klebsiella pneumoniae in Italy and novel insights into the origin and global evolution of its resistance to carbapenem

antibiotics. *Antimicrob Agents Chemother* **59**, 389-396, doi:10.1128/AAC.04224-14 (2015).

364 Perez-Vazquez, M. *et al.* Phylogeny, resistome and mobile genetic elements of emergent OXA-48 and OXA-245 Klebsiella pneumoniae clones circulating in Spain. *J Antimicrob Chemother* **71**, 887-896, doi:10.1093/jac/dkv458 (2016).

365 Damjanova, I. *et al.* Expansion and countrywide dissemination of ST11, ST15 and ST147 ciprofloxacin-resistant CTX-M-15-type β-lactamase-producing Klebsiella pneumoniae epidemic clones in Hungary in 2005—the new 'MRSAs'? **62**, 978-985 (2008).

366 Zhou, K. *et al.* Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum beta-lactamase producing ST15 Klebsiella pneumoniae. *Sci Rep* **6**, 20840, doi:10.1038/srep20840 (2016).

367 Hu, L. *et al.* Emergence of blaNDM-1 among Klebsiella pneumoniae ST15 and novel ST1031 clinical isolates in China. **75**, 373-376 (2013).

368 Markovska, R. *et al.* Clonal dissemination of multilocus sequence type ST15 KPC-2-producing Klebsiella pneumoniae in Bulgaria. *APMIS* **123**, 887-894, doi:10.1111/apm.12433 (2015).

369 Brisse, S. *et al.* Virulent clones of Klebsiella pneumoniae: identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS One* **4**, e4982, doi:10.1371/journal.pone.0004982 (2009).

370 Lam, M. M. C. *et al.* Population genomics of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and rapid global dissemination. *Nat Commun* **9**, 2703, doi:10.1038/s41467-018-05114-7 (2018).

371 Wyres, K. L. *et al.* Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. *PLoS Genet* **15**, e1008114, doi:10.1371/journal.pgen.1008114 (2019).

372 Literacka, E. *et al.* Spread of Klebsiella pneumoniae ST45 Producing GES-5 Carbapenemase or GES-1 Extended-Spectrum beta-Lactamase in Newborns and Infants. *Antimicrob Agents Chemother* **64**, doi:10.1128/AAC.00595-20 (2020).

373 White, A. K. *et al.* High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci U S A* **108**, 13999-14004, doi:10.1073/pnas.1019446108 (2011).

374 Fox, B. C., Devonshire, A. S., Baradez, M. O., Marshall, D. & Foy, C. A. Comparison of reverse transcription-quantitative polymerase chain reaction methods and

platforms for single cell gene expression analysis. *Anal Biochem* **427**, 178-186, doi:10.1016/j.ab.2012.05.010 (2012).

375 Olwagen, C. P., Adrian, P. V. & Madhi, S. A. Performance of the Biomark HD real-time qPCR System (Fluidigm) for the detection of nasopharyngeal bacterial pathogens and Streptococcus pneumoniae typing. *Sci Rep* **9**, 6494, doi:10.1038/s41598-019-42846-y (2019).

376 Dreier, M., Berthoud, H., Shani, N., Wechsler, D. & Junier, P. Development of a High-Throughput Microfluidic qPCR System for the Quantitative Determination of Quality-Relevant Bacteria in Cheese. *Front Microbiol* **11**, 619166, doi:10.3389/fmicb.2020.619166 (2020).

377 Li, X. Z. & Nikaido, H. Efflux-mediated drug resistance in bacteria: an update. *Drugs* **69**, 1555-1623, doi:10.2165/11317030-000000000-00000 (2009).

378 Wand, M. E., Jamshidi, S., Bock, L. J., Rahman, K. M. & Sutton, J. M. SmvA is an important efflux pump for cationic biocides in Klebsiella pneumoniae and other Enterobacteriaceae. *Sci Rep* **9**, 1344, doi:10.1038/s41598-018-37730-0 (2019).

379 Tsai, Y. K. *et al.* Klebsiella pneumoniae outer membrane porins OmpK35 and OmpK36 play roles in both antimicrobial resistance and virulence. *Antimicrob Agents Chemother* **55**, 1485-1493, doi:10.1128/AAC.01275-10 (2011).

380 Sridhar, S. *et al.* Inhibitory Concentrations of Ciprofloxacin Induce an Adaptive Response Promoting the Intracellular Survival of Salmonella enterica Serovar Typhimurium. *mBio* **12**, e0109321, doi:10.1128/mBio.01093-21 (2021).

381 Abdul Rahim, N. *et al.* Transcriptomic responses of a New Delhi metallo-beta-lactamase-producing Klebsiella pneumoniae isolate to the combination of polymyxin B and chloramphenicol. *Int J Antimicrob Agents* **56**, 106061, doi:10.1016/j.ijantimicag.2020.106061 (2020).

382 Guilhen, C. *et al.* Transcriptional profiling of Klebsiella pneumoniae defines signatures for planktonic, sessile and biofilm-dispersed cells. *BMC Genomics* **17**, 237, doi:10.1186/s12864-016-2557-x (2016).

383 Low, Y. M. *et al.* Elucidating the survival and response of carbapenem resistant Klebsiella pneumoniae after exposure to imipenem at sub-lethal concentrations. *Pathog Glob Health* **112**, 378-386, doi:10.1080/20477724.2018.1538281 (2018).

384 Uranga, L. A., Reyes, E. D., Patidar, P. L., Redman, L. N. & Lusetti, S. L. The cohesin-like RecN protein stimulates RecA-mediated recombinational repair of DNA double-strand breaks. *Nat Commun* **8**, 15282, doi:10.1038/ncomms15282 (2017).

385 Erickson, H. P., Anderson, D. E. & Osawa, M. FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiol Mol Biol Rev* **74**, 504-528, doi:10.1128/MMBR.00021-10 (2010).

386 Yao, Z. & Carballido-Lopez, R. Fluorescence imaging for bacterial cell biology: from localization to dynamics, from ensembles to single molecules. *Annu Rev Microbiol* **68**, 459-476, doi:10.1146/annurev-micro-091213-113034 (2014).

387 Bruchmann, S. *et al.* Deep transcriptome profiling of clinical Klebsiella pneumoniae isolates reveals strain and sequence type-specific adaptation. *Environ Microbiol* **17**, 4690-4710, doi:10.1111/1462-2920.13016 (2015).

388 Zhang, Y. *et al.* Evolution of hypervirulence in carbapenem-resistant Klebsiella pneumoniae in China: a multicentre, molecular epidemiological analysis. *J Antimicrob Chemother* **75**, 327-336, doi:10.1093/jac/dkz446 (2020).

389 World Health Organization. *WHO report on surveillance of antibiotic consumption: 2016-2018 early implementation.*, Available at: https://www.who.int/medicines/areas/rational_use/oms-amr-amc-report-2016-2018/en/ (2018).

390 BBC News. *Flu outbreak shuts wards at Addenbrooke's Hospital*, Available at: https://www.bbc.co.uk/news/uk-england-cambridgeshire-38788086 (2017).

391 Huzly, D., Kurz, S., Ebner, W., Dettenkofer, M. & Panning, M. Characterisation of nosocomial and community-acquired influenza in a large university hospital during two consecutive influenza seasons. *J Clin Virol* **73**, 47-51, doi:10.1016/j.jcv.2015.10.016 (2015).

392 Taylor, G. *et al.* Healthcare-associated influenza in Canadian hospitals from 2006 to 2012. *Infect Control Hosp Epidemiol* **35**, 169-175, doi:10.1086/674858 (2014).

393 Cox, N. J. & Subbarao, K. Influenza. *Lancet* **354**, 1277-1282, doi:10.1016/S0140-6736(99)01241-6 (1999).

394 Reacher, M. *et al.* Influenza-associated mortality in hospital care: a retrospective cohort study of risk factors and impact of oseltamivir in an English teaching hospital, 2016 to 2017. *Euro Surveill* **24**, doi:10.2807/1560-7917.ES.2019.24.44.1900087 (2019).

395 Davis, S. *et al.* Diagnostic accuracy and cost analysis of the Alere i Influenza A&B near-patient test using throat swabs. *J Hosp Infect* **97**, 301-309, doi:10.1016/j.jhin.2017.05.017 (2017).

396 Pebody, R. G. *et al.* Uptake and effectiveness of influenza vaccine in those aged 65 years and older in the United Kingdom, influenza seasons 2010/11 to 2016/17. *Euro Surveill* **23**, doi:10.2807/1560-7917.ES.2018.23.39.1800092 (2018).

397 Kissling, E., Rondy, M. & team, I. M. I. M. s. Early 2016/17 vaccine effectiveness estimates against influenza A(H3N2): I-MOVE multicentre case control studies at primary care and hospital levels in Europe. *Euro Surveill* **22**, doi:10.2807/1560-7917.ES.2017.22.7.30464 (2017).

398 Qalla-Widmer, L. *et al.* Nosocomial influenza in south-western Swiss hospitals during two seasonal epidemics: an observational study. *J Hosp Infect* **109**, 115-122, doi:10.1016/j.jhin.2020.12.020 (2021).

399 Venier, A. G. Root cause analysis to support infection control in healthcare premises. *J Hosp Infect* **89**, 331-334, doi:10.1016/j.jhin.2014.12.003 (2015).

400 Inkster, T. & Cuddihy, J. Duty of candour and communication during an infection control incident in a paediatric ward of a Scottish hospital: how can we do better? *J Med Ethics*, doi:10.1136/medethics-2020-106862 (2021).

401 Blackburn, R. M. *et al.* Nosocomial transmission of influenza: A retrospective cross-sectional study using next generation sequencing at a hospital in England (2012-2014). *Influenza Other Respir Viruses* **13**, 556-563, doi:10.1111/irv.12679 (2019).

402 Wilson, K. E. *et al.* Nosocomial outbreak of influenza A H3N2 in an inpatient oncology unit related to health care workers presenting to work while ill. *Am J Infect Control* **47**, 683-687, doi:10.1016/j.ajic.2018.10.024 (2019).

403 Godoy, P. *et al.* Hospital-acquired influenza infections detected by a surveillance system over six seasons, from 2010/2011 to 2015/2016. *BMC Infect Dis* **20**, 80, doi:10.1186/s12879-020-4792-7 (2020).

404 Meinel, D. M. *et al.* Whole genome sequencing identifies influenza A H3N2 transmission and offers superior resolution to classical typing methods. *Infection* **46**, 69-76, doi:10.1007/s15010-017-1091-3 (2018).

405 MacFadden, D. R. *et al.* Use of Genome Sequencing to Define Institutional Influenza Outbreaks, Toronto, Ontario, Canada, 2014-15. *Emerg Infect Dis* **24**, 492-497, doi:10.3201/eid2403.171499 (2018).

406 Heneghan, C. J. *et al.* Neuraminidase inhibitors for influenza: a systematic review and meta-analysis of regulatory and mortality data. *Health Technol Assess* **20**, 1-242, doi:10.3310/hta20420 (2016).

407 Leekha, S. *et al.* Duration of influenza A virus shedding in hospitalized patients and implications for infection control. *Infect Control Hosp Epidemiol* **28**, 1071-1076, doi:10.1086/520101 (2007).

408 Kmietowicz, Z. Study claiming Tamiflu saved lives was based on "flawed" analysis. *BMJ* **348**, g2228, doi:10.1136/bmj.g2228 (2014).

409 Kmietowicz, Z. Critics attack chief medical officer's advice to use antivirals for flu. *BMJ* **348**, g1496, doi:10.1136/bmj.g1496 (2014).

410 Su, S. *et al.* Comparing clinical characteristics between hospitalized adults with laboratory-confirmed influenza A and B virus infection. *Clin Infect Dis* **59**, 252-255, doi:10.1093/cid/ciu269 (2014).

411 Clark, T. W. *et al.* Clinical impact of a routine, molecular, point-of-care, test-and-treat strategy for influenza in adults admitted to hospital (FluPOC): a multicentre, open-label, randomised controlled trial. *Lancet Respir Med* **9**, 419-429, doi:10.1016/S2213-2600(20)30469-0 (2021).

412 Peaper, D. R. *et al.* Clinical impact of rapid influenza PCR in the adult emergency department on patient management, ED length of stay, and nosocomial infection rate. *Influenza Other Respir Viruses* **15**, 254-261, doi:10.1111/irv.12800 (2021).

413 Smielewska, A. *et al.* Unrecognised Outbreak: Human parainfluenza virus infections in a pediatric oncology unit. A new diagnostic PCR and virus monitoring system may allow early detection of future outbreaks. *Wellcome Open Res* **3**, 119, doi:10.12688/wellcomeopenres.14732.1 (2018).

414 Bonilla-Aldana, D. K. *et al.* Coronavirus infections reported by ProMED, February 2000-January 2020. *Travel Med Infect Dis* **35**, 101575, doi:10.1016/j.tmaid.2020.101575 (2020).

415 Madoff, L. C. ProMED-mail: an early warning system for emerging diseases. *Clin Infect Dis* **39**, 227-232, doi:10.1086/422003 (2004).

416 Guan, W. J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708-1720, doi:10.1056/NEJMoa2002032 (2020).

417 Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239-1242, doi:10.1001/jama.2020.2648 (2020).

418 Public Health England. *Flu annual report: winter 2019 to 2020* (2020).

419 RECOVERY Collaborative Group *et al.* Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* **384**, 693-704, doi:10.1056/NEJMoa2021436 (2021).

420 Kraemer, M. U. G. *et al.* Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science* **373**, 889-895, doi:10.1126/science.abj0113 (2021).

421 Ramsay, I. *et al.* Investigation of healthcare-associated SARS-CoV-2 infection: Learning outcomes from an investigative process in the initial phase of the pandemic. *(In submission)* (2021).

422 Gardy, J. L. *et al.* Whole-Genome Sequencing of Measles Virus Genotypes H1 and D8 During Outbreaks of Infection Following the 2010 Olympic Winter Games Reveals Viral Transmission Routes. *J Infect Dis* **212**, 1574-1578, doi:10.1093/infdis/jiv271 (2015).

423 Koser, C. U. *et al.* Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* **366**, 2267-2275, doi:10.1056/NEJMoa1109910 (2012).

424 Hamilton, W. L. *et al.* Applying prospective genomic surveillance to support investigation of hospital-onset COVID-19. *Lancet Infect Dis* **21**, 916-917, doi:10.1016/S1473-3099(21)00251-6 (2021).

425 Illingworth, C. J. R. *et al.* A2B-COVID: A method for evaluating potential SARS-CoV-2 transmission events. *medRxiv*, 2020.2010.2026.20219642, doi:10.1101/2020.10.26.20219642 (2020).

426 Illingworth, C. J. *et al.* Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections. *Elife* **10**, doi:10.7554/eLife.67308 (2021).

427 Hamilton, W. L. *et al.* Genomic epidemiology of COVID-19 in care homes in the east of England. *Elife* **10**, doi:10.7554/eLife.64618 (2021).

428 Cooper, D. J. *et al.* A prospective study of risk factors associated with seroprevalence of SARS-CoV-2 antibodies in healthcare workers at a large UK teaching hospital. *medRxiv*, 2020.2011.2003.20220699, doi:10.1101/2020.11.03.20220699 (2020).

429 Jones, N. K. *et al.* Single-dose BNT162b2 vaccine protects against asymptomatic SARS-CoV-2 infection. *Elife* **10**, doi:10.7554/eLife.68808 (2021).

430 Evans, S. *et al.* The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. *Philos Trans R Soc Lond B Biol Sci* **376**, 20200268, doi:10.1098/rstb.2020.0268 (2021).

431 Steel, K. & Fordham, E. *Coronavirus (COVID-19) Infection Survey pilot: England, Wales and Northern Ireland, 9 October 2020, Office for National Statistics*, Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditio

nsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/englandwalesandnorthernireland9october2020 (2020).

432 Byrne, A. W. *et al.* Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open* **10**, e039856, doi:10.1136/bmjopen-2020-039856 (2020).

433 Wolfel, R. *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465-469, doi:10.1038/s41586-020-2196-x (2020).

434 Meng, B. *et al.* Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep* **35**, 109292, doi:10.1016/j.celrep.2021.109292 (2021).

435 Hodcroft, E. B. *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **595**, 707-712, doi:10.1038/s41586-021-03677-y (2021).

436 Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266-269, doi:10.1038/s41586-021-03470-x (2021).

437 Aggarwal, D. *et al.* Genomic epidemiology of SARS-CoV-2 in the University of Cambridge identifies dynamics of transmission. *Nature Communications* **(In press)** (2021).

438 Singanayagam, A. *et al.* Duration of infectiousness and correlation with RT-PCR cycle threshold values in cases of COVID-19, England, January to May 2020. *Euro Surveill* **25**, doi:10.2807/1560-7917.ES.2020.25.32.2001483 (2020).

439 Sehanobish, E. *et al.* COVID-19-Induced Anosmia and Ageusia Are Associated With Younger Age and Lower Blood Eosinophil Counts. *Am J Rhinol Allergy*, 19458924211004800, doi:10.1177/19458924211004800 (2021).

440 von Bartheld, C. S., Hagen, M. M. & Butowt, R. Prevalence of Chemosensory Dysfunction in COVID-19 Patients: A Systematic Review and Meta-analysis Reveals Significant Ethnic Differences. *ACS Chem Neurosci* **11**, 2944-2961, doi:10.1021/acschemneuro.0c00460 (2020).

441 Warne, B. *et al.* Feasibility and efficacy of mass testing for SARS-CoV-2 in a UK university using swab pooling and PCR. *Research Square*, doi:10.21203/rs.3.rs-520626/v1 (2021).

442 Enright, J. *et al.* SARS-CoV-2 infection in UK university students: Lessons from September-December 2020 and modelling insights for future student return. *INI Preprint Series* (2021). <https://www.newton.ac.uk/files/preprints/ni20004.pdf>.

443 Berke, E. M. *et al.* Pooling in a Pod: A Strategy for COVID-19 Testing to Facilitate a Safe Return to School. *Public Health Rep*, 333549211045816, doi:10.1177/00333549211045816 (2021).

444 Cox C *et al.* Developing an ethical framework for asymptomatic COVID-19 testing programmes in higher education institutions [version 1; peer review: awaiting peer review]. *Wellcome Open Res* **6** (2021).

445 Deeks, J. J., Brookes, A. J. & Pollock, A. M. Operation Moonshot proposals are scientifically unsound. *BMJ* **370**, m3699, doi:10.1136/bmj.m3699 (2020).

446 Mina, M. J., Peto, T. E., Garcia-Finana, M., Semple, M. G. & Buchan, I. E. Clarifying the evidence on SARS-CoV-2 antigen rapid tests in public health responses to COVID-19. *Lancet* **397**, 1425-1427, doi:10.1016/S0140-6736(21)00425-6 (2021).

447 Pavelka, M. *et al.* The impact of population-wide rapid antigen testing on SARS-CoV-2 prevalence in Slovakia. *Science*, doi:10.1126/science.abf9648 (2021).

448 Wolf, A., Hulmes, J. & Hopkins, S. *Lateral flow device specificity in phase 4 (post marketing) surveillance*, Available at: https://www.gov.uk/government/publications/lateral-flow-device-specificity-in-phase-4-post-marketing-surveillance (2021).

449 Scientific Advisory Group for Emergencies. *Innova Lateral Flow SARS-CoV-2 Antigen test accuracy in Liverpool Pilot: Preliminary Data, 26 November 2020*, Available at: https://www.gov.uk/government/publications/innova-lateral-flow-sars-cov-2-antigen-test-accuracy-in-liverpool-pilot-preliminary-data-26-november-2020 (2020).

450 Peto, T. COVID-19: Rapid Antigen detection for SARS-CoV-2 by lateral flow assay: a national systematic evaluation for mass-testing. *medRxiv*, 2021.2001.2013.21249563, doi:10.1101/2021.01.13.21249563 (2021).

451 He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* **26**, 672-675, doi:10.1038/s41591-020-0869-5 (2020).

452 Long, Q. X. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat Med* **26**, 1200-1204, doi:10.1038/s41591-020-0965-6 (2020).

453 Muller, N. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 Outbreak Related to a Nightclub, Germany, 2020. *Emerg Infect Dis* **27**, 645-648, doi:10.3201/eid2702.204443 (2020).

454 Choi, H., Cho, W., Kim, M. H. & Hur, J. Y. Public Health Emergency and Crisis Management: Case Study of SARS-CoV-2 Outbreak. *Int J Environ Res Public Health* **17**, doi:10.3390/ijerph17113984 (2020).

455 Smith, L. E. *et al.* Adherence to the test, trace, and isolate system in the UK: results from 37 nationally representative surveys. *BMJ* **372**, n608, doi:10.1136/bmj.n608 (2021).

456 Seemann, T. *et al.* Tracking the COVID-19 pandemic in Australia using genomics. *Nat Commun* **11**, 4376, doi:10.1038/s41467-020-18314-x (2020).

457 Scientific Advisory Group for Emergencies. *Genomic epidemiology of SARS-CoV-2 in the University of Cambridge identifies dynamics of transmission: an interim report. 2020. Available from: https://www.gov.uk/government/publications/genomic-epidemiology-of-sars-cov-2-in-the-university-of-cambridge-identifies-dynamics-of-transmission-an-interim-report-10-december-2020*.

458 Matheson, N. J., Warne, B., Weekes, M. P. & Maxwell, P. H. Mass testing of university students for covid-19. *BMJ* **375**, n2388, doi:10.1136/bmj.n2388 (2021).

459 Chia, P. Y. *et al.* Virological and serological kinetics of SARS-CoV-2 Delta variant vaccine-breakthrough infections: a multi-center cohort study. *medRxiv*, 2021.2007.2028.21261295, doi:10.1101/2021.07.28.21261295 (2021).

460 Scientific Advisory Group for Emergencies. Vaccine Effectiveness Expert Panel: Vaccine effectiveness table, 27 August 2021. Available from: https://www.gov.uk/government/publications/veep-vaccine-effectiveness-table-7-september-2021 (2021).

461 Torjesen, I. Covid-19: Vaccine centres operating at 30% capacity as young people stay away. *BMJ* **374**, n1808, doi:10.1136/bmj.n1808 (2021).

462 Tartof, S. Y. *et al.* Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. *Lancet* **398**, 1407-1416, doi:10.1016/S0140-6736(21)02183-8 (2021).

463 Rodrigues, C. *et al.* Description of Klebsiella africanensis sp. nov., Klebsiella variicola subsp. tropicalensis subsp. nov. and Klebsiella variicola subsp. variicola subsp. nov. *Res Microbiol* **170**, 165-170, doi:10.1016/j.resmic.2019.02.003 (2019).

Appendix A Outputs during clinical research fellowship

A.1 Publications

Below is a list of all publications, preprints and papers submitted to peer-reviewed journals during the course of my fellowship (*denotes joint first authorship):

Sridhar, S.*, Forrest, S.*, **Warne, B**., *et al.* (2021) High-Content Imaging to Phenotype Antimicrobial Effects on Individual Bacteria at Scale. *mSystems*. **6**(3).

Reacher, M. **Warne, B**., Reeve, L. *et al.* (2019) Influenza-associated mortality in hospital care: a retrospective cohort study of risk factors and impact of oseltamivir in an English teaching hospital, 2016 to 2017. *Euro Surveill.* **24** (44)

Meredith, L. W.*, Hamilton, W.*, **Warne, B**., *et al.* (2020) Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis.* **20**, 1263-1271

Warne, B., Enright, J., Metaxaki, M., *et al.* (2021) Efficacy of mass testing for SARS-CoV-2 in a UK university using swab pooling and PCR. *Research Square.* (Preprint) doi:10.21203/rs.3.rs-520626/v1

Aggarwal, D.*, **Warne, B**.*, Jahun, A., *et al.* (2021) Genomic epidemiology of SARS-CoV-2 in the University of Cambridge identifies dynamics of transmission. *Nat Commun.* (In press)

Matheson, N.J., **Warne, B.**, Weekes, M.P., Maxwell, P.H. (2021) Mass testing of university students for covid-19. BMJ. Oct 1;375:n2388. doi: 10.1136/bmj.n2388.

Macleod, C.K., Khokhar, F.A., **Warne, B.**, *et al.* (2021) Rapid Whole Genome Sequencing of Serotype K1 Hypervirulent *Klebsiella pneumoniae* from an Undocumented Chinese Migrant. Case Rep Infect Dis. 2021:6638780.

Collier, D.A., Assennato, S.M., **Warne, B**., *et al*. (2020) Point of Care Nucleic Acid Testing for SARS-CoV-2 in Hospitalized Patients: A Clinical Validation Trial and Implementation Study. Cell Rep Med. 2020;1(5):100062.

Hamilton, W.L.*, Fieldman, T.*, Jahun, A.*, **Warne, B.**, *et al.* (2021) Applying prospective genomic surveillance to support investigation of hospital-onset COVID-19. Lancet Infect Dis. **21**(7):916-7.

Illingworth, C.J., Hamilton, W.L., **Warne, B.**, *et al.* (2021) Superspreaders drive the largest outbreaks of hospital onset COVID-19 infections. Elife. Aug 24;10:e67308. doi: 10.7554/eLife.67308.

Cooper, D.J., Lear, S., *et al.* including **Warne, B**. (2020) A prospective study of risk factors associated with seroprevalence of SARS-CoV-2 antibodies in healthcare workers at a large UK teaching hospital. medRxiv. 2020.11.03.20220699.

Evans, S., Agnew, E., *et al.* including **Warne, B.** (2021) The impact of testing and infection prevention and control strategies on within-hospital transmission dynamics of COVID-19 in English hospitals. Philos Trans R Soc Lond B Biol Sci. 376(1829):20200268.

Hamilton, W.L., Tonkin-Hill, G., *et al.* including **Warne, B**. (2021) Genomic epidemiology of COVID-19 in care homes in the east of England. Elife. Mar 2;10:e64618. doi: 10.7554/eLife.64618.

Illingworth, C.J., Hamilton, W.L., *et al.* including **Warne, B**. (2020) A2B-COVID: A method for evaluating potential SARS-CoV-2 transmission events. medRxiv 2020.10.26.20219642.

Jones, N.K., Rivett, L., Seaman, S., Samworth, R.J., **Warne, B.**, *et al*. (2021) Single-dose BNT162b2 vaccine protects against asymptomatic SARS-CoV-2 infection. Elife. Apr 8;10:e68808. doi: 10.7554/eLife.68808.

Jones, N.K., Rivett, L., *et al.* including **Warne, B**. (2020) Effective control of SARS-CoV-2 transmission between healthcare workers during a period of diminished community prevalence of COVID-19. Elife. Jun 19;9:e59391. doi: 10.7554/eLife.59391.

Ramsay, I., Sharrocks, K., **Warne, B.**, *et al.* (2021) Investigation of healthcare-associated SARS-CoV-2 infection: Learning outcomes from an investigative process in the initial phase of the pandemic. (In submission).

Cunniffe, N., Guntera, S., *et al.* including **Warne, B.** (2020). How achievable are COVID-19 clinical trial recruitment targets? A UK observational cohort study and trials registry analysis. BMJ Open. 10(10):e044566

Forbes, H., **Warne, B.**, *et al.* (2019) Risk factors for herpes simplex virus type-1 infection and reactivation: Cross-sectional studies among EPIC-Norfolk participants. PLoS One 14(5):e0215553

Bousfield, R., Ramsay, I., **Warne, B.**, *et al*. (2021). A Retrospective Cohort Study of Bacterial Native Vertebral Osteomyelitis and its Management in the UK. Clinical Infection in Practice, 100101.

Mawer, D., Byrne, F., *et al.* including **Warne, B.** (2019) Cross-sectional study of the prevalence, causes and management of hospital-onset diarrhoea. J Hosp Infect. Oct;103(2):200-209

Reacher, M., **Warne, B.**, Verlander, N. Q., *et al.* (2021) Oseltamivir treatment of influenza B and A(H3N2) in hospitalised patients reduces mortality. (In submission)

Appendix B Supplementary imaging methods

B.1 Minimum inhibitory concentrations of isolates used in HCl optimisation

Table 1: **Minimum Inhibitory Concentrations**. MICs were determined by Etest and are presented in µg/mL. * indicates the MIC is above the highest antimicrobial concentration on the Etest; ND indicates that MIC was not determined.

Species	ID	Ampicillin	Azithromycin	Trimethoprim-sulfamethoxazole	Ciprofloxacin	Gentamicin	Rifampicin	Meropenem	Tigecycline	Cefuroxime	Oxacillin	Vancomycin
K. pneumoniae	NCTC 13438	>256*	32	>32*	>32*	6	16	0.125	0.094	>256*	ND	ND
K. pneumoniae	ATCC 43816	64	2	0.5	0.16	1.5	>32*	0.125	0.5	1.5	ND	ND
S. Typhimurium	NCTC 13347	0.5	2	0.25	0.012	2	16	0.125	0.19	3	ND	ND
S. Typhimurium	NCTC 13348	>256*	3	0.25	0.012	3	24	0.064	0.19	3	ND	ND
S. aureus	NCTC 6571	ND	ND	0.125	0.094	2	ND	ND	ND	ND	0.19	1
S. aureus	ATCC 29213	ND	ND	0.19	0.25	2	ND	ND	ND	ND	0.38	1.5

B.2 Plate coating conditions

Condition	Manufacturer	Catalogue Number	Final Concentration	Diluent	Number of washes	Wash buffer	Volume per well
Non-coated	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Thick Collagen	Thermo Fisher	A10483	2mg/ml	H_2O, 1N NaOH, 10x PBS	1	PBS	30µl
Thin Collagen	Thermo Fisher	A10483	50µg/ml	20mM Acetic Acid	3	PBS	50µl
Matrigel	Corning	536231	0.5mg/ml	PBS	1	PBS	50µl
Vitronectin	Stem Cell Technologies	7180	0.01mg/ml	Dilution buffer	1	Dilution buffer	50µl
Fibronectin	Thermo Fisher	354008	50µg/ml	PBS	1	H2O	50µl
Cell-Tak	Corning	354240	3.5mg/ml	0.1N Sodium bicarbonate	1	H2O	50µl
Laminin	Sigma	L4544	5µg/ml	HBSS	1	HBSS	30µl

B.3 Gram-negative image analysis pipeline

Bacterial analysis pipelines for Gram-negative rods using Harmony v4.9 (PerkinElmer)

Input Image	**Input**		
	Flatfield Correction: Basic Brightfield Correction **Stack Processing**: Maximum Projection (for non-adherent isolates, individual planes were analysed) Min. Global Binning: Dynamic		
Filter Image	Input	Method	Output
	Channel : FM4-64	**Method** : Texture SER Filter : SER Ridge Scale : 1 px Normalization by : Kernel	Output Image : SER Ridge
Find Image Region	Input	Method	Output
	Channel: SER Ridge **ROI**: None	Method : Common Threshold Threshold : 0.4 Split into Objects Area : > 100 px²	Output Population : Image Region Output Region : Image Region
Calculate Intensity Properties	Input	Method	Output
	Channel : SER Ridge **Population** : Image Region **Region** : Image Region	**Method** : Standard Mean	Property Prefix : Intensity Image Region SER Ridge
Select Population	Input	Method	Output
	Population : Image Region	**Method** Filter by Property Intensity Image Region SER Ridge Mean : > 0.02	Output Population : Image Region Selected
Select Region	Input	Method	Output
	Population : Image Region Selected **Region** : Image Region	**Method** : Resize Region [µm/px] Outer Border : -4µm Restrictive Population : None Restrictive Region : Keep Image Border Inner Border : INF µm	Output Population : Image Region Resized
Select Region (2)	Input	Method	Output
	Population : Image Region Selected **Region** : Image Region Resized	**Method** : Standard Border Filled Region : INF µm²	Property Prefix : Image Region Resized Filled
Select Region (3)	Input	Method	Output
	Population : Image Region Selected **Region** : Image Region Resized Filled	**Method** : Resize Region [µm/px] Outer Border : 4 px Restrictive Population : None Restrictive Region : Keep Image Border Inner Border : INF px	Output Population : Image Region Resized Filled Resized
Modify Population	Input	Method	Output
	Population : Image Region Selected Region : Image Region Resized Filled Resized	**Method** : Cluster by Distance Distance : 0 px Area : > 0 px²	Output Population : Modified Image Region Selected Output Region : Modified Image Region
Select Population (2)	Input	Method	Output

	Population : Modified Image Region Selected	**Method** : Common Filters Remove Border Objects Region : Modified Image Region	Output Population : Modified Image Region Selected Border Removed
Calculate Image	Input	Method	Output
		Method : By Formula Formula : 100*(A-300)+100*(B-300) Channel A : DAPI Channel B : FM4-64 Negative Values : Set to Zero Undefined Values : Set to Local Average	Output Image : Calculated Image
Find Spots	Input	Method	Output
	Channel : Calculated Image **ROI** : Modified Image Region Selected Border Removed **ROI Region** : Modified Image Region	Method : D Detection Sensitivity : 0.5 Splitting Sensitivity : 0.1 Background Correction : 0.5 Calculate Spot Properties	Output Population : Spots
Calculate Morphology Properties	Input	Method	Output
	Population : Spots **Region** : Spot	**Method** : Standard Area Roundness	Property Prefix : Spot
Select Population (3)	Input	Method	Output
	Population : Spot	**Method** Filter by Property Spot Area [px²] : > 1	Output Population : bacteria
Calculate Morphology Properties (2)	Input	Method	Output
	Population : bacteria **Region** : Spot	**Method** : Standard Area Roundness Width Length Ratio Width to Length	Property Prefix : bacteria
Calculate Morphology Properties (3)	Input	Method	Output
	Population : bacteria **Region** : Spot	**Method** : STAR Channel : DAPI Symmetry Threshold Compactness Axial Radial Profile Profile Width : 3 px Sliding Parabola Curvature : 10 Texture SER Scale : 1 px Normalization by : Kernel	Property Prefix : Bacteria DAPI
Calculate Morphology Properties (4)	Input	Method	Output

	Population : bacteria **Region** : Spot	**Method** : STAR Channel : FM4-64 Symmetry Threshold Compactness Axial Radial Profile Profile Width : 3 px Sliding Parabola Curvature : 10 Texture SER Scale : 1 px Normalization by : Kernel	Property Prefix : Bacteria FM4-64
Calculate Morphology Properties (5)	Input	Method	Output
	Population : bacteria **Region** : Spot	**Method** : STAR Channel : SYTOX green Symmetry Threshold Compactness Axial Radial Profile Profile Width : 3 px Sliding Parabola Curvature : 10 Texture SER Scale : 1 px Normalization by : Kernel	Property Prefix : Bacteria SYTOX green
Calculate Intensity Properties (3)	Input	Method	Output
	Channel : DAPI Population : bacteria **Region** : Spot	**Method** : Standard Mean Standard Deviation	Property Prefix : Intensity Spot DAPI
Calculate Intensity Properties (4)	Input	Method	Output
	Channel : FM4-64 Population : bacteria **Region** : Spot	**Method** : Standard Mean Standard Deviation	Property Prefix : Intensity Spot FM4-64 green
Calculate Intensity Properties (5)	Input	Method	Output
	Channel : SYTOX green Population : bacteria **Region** : Spot	**Method** : Standard Mean Standard Deviation	Property Prefix : Intensity Spot SYTOX green
Select Population (4)	Input	Method	Output
	Population : bacteria	**Method** Linear Classifier Number of Classes : 3	Output Population A : Single Cells Output Population B : Dividing Cells Output Population C : Other

List of Output parameters

Population : bacteria
- Number of Objects
- Intensity Spot DAPI Mean : Mean
- Intensity Spot FM4-64 Mean : Mean
- Intensity Spot SYTOX green Mean : Mean

Population : Single Cells
- Number of Objects
- Relative Spot Intensity : Mean+StdDev
- Corrected Spot Intensity : Mean+StdDev
- Uncorrected Spot Peak Intensity : Mean+StdDev
- Spot Contrast : Mean+StdDev
- Spot Background Intensity : Mean+StdDev
- Spot Area [px²] : Mean+StdDev
- Region Intensity : Mean+StdDev
- Spot to Region Intensity : Mean+StdDev
- Spot Area [µm²] : Mean+StdDev
- Spot Roundness : Mean+StdDev
- bacteria Area [µm²] : Mean+StdDev
- bacteria Roundness : Mean+StdDev
- bacteria Width [µm] : Mean+StdDev
- bacteria Length [µm] : Mean+StdDev
- bacteria Ratio Width to Length : Mean+StdDev
- STAR Properties (Bacteria DAPI) : Mean+StdDev
- STAR Properties (Bacteria FM4-64) : Mean+StdDev
- STAR Properties (Bacteria Sytox) : Mean+StdDev
- Intensity Spot DAPI Mean : Mean+StdDev
- Intensity Spot FM4-64 Mean : Mean+StdDev
- Intensity Spot SYTOX green Mean : Mean+StdDev
- Intensity Spot SYTOX green StdDev : Mean+StdDev
- Regression A-B : Mean+StdDev

Population : Dividing Cells
- Number of Objects
- Population : Other
- Number of Objects

B.4 Gram-positive image analysis pipeline

Bacterial analysis pipelines for Gram-positive cocci using Harmony v4.9 (PerkinElmer)

Input Image	**Input**		
	Flatfield Correction: Basic Brightfield Correction **Stack Processing**: Maximum Projection Min. Global Binning: Dynamic		
Calculate Image	Input	Method	Output
		Method : By Formula Formula : (100*(A-300))+(100*(B-300)) Channel A : FM4-64 Channel B : DAPI Negative Values : Set to Zero Undefined Values : Set to Local Average	Output Image : Calculated Image
Find Image Region	Input	Method	Output
	Channel: Calculated Image **ROI**: None	Method : Common Threshold Threshold : 0.56 Split into Objects Area : > 0 px²	Output Population : Image Region Output Region : Image Region
Calculate Image (2)	Input	Method	Output
		Method : By Formula Formula : 100*(A-300) Channel A : DAPI Negative Values : Set to Zero Undefined Values : Set to Local Average	Output Image : Calculated Image (2)
Find Spots	Input	Method	Output
	Channel : DAPI **ROI** : Image Region **ROI Region** : Image Region	Method : D Detection Sensitivity : 0.8 Splitting Sensitivity : 0.122 Background Correction : 0.5 Calculate Spot Properties	Output Population : Spots
Select Region	Input	Method	Output
	Population : Spots Region : Spot	**Method** : Resize Region [µm/px] Outer Border : -95 µm Restrictive Population : Image Region Restrictive Region : Image Region Inner Border : INF µm	Output Population : Spots Resized
Calculate Morphology Properties	Input	Method	Output
	Population : Spots **Region** : Spot Resized	**Method** : Standard Area Roundness	Property Prefix : Spot Resized
Select Population	Input	Method	Output
	Population : Spots	**Method** : Filter by Property Spot Resized Area [µm²] : > 0.5	Output Population : bacteria

Calculate Morphology Properties (2)	Input	Method	Output
	Population : bacteria **Region** : Spot Resized	**Method** : Standard Area Roundness Width Length Ratio Width to Length	Property Prefix : bacteria
Calculate Morphology Properties (3)	Input	Method	Output
	Population : bacteria **Region** : Spot Resized	**Method** : STAR Channel : DAPI Symmetry Threshold Compactness Axial Radial Profile Profile Width : 3 px Sliding Parabola Curvature : 10 Texture SER Scale : 1 px Normalization by : Kernel	Property Prefix : Bacteria DAPI
Calculate Morphology Properties (4)	Input	Method	Output
	Population : bacteria **Region** : Spot Resized	**Method** : STAR Channel : FM4-64 Symmetry Threshold Compactness Axial Radial Profile Profile Width : 3 px Sliding Parabola Curvature : 10 Texture SER Scale : 1 px Normalization by : Kernel	Property Prefix : Bacteria FM4-64
Calculate Morphology Properties (5)	Input	Method	Output
	Population : bacteria **Region** : Spot Resized	**Method** : STAR Channel : SYTOX green Symmetry Threshold Compactness Axial Radial Profile Profile Width : 3 px Sliding Parabola Curvature : 10 Texture SER Scale : 1 px	Property Prefix : Bacteria SYTOX green

		Normalization by : Kernel	
Calculate Intensity Properties	Input	Method	Output
	Channel : DAPI Population : bacteria **Region** : Spot Resized	**Method** : Standard Mean	Property Prefix : Intensity Spot Resized DAPI
Calculate Intensity Properties (2)	Input	Method	Output
	Channel : FM4-64 Population : bacteria **Region** : Spot Resized	**Method** : Standard Mean	Property Prefix : Intensity Spot Resized FM4-64
Calculate Intensity Properties (3)	Input	Method	Output
	Channel : SYTOX green Population : bacteria **Region** : Spot Resized	**Method** : Standard Mean	Property Prefix : Intensity Spot Resized SYTOX green
Select Population (2)	Input	Method	Output
	Population : bacteria	**Method** : Common Filters Remove Border Objects Region : Spot	Output Population : bacteria Selected

Method : List of Outputs
Population : bacteria
Number of Objects

Population : bacteria Selected
Number of Objects
Relative Spot Intensity : Mean+StdDev
Corrected Spot Intensity : Mean+StdDev
Uncorrected Spot Peak Intensity : Mean+StdDev
Spot Contrast : Mean+StdDev
Spot Background Intensity : Mean+StdDev
Spot Area [px²] : Mean+StdDev
Region Intensity : Mean+StdDev
Spot to Region Intensity : Mean+StdDev
Spot Resized Area [µm²] : Mean+StdDev
Spot Resized Roundness : Mean+StdDev
bacteria Area [µm²] : Mean+StdDev
bacteria Roundness : Mean+StdDev
bacteria Width [µm] : Mean+StdDev
bacteria Length [µm] : Mean+StdDev
bacteria Ratio Width to Length : Mean+StdDev
STAR Properties (Bacteria DAPI) : Mean+StdDev
STAR Properties (Bacteria FM4-64) : Mean+StdDev
STAR Properties (Bacteria Sytox) : Mean+StdDev
Intensity Spot Resized SYTOX green Mean : Mean+StdDev
Intensity Spot Resized SYTOX green StdDev : Mean+StdDev
Intensity Spot Resized DAPI Mean : Mean+StdDev
Intensity Spot Resized FM4-64 Mean : Mean+StdDev

Method : Standard Output
bacteria - Relative Spot Intensity : Mean
Output Name : bacteria - Relative Spot Intensity - Mean per Well

Object Results
Population : bacteria : ALL
Population : Spots : None
Population : Image Region : None
Population : bacteria Selected : Use Selected Well Results

B.5 Calculated morphological properties used in analysis

Full list of variables generated in Harmony and used in TIBCO Spotfire for principal component analysis. All mean and standard deviation (StdDev) values are per well for objects classified as single cells using the linear classifier described in **Chapter 4**.

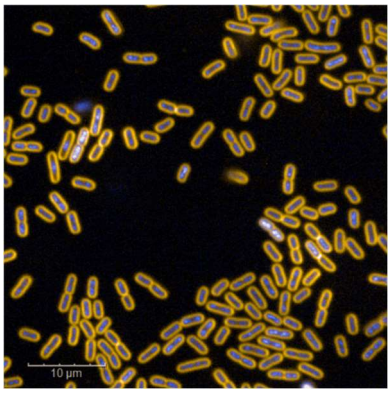
Number of Objects
Bacteria Area [µm²] - Mean & StdDev
Bacteria Area [µm²] - Sum
Bacteria Area [µm²] - Max
Bacteria Area [µm²] - Min
Bacteria Area [µm²] - Median
Bacteria Roundness - Mean & StdDev
Bacteria Width [µm] - Mean & StdDev
Bacteria Length [µm] - Mean & StdDev
Bacteria Ratio Width to Length - Mean & StdDev
Bacteria DAPI Symmetry 02 - Mean & StdDev
Bacteria DAPI Symmetry 03 - Mean & StdDev
Bacteria DAPI Symmetry 04 - Mean & StdDev
Bacteria DAPI Symmetry 05 - Mean & StdDev
Bacteria DAPI Symmetry 12 - Mean & StdDev
Bacteria DAPI Symmetry 13 - Mean & StdDev
Bacteria DAPI Symmetry 14 - Mean & StdDev
Bacteria DAPI Symmetry 15 - Mean & StdDev
Bacteria DAPI Threshold Compactness 30% - Mean & StdDev
Bacteria DAPI Threshold Compactness 40% - Mean & StdDev
Bacteria DAPI Threshold Compactness 50% - Mean & StdDev
Bacteria DAPI Threshold Compactness 60% - Mean & StdDev
Bacteria DAPI Axial Small Length - Mean & StdDev
Bacteria DAPI Axial Length Ratio - Mean & StdDev
Bacteria DAPI Radial Mean & StdDev - Mean & StdDev
Bacteria DAPI Radial Relative Deviation - Mean & StdDev
Bacteria DAPI Profile 1/2 - Mean & StdDev
Bacteria DAPI Profile 2/2 - Mean & StdDev
Bacteria FM4-64 Symmetry 02 - Mean & StdDev
Bacteria FM4-64 Symmetry 03 - Mean & StdDev
Bacteria FM4-64 Symmetry 04 - Mean & StdDev
Bacteria FM4-64 Symmetry 05 - Mean & StdDev
Bacteria FM4-64 Symmetry 12 - Mean & StdDev
Bacteria FM4-64 Symmetry 13 - Mean & StdDev
Bacteria FM4-64 Symmetry 14 - Mean & StdDev
Bacteria FM4-64 Symmetry 15 - Mean & StdDev
Bacteria FM4-64 Threshold Compactness 30% - Mean & StdDev
Bacteria FM4-64 Threshold Compactness 40% - Mean & StdDev
Bacteria FM4-64 Threshold Compactness 50% - Mean & StdDev
Bacteria FM4-64 Threshold Compactness 60% - Mean & StdDev
Bacteria FM4-64 Axial Small Length - Mean & StdDev
Bacteria FM4-64 Axial Length Ratio - Mean & StdDev
Bacteria FM4-64 Radial Mean & StdDev - Mean & StdDev
Bacteria FM4-64 Radial Relative Deviation - Mean & StdDev
Bacteria FM4-64 Profile 1/2 - Mean & StdDev
Bacteria FM4-64 Profile 2/2 - Mean & StdDev
Bacteria SYTOX Symmetry 02 - Mean & StdDev
Bacteria SYTOX Symmetry 03 - Mean & StdDev
Bacteria SYTOX Symmetry 04 - Mean & StdDev
Bacteria SYTOX Symmetry 05 - Mean & StdDev
Bacteria SYTOX Symmetry 12 - Mean & StdDev
Bacteria SYTOX Symmetry 13 - Mean & StdDev
Bacteria SYTOX Symmetry 14 - Mean & StdDev
Bacteria SYTOX Symmetry 15 - Mean & StdDev
Bacteria SYTOX Threshold Compactness 30% - Mean & StdDev
Bacteria SYTOX Threshold Compactness 40% - Mean & StdDev
Bacteria SYTOX Threshold Compactness 50% - Mean & StdDev
Bacteria SYTOX Threshold Compactness 60% - Mean & StdDev
Bacteria SYTOX Axial Small Length - Mean & StdDev
Bacteria SYTOX Axial Length Ratio - Mean & StdDev
Bacteria SYTOX Radial Mean & StdDev - Mean & StdDev
Bacteria SYTOX Radial Relative Deviation - Mean & StdDev
Bacteria SYTOX Profile 1/2 - Mean & StdDev
Bacteria SYTOX Profile 2/2 - Mean & StdDev
Intensity Bacteria SYTOX green Mean & StdDev - Mean & StdDev
Regression A-B - Mean & StdDev

B.6 Morphological properties that distinguish treated and untreated bacteria

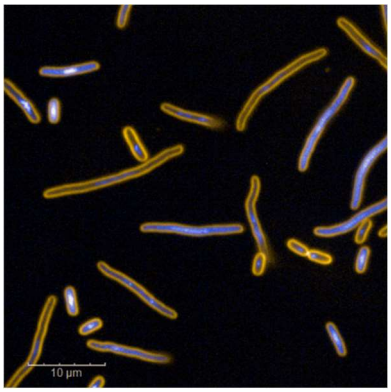
Morphological properties that distinguish untreated cells of *K. pneumoniae*, from those treated with ciprofloxacin (**Table B.6.1**), meropenem (**table B.6.2**) and tigecycline (**Table B.6.3**) at MIC for 2 hours. Z' values closest to 1 demonstrate the most substantial differences between treated and untreated bacteria; any value between 0.6 and 1 is considered a reliable feature for distinguishing positive and negative controls. Only the highest Z' scores are shown. Representative confocal images of treated and untreated bacteria that contribute to the Z' values are given alongside each table.

Table B.6.1 – Ciprofloxacin

Stain	Morphological property	Z' Score
FM4-64	Symmetry 14 - mean per well	0.955
FM4-64	Symmetry 04 - mean per well	0.953
DAPI	Threshold compactness 40% - StdDev	0.953
SYTOX	Symmetry 14 - mean per well	0.947
FM4-64	Symmetry 02 - mean per well	0.946
Bacteria	Width to length ratio	0.944
SYTOX	Threshold compactness 50% - StdDev	0.940
SYTOX	Symmetry 02 - mean per well	0.940
SYTOX	Axial Length Ratio - mean	0.938
SYTOX	Symmetry 04 - mean per well	0.937
FM4-64	Axial Length Ratio - mean	0.936
SYTOX	Symmetry 12 - mean	0.934
FM4-64	Radial Relative Deviation - mean	0.928
FM4-64	Symmetry 12 - mean	0.926
Bacteria	Area - median	0.924



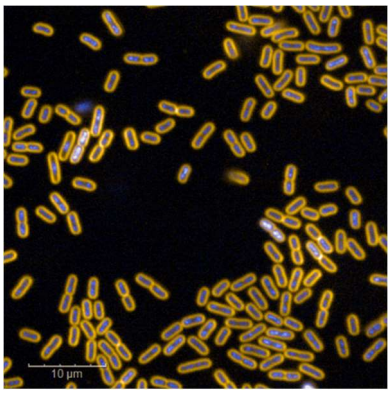
Untreated



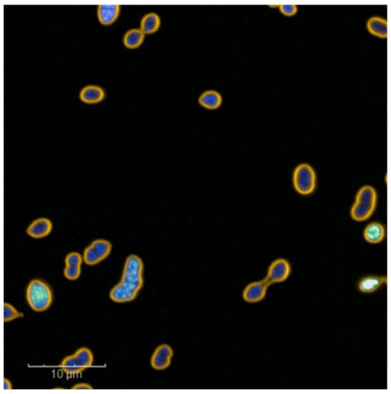
Ciprofloxacin

Table B.6.2 – Meropenem

Stain	Morphological property	Z' Score
Bacteria	Width to length - mean	0.952
FM4-64	Axial length ratio - mean	0.950
SYTOX	Axial length ratio - mean	0.942
FM4-64	Symmetry 12 - mean	0.939
SYTOX	Symmetry 12 - mean	0.937
Bacteria	Width - mean	0.930
SYTOX	Symmetry 02 - mean	0.928
FM4-64	Symmetry 02 - mean	0.922
FM4-64	Axial small length - mean	0.908
DAPI	Axial length ratio - mean	0.907
DAPI	Symmetry 12 - mean	0.899
DAPI	Profile 1/2 - std dev	0.898
Bacteria	Area - median	0.892
SYTOX	Radial Relative Deviation - mean	0.890
DAPI	Symmetry 02 - mean	0.888



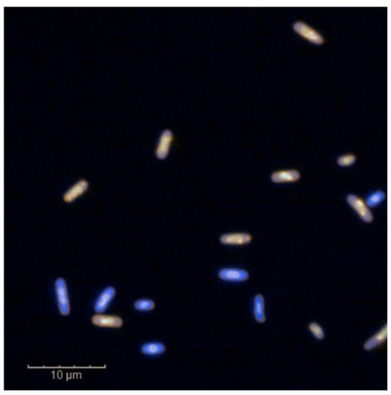
Untreated



Meropenem

Table B.6.3 – Tigecycline

Feature		Z' Score
FM4-64	Threshold Compactness 50% - mean	0.914
FM4-64	Threshold Compactness 50% - std dev	0.884
SYTOX	Threshold Compactness 50% - std dev	0.884
DAPI	Radial Relative Deviation - mean	0.872
DAPI	Symmetry 02 - std dev	0.822
DAPI	Profile 2/2 - mean	0.818
DAPI	Symmetry 05 - mean	0.785
FM4-64	Threshold compactness 60% - mean	0.780
DAPI	Threshold compactness 60% - mean	0.774
DAPI	Profile 1/2	0.768
SYTOX	Threshold compactness 60% - std dev	0.767
DAPI	Symmetry 03 - mean	0.764
Bacteria	Area -median	0.754



Tigecycline

Appendix C Supplementary antimicrobial susceptibility results

C.1 Minimum inhibitory concentrations

MICs of isolates KC006 and KC036, as determined by Etest (bioMérieux), except where stated.

Antimicrobial	KC006 MIC (µg/mL)	KC036 MIC (µg/mL)
Amikacin	3	1
Aztreonam	0.125	0.38
Cefoxitin	6	3
Cefuroxime	3	2
Chloramphenicol	64	2
Ciprofloxacin	>32*	0.03
Colistin	6	3
Co-trimoxazole	>32	0.047
Ertapenem	0.023	0.016
Fosfomycin	2	3
Gentamicin	16	0.38
Kanamycin	8	1
Meropenem	0.064	0.064
Nalidixic Acid	>256	2
Oxacillin	>256	>256
Polymixin B	1.0	0.75
Rifampicin	>32	>32
Tetracycline	>256	>256
Tigecycline	1.0	0.38
Trimethoprim	>32	0.25

*The MIC of ciprofloxacin was determined to be 64 µg/mL using both broth and agar dilution.

Appendix D Supplementary RNAseq data

D.1 Differential expression – KC006 in the presence of different antimicrobials

Results are ordered by log_2 fold change in expression. Only upregulated transcripts encoding predicted coding regions are included. Genes encoded on a plasmid are highlighted in red. Gene product details are derived from annotations from Prokka.

Table D.1 – Differentially expressed genes of KC006 exposed to ciprofloxacin at MIC after 2 hours incubation (repeat experiment)

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
05214	7.199466	1.86E-16	unknown	hypothetical protein
05213	7.198758	1.8E-15	unknown	hypothetical protein
01153	6.982713	4.36E-82	unknown	toxic peptide TisB
05104	6.979219	9.05E-51	dinI_2	virulence protein,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
02336	6.630775	3.36E-68	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
04860	6.22071	4.45E-78	sulA	cell division inhibitor,Cell division inhibitor SulA,SOS cell division inhibitor,SOS-response cell division inhibitor%2C blocks FtsZ ring formation,cell division inhibitor SulA,Cell division inhibitor SulA
05105	6.10479	3.07E-52	umuD_1	Error-prone repair protein UmuD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
05106	6.062254	3.94E-89	umuC_1	Error-prone%2C lesion bypass DNA polymerase V (UmuC),hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
03927	5.711942	1.65E-72	yebE	membrane protein,Inner membrane protein yebE,Protein of unknown function (DUF533)
04992	5.594039	3.44E-68	dinI_1	DNA-damage-inducible protein I,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
03928	5.410026	1.72E-77	unknown	LexA family transcriptional regulator,DNA damage-inducible protein YebG,Uncharacterized protein conserved in bacteria,YebG protein
05200	5.342517	1.81E-52	unknown	putative dopa decarboxylase protein remnant
05221	5.284752	1.37E-67	umuD_2	umuDC operon protein-like protein,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01152	5.22839	2.39E-66	unknown	putative acyltransferase,Acetyltransferase (GNAT) family
00574	5.195579	2.6E-100	nrdD	anaerobic ribonucleoside triphosphate reductase,Anaerobic ribonucleoside-triphosphate reductase,anaerobic ribonucleoside triphosphate reductase,anaerobic ribonucleoside-triphosphate reductase,Glycine radical
05199	5.189876	2.29E-42	unknown	hypothetical protein
02257	5.127244	4.4E-87	recA	recombinase A,Recombinase A,recombinase A,RecA-superfamily ATPases implicated in signal transduction,protein RecA,recA bacterial DNA recombination protein
02645	5.118826	7.3E-110	nrdA	ribonucleotide reductase of class Ia,Ribonucleoside-diphosphate reductase 1 subunit alpha,ribonucleotide-diphosphate reductase subunit alpha,ribonucleoside-diphosphate reductase%2C alpha subunit,Ribonucleotide reductase%2C barrel domain
00576	5.064053	2.34E-70	relE	RelE/StbE replicon stabilization toxin,mRNA interferase RelE,addiction module toxin%2C RelE/StbE family,Plasmid stabilisation system protein
05195	4.889964	3.06E-54	unknown	Domain of unknown function (DUF932)
02258	4.879651	2.06E-44	recX	Regulatory protein RecX,Regulatory protein recX,recombination regulator RecX,Uncharacterized protein conserved in bacteria,RecX family
00718	4.815168	1.07E-52	unknown	alkylphosphonate utilization operon protein PhnA,hypothetical protein,putative alkylphosphonate utilization operon protein PhnA,PhnA protein
00575	4.718301	1.7E-111	pflA_1	ribonucleotide reductase of class III (anaerobic),Pyruvate formate-lyase 1-activating enzyme,anaerobic ribonucleotide reductase-activating protein,anaerobic ribonucleoside-triphosphate reductase activating protein
05222	4.618845	1.56E-92	umuC_2	DNA polymerase V subunit UmuC,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
04901	4.421409	2.8E-22	unknown	Regulatory phage protein cox
05212	4.360529	6.76E-78	unknown	Domain of unknown function (DUF3560)
01137	4.308103	7.04E-10	ibpB	16 kDa heat shock protein B,16 kDa heat shock protein B,heat shock chaperone IbpB,Hsp20/alpha crystallin family
02644	4.257968	5.5E-100	nrdB	ribonucleotide-diphosphate reductase subunit beta,Ribonucleoside-diphosphate reductase 1 subunit beta,ribonucleotide-diphosphate reductase subunit beta,Ribonucleotide reductase%2C beta subunit,ribonucleoside-diphosphate reductase%2C class 1b%2C beta subunit,Ribonucleotide reductase%2C small chain
02643	4.254934	1.23E-87	ascD	Ferredoxin,CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydrase reductase,2Fe-2S ferredoxin YfaE,Uncharacterized Fe-S protein,ferredoxin [2Fe-2S],2Fe-2S iron-sulfur cluster binding domain
01238	4.238122	1.19E-50	yicG	inner membrane protein,hypothetical protein,hypothetical protein,Predicted membrane protein,UPF0126 domain
05107	4.193491	3.65E-71	dmlR_29	LysR family transcriptional regulator,D-malate degradation protein R,transcriptional regulator,ABC-type uncharacterized transport system%2C periplasmic component,putative choline sulfate-utilization transcription factor,LysR substrate binding domain
05198	3.871089	1.36E-28	unknown	hypothetical protein
00778	3.858135	1.8E-149	ssb_1	ssDNA-binding protein,Helix-destabilizing protein,single-stranded DNA-binding protein,single-stranded DNA-binding protein,Single-strand binding protein family
05108	3.814859	4.17E-34	butA	3-oxoacyl-ACP reductase,Diacetyl reductase [(S)-acetoin forming],short chain dehydrogenase,Uncharacterized conserved protein,3-oxoacyl-[acyl-carrier-protein] reductase,short chain dehydrogenase
04622	3.814296	3.23E-42	qseD_2	transcriptional regulator LysR,Quorum-sensing regulator protein D,cell density-dependent motility repressor,pca operon transcription factor PcaQ,Bacterial regulatory helix-turn-helix protein%2C lysR family
00085	3.728285	2.9E-114	dinB	DNA polymerase IV,DNA polymerase IV,DNA polymerase IV,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
03410	3.676578	4.21E-27	unknown	ABC transporter substrate-binding protein,hypothetical protein,2-aminoethylphosphonate ABC transporter substrate-binding protein,ABC-type thiamine transport system%2C periplasmic component,putative 2-aminoethylphosphonate ABC transporter%2C periplasmic 2-aminoethylphosphonate-binding protein,Bacterial extracellular solute-binding protein
04902	3.652338	2.1E-12	unknown	hypothetical protein

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01874	3.59292	5.37E-55	endA	DNA-specific endonuclease I,Endonuclease-1 precursor,DNA-specific endonuclease I,Endonuclease I
00577	3.541058	3.12E-37	relB	RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE),Antitoxin RelB,bifunctional antitoxin/transcriptional repressor RelB,addiction module antitoxin%2C RelB/DinJ family,RelB antitoxin
02335	3.496379	3.8E-120	smpA	outer membrane lipoprotein SmpA,Small protein A precursor,outer membrane biogenesis protein BamE,SmpA / OmlA family
05238	3.484951	9.91E-22	unknown	putative plasmid-like protein
00793	3.48376	6.3E-111	lexA	SOS-response repressor and protease LexA,LexA repressor,LexA repressor,repressor LexA,LexA DNA binding domain
01151	3.453934	1.36E-48	unknown	integral membrane protein,Membrane transporters of cations and cationic drugs,phosphonate utilization associated putative membrane protein,EamA-like transporter family
04909	3.43401	9.22E-17	unknown	hypothetical protein
05103	3.415077	4.15E-36	unknown	heptosyltransferase family protein,lipopolysaccharide core biosynthesis protein,lipopolysaccharide heptosyltransferase I,Glycosyltransferase family 9 (heptosyltransferase)
05208	3.401041	6.2E-28	unknown	DNA-binding protein,plasmid partitioning protein,ParB/RepB/Spo0J family partition protein,ParB-like nuclease domain
03446	3.368134	3.44E-68	xylF_2	alpha/beta fold family hydrolase,2-hydroxymuconic semialdehyde hydrolase,acetoin dehydrogenase E2 subunit dihydrolipoyllysine-residue acetyltransferase,2-succinyl-6-hydroxy-2%2C4-cyclohexadiene-1-carboxylate synthase,Alpha/beta hydrolase family
00042	3.348739	3.07E-24	phoE	outer membrane pore protein E,Outer membrane pore protein E precursor,outer membrane phosphoporin protein E,Outer membrane protein (porin),Gram-negative porin
05210	3.29477	8.66E-21	ssb_2	single-strand binding protein,Helix-destabilizing protein,single-stranded DNA-binding protein,single-stranded DNA-binding protein,Single-strand binding protein family
05270	3.277822	0.001634	pcoE	copper-binding protein,Probable copper-binding protein pcoE precursor
04905	3.23699	2.17E-08	unknown	hypothetical protein
02768	3.214826	3.33E-30	dnaK_2	heat shock protein YegD,Heat shock protein 70,putative chaperone,chaperone protein DnaK,Hsp70 protein
00003	3.205821	3.67E-28	unknown	hypothetical protein
05080	3.20209	2.03E-09	unknown	hypothetical protein
05009	3.14389	7.49E-71	fabH	3-oxoacyl-ACP synthase,3-oxoacyl-[acyl-carrier-protein] synthase 3,3-oxoacyl-(acyl carrier protein) synthase III,3-oxoacyl-[acyl-carrier-protein] synthase III,3-oxoacyl-[acyl-carrier-protein] synthase III,3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C terminal
00779	3.111311	4.9E-236	uvrA	excinuclease ABC subunit A,Excinuclease ABC subunit A,excinuclease ABC subunit A,ABC-type transport system involved in cytochrome bd biosynthesis%2C fused ATPase and permease components,excinuclease ABC subunit A,ABC transporter
04784	3.10741	4.56E-39	aqpZ	aquaporin,Bacterial nodulin-like intrinsic protein,aquaporin Z,MIP family channel proteins,Major intrinsic protein
05097	3.10611	2.07E-56	pyrG_2	CTP synthase,CTP synthase,hypothetical protein,CTP synthase,Glutamine amidotransferase class-I
00084	3.088373	6.64E-93	unknown	acetyltransferase YafP,putative acyltransferase

Table D.2 – Differentially expressed genes of KC006 exposed to tigecycline at MIC after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
05349	6.331553	3.3E-124	tetA_2	tetracycline resistance protein,Tetracycline resistance protein%2C class C,bicyclomycin/multidrug efflux system,Arabinose efflux permease,multidrug resistance protein,Major Facilitator Superfamily
05348	6.254361	1.36E-89	yedA_2	drug/metabolite transporter permease,Uncharacterized inner membrane transporter yedA,putative DMT superfamily transporter inner membrane protein,Predicted permease%2C DMT superfamily,carboxylate/amino acid/amine transporter,EamA-like transporter family
04400	5.485947	3.47E-57	ctpF_1	cation-transporting P-type ATPase,Probable cation-transporting ATPase F,magnesium-transporting ATPase MgtA,Uncharacterized protein conserved in bacteria,calcium-transporting P-type ATPase%2C PMR1-type,E1-E2 ATPase
05351	5.474739	6.16E-52	unknown	transposase for transposon Tn1721,Transposase and inactivated derivatives%2C TnpA family,Tn3 transposase DDE domain
03693	5.090437	4.54E-65	hmuT_2	ABC transporter substrate-binding protein,Hemin-binding periplasmic protein hmuT precursor,corrinoid ABC transporter substrate-binding protein,ABC-type Fe3+-citrate transport system%2C periplasmic component,proposed F420-0 ABC transporter%2C periplasmic F420-0 binding protein,Periplasmic binding protein
03401	4.935571	2.47E-14	smvA	methyl viologen resistance protein SmvA,Methyl viologen resistance protein SmvA,methyl viologen resistance protein SmvA,Arabinose efflux permease,drug resistance MFS transporter%2C drug:H+ antiporter-2 (14 Spanner) (DHA2) family,Major Facilitator Superfamily
04834	4.831384	1.95E-82	ompK35	outer membrane protein F,Porin OmpF,outer membrane porin protein C,Outer membrane protein (porin),Gram-negative porin
03400	4.747797	6.13E-28	pobR_2	Pca regulon regulatory protein PcaR,p-hydroxybenzoate hydroxylase transcriptional activator,DNA-binding transcriptional regulator KdgR,beta-ketoadipate pathway transcriptional regulators%2C PcaR/PcaU/PobR family,Bacterial transcriptional regulator
01314	4.744424	5.83E-76	yheS_1	ABC transporter ATP-binding protein,Uncharacterized ABC transporter ATP-binding protein YheS,putative ABC transporter ATP-binding protein,ABC-type transport system involved in cytochrome c biogenesis%2C ATPase component,ATP-binding cassette protein%2C ChvD family,ABC transporter
04399	4.581794	1.89E-13	unknown	hypothetical protein
00703	4.540002	2.96E-27	unknown	LuxR family transcriptional regulator
05248	4.506208	3.74E-05	silE_1	SilE,Silver-binding protein silE precursor
03255	4.456071	9.29E-67	rbsB_3	putative ABC transport system nucleotide binding/ATPase component,D-ribose-binding periplasmic protein precursor,D-ribose transporter subunit RbsB,ABC-type sugar transport system%2C periplasmic component,rhamnose ABC transporter%2C rhamnose-binding protein,Periplasmic binding proteins and sugar binding domain of LacI family
02282	4.3896	1.4E-100	proV_1	L-proline glycine betaine ABC transport system permease ProV,Glycine betaine/L-proline transport ATP-binding protein ProV,glycine betaine transporter ATP-binding subunit,ABC-type proline/glycine betaine transport system%2C ATPase component,glycine betaine/L-proline transport ATP binding subunit,ABC transporter
02283	4.340877	3.3E-69	nrdF	ribonucleotide-diphosphate reductase subunit beta,Ribonucleoside-diphosphate reductase 2 subunit beta,ribonucleotide-diphosphate reductase subunit beta,Ribonucleotide reductase%2C beta subunit,ribonucleoside-diphosphate reductase%2C class 1b%2C beta subunit,Ribonucleotide reductase%2C small chain
03900	4.308611	1.51E-84	rlmA	ribosomal RNA large subunit methyltransferase A,Ribosomal RNA large subunit methyltransferase A,23S rRNA methyltransferase A,biotin biosynthesis protein BioC,Methyltransferase domain
02554	4.307316	2.83E-41	unknown	hypothetical protein
05030	4.230066	3.05E-24	bhsA_5	outer membrane protein,Multiple stress resistance protein BhsA precursor,hypothetical protein,Protein of unknown function (DUF1471)
05085	4.162258	6.66E-59	gltC_7	LysR family transcriptional regulator,HTH-type transcriptional regulator gltC,DNA-binding transcriptional activator XapR,aminoethylphosphonate

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
				catabolism associated LysR family transcriptional regulator,LysR substrate binding domain
03547	4.132192	8.13E-31	unknown	beta-lactamase,ribonuclease Z,Predicted metal-dependent RNase%2C consists of a metallo-beta-lactamase domain and an RNA-binding KH domain,Metallo-beta-lactamase superfamily
03254	4.088482	1.96E-44	rbsK_4	ribokinase,Ribokinase,ribokinase,Uncharacterized protein conserved in bacteria,ribokinase,pfkB family carbohydrate kinase
03253	4.044562	4.71E-38	sgcQ	BtpA family protein,Putative sgc region protein SgcQ,membrane complex biogenesis protein%2C BtpA family,BtpA family
05281	4.034673	0.002445	unknown	IS2 repressor TnpA,IS2 repressor TnpA,Transposase
03694	4.022303	8.63E-38	fecD_1	ABC transporter permease,Iron(III) dicitrate transport system permease protein fecD,iron-hydroxamate transporter permease subunit,ABC-type Fe3+-siderophore transport system%2C permease component,proposed F420-0 ABC transporter%2C permease protein,FecCD transport family
03899	4.002546	2.31E-94	yebN	membrane protein,hypothetical protein,hypothetical protein,putative sporulation protein YtaF,Domain of unknown function DUF
00328	3.908539	2.01E-37	unknown	lipoprotein
03252	3.848264	1.25E-33	unknown	thiaminase II,multifunctional hydroxymethylpyrimidine phosphokinase/4-amino-5-aminomethyl-2-methylpyrimidine hydrolase,TENA/THI-4/PQQC family
04401	3.842741	1.22E-98	mdtE_1	RND efflux membrane fusion protein,Multidrug resistance protein MdtE precursor,multidrug efflux system subunit MdtA,efflux transporter%2C RND family%2C MFP subunit,HlyD family secretion protein
02884	3.838069	1.24E-26	unknown	cold-shock protein
03416	3.822017	3.1E-28	scrY_2	maltoporin,Sucrose porin precursor,maltoporin,Maltoporin (phage lambda and maltose receptor),LamB porin
05192	3.793155	5.08E-15	unknown	conjugal transfer transcriptional regulator TraJ,conjugal transfer transcriptional regulator TraJ
02253	3.79118	3E-17	yusV_1	Zinc ABC transporter ATP-binding protein ZnuC,Probable siderophore transport system ATP-binding protein YusV,iron-dicitrate transporter ATP-binding subunit,ABC-type phosphate/phosphonate transport system%2C ATPase component,anchored repeat-type ABC transporter%2C ATP-binding subunit,ABC transporter
02903	3.771914	9.02E-18	unknown	putative membrane proteins,Predicted chitinase
02793	3.73002	8.77E-45	ydjM	membrane protein,Inner membrane protein ydjM,inner membrane protein,Predicted membrane-bound metal-dependent hydrolase (DUF457)
03197	3.714025	3.71E-16	unknown	hypothetical protein
02281	3.711621	4.43E-88	proW	L-proline glycine betaine ABC transport system permease protein ProW,Glycine betaine/L-proline transport system permease protein proW,glycine betaine transporter membrane protein,ABC-type proline/glycine betaine transport systems%2C periplasmic components,choline ABC transporter%2C permease protein,Binding-protein-dependent transport system inner membrane component
02252	3.601926	1.03E-05	unknown	hypothetical protein
02764	3.568373	2.31E-24	unknown	positive transcription regulator
01970	3.54915	7.54E-40	xerC_2	type 1 fimbriae regulatory protein FimE,Tyrosine recombinase XerC,tyrosine recombinase,Site-specific recombinase XerD,tyrosine recombinase XerC,Phage integrase family
03397	3.513018	2.01E-15	catA	catechol 1%2C2-dioxygenase,Catechol 1%2C2-dioxygenase,Protocatechuate 3%2C4-dioxygenase beta subunit,catechol 1%2C2-dioxygenase,Dioxygenase
02438	3.503432	8.28E-15	unknown	hypothetical protein
03196	3.5025	2.95E-15	besA	hydrolase,Ferri-bacillibactin esterase BesA,Predicted hydrolase of the alpha/beta superfamily
02254	3.499586	2.44E-15	mntB_3	Zinc ABC transporter,Manganese transport system membrane protein mntB,high-affinity zinc transporter membrane component,anchored repeat-type ABC transporter%2C permease subunit,ABC 3 transport family
04007	3.455441	1.1E-39	nhaK_3	Na+/H+ antiporter,Sodium%2C potassium%2C lithium and rubidium/H(+) antiporter,NhaP-type Na+/H+ and K+/H+ antiporters,Na+/H+ antiporter,Sodium/hydrogen exchanger family
03550	3.417344	1.43E-34	unknown	response regulator receiver protein
03183	3.4061	2.47E-32	hipB_1	antitoxin,Antitoxin HipB,antitoxin HipB,Uncharacterized conserved small protein,transcriptional regulator%2C y4mF family,Helix-turn-helix
04877	3.398464	1.18E-37	ycgG_4	putative protease,hypothetical protein,phage resistance protein,Predicted signal transduction protein containing sensor and EAL domains,EAL domain

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01981	3.388537	7.15E-19	yhjH_1	cyclic diguanylate phosphodiesterase (EAL) domain-containing protein,Cyclic di-GMP phosphodiesterase YhjH,cyclic-di-GMP phosphodiesterase,Predicted signal transduction protein containing sensor and EAL domains,EAL domain
05199	3.386374	1.13E-19	unknown	hypothetical protein
04030	3.384755	2.13E-28	srfAD	YbtT,Surfactin synthase thioesterase subunit,Thioesterase domain
03495	3.376047	1.16E-50	rnfA_2	electron transport complex protein RnfA,Electron transport complex protein rnfA,electron transport complex protein RsxA,Predicted NADH:ubiquinone oxidoreductase%2C subunit RnfA,electron transport complex%2C RnfABCDGE type%2C A subunit,Rnf-Nqr subunit%2C membrane protein
04712	3.353373	2.06E-57	ybiP	putative enzyme,Putative phosphoethanolamine transferase ybiP,hypothetical protein,Predicted membrane-associated%2C metal-dependent hydrolase,Sulfatase
05200	3.34369	6.95E-22	unknown	putative dopa decarboxylase protein remnant
02765	3.337606	2.9E-08	unknown	hypothetical protein
03327	3.334584	1.37E-59	ddc	L-2%2C4-diaminobutyrate decarboxylase,L-2%2C4-diaminobutyrate decarboxylase,L-tyrosine decarboxylase,4-aminobutyrate aminotransferase and related aminotransferases,putative pyridoxal-dependent aspartate 1-decarboxylase,Pyridoxal-dependent decarboxylase conserved domain
03447	3.332842	1.83E-24	unknown	Protein of unknown function (DUF2566)
03586	3.309285	7.81E-35	ytcD	transcriptional regulator,Uncharacterized HTH-type transcriptional regulator ytcD,HxlR-like helix-turn-helix
00792	3.295887	8.15E-41	dinF	DNA-damage-inducible SOS response protein,Multidrug export protein mepA,DNA-damage-inducible SOS response protein,MATE efflux family protein,MatE
03816	3.292083	3.66E-07	unknown	hypothetical protein
01644	3.277713	1.5E-62	yhbE	drug/metabolite transporter permease,Uncharacterized inner membrane transporter yhbE,Predicted permeases,carboxylate/amino acid/amine transporter,EamA-like transporter family

Table D.3 – Differentially expressed genes of KC006 exposed to chloramphenicol at MIC after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
03693	5.185805	2.08E-67	hmuT_2	ABC transporter substrate-binding protein,Hemin-binding periplasmic protein hmuT precursor,corrinoid ABC transporter substrate-binding protein,ABC-type Fe3+-citrate transport system%2C periplasmic component,proposed F420-0 ABC transporter%2C periplasmic F420-0 binding protein,Periplasmic binding protein
04834	4.958863	8.98E-87	ompF	outer membrane protein F,Porin OmpF,outer membrane porin protein C,Outer membrane protein (porin),Gram-negative porin
01314	4.870638	5.57E-80	yheS_1	ABC transporter ATP-binding protein,Uncharacterized ABC transporter ATP-binding protein YheS,putative ABC transporter ATP-binding protein,ABC-type transport system involved in cytochrome c biogenesis%2C ATPase component,ATP-binding cassette protein%2C ChvD family,ABC transporter
04400	4.730585	8.76E-43	ctpF_1	cation-transporting P-type ATPase,Probable cation-transporting ATPase F,magnesium-transporting ATPase MgtA,Uncharacterized protein conserved in bacteria,calcium-transporting P-type ATPase%2C PMR1-type,E1-E2 ATPase
03900	4.53064	2.64E-93	rlmA	ribosomal RNA large subunit methyltransferase A,Ribosomal RNA large subunit methyltransferase A,23S rRNA methyltransferase A,biotin biosynthesis protein BioC,Methyltransferase domain
05085	4.333969	1.12E-63	gltC_7	LysR family transcriptional regulator,HTH-type transcriptional regulator gltC,DNA-binding transcriptional activator XapR,aminoethylphosphonate catabolism associated LysR family transcriptional regulator,LysR substrate binding domain
05281	4.243707	0.001542	unknown	IS2 repressor TnpA,IS2 repressor TnpA,Transposase
05104	4.236494	1.03E-19	dinI_2	virulence protein,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
03694	4.207726	3.32E-41	fecD_1	ABC transporter permease,Iron(III) dicitrate transport system permease protein fecD,iron-hydroxamate transporter permease subunit,ABC-type Fe3+-siderophore transport system%2C permease component,proposed F420-0 ABC transporter%2C permease protein,FecCD transport family
02793	4.102538	6.78E-54	ydjM	membrane protein,Inner membrane protein ydjM,inner membrane protein,Predicted membrane-bound metal-dependent hydrolase (DUF457)
03899	4.08562	3.13E-98	yebN	membrane protein,hypothetical protein,hypothetical protein,putative sporulation protein YtaF,Domain of unknown function DUF
02283	4.051402	3.3E-60	nrdF	ribonucleotide-diphosphate reductase subunit beta,Ribonucleoside-diphosphate reductase 2 subunit beta,ribonucleotide-diphosphate reductase subunit beta,Ribonucleotide reductase%2C beta subunit,ribonucleoside-diphosphate reductase%2C class 1b%2C beta subunit,Ribonucleotide reductase%2C small chain
03255	4.035432	8.23E-55	rbsB_3	putative ABC transport system nucleotide binding/ATPase component,D-ribose-binding periplasmic protein precursor,D-ribose transporter subunit RbsB,ABC-type sugar transport system%2C periplasmic component,rhamnose ABC transporter%2C rhamnose-binding protein,Periplasmic binding proteins and sugar binding domain of LacI family
00792	4.018845	3.96E-60	dinF	DNA-damage-inducible SOS response protein,Multidrug export protein mepA,DNA-damage-inducible SOS response protein,MATE efflux family protein,MatE
05351	3.935913	2.3E-27	unknown	transposase for transposon Tn1721,Transposase and inactivated derivatives%2C TnpA family,Tn3 transposase DDE domain
05103	3.900894	1.17E-49	unknown	heptosyltransferase family protein,lipopolysaccharide core biosynthesis protein,lipopolysaccharide heptosyltransferase I,Glycosyltransferase family 9 (heptosyltransferase)
03547	3.888385	2.38E-27	unknown	beta-lactamase,ribonuclease Z,Predicted metal-dependent RNase%2C consists of a metallo-beta-lactamase domain and an RNA-binding KH domain,Metallo-beta-lactamase superfamily
05030	3.838526	3.7E-20	bhsA_5	outer membrane protein,Multiple stress resistance protein BhsA precursor,hypothetical protein,Protein of unknown function (DUF1471)
03927	3.82519	1.46E-33	yebE	membrane protein,Inner membrane protein yebE,Protein of unknown function (DUF533)

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
03401	3.784435	6.94E-09	smvA	methyl viologen resistance protein SmvA,Methyl viologen resistance protein SmvA,methyl viologen resistance protein SmvA,Arabinose efflux permease,drug resistance MFS transporter%2C drug:H+ antiporter-2 (14 Spanner) (DHA2) family,Major Facilitator Superfamily
03400	3.70673	1.87E-17	pobR_2	Pca regulon regulatory protein PcaR,p-hydroxybenzoate hydroxylase transcriptional activator,DNA-binding transcriptional regulator KdgR,beta-ketoadipate pathway transcriptional regulators%2C PcaR/PcaU/PobR family,Bacterial transcriptional regulator
02803	3.679883	1.68E-29	unknown	iron-uptake factor PiuC,PKHD-type hydroxylase Sbal_3634,Fe(II)-dependent oxygenase superfamily protein,Uncharacterized iron-regulated protein
04320	3.663811	2.86E-09	adiC	Lysine/cadaverine antiporter membrane protein CadB,Arginine/agmatine antiporter,lysine/cadaverine antiporter,Amino acid transporters,transporter%2C basic amino acid/polyamine antiporter (APA) family,Amino acid permease
03253	3.644795	4.98E-31	sgcQ	BtpA family protein,Putative sgc region protein SgcQ,membrane complex biogenesis protein%2C BtpA family,BtpA family
03254	3.608809	9.22E-35	rbsK_4	ribokinase,Ribokinase,ribokinase,Uncharacterized protein conserved in bacteria,ribokinase,pfkB family carbohydrate kinase
01159	3.555718	1.43E-62	degS_1	integral membrane sensor signal transduction histidine kinase,Sensor protein degS,sensory histidine kinase UhpB,Signal transduction histidine kinase,MASE1
04030	3.5407	6.83E-31	srfAD	YbtT,Surfactin synthase thioesterase subunit,Thioesterase domain
02336	3.53523	2.02E-20	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
01158	3.483558	4.48E-57	uhpA_1	transcriptional regulatory protein UhpA,Transcriptional regulatory protein uhpA,DNA-binding transcriptional activator UhpA,Response regulator,nitrogen regulation protein NR(I),Response regulator receiver domain
04399	3.469986	4.31E-08	unknown	hypothetical protein
04877	3.453792	1.02E-38	ycgG_4	putative protease,hypothetical protein,phage resistance protein,Predicted signal transduction protein containing sensor and EAL domains,EAL domain
01152	3.438023	8.2E-30	unknown	putative acyltransferase,Acetyltransferase (GNAT) family
04029	3.436641	3.4E-29	ybtU	thiazolinyl-S-HMWP1 reductase,Oxidoreductase (NAD-binding)%2C involved in siderophore biosynthesis,thiazolinyl imide reductase,Oxidoreductase family%2C NAD-binding Rossmann fold
04712	3.427734	8.91E-60	ybiP	putative enzyme,Putative phosphoethanolamine transferase ybiP,hypothetical protein,Predicted membrane-associated%2C metal-dependent hydrolase,Sulfatase
02687	3.412005	1.27E-67	setB	sugar efflux transporter B,Sugar efflux transporter B,sugar efflux transporter B,Arabinose efflux permease,sugar efflux transporter,Major Facilitator Superfamily
02282	3.409593	6.38E-61	proV_1	L-proline glycine betaine ABC transport system permease ProV,Glycine betaine/L-proline transport ATP-binding protein ProV,glycine betaine transporter ATP-binding subunit,ABC-type proline/glycine betaine transport system%2C ATPase component,glycine betaine/L-proline transport ATP binding subunit,ABC transporter
01644	3.394814	5.67E-67	yhbE	drug/metabolite transporter permease,Uncharacterized inner membrane transporter yhbE,Predicted permeases,carboxylate/amino acid/amine transporter,EamA-like transporter family
03252	3.382691	3.85E-26	unknown	thiaminase II,multifunctional hydroxymethylpyrimidine phosphokinase/4-amino-5-aminomethyl-2-methylpyrimidine hydrolase,TENA/THI-4/PQQC family
02804	3.367348	2.58E-68	fiu	ferrichrome-iron receptor,TonB-dependent receptor Fiu,catecholate siderophore receptor Fiu,Outer membrane receptor for monomeric catechols,TonB-dependent siderophore receptor,TonB dependent receptor
03586	3.31986	7.24E-35	ytcD	transcriptional regulator,Uncharacterized HTH-type transcriptional regulator ytcD,HxlR-like helix-turn-helix
03695	3.277697	5.28E-27	yusV_3	iron(III) dicitrate transport ATP-binding protein,Probable siderophore transport system ATP-binding protein YusV,iron-dicitrate transporter ATP-binding subunit,ABC-type enterochelin transport system%2C ATPase component,proposed F420-0 ABC transporter%2C ATP-binding protein,ABC transporter
04879	3.269746	1.64E-14	unknown	iron dicitrate ABC transporter ATP-binding protein
01981	3.266941	1.49E-17	yhjH_1	cyclic diguanylate phosphodiesterase (EAL) domain-containing protein,Cyclic di-GMP phosphodiesterase YhjH,cyclic-di-GMP phosphodiesterase,Predicted signal transduction protein containing sensor and EAL domains,EAL domain

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
04396	3.255678	2.72E-37	unknown	outer membrane protein romA,metal-dependent hydrolase,Beta-lactamase superfamily domain
00479	3.184359	1.34E-12	unknown	transcriptional regulator,Ketosteroid isomerase-related protein,SnoaL-like domain
02113	3.175575	1.36E-20	fucU	L-fucose mutarotase,L-fucose mutarotase,L-fucose mutarotase,ABC-type ribose transport system%2C auxiliary component,RbsD / FucU transport protein family
02861	3.169823	3.19E-61	norB	transporter,Quinolone resistance protein norB,methyl viologen resistance protein SmvA,Arabinose efflux permease,drug resistance MFS transporter%2C drug:H+ antiporter-2 (14 Spanner) (DHA2) family,Major Facilitator Superfamily
01941	3.152979	1.17E-17	hsrA_2	transporter,High-copy suppressor of rspA,putative transporter,Arabinose efflux permease,drug resistance MFS transporter%2C drug:H+ antiporter-2 (14 Spanner) (DHA2) family,Major Facilitator Superfamily
03183	3.148529	1.17E-27	hipB_1	antitoxin,Antitoxin HipB,antitoxin HipB,Uncharacterized conserved small protein,transcriptional regulator%2C y4mF family,Helix-turn-helix
04401	3.148079	4.23E-66	mdtE_1	RND efflux membrane fusion protein,Multidrug resistance protein MdtE precursor,multidrug efflux system subunit MdtA,efflux transporter%2C RND family%2C MFP subunit,HlyD family secretion protein
05098	3.144093	1.31E-45	yeaN_2	cyanate transporter CynX family,Inner membrane transport protein YeaN,putative cyanate transporter,cyanate transporter,Major Facilitator Superfamily
02884	3.143588	4.22E-18	unknown	cold-shock protein
05199	3.128889	8.51E-17	unknown	hypothetical protein
01737	3.112726	7.81E-57	unknown	B3/4 domain-containing protein,B3/4 domain
02389	3.093449	1.11E-48	yfhL	4Fe-4S ferredoxin,Uncharacterized ferredoxin-like protein yfhL,ferredoxin,archaeoflavoprotein%2C MJ0208 family,4Fe-4S dicluster domain
03416	3.054175	2.24E-18	scrY_2	maltoporin,Sucrose porin precursor,maltoporin,Maltoporin (phage lambda and maltose receptor),LamB porin
05200	3.040068	3.44E-18	unknown	putative dopa decarboxylase protein remnant
00321	3.001855	4.82E-23	yaaH	YaaH protein,Inner membrane protein yaaH,hypothetical protein,GPR1/FUN34/yaaH family
02929	3.001173	6.83E-25	metN_3	ABC transporter%2C ATP-binding protein,Methionine import ATP-binding protein MetN,DL-methionine transporter ATP-binding subunit,ABC-type oligopeptide transport system%2C ATPase component,D-methionine ABC transporter%2C ATP-binding protein,ABC transporter
03587	2.999562	3.47E-60	unknown	major facilitator family transporter,multidrug resistance protein,Major Facilitator Superfamily

Table D.4 – Differentially expressed genes of KC006 exposed to meropenem at MIC after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
04230	3.432722	5.52E-08	glnK	nitrogen regulatory protein P-II,Nitrogen regulatory protein P-II 2,nitrogen regulatory protein P-II 2,Nitrogen regulatory protein P-II
02931	3.187846	3.32E-06	tyrB_2	aromatic amino acid aminotransferase,Aromatic-amino-acid aminotransferase,aromatic amino acid aminotransferase,histidinol-phosphate transaminase,Aminotransferase class I and II
04321	3.050473	0.000196	cadA	lysine decarboxylase 1,Lysine decarboxylase%2C inducible,lysine decarboxylase CadA,Orn/Lys/Arg decarboxylase%2C major domain
02932	3.034143	0.000166	unknown	membrane protein,Uncharacterized BCR%2C YitT family COG1284
05254	2.981431	0.043708	cusB_3	copper/silver efflux system membrane fusion protein CusB,Cation efflux system protein CusB precursor,copper/silver efflux system membrane fusion protein CusB
05253	2.928866	0.032127	cusB_2	copper/silver efflux system membrane fusion protein CusB,Cation efflux system protein CusB precursor,copper/silver efflux system membrane fusion protein CusB
05251	2.798333	0.035904	silC	copper/silver efflux system outer membrane protein CusC,Cation efflux system protein CusC precursor,copper/silver efflux system outer membrane protein CusC,efflux transporter%2C outer membrane factor (OMF) lipoprotein%2C NodT family,Outer membrane efflux protein
04005	2.644231	6.68E-06	oxyR_5	nitrogen assimilation transcriptional regulator,Morphology and auto-aggregation control protein,nitrogen assimilation transcriptional regulator,pca operon transcription factor PcaQ,LysR substrate binding domain
05258	2.199033	0.045538	cusA_5	SilA,Cation efflux system protein CusA,multidrug efflux system subunit MdtC,heavy metal efflux pump%2C CzcA family,AcrB/AcrD/AcrF family
01087	2.189431	4.61E-17	asnA	asparagine synthetase AsnA,Aspartate--ammonia ligase,asparagine synthetase AsnA,aspartate--ammonia ligase,Aspartate-ammonia ligase
03325	2.072717	0.027506	unknown	hydroxymethylpyrimidine ABC transporter substrate-binding protein,Putative thiamine biosynthesis protein HI_0357,ABC-type taurine transport system%2C periplasmic component,ABC transporter%2C substrate-binding protein%2C aliphatic sulfonates family,NMT1/THI5 like
04320	2.062901	0.010593	adiC	Lysine/cadaverine antiporter membrane protein CadB,Arginine/agmatine antiporter,lysine/cadaverine antiporter,Amino acid transporters,transporter%2C basic amino acid/polyamine antiporter (APA) family,Amino acid permease
04231	2.020688	0.000232	amtB	ammonium transporter,Ammonia transporter,ammonium transporter,Ammonia permease,ammonium transporter,Ammonium Transporter Family
03322	2.002974	0.002129	tenA	thiaminase II,Thiaminase-2,multifunctional hydroxymethylpyrimidine phosphokinase/4-amino-5-aminomethyl-2-methylpyrimidine hydrolase,TENA/THI-4/PQQC family
02741	1.960975	0.015842	thiD	phosphomethylpyrimidine kinase,Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase,bifunctional hydroxy-methylpyrimidine kinase/ hydroxy-phosphomethylpyrimidine kinase,Thiamine monophosphate synthase,phosphomethylpyrimidine kinase,Phosphomethylpyrimidine kinase
05089	1.906898	6.84E-05	htrB_3	putative lipid A biosynthesis acyltransferase,Lipid A biosynthesis lauroyl acyltransferase,lipid A biosynthesis palmitoleoyl acyltransferase,Lauroyl/myristoyl acyltransferase,lipid A biosynthesis lauroyl (or palmitoleoyl) acyltransferase,Bacterial lipid A biosynthesis acyltransferase
04707	1.905239	5.35E-07	glnP	glutamate transport membrane-spanning protein,Glutamine transport system permease protein glnP,glutamine ABC transporter permease protein,ABC-type amino acid transport system%2C permease component,ectoine/hydroxyectoine ABC transporter%2C permease protein EhuC,Binding-protein-dependent transport system inner membrane component
04553	1.904683	6.07E-14	asnB	asparagine synthetase B,Asparagine synthetase B [glutamine-hydrolyzing],asparagine synthetase B,Asparagine synthase (glutamine-hydrolyzing),asparagine synthase (glutamine-hydrolyzing),Asparagine synthase
03401	1.902375	0.03203	smvA	methyl viologen resistance protein SmvA,Methyl viologen resistance protein SmvA,methyl viologen resistance protein SmvA,Arabinose efflux permease,drug resistance MFS transporter%2C drug:H+ antiporter-2 (14 Spanner) (DHA2) family,Major Facilitator Superfamily

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
03323	1.824574	0.031764	ssuB_1	ABC transporter ATP-binding protein,Aliphatic sulfonates import ATP-binding protein SsuB,aliphatic sulfonates transport ATP-binding subunit,ABC-type spermidine/putrescine transport systems%2C ATPase components,choline ABC transporter%2C ATP-binding protein,ABC transporter
04706	1.799771	1.11E-08	glnQ_2	glutamate transport ATP-binding protein,Glutamine transport ATP-binding protein GlnQ,glutamine ABC transporter ATP-binding protein,ABC-type histidine transport system%2C ATPase component,ectoine/hydroxyectoine ABC transporter%2C ATP-binding protein EhuA,ABC transporter
01054	1.779093	0.000302	glnA	glutamine synthetase,Glutamine synthetase,glutamine synthetase,Predicted metal-dependent hydrolase of the TIM-barrel fold,glutamine synthetase%2C type I,Glutamine synthetase%2C catalytic domain
00849	1.732487	0.018371	thiS	sulfur carrier protein ThiS,Thiamine biosynthesis protein ThiS,sulfur carrier protein ThiS,thiamine biosynthesis protein ThiS,ThiS family
02930	1.694238	0.006426	metI_3	methionine ABC transporter permease,D-methionine transport system permease protein metI,DL-methionine transporter permease subunit,ABC-type phosphate transport system%2C permease component,phosphonate ABC transporter%2C permease protein PhnE,Binding-protein-dependent transport system inner membrane component
01787	1.67546	0.004034	unknown	transmembrane protein,Permease for cytosine/purines%2C uracil%2C thiamine%2C allantoin
05027	1.65082	6.68E-06	unknown	17 kDa surface antigen,hypothetical protein,Uncharacterized protein conserved in bacteria,Glycine zipper 2TM domain
05143	1.580588	0.013323	astC_2	succinylornithine transaminase,Succinylornithine transaminase,bifunctional succinylornithine transaminase/acetylornithine transaminase,Ornithine/acetylornithine aminotransferase,succinylornithine transaminase family,Aminotransferase class-III
04708	1.537804	7.97E-05	glnH_2	glutamine ABC transporter periplasmic protein,Glutamine-binding periplasmic protein precursor,glutamine ABC transporter periplasmic protein,Response regulator containing a CheY-like receiver domain and a GGDEF domain,lysine-arginine-ornithine-binding periplasmic protein,Bacterial extracellular solute-binding proteins%2C family 3
03441	1.533531	0.031764	yncA_2	hypothetical acetyltransferase,N-acyltransferase YncA,Uncharacterized protein conserved in bacteria,pseudaminic acid biosynthesis N-acetyl transferase,Acetyltransferase (GNAT) family
04893	1.459235	0.017405	unknown	hypothetical protein
00300	1.446784	0.02394	ygdR_1	lipoprotein,Uncharacterized lipoprotein ygdR precursor,Bacterial protein of unknown function (DUF903)
04539	1.413404	4.77E-05	gltI	glutamate and aspartate transporter subunit,Glutamate/aspartate periplasmic-binding protein precursor,glutamate and aspartate transporter subunit,lysine-arginine-ornithine-binding periplasmic protein,Bacterial extracellular solute-binding proteins%2C family 3
02934	1.411064	0.039017	aroP_2	aromatic amino acid transport protein AroP,General aromatic amino acid permease,aromatic amino acid transporter,Amino acid transporters,GABA permease,Amino acid permease
03324	1.402612	0.04793	ssuC_1	hydroxymethylpyrimidine ABC transporter transmembrane protein,Putative aliphatic sulfonates transport permease protein ssuC,alkanesulfonate transporter permease subunit,ABC-type anion transport system%2C duplicated permease component,choline ABC transporter%2C permease protein,Binding-protein-dependent transport system inner membrane component
05145	1.355016	0.041192	astD_2	succinylglutamic semialdehyde dehydrogenase,N-succinylglutamate 5-semialdehyde dehydrogenase,succinylglutamic semialdehyde dehydrogenase,Delta 1-pyrroline-5-carboxylate dehydrogenase,succinylglutamic semialdehyde dehydrogenase,Aldehyde dehydrogenase family
03918	1.352585	0.006407	unknown	hypothetical protein
04481	1.279363	0.022687	codA_3	cytosine deaminase,Cytosine deaminase,cytosine deaminase,Cytosine deaminase and related metal-dependent hydrolases,imidazolonepropionase,Amidohydrolase family
03495	1.273191	5.37E-06	rnfA_2	electron transport complex protein RnfA,Electron transport complex protein rnfA,electron transport complex protein RsxA,Predicted NADH:ubiquinone oxidoreductase%2C subunit RnfA,electron transport complex%2C RnfABCDGE type%2C A subunit,Rnf-Nqr subunit%2C membrane protein
04537	1.261671	0.002816	gltK_2	glutamate Aspartate transport system permease protein GltJ,Glutamate/aspartate transport system permease protein gltK,glutamate/aspartate transport system permease GltK,ABC-type amino acid

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
				transport system%2C permease component,ectoine/hydroxyectoine ABC transporter%2C permease protein EhuC,Binding-protein-dependent transport system inner membrane component
02843	1.216478	0.005237	unknown	Osmotically inducible lipoprotein B,lipoprotein,Glycine zipper 2TM domain
02378	1.20592	1.88E-08	rpoE	RNA polymerase sigma factor RpoE,Sigma-24,RNA polymerase sigma factor AlgU,RNA polymerase sigma factor RpoE,Sigma-70%2C region 4
04536	1.175896	0.000411	gltK_1	glutamate Aspartate transport system permease protein GltK,Glutamate/aspartate transport system permease protein gltK,glutamate/aspartate transport system permease GltK,ABC-type amino acid transport system%2C permease component,ectoine/hydroxyectoine ABC transporter%2C permease protein EhuC,Binding-protein-dependent transport system inner membrane component
03190	1.174429	0.001421	unknown	hypothetical protein
01799	1.136831	1.92E-05	unknown	arylsulfatase,SH3 domain-containing protein,Uncharacterized protein with a bacterial SH3 domain homologue,SH3 domain protein
05249	1.13567	0.006503	cusS_2	sensor kinase CusS,Sensor kinase CusS,sensor kinase CusS,Predicted periplasmic ligand-binding sensor domain,heavy metal sensor kinase,Histidine kinase-%2C DNA gyrase B-%2C and HSP90-like ATPase
02381	1.125758	1.69E-07	rseC	Sigma factor RpoE regulatory protein RseC,Sigma-E factor regulatory protein rseC,SoxR reducing system protein RseC,Positive regulator of sigma E activity,Positive regulator of sigma(E)%2C RseC/MucC
03494	1.096698	0.000124	rnfA_1	electron transport complex protein RnfA,Electron transport complex protein rnfA,electron transport complex protein RsxA,Predicted NADH:ubiquinone oxidoreductase%2C subunit RnfA,electron transport complex%2C RnfABCDGE type%2C A subunit,Rnf-Nqr subunit%2C membrane protein
04995	1.090401	2.83E-05	yceL	multidrug resistance protein MdtH,Multidrug resistance protein MdtH,multidrug resistance protein MdtH,Arabinose efflux permease,multidrug resistance protein,Major Facilitator Superfamily
00043	1.087747	0.00018	spr_1	lipoprotein nlpC,Probable endopeptidase Spr precursor,outer membrane lipoprotein,putative phage cell wall peptidase%2C NlpC/P60 family,NlpC/P60 family
03508	1.084462	0.002553	mliC	lysozyme inhibitor,Membrane-bound lysozyme inhibitor of C-type lysozyme precursor,lysozyme inhibitor,Membrane-bound lysozyme-inhibitor of c-type lysozyme
04535	1.077935	0.000216	artM_4	glutamate Aspartate transport ATP-binding protein GltL,Arginine transport ATP-binding protein ArtM,glutamine ABC transporter ATP-binding protein,ABC-type histidine transport system%2C ATPase component,ectoine/hydroxyectoine ABC transporter%2C ATP-binding protein EhuA,ABC transporter
02292	1.072352	1.6E-06	alaE	inner membrane protein,L-alanine exporter AlaE,Protein of unknown function (DUF1144)
02168	1.06981	0.0005	unknown	Protein of unknown function (DUF3142)
01055	1.057552	1.07E-05	glnL	nitrogen regulation protein NR(II),Nitrogen regulation protein NR(II),nitrogen regulation protein NR(II),Signal transduction histidine kinase%2C nitrogen specific,phosphate regulon sensor kinase PhoR,Histidine kinase-%2C DNA gyrase B-%2C and HSP90-like ATPase
04136	1.053044	0.030728	yaiY	inner membrane protein,Inner membrane protein yaiY,Protein of unknown function (DUF2755)
03498	1.042584	0.000277	nqrB_2	electron transport complex protein RnfD,Na(+)-translocating NADH-quinone reductase subunit B,electron transport complex protein RnfD,Predicted NADH:ubiquinone oxidoreductase%2C subunit RnfD,electron transport complex%2C RnfABCDGE type%2C D subunit,NQR2%2C RnfD%2C RnfE family
04457	1.04155	0.001896	unknown	bacteriophage protein,Plasmid stabilisation system protein
03885	1.03564	0.000205	slp	outer membrane protein Slp,Outer membrane protein slp precursor,Starvation-inducible outer membrane lipoprotein,outer membrane lipoprotein%2C Slp family,Outer membrane lipoprotein Slp family
04482	1.026685	0.030728	unknown	transcriptional regulatory protein,Sugar diacid utilization regulator,Purine catabolism regulatory protein-like family
03496	1.013297	5.52E-08	rnfB	electron transport complex protein RnfB,Nitrogen fixation protein rnfB,electron transport complex protein RnfB,Dihydroorotate dehydrogenase,electron transport complex%2C RnfABCDGE type%2C B subunit,Putative Fe-S cluster

Table D.5 – Differentially expressed genes of KC006 exposed to gentamicin at MIC after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01137	5.44	9.08E-13	ibpB	16 kDa heat shock protein B,16 kDa heat shock protein B,heat shock chaperone IbpB,Hsp20/alpha crystallin family
00315	4.17	0.000296	nhaA	Na+/H+ antiporter NhaA type,Sodium/proton antiporter nhaA,pH-dependent sodium/proton antiporter,Protein-disulfide isomerase,Na+/H+ antiporter NhaA,Na+/H+ antiporter 1
01008	2.84	0.001362	cpxP	P pilus assembly/Cpx signaling pathway,Periplasmic protein CpxP precursor,periplasmic repressor CpxP,P pilus assembly/Cpx signaling pathway%2C periplasmic inhibitor/zinc-resistance associated protein,LTXXQ motif
04484	2.25	0.045114	mtnB	methylthioribulose-1-phosphate dehydratase,Methylthioribulose-1-phosphate dehydratase,methylthioribulose-1-phosphate dehydratase,Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases,methylthioribulose-1-phosphate dehydratase,Class II Aldolase and Adducin N-terminal domain
00314	2.04	0.006877	nhaR	LysR family transcriptional regulator,Na(+)/H(+) antiporter regulatory protein,transcriptional activator NhaR,pca operon transcription factor PcaQ,Bacterial regulatory helix-turn-helix protein%2C lysR family
01224	1.78	0.045114	cdh	CDP-diacylglycerol pyrophosphatase,CDP-diacylglycerol pyrophosphatase,CDP-diacylglycerol pyrophosphatase,CDP-diacylglycerol pyrophosphatase,CDP-diacylglycerol pyrophosphatase
05089	1.37	0.049917	htrB_3	putative lipid A biosynthesis acyltransferase,Lipid A biosynthesis lauroyl acyltransferase,lipid A biosynthesis palmitoleoyl acyltransferase,Lauroyl/myristoyl acyltransferase,lipid A biosynthesis lauroyl (or palmitoleoyl) acyltransferase,Bacterial lipid A biosynthesis acyltransferase
00565	1.13	0.029497	mgtA	magnesium-transporting ATPase MgtA,Magnesium-transporting ATPase%2C P-type 1,magnesium-transporting ATPase MgtA,Uncharacterized protein conserved in bacteria,magnesium-translocating P-type ATPase,E1-E2 ATPase
00690	1.08	0.048248	fxsA	FxsA protein,Suppressor of F exclusion of phage T7,phage T7 F exclusion suppressor FxsA,Protein affecting phage T7 exclusion by the F plasmid,FxsA cytoplasmic membrane protein
01650	0.90	0.002849	rrmJ	23S rRNA methyltransferase J,Ribosomal RNA large subunit methyltransferase E,23S rRNA methyltransferase J,Predicted RNA-binding protein%2C contains TRAM domain,ribosomal RNA large subunit methyltransferase J,FtsJ-like methyltransferase
01799	0.89	0.013875	unknown	arylsulfatase,SH3 domain-containing protein,Uncharacterized protein with a bacterial SH3 domain homologue,SH3 domain protein
01644	0.83	0.007633	yhbE	drug/metabolite transporter permease,Uncharacterized inner membrane transporter yhbE,Predicted permeases,carboxylate/amino acid/amine transporter,EamA-like transporter family
02378	0.83	0.002849	rpoE	RNA polymerase sigma factor RpoE,Sigma-24,RNA polymerase sigma factor AlgU,RNA polymerase sigma factor RpoE,Sigma-70%2C region 4
02379	0.67	0.002849	rseA	anti-RNA polymerase sigma factor SigE,Sigma-E factor negative regulatory protein,anti-RNA polymerase sigma factor SigE,Anti sigma-E protein RseA%2C N-terminal domain
03951	0.66	0.032896	cutC	cytoplasmic copper homeostasis protein cutC,Copper homeostasis protein CutC,copper homeostasis protein CutC,Uncharacterized protein involved in copper resistance,CutC family
01120	0.65	0.049917	unknown	RNase P
01651	0.65	0.027891	hflB	ATP-dependent metalloprotease,ATP-dependent zinc metalloprotease FtsH,ATP-dependent metalloprotease,Ribosome-associated chaperone zuotin,ATP-dependent metallopeptidase HflB,Peptidase family M41
02111	0.64	0.001362	rlmM	RNA 2'-O-ribose methyltransferase mtfA,Ribosomal RNA large subunit methyltransferase M,putative 23S rRNA C2498 ribose 2'-O-ribose methyltransferase,FtsJ-like methyltransferase

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01466	0.59	0.028464	hslO	33 kDa chaperonin (Heat shock protein 33) (HSP33),Heat shock protein 33,Hsp33-like chaperonin,Phosphoenolpyruvate carboxykinase (ATP),Hsp33 protein
00144	0.57	0.016361	yaeL	zinc metallopeptidase RseP,Regulator of sigma E protease,zinc metallopeptidase RseP,Translation elongation factor Ts,RIP metalloprotease RseP,Peptidase family M50
00889	0.51	0.032896	tatD	Deoxyribonuclease TatD,Deoxyribonuclease TatD,DNase TatD,Mg-dependent DNase,hydrolase%2C TatD family,TatD related DNase
01862	0.50	0.029497	unknown	hypothetical protein,Protein of unknown function (DUF2884)
04076	0.50	0.026039	cobT_2	nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase,Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase,nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase,nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase,Phosphoribosyltransferase
01797	0.45	0.048248	uppP	undecaprenyl pyrophosphate phosphatase,Undecaprenyl-diphosphatase,undecaprenyl pyrophosphate phosphatase,undecaprenyl-diphosphatase UppP,Bacitracin resistance protein BacA
04107	0.45	0.048248	kpsM	ABC transporter permease,Polysialic acid transport protein kpsM,ABC-2 type transporter
00890	0.44	0.016361	tatC	twin-arginine translocation protein TatC,Sec-independent protein translocase protein TatC,twin-arginine protein translocation system subunit TatC,twin arginine-targeting protein translocase TatC,Sec-independent protein translocase protein (TatC)
04603	0.34	0.023216	tolA_2	cell envelope integrity inner membrane protein TolA,hypothetical protein,cell envelope integrity inner membrane protein TolA,Membrane protein involved in colicin uptake,protein TolA,TolA C-terminal

Table D.6 – Differentially expressed genes of KC006 exposed to heat at 45°C after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
04630	3.46	8.45E-13	galM_2	aldose 1-epimerase,Aldose 1-epimerase,galactose-1-epimerase,galactose mutarotase,Aldose 1-epimerase
04631	3.43	1.69E-12	galK	galactokinase,Galactokinase,galactokinase,Galactokinase,galactokinase,Galactokinase galactose-binding signature
01026	3.14	1.5E-06	lacY_1	galactoside permease,Lactose-proton symport,galactoside permease,Uncharacterized protein conserved in bacteria,oligosaccharide:H+ symporter,LacY proton/sugar symporter
01027	3.01	6.31E-06	rafA	alpha-galactosidase,Alpha-galactosidase,Alpha-galactosidase,Melibiase
01876	2.98	2.1E-05	galP	Arabinose-proton symporter,Galactose transporter,D-xylose transporter XylE,Transcription initiation factor TFIID%2C subunit TAF9 (also component of histone acetyltransferase SAGA),MFS transporter%2C sugar porter (SP) family,Sugar (and other) transporter
00455	2.70	2.29E-23	trpS2	tryptophanyl-tRNA synthetase,Tryptophan--tRNA ligase 2,tryptophanyl-tRNA synthetase II,Tryptophanyl-tRNA synthetase,tryptophan--tRNA ligase,tRNA synthetases class I (W and Y)
04083	2.65	2.31E-37	fusA_2	translation elongation factor G,Elongation factor G,elongation factor G,GTPases - translation elongation factors,translation elongation factor G,Elongation factor Tu GTP binding domain
04632	2.62	3.12E-10	galT	galactose-1-phosphate uridylyltransferase,Galactose-1-phosphate uridylyltransferase,galactose-1-phosphate uridylyltransferase,galactose-1-phosphate uridylyltransferase,Galactose-1-phosphate uridyl transferase%2C N-terminal domain
01080	2.61	2.53E-10	rbsB_2	ribose ABC transport system,D-ribose-binding periplasmic protein precursor,D-ribose transporter subunit RbsB,ABC-type sugar transport system%2C periplasmic component,D-xylose ABC transporter%2C D-xylose-binding protein,Periplasmic binding proteins and sugar binding domain of LacI family
02802	2.54	7.01E-05	ompW	outer membrane protein W,Outer membrane protein W precursor,outer membrane protein W,Outer membrane protein W,OmpW family
01083	2.46	4.07E-08	rbsD	ribose ABC transport system,D-ribose pyranase,D-ribose pyranase,ABC-type ribose transport system%2C auxiliary component,RbsD / FucU transport protein family
01082	2.40	1.39E-09	rbsA_4	ribose ABC transport system,Ribose import ATP-binding protein RbsA,D-ribose transporter ATP binding protein,Predicted periplasmic lipoprotein,D-xylose ABC transporter%2C ATP-binding protein,ABC transporter
01081	2.36	2.39E-07	rbsC_3	ribose ABC transport system,Ribose transport system permease protein rbsC,ribose ABC transporter permease protein,Predicted ABC-type sugar transport system%2C permease component,urea ABC transporter%2C permease protein UrtB,Branched-chain amino acid transport system / permease component
01762	2.35	1.94E-07	dhaK_2	phosphoenolpyruvate-dihydroxyacetone phosphotransferase,PTS-dependent dihydroxyacetone kinase%2C dihydroxyacetone-binding subunit dhaK,dihydroxyacetone kinase subunit DhaK,Transcriptional regulator%2C contains sigma factor-related N-terminal domain,dihydroxyacetone kinase%2C DhaK subunit,Dak1 domain
00599	2.34	1.06E-07	ytfQ	periplasmic binding protein/LacI transcriptional regulator,ABC transporter periplasmic-binding protein ytfQ precursor,D-ribose transporter subunit RbsB,ABC-type sugar transport system%2C periplasmic component,D-xylose ABC transporter%2C D-xylose-binding protein,Periplasmic binding proteins and sugar binding domain of LacI family
03217	2.33	0.005245	fucP_2	fucose/glucose/galactose permease,L-fucose permease,L-fucose transporter,Arabinose efflux permease,L-fucose:H+ symporter permease,Major Facilitator Superfamily
03216	2.25	2.13E-05	unknown	putative glyoxalase/bleomycin resistance protein/dioxygenas,Glyoxalase-like domain
01137	2.20	0.00724	ibpB	16 kDa heat shock protein B,16 kDa heat shock protein B,heat shock chaperone IbpB,Hsp20/alpha crystallin family
01761	2.20	6.69E-07	dhaL	phosphoenolpyruvate-dihydroxyacetone phosphotransferase,PTS-dependent dihydroxyacetone kinase%2C ADP-binding subunit dhaL,dihydroxyacetone kinase subunit DhaL,dihydroxyacetone kinase%2C L subunit,DAK2 domain
01760	2.17	7.67E-07	dhaM	phosphoenolpyruvate-dihydroxyacetone phosphotransferase,PTS-dependent dihydroxyacetone kinase%2C phosphotransferase subunit dhaM,dihydroxyacetone kinase subunit M,Uncharacterized protein conserved in bacteria,dihydroxyacetone kinase%2C phosphotransfer subunit,PTS system fructose IIA component

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
04643	2.15	2.99E-13	unknown	hypothetical protein
04644	2.10	9.91E-14	unknown	two-component-system connector protein YcgZ
05318	2.05	4.82E-06	lacY_6	galactoside permease,Lactose-proton symport,galactoside permease,Uncharacterized protein conserved in bacteria,oligosaccharide:H+ symporter,LacY proton/sugar symporter
02745	2.02	4.9E-06	rbtD	ribitol 2-dehydrogenase,Ribitol 2-dehydrogenase,3-ketoacyl-(acyl-carrier-protein) reductase,Uncharacterized conserved protein,3-hydroxybutyrate dehydrogenase,short chain dehydrogenase
01087	2.01	4.77E-15	asnA	asparagine synthetase AsnA,Aspartate--ammonia ligase,asparagine synthetase AsnA,aspartate--ammonia ligase,Aspartate-ammonia ligase
03300	1.98	3.62E-08	yybR_1	HxlR family transcriptional regulator,Uncharacterized HTH-type transcriptional regulator yybR,HxlR-like helix-turn-helix
05317	1.96	7.49E-08	lacZ_2	beta-D-galactosidase,Beta-galactosidase,beta-D-galactosidase,Tryptophan synthase alpha chain,Glycosyl hydrolases family 2%2C TIM barrel domain
02803	1.95	9.96E-08	unknown	iron-uptake factor PiuC,PKHD-type hydroxylase Sbal_3634,Fe(II)-dependent oxygenase superfamily protein,Uncharacterized iron-regulated protein
03302	1.92	4.29E-16	fabF_2	3-oxoacyl-ACP synthase,3-oxoacyl-[acyl-carrier-protein] synthase 2,3-oxoacyl-(acyl carrier protein) synthase II,Enoyl reductase domain of yeast-type FAS1,beta-ketoacyl-acyl-carrier-protein synthase II,Beta-ketoacyl synthase%2C N-terminal domain
01975	1.87	4.45E-08	unknown	outer membrane protein,Protein of unknown function (DUF1471)
03301	1.84	2.83E-09	fabF_1	3-oxoacyl-ACP synthase,3-oxoacyl-[acyl-carrier-protein] synthase 2,3-oxoacyl-(acyl carrier protein) synthase II,beta-ketoacyl-acyl-carrier-protein synthase II,Beta-ketoacyl synthase%2C N-terminal domain
04633	1.83	1.06E-07	galE	UDP-galactose-4-epimerase,UDP-glucose 4-epimerase,UDP-galactose-4-epimerase,Putative NADH-flavin reductase,UDP-glucose 4-epimerase,NAD dependent epimerase/dehydratase family
03203	1.82	8.52E-12	yidP_2	HTH-type transcriptional regulator YidP,Uncharacterized HTH-type transcriptional regulator yidP,DNA-binding transcriptional repressor MngR,Transcriptional regulators,trehalose operon repressor,UTRA domain
00560	1.78	3.25E-07	frlB_2	phosphosugar isomerase,Fructosamine deglycase frlB,fructoselysine-6-P-deglycase,Predicted phosphosugar isomerases,glutamine-fructose-6-phosphate transaminase (isomerizing),SIS domain
02374	1.74	1.8E-05	grcA	pyruvate formate-lyase,Autonomous glycyl radical cofactor,autonomous glycyl radical cofactor GrcA,formate acetyltransferase,Glycine radical
03218	1.73	0.047712	rbsK_3	putative ribokinase,Ribokinase,ribokinase,ADP-heptose synthase%2C bifunctional sugar kinase/adenylyltransferase,ribokinase,pfkB family carbohydrate kinase
04642	1.73	4.54E-09	ariR	two-component-system connector protein AriR,Probable two-component-system connector protein AriR,Biofilm development protein YmgB/AriR
03230	1.72	4.19E-11	marA_3	transcriptional regulator,Multiple antibiotic resistance protein marA,transcriptional activator FtrA,DNA gyrase inhibitor,4-hydroxyphenylacetate catabolism regulatory protein HpaA,Helix-turn-helix domain
05030	1.69	0.00048	bhsA_5	outer membrane protein,Multiple stress resistance protein BhsA precursor,hypothetical protein,Protein of unknown function (DUF1471)
02708	1.69	4.35E-10	yohJ	LrgA family protein,hypothetical protein,hypothetical protein,Putative effector of murein hydrolase LrgA,LrgA family
01079	1.67	7.06E-08	rbsK_2	ribokinase,Ribokinase,ribokinase,Uncharacterized protein conserved in bacteria,ribokinase,pfkB family carbohydrate kinase
00687	1.67	0.004556	groL	Heat shock protein 60 family chaperone GroEL,hypothetical protein,chaperonin GroEL,chaperonin GroL,TCP-1/cpn60 chaperonin family
00688	1.62	0.026274	groS	Heat shock protein 60 family co-chaperone GroES,hypothetical protein,co-chaperonin GroES,Chaperonin 10 Kd subunit
04694	1.62	0.002303	fabG_12	putative short-chain dehydrogenase/reductase SDR,3-oxoacyl-[acyl-carrier-protein] reductase FabG,3-ketoacyl-(acyl-carrier-protein) reductase,Uncharacterized conserved protein,3-oxoacyl-[acyl-carrier-protein] reductase,short chain dehydrogenase
01763	1.61	2.63E-06	dhaD	glycerol dehydrogenase,Glycerol dehydrogenase,glycerol dehydrogenase,Glycerol dehydrogenase and related enzymes,lactaldehyde reductase,Iron-containing alcohol dehydrogenase
00592	1.61	0.049655	unknown	thioredoxin-like protein
04084	1.61	1.59E-08	unknown	transcriptional regulator,Helix-turn-helix domain
00574	1.57	4.59E-09	nrdD	anaerobic ribonucleoside triphosphate reductase,Anaerobic ribonucleoside-triphosphate reductase,anaerobic ribonucleoside triphosphate reductase,anaerobic ribonucleoside-triphosphate reductase,Glycine radical
02299	1.56	3.41E-09	yafC	LysR family transcriptional regulator,D-malate degradation protein R,LysR family transcriptional regulator,ABC-type uncharacterized transport system%2C periplasmic

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
				component,putative choline sulfate-utilization transcription factor,LysR substrate binding domain
00576	1.56	1.57E-06	relE	RelE/StbE replicon stabilization toxin,mRNA interferase RelE,addiction module toxin%2C RelE/StbE family,Plasmid stabilisation system protein
03488	1.56	2.01E-10	ynfM_2	major facilitator transporter,Inner membrane transport protein ynfM,putative arabinose transporter,Arabinose efflux permease,MFS transporter%2C aromatic acid:H+ symporter (AAHS) family,Major Facilitator Superfamily
02701	1.56	0.00321	mglB_1	galactose ABC transporter substrate-binding protein,D-galactose/ D-glucose-binding protein,methyl-galactoside ABC transporter galactose-binding periplasmic protein MglB,ABC-type sugar transport system%2C periplasmic component,D-xylose ABC transporter%2C D-xylose-binding protein,Periplasmic binding proteins and sugar binding domain of LacI family
02879	1.53	0.000483	dmlR_12	LysR family transcriptional regulator,D-malate degradation protein R,LysR family transcriptional regulator,ABC-type uncharacterized transport system%2C periplasmic component,putative choline sulfate-utilization transcription factor,LysR substrate binding domain
01927	1.52	1.87E-12	yqfA	membrane protein hemolysin III,hypothetical protein,hemolysin,Predicted membrane protein%2C hemolysin III homolog,channel protein%2C hemolysin III family,Haemolysin-III related
03129	1.49	0.005303	marA_2	multiple antibiotic resistance protein MarA,Multiple antibiotic resistance protein marA,DNA-binding transcriptional activator MarA,DNA gyrase inhibitor,YSIRK-targeted surface antigen transcriptional regulator,Helix-turn-helix domain
02707	1.49	2.08E-07	yohK_2	LrgA-associated membrane protein LrgB,Inner membrane protein yohK,hypothetical protein,Cytidylyltransferase,conserved hypothetical protein,LrgB-like family
02404	1.48	3.02E-10	mlc_1	NAGC-like transcriptional regulator,Making large colonies protein,N-acetylmannosamine kinase,ROK family protein (putative glucokinase),ROK family
03113	1.46	5.55E-07	lacI_1	lac operon transcriptional repressor,Lactose operon repressor,lac repressor,ABC-type sugar transport system%2C periplasmic component,catabolite control protein A,Bacterial regulatory proteins%2C lacI family
01453	1.46	0.009936	gntT	High-affinity gluconate transporter GntT,Gnt-I system,high-affinity gluconate transporter,H+/gluconate symporter and related permeases,transporter%2C gluconate:H+ symporter (GntP) family,GntP family permease
01595	1.44	0.000321	ycfR_2	protein YcfR,Multiple stress resistance protein BhsA precursor,hypothetical protein,Protein of unknown function (DUF1471)

D.2 Differential expression – Different isolates in the presence of ciprofloxacin

Results are ordered by log_2 fold change in expression. Only upregulated transcripts encoding predicted coding regions are included. Genes encoded on a plasmid are highlighted in red. Gene product details are derived from annotations from Prokka.

Table D.7 – Differentially expressed genes of KC036 exposed to ciprofloxacin after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01190	2.636683	4E-26	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
00032	2.202184	1.17E-07	unknown	toxic peptide TisB
01577	2.084433	5.75E-05	unknown	IS2 repressor TnpA,IS2 repressor TnpA,Transposase
01575	2.051255	2.02E-10	umuD_1	DNA polymerase V subunit UmuD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
01576	2.009286	4.64E-11	umuC_1	Error-prone,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
00031	1.957746	9.05E-13	unknown	putative acyltransferase,Acetyltransferase (GNAT) family
01115	1.89756	7.11E-15	recA	recombinase A,Recombinase A,recombinase A,RecA-superfamily ATPases implicated in signal transduction,protein RecA,recA bacterial DNA recombination protein
03066	1.796562	5.37E-06	umuC_3	Error-prone%2C lesion bypass DNA polymerase V (UmuC),hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
04307	1.783931	5.2E-17	polB	DNA polymerase II,DNA polymerase II,DNA polymerase II,DNA polymerase (pol2),DNA polymerase family B
04115	1.736501	2.2E-15	dinB	DNA polymerase IV,DNA polymerase IV,DNA polymerase IV,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
04808	1.728324	7.11E-15	lexA_2	SOS-response repressor and protease LexA,LexA repressor,LexA repressor,repressor LexA,LexA DNA binding domain
03067	1.59924	0.013223	umuD_2	Error-prone repair protein UmuD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
05359	1.57512	3.9E-06	umuD_3	Protein impA,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
05360	1.575101	2.36E-11	umuC_4	DNA polymerase V subunit UmuC,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
01915	1.532369	0.000255	yebE	membrane protein,Inner membrane protein yebE,Protein of unknown function (DUF533)
01158	1.524011	0.039837	fimA_1	fimbrial protein,Type-1A pilin,fimbrial protein,P pilus assembly protein%2C pilin FimA,Fimbrial protein
01914	1.52095	0.001381	unknown	LexA family transcriptional regulator,DNA damage-inducible protein YebG,Uncharacterized protein conserved in bacteria,YebG protein

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
04807	1.512591	8.95E-06	dinF	DNA-damage-inducible SOS response protein,Multidrug export protein mepA,DNA-damage-inducible SOS response protein,MATE efflux family protein,MatE
01116	1.461031	1.2E-05	recX	Regulatory protein RecX,Regulatory protein recX,recombination regulator RecX,Uncharacterized protein conserved in bacteria,RecX family
01741	1.431275	0.000588	sbmC	DNA gyrase inhibitory protein,DNA gyrase inhibitor,DNA gyrase inhibitor,DNA gyrase inhibitor,Bacterial transcription activator%2C effector binding domain
03454	1.405432	1.95E-08	dinG_2	ATP-dependent DNA helicase DinG,Probable ATP-dependent helicase dinG homolog,ATP-dependent DNA helicase DinG,putative DnaQ family exonuclease/DinG family helicase,DEAD/DEAH box helicase
04794	1.38585	1.73E-08	uvrA	excinuclease ABC subunit A,Excinuclease ABC subunit A,excinuclease ABC subunit A,ABC-type transport system involved in cytochrome bd biosynthesis%2C fused ATPase and permease components,excinuclease ABC subunit A,ABC transporter
02750	1.377804	5.96E-06	xylF_2	alpha/beta fold family hydrolase,2-hydroxymuconic semialdehyde hydrolase,acetoin dehydrogenase E2 subunit dihydrolipoyllysine-residue acetyltransferase,2-succinyl-6-hydroxy-2%2C4-cyclohexadiene-1-carboxylate synthase,Alpha/beta hydrolase family
03214	1.365433	0.000346	dinI_2	DNA-damage-inducible protein I,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
03297	1.303139	3.03E-05	sulA	cell division inhibitor,Cell division inhibitor SulA,SOS cell division inhibitor,SOS-response cell division inhibitor%2C blocks FtsZ ring formation,cell division inhibitor SulA,Cell division inhibitor SulA
05131	1.291159	6.22E-05	xerD_2	site-specific recombinase%2C phage integrase family,Tyrosine recombinase XerD,site-specific tyrosine recombinase XerD,Site-specific recombinase XerD,tyrosine recombinase XerD,Phage integrase family
05058	1.284433	1.89E-05	aes	lipase%2C GDXG family,Acetyl esterase,acetyl esterase,alpha/beta hydrolase fold
04815	1.240574	0.014481	malK_3	maltose/maltodextrin transport ATP-binding protein MalK,Maltose/maltodextrin import ATP-binding protein MalK,maltose/maltodextrin transporter ATP-binding protein,ABC-type spermidine/putrescine transport systems%2C ATPase components,putative 2-aminoethylphosphonate ABC transporter%2C ATP-binding protein,ABC transporter
01742	1.199383	0.011876	yeeA	inner membrane protein,Inner membrane protein yeeA,Fusaric acid resistance protein family
00270	1.196328	0.039417	arnC	Polymyxin resistance protein ArnC,Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase,undecaprenyl phosphate 4-deoxy-4-formamido-L-arabinose transferase,Predicted glycosyl hydrolase,glycosyltransferase,Glycosyl transferase family 2
04593	1.157962	0.013739	nrdD	anaerobic ribonucleoside triphosphate reductase,Anaerobic ribonucleoside-triphosphate reductase,anaerobic ribonucleoside triphosphate reductase,anaerobic ribonucleoside-triphosphate reductase,Glycine radical
00269	1.134554	0.026255	arnB_1	UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase,UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase,UDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase,Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis,UDP-4-keto-6-deoxy-N-acetylglucosamine 4-aminotransferase,DegT/DnrJ/EryC1/StrS aminotransferase family
04114	1.077349	0.000588	unknown	acetyltransferase YafP,putative acyltransferase
03023	0.967178	0.048072	cho	excinuclease cho (excinuclease ABC alternative C subunit),Excinuclease cho,nucleotide excision repair endonuclease,Nuclease subunit of the excinuclease complex,excinuclease ABC subunit C
03825	0.959946	0.014355	unknown	hypothetical protein
04818	0.944362	0.039837	malG_2	maltose/maltodextrin ABC transporter,Maltose transport system permease protein malG,maltose transporter permease,ABC-type maltose transport systems%2C permease component,NifC-like ABC-type porter,Binding-protein-dependent transport system inner membrane component
01901	0.936969	0.009134	ruvA	Holliday junction DNA helicase RuvA,Holliday junction ATP-dependent DNA helicase RuvA,Holliday junction DNA helicase RuvA,Holliday junction DNA helicase RuvA,RuvA N terminal domain
01490	0.876135	0.024754	nrdA	ribonucleotide reductase of class Ia,Ribonucleoside-diphosphate reductase 1 subunit alpha,ribonucleotide-diphosphate reductase subunit alpha,ribonucleoside-diphosphate reductase%2C alpha subunit,Ribonucleotide reductase%2C barrel domain

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01902	0.866691	0.008247	ruvB	Holliday junction DNA helicase RuvB,Holliday junction ATP-dependent DNA helicase RuvB,Holliday junction DNA helicase RuvB,Holliday junction resolvasome%2C helicase subunit,Holliday junction DNA helicase RuvB,Holliday junction DNA helicase ruvB N-terminus
04925	0.79885	0.003667	uvrD	DNA-dependent helicase II,DNA helicase II,DNA-dependent helicase II,Superfamily I DNA and RNA helicases,DNA helicase II,UvrD/REP helicase
04793	0.66485	0.027581	ssb_2	ssDNA-binding protein,Helix-destabilizing protein,single-stranded DNA-binding protein,single-stranded DNA-binding protein,Single-strand binding protein family

Table D.8 – Differentially expressed genes of KC153 exposed to ciprofloxacin after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
01182	2.359577	1.13E-25	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
03034	2.104344	0.000186	unknown	hypothetical protein
02937	2.005459	0.001729	umuD_2	Error-prone repair protein UmuD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
05163	1.88506	1.24E-13	ceaC	Ccl,Colicin-E3,Colicin-like bacteriocin tRNase domain
05075	1.797335	0.000169	umuD_4	umuDC operon protein-like protein,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
02938	1.778522	0.003604	dinI_2	virulence protein,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
05076	1.743257	5.44E-09	umuC_4	DNA polymerase V subunit UmuC,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
00031	1.711177	5.73E-05	unknown	toxic peptide TisB
05164	1.701325	1.98E-11	imm	cloacin immunity protein,Microcin-E3 immunity protein,Cloacin immunity protein
01770	1.594775	1.53E-07	unknown	LexA family transcriptional regulator,DNA damage-inducible protein YebG,Uncharacterized protein conserved in bacteria,YebG protein
03036	1.495699	0.049407	unknown	PRTRC system ParB family protein
03021	1.451615	0.043131	rusA	bacteriophage V crossover junction endodeoxyribonuclease,Crossover junction endodeoxyribonuclease rusA,endodeoxyribonuclease RUS,Holliday junction resolvase,Endodeoxyribonuclease RusA
02936	1.445177	0.000327	umuC_2	Error-prone%2C lesion bypass DNA polymerase V (UmuC),hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
02809	1.396066	0.000358	dinI_1	DinI family protein,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
03205	1.395775	1.29E-06	sulA	cell division inhibitor,Cell division inhibitor SulA,SOS cell division inhibitor,SOS-response cell division inhibitor%2C blocks FtsZ ring formation,cell division inhibitor SulA,Cell division inhibitor SulA
03114	1.361824	0.004872	dinI_3	DNA-damage-inducible protein I,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
01771	1.293665	0.001501	yebE	membrane protein,Inner membrane protein yebE,Protein of unknown function (DUF533)
02765	1.227759	0.048699	umuD_1	DNA polymerase V subunit UmuD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
04659	1.225697	2.84E-08	lexA_2	SOS-response repressor and protease LexA,LexA repressor,LexA repressor,repressor LexA,LexA DNA binding domain
01085	1.204382	7.01E-09	recA	recombinase A,Recombinase A,recombinase A,RecA-superfamily ATPases implicated in signal transduction,protein RecA,recA bacterial DNA recombination protein

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
02979	1.187129	0.002287	umuC_3	Error-prone,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
03966	1.181129	3.8E-10	dinB	DNA polymerase IV,DNA polymerase IV,DNA polymerase IV,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
02827	1.122457	1.78E-07	recE	putative exodeoxyribonuclease VIII,Exodeoxyribonuclease 8,exonuclease VIII,Enterobacterial exodeoxyribonuclease VIII
02828	1.068735	0.000919	unknown	recombinase%2C RecT family,RecT family
02820	1.043763	0.001129	unknown	Protein of unknown function (DUF1019)
02819	1.042327	0.029035	unknown	ATP-dependent protease ATP-binding subunit ClpX,ClpX C4-type zinc finger
03033	1.02674	0.008683	unknown	Uncharacterized conserved protein,Uncharacterized conserved protein (DUF2303)
01086	1.014191	0.042817	recX	Regulatory protein RecX,Regulatory protein recX,recombination regulator RecX,Uncharacterized protein conserved in bacteria,RecX family
02892	0.987197	0.029249	cho	excinuclease cho (excinuclease ABC alternative C subunit),Excinuclease cho,nucleotide excision repair endonuclease,Nuclease subunit of the excinuclease complex,excinuclease ABC subunit C
02207	0.967062	0.001263	unknown	alpha/beta fold family hydrolase,acetoin dehydrogenase E2 subunit dihydrolipoyllysine-residue acetyltransferase,2-succinyl-6-hydroxy-2%2C4-cyclohexadiene-1-carboxylate synthase,Alpha/beta hydrolase family
04658	0.865515	0.019317	dinF	DNA-damage-inducible SOS response protein,Multidrug export protein mepA,DNA-damage-inducible SOS response protein,MATE efflux family protein,MatE
02815	0.796768	0.047818	unknown	PRTRC system ParB family protein
04902	0.77659	0.019774	aes	lipase%2C GDXG family,Acetyl esterase,acetyl esterase,alpha/beta hydrolase fold
03361	0.756131	0.009266	dinG_2	ATP-dependent DNA helicase DinG,Probable ATP-dependent helicase dinG homolog,ATP-dependent DNA helicase DinG,putative DnaQ family exonuclease/DinG family helicase,DEAD/DEAH box helicase
04645	0.745031	9.94E-05	uvrA	excinuclease ABC subunit A,Excinuclease ABC subunit A,excinuclease ABC subunit A,ABC-type transport system involved in cytochrome bd biosynthesis%2C fused ATPase and permease components,excinuclease ABC subunit A,ABC transporter

Table D.9 – Differentially expressed genes of KC154 exposed to ciprofloxacin after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
02993	1.552272	0.010584	unknown	secreted protein,Protein of unknown function (DUF1482)
06044	1.495848	0.00214	unknown	conjugal transfer protein
06037	1.426044	0.000579	unknown	mobC protein
06039	1.419088	0.000241	unknown	putative dopa decarboxylase protein remnant
06043	1.417205	1.48E-05	ptlE	inner membrane protein VirB8,Pertussis toxin liberation protein E,type IV secretion system protein VirB8,VirB8 protein
04360	1.350463	0.009331	unknown	Protein of unknown function (DUF3164)
06038	1.273648	0.001051	unknown	MobB protein,Type IV secretory pathway%2C VirD4 components,conjugative coupling factor TraD%2C SXT/TOL subfamily,Type IV secretory system Conjugative DNA transfer
06040	1.231034	0.000125	unknown	VirB11,Type IV secretion system protein virB11,type IV secretion system ATPase VirB11,Flp pilus assembly protein%2C ATPase CpaF,P-type DNA transfer ATPase VirB11,Type II/IV secretion system protein
06051	1.216707	0.022998	unknown	hypothetical protein

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
06052	1.184357	0.0046	unknown	ATP/GTP-binding protein remnant
01214	1.105364	0.000241	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
06047	1.087426	0.002979	virB5	type IV secretion system VirB5 component,P-type DNA transfer protein VirB5,Type IV secretion system proteins
06054	0.942822	0.0046	rfaH_2	transcriptional activator RfaH,Transcriptional activator rfaH,transcriptional activator RfaH,transcriptional activator RfaH,Transcription termination factor nusG
06045	0.92255	0.022239	virB6	type IV secretion system VirB6 component,TrbL/VirB6 plasmid conjugal transfer protein
03208	0.899627	0.0046	unknown	Uncharacterized conserved protein,Uncharacterized conserved protein (DUF2303)
06019	0.734566	0.009331	umuC_6	DNA-directed DNA polymerase,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
06018	0.71697	0.049826	umuD_9	umuDC operon protein-like protein,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like

Table D.10 – Differentially expressed genes of KC155 exposed to ciprofloxacin after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
05040	3.672125	3.23E-48	qnrS1	QnrS1,Type III effector pipB2,secreted effector protein PipB2,Uncharacterized protein conserved in bacteria,Pentapeptide repeats (8 copies)
01152	1.425531	1.98E-14	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
05148	1.033647	7.89E-09	unknown	ThiF family protein,thiamine biosynthesis protein ThiF,thiazole biosynthesis adenylyltransferase ThiF,ThiF family
05149	1.011184	7.89E-09	unknown	hypothetical protein
05225	0.910383	0.003045	umuD_4	UmnD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
05150	0.855484	0.001625	unknown	hypothetical protein
05195	0.850964	0.049959	unknown	site-specific recombinase%2C phage integrase family,Site-specific recombinase XerD,Phage integrase family
05353	0.84594	0.02179	umuD_5	umuDC operon protein-like protein,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
05226	0.815627	0.021532	unknown	putative DinI-like damage-inducible protein,DNA damage-inducible protein I,DinI-like family
05224	0.735235	0.001625	umuC_5	UmuC,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
01077	0.715976	5.7E-05	recA	recombinase A,Recombinase A,recombinase A,RecA-superfamily ATPases implicated in signal transduction,protein RecA,recA bacterial DNA recombination protein
04928	0.699602	0.042096	aes	lipase%2C GDXG family,Acetyl esterase,acetyl esterase,alpha/beta hydrolase fold
05354	0.683659	0.031844	umuC_6	DNA-directed DNA polymerase,hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
01675	0.616762	0.049959	yeeA	inner membrane protein,Inner membrane protein yeeA,Fusaric acid resistance protein family
04663	0.542766	0.049959	lexA	SOS-response repressor and protease LexA,LexA repressor,LexA repressor,repressor LexA,LexA DNA binding domain

Table D.11 – Differentially expressed genes of ATCC 43816 exposed to ciprofloxacin after 2 hours incubation

Gene ID	Log_2 fold change	Adjusted *P* value	Gene	Product
04467	4.1874	3E-122	recN	DNA repair protein RecN,Recombination protein N,recombination and repair protein,Predicted ATPase,DNA repair protein RecN,RecF/RecN/SMC N terminal domain
03307	4.142984	3.92E-57	unknown	toxic peptide TisB
05077	3.814781	1.07E-48	unknown	hypothetical protein
05071	3.78445	1.05E-50	unknown	putative glycoside hydrolase,Predicted chitinase,Chitinase class I
05073	3.78178	5.26E-44	unknown	Protein of unknown function (DUF2560)
05060	3.776608	2.7E-81	rdgC_2	DNA recombination-dependent growth factor C,Recombination-associated protein rdgC,recombination associated protein,Putative exonuclease%2C RdgC
05090	3.72101	8.75E-34	unknown	alkanesulfonate ABC transporter permease
05082	3.705434	1.07E-48	unknown	phage conserved hypothetical protein%2C phiE125 gp8 family,Phage gp6-like head-tail connector protein
05076	3.702059	1.45E-40	unknown	hypothetical protein
05075	3.679522	2.67E-51	unknown	HNH endonuclease
05072	3.678538	3.61E-31	unknown	hypothetical protein
05059	3.6414	4.41E-53	unknown	hypothetical protein
05057	3.628489	8.61E-38	unknown	hypothetical protein
05084	3.624226	9.61E-39	unknown	phage protein%2C HK97 gp10 family,Bacteriophage protein of unknown function (DUF646)
05087	3.615942	9.1E-53	unknown	hypothetical protein
05086	3.611059	7.31E-44	unknown	hypothetical protein
05058	3.601582	3.47E-60	unknown	PRTRC system ParB family protein
04959	3.572649	6.16E-47	sbmC	DNA gyrase inhibitory protein,DNA gyrase inhibitor,DNA gyrase inhibitor,DNA gyrase inhibitor,Bacterial transcription activator%2C effector binding domain
05080	3.543561	7.22E-52	unknown	phage prohead protease,phage prohead protease%2C HK97 family,Caudovirus prohead protease
05085	3.537917	1.46E-41	unknown	Protein of unknown function (DUF3168)
05063	3.533294	2.95E-58	unknown	Replication protein O
05053	3.502798	5.87E-35	unknown	hypothetical protein
05052	3.459461	1.95E-39	unknown	hypothetical protein
05078	3.449485	4.34E-51	unknown	phage terminase large subunit,Phage terminase-like protein%2C large subunit,Phage Terminase
05081	3.443372	3.92E-57	unknown	phage major capsid protein,Predicted phage phi-C31 gp36 major capsid-like protein,phage major capsid protein%2C HK97 family,Phage capsid family
05089	3.425065	7.51E-46	unknown	Prophage tail length tape measure,Phage-related minor tail protein,phage tail tape measure protein%2C lambda family,Lambda phage tail tape-measure protein (Tape_meas_lam_C)
05056	3.422578	9.54E-49	unknown	hypothetical protein
05062	3.416409	2.39E-43	unknown	hypothetical protein
05054	3.410489	1.6E-57	unknown	RelE family toxin-antitoxin system
05070	3.369539	5.18E-41	unknown	membrane protein,Protein of unknown function (DUF754)
05055	3.368292	6.11E-64	unknown	hypothetical protein
05067	3.347129	4.6E-30	unknown	Protein of unknown function (DUF1133)
05083	3.345889	4.2E-42	unknown	head-tail adaptor protein,Bacteriophage head-tail adaptor,putative phage head-tail adaptor,Phage head-tail joining protein
05051	3.338181	6.09E-52	unknown	hypothetical protein
05079	3.32015	1.87E-42	unknown	phage portal protein,phage portal protein%2C HK97 family,Phage portal protein
05069	3.302119	3.82E-38	unknown	putative prophage membrane protein

05064	3.286387	4.37E-47	dam_3	DNA adenine methylase,DNA adenine methylase,DNA adenine methylase,Site-specific DNA methylase,DNA adenine methylase,D12 class N6 adenine-specific DNA methyltransferase
00079	3.264353	5.47E-40	unknown	LexA family transcriptional regulator,DNA damage-inducible protein YebG,Uncharacterized protein conserved in bacteria,YebG protein
05091	3.251202	2.01E-29	unknown	ArsR family transcriptional regulator,Domain of unknown function (DUF1833)
05068	3.210162	1.33E-32	unknown	hypothetical protein
00080	3.156311	3.43E-26	yebE	membrane protein,Inner membrane protein yebE,Protein of unknown function (DUF533)
00008	3.154895	1.6E-33	unknown	hypothetical protein
05088	3.127367	1.32E-32	unknown	hypothetical protein
05065	3.108743	5.34E-35	rusA	crossover junction endodeoxyribonuclease,Crossover junction endodeoxyribonuclease rusA,endodeoxyribonuclease RUS,Holliday junction resolvase,Endodeoxyribonuclease RusA
05066	3.040779	1.87E-42	unknown	Protein of unknown function (DUF968)
00005	2.991819	3.64E-06	unknown	hypothetical protein
05092	2.985392	1.32E-27	unknown	nitrite transporter
05093	2.926453	3.94E-33	unknown	putative kinase
00009	2.907132	6.96E-24	unknown	Error-prone repair protein UmuD
05050	2.896328	2.04E-26	unknown	hypothetical protein
01585	2.88921	5.37E-19	unknown	Protein of unknown function (DUF2732)
01584	2.850345	5.53E-29	dam_1	DNA adenine methylase,DNA adenine methylase,DNA adenine methylase,Site-specific DNA methylase,DNA adenine methylase,D12 class N6 adenine-specific DNA methyltransferase
01236	2.81779	1.18E-15	umuD	Error-prone repair protein UmuD,hypothetical protein,DNA polymerase V subunit UmuD,repressor LexA,Peptidase S24-like
04392	2.809766	4.02E-53	recA	recombinase A,Recombinase A,recombinase A,RecA-superfamily ATPases implicated in signal transduction,protein RecA,recA bacterial DNA recombination protein
04960	2.773714	2.41E-39	yeeA	inner membrane protein,Inner membrane protein yeeA,Fusaric acid resistance protein family
01345	2.610563	3.08E-21	dinI_2	DNA-damage-inducible protein I,DNA-damage-inducible protein I,DNA damage-inducible protein I,DinI-like family
03204	2.556991	1.91E-22	aes	lipase%2C GDXG family,Acetyl esterase,acetyl esterase,alpha/beta hydrolase fold
01235	2.516183	7.26E-17	umuC	Error-prone%2C lesion bypass DNA polymerase V (UmuC),hypothetical protein,DNA polymerase V subunit UmuC,Nucleotidyltransferase/DNA polymerase involved in DNA repair,impB/mucB/samB family
03306	2.513942	7.19E-19	unknown	putative acyltransferase,Acetyltransferase (GNAT) family
01586	2.478546	1.26E-18	unknown	hypothetical protein

Appendix E *Klebsiella pneumoniae* collection

E.1 Collection of *K. pneumoniae* isolates used in HCI

Isolate	Original collection	Isolate metadata				Genotype		MIC (µg/mL)			Rationale for inclusion
		Country	Year	Source	Type	Species	ST	Cipro	Mero	Tige	
KC001	Case Report	UK	2019	Human	Blood	Kp1	ST23	0.032	0.012	0.25	A
KC002	Reference	Unknown	Unknown	Unknown	Unknown	Kp1	ST493	0.032	0.016	0.75	B
KC003	CUH surveillance	UK	2016	Human	Blood	Kp4	ST2366-1LV	0.032	0.016	0.25	C
KC004	CUH surveillance	UK	2016	Human	Blood	Kp2	ST1466-1LV	0.032	0.016	0.25	C
KC005	CUH surveillance	UK	2016	Human	Blood	Kp1	ST268	0.016	0.023	0.38	D
KC006	CUH surveillance	UK	2016	Human	Blood	Kp1	ST45	64	0.016	1	E,F
KC007	CUH surveillance	UK	2016	Human	Stool/rectal swab	Kp1	ST78	>32	>32	1.5	E,J
KC008	CUH surveillance	UK	2016	Human	Blood	Kp1	ST14-2LV	>32	16	0.75	E,J
KC009	CUH surveillance	UK	2016	Human	Blood	Kp1	ST29	0.016	0.032	0.19	G
KC010	CUH surveillance	UK	2016	Human	Wound swab	Kp1	ST78	>32	>32	3	E,J
KC011	CUH surveillance	UK	2016	Environmental	Hospital wastewater	Kp1	ST78	>32	>32	2	E,H,J
KC012	CUH surveillance	UK	2016	Human	Blood	Kp1	ST23	0.25	0.012	1.5	A
KC013	CUH surveillance	UK	2016	Human	Blood	Kp3	ST4312	0.25	0.012	0.38	C
KC014	CUH surveillance	UK	2016	Human	Blood	Kp3	ST4364	0.032	0.023	0.25	C
KC015	CUH surveillance	UK	2016	Human	Blood	Kp3	ST1511	0.016	0.012	0.25	C
KC016	CUH surveillance	UK	2016	Human	Blood	Kp3	ST1219	0.032	0.012	0.38	C
KC017	CUH surveillance	UK	2016	Human	Blood	Kp4	ST414	0.064	0.012	0.19	C
KC018	CUH surveillance	UK	2016	Human	Blood	Kp1	ST76	0.016	0.012	0.19	I
KC019	CUH surveillance	UK	2016	Human	Blood	Kp1	ST45	0.012	0.008	0.38	E,F
KC020	CUH surveillance	UK	2016	Human	Blood	Kp4	ST3387	0.064	0.012	0.25	C
KC021	CUH surveillance	UK	2016	Human	Blood	Kp3	ST1181	0.032	0.047	0.38	C
KC022	CUH surveillance	UK	2016	Human	Skin	Kp3	ST1023-1LV	0.016	0.016	0.125	C
KC023	CUH surveillance	UK	2016	Human	Blood	Kp3	ST2228	0.032	0.012	0.25	C
KC024	CUH surveillance	UK	2016	Human	Blood	Kp1	ST15	12	0.032	0.75	E
KC025	CUH surveillance	UK	2016	Human	Blood	Kp4	ST2855-4LV	0.032	0.012	0.25	C
KC026	CUH surveillance	UK	2016	Human	Blood	Kp1	ST2806	0.032	0.012	0.25	E
KC027	CUH surveillance	UK	2016	Human	Stool/rectal swab	Kp1	ST78	>32	>32	3	E,J
KC028	CUH surveillance	UK	2016	Human	Blood	Kp1	ST11	>32	0.008	1	E
KC029	CUH surveillance	UK	2016	Human	Blood	Kp1	ST17	0.032	0.008	0.5	E
KC030	CUH surveillance	UK	2016	Human	Blood	Kp3	ST938	0.023	0.016	0.38	C
KC031	CUH surveillance	UK	2016	Human	Blood	Kp1	ST42	0.023	0.008	0.5	I

Isolate	Original collection	Isolate metadata				Genotype		MIC (µg/mL)			Rationale for inclusion
		Country	Year	Source	Type	Species	ST	Cipro	Mero	Tige	
KC032	CUH surveillance	UK	2016	Human	Blood	Kp1	ST66	0.023	0.023	0.19	A
KC033	CUH surveillance	UK	2016	Human	Blood	Kp1	ST307	16	0.023	0.25	E
KC034	CUH surveillance	UK	2016	Human	Blood	Kp1	ST37	0.023	0.12	0.38	G
KC035	CUH surveillance	UK	2016	Human	Blood	Kp2	ST3212	0.032	0.023	0.38	C
KC036	CUH surveillance	UK	2016	Human	Blood	Kp1	ST45	0.032	0.023	0.38	E,F
KC037	EoE OneHealth	UK	2014	Animal	Animal	Kp1	ST661	0.016	0.008	0.19	D
KC038	EoE OneHealth	UK	2014	Environmental	Wastewater	Kp1	ST20	0.032	0.012	0.5	G,H
KC039	EoE OneHealth	UK	2014	Environmental	Wastewater	Kp1	ST3008	0.016	0.023	0.25	I
KC040	EoE OneHealth	UK	2015	Environmental	Hospital bedspace	Kp1	ST23	0.032	0.016	0.25	A
KC041	EoE OneHealth	UK	2015	Environmental	Hospital bedspace	Kp1	ST23	0.032	0.023	0.19	A
KC042	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST20	0.25	0.012	0.19	G
KC043	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST37	0.016	0.012	0.75	G
KC044	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST29	0.023	0.012	0.5	G
KC045	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST661	0.023	0.016	0.25	D
KC046	EoE OneHealth	UK	2014	Environmental	Wastewater	Kp1	ST45	1.5	0.023	0.38	E,F
KC047	EoE OneHealth	UK	2014	Environmental	Hospital wastewater	Kp1	ST1436	1	0.012	0.38	H
KC048	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST2004	0.19	0.016	0.125	I
KC049	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST268	2	0.016	0.75	D
KC050	EoE OneHealth	UK	2015	Human	Stool/rectal swab	Kp1	ST268	2	0.023	0.75	D
KC051	EoE OneHealth	UK	2015	Environmental	Hospital wastewater	Kp1	ST258	>32	>32	1	E,J
KC052	EoE OneHealth	UK	2015	Environmental	Hospital bedspace	Kp1	ST307	>32	0.016	0.5	E,H
KC053	EoE OneHealth	UK	2015	Environmental	Hospital wastewater	Kp1	ST874	>32	0.023	0.38	E,H
KC054	EoE OneHealth	UK	2015	Environmental	Hospital bathroom	Kp1	ST268	2	0.016	0.5	D
KC055	EoE OneHealth	UK	2014	Human	Blood	Kp1	ST307	>32	0.016	0.75	E
KC056	BSAC	UK	2001	Human	Blood	Kp1	ST152	0.25	0.016	2	I
KC057	BSAC	UK	2002	Human	Blood	Kp1	ST448-1LV	1.5	0.016	1.5	I
KC058	BSAC	UK	2002	Human	Blood	Kp1	ST20	0.5	0.004	1	G
KC059	BSAC	UK	2002	Human	Blood	Kp1	ST35	1	0.012	1.5	I
KC060	BSAC	UK	2003	Human	Blood	Kp1	ST101	>32	0.016	1.5	E
KC061	BSAC	UK	2003	Human	Blood	Kp1	ST147	>32	0.016	0.38	E,I
KC062	BSAC	UK	2003	Human	Blood	Kp1	ST1	>32	0.012	4	I
KC063	BSAC	UK	2003	Human	Blood	Kp1	ST490	>32	0.006	4	I
KC064	BSAC	Ireland	2004	Human	Blood	Kp1	ST14	0.25	0.016	1	E
KC065	BSAC	UK	2004	Human	Blood	Kp1	ST1	>32	0.012	0.75	I
KC066	BSAC	UK	2004	Human	Blood	Kp1	ST17	0.032	0.023	0.38	E
KC067	BSAC	UK	2004	Human	Blood	Kp1	ST147	>32	0.047	0.5	E,I
KC068	BSAC	UK	2004	Human	Blood	Kp1	ST45	0.016	0.032	0.5	E,F
KC069	BSAC	UK	2004	Human	Blood	Kp1	ST15	>32	0.032	0.25	E
KC070	BSAC	Ireland	2005	Human	Blood	Kp1	ST76	0.75	0.016	1.5	I
KC071	BSAC	UK	2005	Human	Blood	Kp1	ST101	>32	0.047	2	E,I

Isolate	Original collection	Isolate metadata				Genotype		MIC (µg/mL)			Rationale for inclusion
		Country	Year	Source	Type	Species	ST	Cipro	Mero	Tige	
KC072	BSAC	UK	2005	Human	Blood	Kp1	ST101	>32	0.008	1.5	E,I
KC073	BSAC	UK	2005	Human	Blood	Kp1	ST3697	>32	0.032	1	G,I
KC074	BSAC	UK	2005	Human	Blood	Kp1	ST340	>32	0.047	0.5	G
KC075	BSAC	Ireland	2005	Human	Blood	Kp1	ST108-2LV	0.047	0.003	0.38	I
KC076	BSAC	UK	2005	Human	Blood	Kp3	ST925	0.5	0.016	0.38	C,I
KC077	BSAC	UK	2005	Human	Blood	Kp1	ST11	>32	0.016	0.75	E
KC078	BSAC	UK	2006	Human	Blood	Kp1	ST101	>32	0.023	0.5	E
KC079	BSAC	UK	2006	Human	Blood	Kp1	ST16	>32	0.008	1.5	E
KC080	BSAC	UK	2006	Human	Blood	Kp1	ST42	0.023	0.016	0.5	I
KC081	BSAC	UK	2006	Human	Blood	Kp1	ST874	>32	0.032	0.38	E,I
KC082	BSAC	Ireland	2007	Human	Blood	Kp1	ST11	>32	0.023	0.5	E
KC083	BSAC	UK	2007	Human	Blood	Kp1	ST16	>32	0.016	1	E
KC084	BSAC	UK	2007	Human	Blood	Kp1	ST37	0.032	0.032	0.25	G
KC085	BSAC	UK	2007	Human	Blood	Kp1	ST340	>32	0.032	0.75	G
KC086	BSAC	UK	2007	Human	Blood	Kp1	ST15	>32	0.032	3	E,I
KC087	BSAC	UK	2007	Human	Blood	Kp1	ST15	2	0.016	0.75	E,I
KC088	BSAC	UK	2008	Human	Blood	Kp1	ST1380	0.25	0.016	1	I
KC089	BSAC	Ireland	2008	Human	Blood	Kp1	ST15	6	0.012	0.5	E,I
KC090	BSAC	UK	2008	Human	Blood	Kp1	ST101	0.032	0.012	1	E,I
KC091	BSAC	UK	2008	Human	Blood	Kp1	ST2004	>32	0.016	1.5	I
KC092	BSAC	UK	2008	Human	Blood	Kp1	ST1180	0.023	0.023	0.5	I
KC093	BSAC	UK	2008	Human	Blood	Kp1	ST14	0.023	0.008	0.38	E
KC094	BSAC	UK	2008	Human	Blood	Kp1	ST5	0.125	0.012	3	A
KC095	BSAC	UK	2008	Human	Blood	Kp1	ST45	0.023	0.016	0.5	E,F
KC096	BSAC	UK	2008	Human	Blood	Kp1	ST874	>32	0.012	0.25	E,I
KC097	BSAC	UK	2009	Human	Blood	Kp4	ST334	1	0.012	4	C,I
KC098	BSAC	UK	2009	Human	Blood	Kp1	ST3697	>32	0.047	1	G,I
KC099	BSAC	UK	2009	Human	Blood	Kp1	ST15	>32	0.023	0.5	E
KC100	BSAC	UK	2010	Human	Blood	Kp1	ST3164	0.38	0.016	4	I
KC101	BSAC	UK	2010	Human	Blood	Kp1	ST35	0.016	0.032	0.5	I
KC102	BSAC	UK	2010	Human	Blood	Kp1	ST976	0.5	0.016	1.5	I
KC103	BSAC	UK	2010	Human	Blood	Kp1	ST29	0.5	0.032	2	G,I
KC104	BSAC	Ireland	2010	Human	Blood	Kp1	ST17	0.25	0.016	3	E
KC105	BSAC	UK	2011	Human	Blood	Kp1	ST307	>32	0.016	1	E
KC106	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST147	>32	>32	1.5	E,J
KC107	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST15	>32	>32	0.5	E,J
KC108	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST147	>32	>32	0.5	E,J
KC110	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST147	0.25	0.023	0.25	E,I
KC111	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST656	>32	>32	0.75	G,J
KC112	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST29	0.38	0.023	1.5	G
KC113	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST15	>32	0.018	0.25	E
KC114	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST15	>32	1	0.38	E,J

Isolate	Original collection	Isolate metadata				Genotype		MIC (µg/mL)			Rationale for inclusion
		Country	Year	Source	Type	Species	ST	Cipro	Mero	Tige	
KC117	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp4	ST334	0.19	0.023	0.19	C
KC118	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST101	0.5	0.023	0.5	E,I
KC119	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST307	>32	0.25	0.25	E,J
KC120	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST147	>32	0.012	0.38	E
KC121	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST37	0.25	0.016	0.25	G
KC122	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST661	0.25	0.023	0.25	D
KC123	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST16	>32	>32	0.75	E,J
KC124	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST307	0.023	0.012	0.19	E
KC125	Hanoi ICU	Vietnam	2017	Human	Pus swab	Kp1	ST307	>32	3	0.75	E,J
KC128	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST11	>32	>32	0.5	E,J
KC129	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST656	>32	>32	0.75	G,J
KC130	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST17	0.25	0.094	0.75	E
KC131	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST15	>32	>32	0.25	E,J
KC132	Hanoi ICU	Vietnam	2017	Human	Stool/rectal swab	Kp1	ST656	>32	>32	0.5	A,G,J
KC133	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST45	1.5	0.032	0.5	E,F
KC134	Hanoi ICU	Vietnam	2017	Human	Sputum/tracheal aspirate	Kp1	ST20	0.5	0.016	0.25	G
KC135	Global collection	Spain	2011	Human	Sputum/tracheal aspirate	Kp1	ST11	>32	0.75	1.5	E,J
KC136	Global collection	Australia	2001	Human	Respiratory tract	Kp3	ST695	0.032	0.016	0.25	C
KC137	Global collection	Australia	2001	Human	Urinary tract	Kp4	ST2119	0.023	0.012	0.25	C
KC138	Global collection	Australia	2002	Human	Respiratory tract	Kp1	ST14	0.25	0.008	4	E
KC139	Global collection	Australia	2002	Human	Respiratory tract	Kp3	ST1980	0.016	0.016	0.19	C
KC140	Global collection	Australia	2002	Human	Urinary tract	Kp3	ST695	0.032	0.016	0.25	C
KC141	Global collection	Australia	2002	Human	Blood	Kp3	ST2122	4	0.016	0.19	C
KC142	Global collection	USA	2007	Human	Intestinal biopsy	Kp1	ST15-1LV	0.032	0.016	0.75	E
KC143	Global collection	USA	2007	Human	Intestinal biopsy	Kp4	ST283-1LV	0.012	0.016	0.25	C
KC144	Reference	Unknown	Unknown	Unknown	Unknown	Kp1	ST66	0.032	0.012	0.19	B
KC145	Reference	Unknown	Unknown	Unknown	Unknown	Kp1		0.032	0.008	0.094	B
KC146	Reference	Australia	2001-2002	Human	Urinary tract	Kp1	ST2121	0.023	0.016	0.19	B
KC147	Reference	Australia	2001-2002	Human	Urinary tract	Kp1		0.023	0.004	0.19	B
KC148	Reference	USA	1911	Unknown	Unknown	Kp1	ST375	0.006	0.008	0.064	B
KC149	Reference	Taiwan	1996-2001	Human	Blood	Kp1	ST23	0.032	0.016	0.25	A,B
KC150	Reference	USA	2011	Human	Skin	Kp1	ST258	>32	0.125	4	B,E,J
KC151	Reference	UK	2012	Human	Wound swab	Kp1	ST258	>32	>32	0.75	B,E,J
KC152	Reference	UK	2007	Human	Blood	Kp1	ST258	>32	>32	0.75	B,E,J
KC153	CUH surveillance	UK	2016	Human	Stool/rectal swab	Kp1	ST251	1	0.032	0.5	F
KC154	CUH surveillance	UK	2016	Human	Blood	Kp1	ST240	8	0.008	1	F
KC155	CUH surveillance	UK	2016	Human	Blood	Kp1	ST4023	0.38	0.19	0.5	F
KC156	Ireland CPE	Ireland	2015	Human	Urinary tract	Kp1	ST307	>32	2	0.75	E,J
KC157	Ireland CPE	Ireland	2015	Human	Stool/rectal swab	Kp1	ST15	>32	0.5	1	E,J

Isolate	Original collection	Isolate metadata				Genotype		MIC (µg/mL)			Rationale for inclusion
		Country	Year	Source	Type	Species	ST	Cipro	Mero	Tige	
KC158	Ireland CPE	Ireland	2016	Human	Stool/rectal swab	Kp4	ST1308	0.023	0.5	0.19	C,J
KC159	Ireland CPE	Ireland	2016	Human	Skin	Kp1	ST37	0.5	0.5	0.125	G,J
KC160	Ireland CPE	Ireland	2015	Human	Blood	Kp1	ST147	>32	>32	0.75	E,J
KC161	Ireland CPE	Ireland	2016	Human	Stool/rectal swab	Kp2	ST3516	0.38	0.75	1.5	C,J
KC162	Ireland CPE	Ireland	2016	Human	Stool/rectal swab	Kp1	ST336	0.094	1	0.25	J
KC163	Ireland CPE	Ireland	2016	Human	Stool/rectal swab	Kp1	ST45	0.023	0.75	0.38	E,F,J
KC164	Ireland CPE	Ireland	2015	Human	Blood	Kp1	ST23	4	1	0.75	A,J
KC165	Ireland CPE	Ireland	2014	Human	Urinary tract	Kp1	ST11	>32	2	3	E,J
KC166	Ireland CPE	Ireland	2015	Human	Blood	Kp1	ST14	0.25	0.5	1	E,J
KC167	Ireland CPE	Ireland	2015	Human	Wound swab	Kp4	ST1859	>32	0.75	1.5	C,J
KC168	Ireland CPE	Ireland	2015	Human	Ileostomy swab	Kp1	ST133	0.125	0.38	1	J
KC169	Ireland CPE	Ireland	2014	Human	Stool/rectal swab	Kp1	ST512	>32	>32	0.38	E,J
KC170	Ireland CPE	Ireland	2015	Human	Unknown	Kp1	ST45	0.38	0.19	0.38	E,F,J
KC171	Ireland CPE	Ireland	2014	Human	Urinary tract	Kp1	ST111	2	8	1	J
KC172	Ireland CPE	Ireland	2016	Human	Urinary tract	Kp1	ST37	3	0.25	1	G,J
KC173	Ireland CPE	Ireland	2014	Human	Stool/rectal swab	Kp3	ST3517	0.047	12	1.5	C,J
KC174	Ireland CPE	Ireland	2014	Human	Urinary tract	Kp1	ST101	>32	1	1.5	E,J
KC175	Ireland CPE	Ireland	2014	Human	Unknown	Kp1	ST3518	>32	12	1	J
KC176	Ireland CPE	Ireland	2017	Human	Urinary tract	Kp1	ST307	>32	12	0.75	E,J
KC177	Ireland CPE	Ireland	2013	Human	Sputum/tracheal aspirate	Kp1	ST258	>32	>32	0.75	E,J
KC178	Ireland CPE	Ireland	2013	Human	Urinary tract	Kp1	ST133	4	0.75	0.25	J
KC179	Ireland CPE	Ireland	2013	Human	Urinary tract	Kp1	ST15	0.5	1	0.25	E,J
KC180	Ireland CPE	Ireland	2014	Human	Peritoneal Fluid	Kp1	ST258	>32	8	1.5	E,J

ST = multilocus sequence type; MIC = minimum inhibitory concentration; cipro = ciprofloxacin; mero = meropenem; tige = tigecycline; CUH = Cambridge University Hospitals; CPE = carbapenemase producing Enterobacteriaceae; ICU = intensive care unit; EoE = East of England; BSAC = British Society for Antimicrobial Chemotherapy.

Legend for Klebsiella species (phylogroups as described in Rodrigues *et al.*[463])

Kp1 – *Klebsiella pneumoniae* sensu stricto

Kp2 – *Klebsiella quasipneumoniae* subsp. *quasipneumoniae*

Kp3 – *Klebsiella variicola* subsp. *variicola*

Kp4 – *Klebsiella quasipneumoniae* subsp. *similipneumoniae*

Legend for inclusion rationale:

A. Hypervirulent isolate, or isolate belonging to a sequence type associated with hypervirulence
B. Reference isolate, well characterised and commonly used outside this study
C. *Klebsiella pneumoniae* subspecies that are not represented elsewhere, namely *K. quasipneumoniae* subsp. *quasipneumoniae*, *K. quasipneumoniae* subsp. *similipneumoniae* and *K. variicola*
D. Sequence type identified in both human and environmental samples in Ludden *et al.* OneHealth study[360]
E. Sequence type (or closely related sequence type) associated with clinical outbreaks
F. Previous work on this isolate/sequence type during this study using RNAseq
G. Sequence type commonly represented in studies used to produce this collection
H. Environmental sample
I. High/low ciprofloxacin MIC in comparison to that expected from common fluoroquinolone resistance genes identified using Kleborate
J. Presence of carbapenemase resistance genes of clinical significance

Appendix F Supplementary influenza data

F.1 Re-classification of hospital onset infection based on clinical review

25 cases* originally designated as HAI (based on time from admission to time of swabbing ≥48 hours) were re-designated CAI after clinical review:

- 5 cases were admitted for other reasons (intentional overdose, decompensated heart failure, decompensated chronic liver disease, sciatica, fall) and had co-incident influenza
- 19 cases were admitted with symptoms consistent with influenza but there was a delay in the recognition of influenza as a possible cause, and therefore delay in requesting respiratory virus swabs
- 4 cases had documented respiratory viruses as a possible cause, but there was a delay in the requesting of respiratory virus swabs
- 3 cases had delays in collection of respiratory swabs >24 hours after they were requested
- 5 cases had a sample collected <48 hours from admission, but there was a delay in the sample being booked into the laboratory
- 1 case had a sample collected promptly after admission, but it was never booked into the laboratory
- 1 case had the incorrect admission date automatically extracted from the hospital record when generating the original 48 hour cut-off

*In some cases multiple factors contributed to delays in sample collection and processing

6 cases originally designated as CAI (based on time from admission to time of swabbing <48 hours) were re-designated HAI after clinical review:

- 2 cases were discharged from hospital and re-admitted with influenza symptoms within 48 hours of discharge
- 3 cases were transferred from other hospitals
- 1 case was asymptomatic on admission, but admitted into a bay with known influenza positive cases, became symptomatic and was swabbed before the 48 hour threshold

There were 8 additional cases that were collected in <48 hours for which genomic data shows clustering in outpatient settings. One further case was originally designated as HAI, but was likely acquired in outpatients given the genomic sequence similarity to another patient with the same outpatient clinic appointment time.

F.2 Case report form for mortality analysis

Please enter the case number for the below individual on each page of the survey.
* Indicates items provided in the table of cases

*MRN											

1. *First name:_____ *Family Name:_____

2.

Male	
Female	

3.

Tick if Pregnant	
Tick if ≤2 weeks post-partum	

4. Date of birth

d	d	m	m	y	y	y	y

5. Date of completing this questionnaire

d	d	m	m	y	y	y	y

6. Initials of investigator completing this questionnaire: _____

7. GP Name:_____

8. Date first positive Flu specimen collected

d	d	m	m	y	y	y	y

9. Oseltamivir prescribed

Yes	No

If yes, Standard therapy completed (e.g. 75mg P.O.bid for 5 days)

If no, explain why not……

10. Date of admission

d	d	m	m	y	y	y	y

Time of admission

24 hr

h	h	m	m

11. From where admitted

Own home	
Residential care	
Another hospital	
Other	

If other specify_____

12. Expired during admission

Yes	No

13. Date **of end of** admission

(date of discharge or expiry)

d	d	m	m	y	y	y	y

14. Reason for admission and active issues at time of swabbing

1	
2	
3	
4	
5	

15. Co-morbidities – Underlying respiratory disease

	Yes	No	Unknown
Current smoker			
Ex-smoker			
Asthma			
COPD			
Bronchiectasis			
Interstitial Lung disease			
Respiratory failure			
Long term oxygen therapy			
Other (please specify):			

16. Active conditions on admission: summarise in tick boxes the active conditions in question 15 above

Diagnostic Group	**Code**	**Tick**
Cardiovascular	1	
Hypertension	2	
Diabetes mellitus	3	
Respiratory	4	
Gastrointestinal	5	
Renal	6	
Urinary tract	7	
Genital	8	
CNS	9	
Psychiatric	10	
Metabolic other than diabetes	11	
Endocrine	12	
Trauma	13	
Malignancy	14	
Skin	15	
Muscular skeletal	16	
Haematology	17	
Infection	18	
Excessive alcohol use	19	
Other……	20	
Surgery during admission†	21	

Charlson Index (see guide for definitions)	**Tick**
Myocardial infarct	
Congestive heart failure	
Peripheral vascular disease/AAA <5cm	
Cerebrovascular disease	
Dementia	
Connective tissue disease	
Peptic ulcer	
Mild liver disease	
Moderate or severe liver disease	
Diabetes without end organ damage	
Diabetes with end organ damage	
Hemiplegic	
Moderate or severe kidney disease	
Tumour without metastasis	
Tumour with metastasis	
Leukaemia	
Lymphoma	
AIDS	
Obesity – BMI ≥40	

†Describe surgery_____

17. Immune suppressed on admission
If yes, tick all categories of immunes suppression which apply

Yes	No	Unknown

Severe primary immunodeficiency	
Steroids ≥20 mg prednisolone daily or equivalent for >1 month (1 mg/kg for children <20kg); OR ≥40mg prednisolone per day for >1 week) AND for at least three months after treatment has stopped	
Chemotherapy within last 6 months	
Generalised radiotherapy within last 6 months	
Organ transplant and on immune suppressants	
Bone marrow transplant recipients currently receiving immunosuppressive treatment or who received it within the last 12 months	
Patients with current graft-versus-host disease	
HIV infection (any stage)	
Currently or recently (within 6 months) on other types of immunosuppressive therapy or where the patient's specialist regards them as severely immunosuppressed	
Asplenia or other splenic dysfunction	
Other category of immune suppression (please specify)	

18. Clinical features at admission for CAI and close to first positive result for HAI (if date of onset is not identifiable from notes.)

Symptoms	**Yes**	**No**
Fever (subjective)		
Chills		
Rigors		
Cough		
Nasal discharge		
Sore throat		
Shortness of breath		
Chest pain		
Auscultation sounds (Rhonchi, rales, wheeze, crepitations)		
Palpable cervical lymph nodes		
Headache		
Photophobia		
Conjunctivitis		
Lacrimation		
Pain with eye movement		
Malaise		
Lethargic		
Myalgia		
Arthralgia		
Nausea		
Anorexia		
Abdominal pain		
Vomiting		
Diarrhoea		
Rash		
Altered mental state/confusion		

Observations	
Highest temperature around the time of sample collection	°C
Systolic blood pressure at time of sample collection	mmHg
Oxygen saturations on room air around time of sample collection	%
Supplementary oxygen required around time of sample collection (Y/N)	

Clinical features not in the list on the left:

19. Respiratory requirements

	Yes	No
CPAP (continuous positive airway pressure)		
NIV (non-invasive ventilation)		
Invasive Ventilation		

20. CAI/HAI: _____

21. Date of onset of clinical features (if identifiable from Epic notes): ----------

22. Pre-admission:

	Yes	**No**	**Dates (if documented)**
GP consultation			
Course of antibiotics			
Course of antivirals			

23. Investigations – earliest blood test results during this admission for CAI, or at time of sample collection for HAI

	Value	Not done
Haemoglobin		
Total White cell count		
Lymphocyte count		
CRP		
Creatinine		
Urea		
Glucose		

	Yes	No	CT/CXR	Not done
Any reported radiologic changes from chest imaging taken around time of resp virus sample collection suggesting pneumonia, e.g. consolidation				

Microbiology Cultures:

Organism	Specimen Site	Date

24. Norovirus PCR: positive/ Negative/ Not done

25. Receipt of 2017/18 influenza vaccination

Yes	No	Date

F.3 Single variable analysis of factors associated with mortality in patients with confirmed influenza, 2016-18

Original analysis and table produced by Neville Verlander. Total number of patients = 1,048.

Variable	Category or measure	Expired	Not expired	OR (95% CI)	*P*-value
Season	2016 to 2017 2017 to 2018	31 52	287 678	1.00 0.71 (0.45, 1.13)	**0.16**
Age at positive specimen (years)	Minimum 25th centile Median 75th centile Maximum	17 73 84 87 102	0 53 74 84 101	1.04 (1.02, 1.06)	**<0.001**
Gender	Female Male	37 46	523 442	1.00 1.47 (0.94, 2.31)	**0.09**
Pregnancy	Pregnant Not pregnant Male	0 37 46	15 508 442	0.00 (n.e.) 1.00 1.43 (0.91, 2.24)	0.09
Oseltamivir course completed	Not given Non-standard course Standard course	8 27 48	96 140 729	1.00 2.31 (1.01, 5.31) 0.79 (0.36, 1.72)	**<0.001**
Days between admission and first dose of oseltamivir	Minimum 25th centile Median 75th centile Maximum	-3 2 2 5 36	-3 2 3 5 36	1.04 (0.94, 1.07)	**0.99**
Receipt of influenza vaccination	Yes No	34 20	391 353	1.53 (0.87, 2.72) 1.00	**0.14**
Current smoker	Yes No	7 64	101 734	0.79 (0.35, 1.78) 1.00	0.6
Long term oxygen therapy	Yes No	5 78	14 951	4.35 (1.53, 12.4) 1.00	**0.01**
Hypertension	Yes No	39 44	324 641	1.75 (1.12, 2.75) 1.00	**0.02**
Trauma	Yes No	3 80	14 951	2.55 (0.72, 9.05) 1.00	**0.19**
Excessive alcohol use	Yes No	4 79	12 953	4.02 (1.27, 12.8) 1.00	**0.04**
Surgery	Yes No	3 80	16 949	2.22 (0.63, 7.79) 1.00	0.3

F.3: Single variable analysis of factors associated with mortality in patients with confirmed influenza, 2016-18 387

Variable	Category or measure	Expired	Not expired	OR (95% CI)	*P*-value
Immune suppressed	Yes No	11 72	182 782	0.66 (0.34, 1.26) 1.00	0.19
Admitted to intensive care	Yes No	17 65	68 884	3.40 (1.89, 6.12) 1.00	**<0.001**
Influenza subtypes	A(H3N2) A (not typed) B A(H1N1pdm09)	51 2 30 0	559 17 341 48	1.00 1.29 (0.29, 5.74) 0.96 (0.60, 1.54) 0.00 (n.e.)	**0.04**
Radiological evidence of pneumonia	Yes No	40 43	200 765	3.56 (2.25, 5.62) 1.00	**<0.001**
Chemotherapy	Yes No	1 82	61 904	0.18 (0.02, 1.32) 1.00	0.02
Radiotherapy	Yes No	1 82	3 962	3.91 (0.40, 30.0) 1.00	0.3
Organ transplant	Yes No	1 82	14 951	0.83 (0.11, 6.38) 1.00	0.9
Bone marrow transplant	Yes No	0 83	15 950	0.00 (n.e.) 1.00	0.11
Bone marrow transplant 2016/17	Yes No	0 83	4 961	0.00 (n.e.) 1.00	0.4
Bone marrow transplant 2017/18	Yes No	0 83	13 952	0.00 (n.e.) 1.00	0.14
Cycle Threshold (CT) value	Minimum 25th centile Median 75th centile Maximum	16 24 28 32 38	14 25 29 32 41	CF [a]	**0.048**
Temperature degrees Celsius	Minimum 25th centile Median 75th centile Maximum	35.4 37.7 38.3 38.7 40.2	35.3 37.7 38.3 38.8 41.1	0.89 (0.69, 1.16)	0.4
Haemoglobin g/L	Minimum 25th centile Median 75th centile Maximum	61 106 117 130 224	45 110 126 139 184	CF[b]	**0.004**
Total white cell count 10^9/L	Minimum 25th centile Median 75th centile Maximum	1.3 6.2 9.4 13.2 84.5	0.0 4.9 6.8 9.4 122	QF [C]	**<0.001**

Variable	Category or measure	Expired	Not expired	OR (95% CI)	*P*-value
Lymphocyte count 10^9/L	Minimum	0.2	0.0	0.99 (0.88, 1.11)	0.9
	25th centile	0.5	0.5		
	Median	0.8	0.8		
	75th centile	1.3	1.2		
	Maximum	5.1	46.7		
C-reactive protein mg/L	Minimum	3.3	1.0	1.01 (1.00, 1.01)	**<0.001**
	25th centile	27.1	16.4		
	Median	68.0	37.2		
	75th centile	134	76.0		
	Maximum	599	479		
Creatinine mmol/L	2016/2017				**<0.001***
	Minimum	53	7.6	QF[d]	
	25th centile	84.5	65.8		
	Median	106	84.9		
	75th centile	141	114.1		
	Maximum	178	928		
	2017/2018				
	Minimum	45	18		
	25th centile	74.8	61		
	Median	117	78		
	75th centile	146	102		
	Maximum	460	622		
Urea mmol/L	2016/2017				**<0.001***
	Minimum	3.7	1.6	QF[e]	
	25th centile	7.8	4.8		
	Median	9.3	6.4		
	75th centile	11.4	8.8		
	Maximum	18.5	39.2		
	2017/2018				
	Minimum	3.0	1.3		
	25th centile	6.9	4.0		
	Median	9.8	5.6		
	75th centile	14.1	7.7		
	Maximum	41.1	47.3		
Glucose mmol/L	Minimum	3.8	3.3	1.04 (0.98, 1.11)	0.2
	25th centile	6.5	6.0		
	Median	7.8	6.9		
	75th centile	8.8	8.4		
	Maximum	24.0	36.2		
Continuous Positive Airways Pressure	Yes	4	13	3.71 (1.18, 11.6)	**0.046**
	No	79	952	1.00	
Non-invasive ventilation	Yes	7	13	6.74 (2.61, 17.4)	**<0.001**
	No	76	952	1.00	
Invasive ventilation	Yes	9	35	3.23 (1.50, 6.98)	**0.007**
	No	74	930	1.00	

F.3: Single variable analysis of factors associated with mortality in patients with confirmed influenza, 2016-18

Variable	Category or measure	Expired	Not expired	OR (95% CI)	*P*-value
Myocardial infarct	Yes No	17 66	87 878	2.60 (1.46, 4.63)	0.003
Congestive heart failure	Yes No	14 69	78 887	2.31 (1.24, 4.29) 1.00	0.01
Peripheral vascular disease	Yes No	5 78	28 937	2.15 (0.81, 5.71) 1.00	0.16
Cerebro vascular disease	Yes No	18 65	97 868	2.48 (1.41, 4.35) 1.00	0.003
Dementia	Yes No	27 56	113 852	3.64 (2.21, 5.99) 1.00	<0.001
Chronic lung disease	Yes No	35 48	278 687	1.80 (1.14, 2.85) 1.00	0.01
Connective tissue disease	Yes No	4 79	43 922	1.09 (0.38, 3.10) 1.00	0.9
Peptic ulcer	Yes No	3 80	16 949	2.22 (0.63, 7.79) 1.00	0.3
Mild liver disease	Yes No	1 82	6 959	1.95 (0.23, 16.4) 1.00	0.6
Moderate or severe liver disease	Yes No	4 79	18 947	2.66 (0.88, 8.06) 1.00	0.12
Diabetes without end organ damage	Yes No	9 74	103 862	1.02 (0.49, 2.09) 1.00	0.96
Diabetes with end organ damage	Yes No	5 78	47 918	1.25 (0.48, 3.24) 1.00	0.99
Hemiplegic	16/17 Yes No 17/18 Yes No	 1 30 3 49	 13 274 4 674	 0.70 (0.90, 5.56) 1.00 6.85 (1.46, 32.1) 0.66 (0.41, 1.07)	0.03*

Variable	Category or measure	Expired	Not expired	OR (95% CI)	*P*-value
Moderate or severe kidney disease	Yes No	14 69	86 879	2.07 (1.12, 3.84) 1.00	0.03
Tumour without metastasis	Yes No	9 74	79 886	1.36 (0.66, 2.83) 1.00	0.4
Tumour with metastasis	Yes No	2 81	36 929	0.64 (0.15, 2.69) 1.00	0.5
Leukaemia	Yes No	2 81	50 915	0.45 (0.11, 1.89) 1.00	0.22
Lymphoma	Yes No	1 82	33 932	0.34 (0.05, 2.55) 1.00	0.22
Obesity	Yes No	1 82	20 945	0.58 (0.08, 4.35) 1.00	0.6
Non- age adjusted Charlson co-morbidity index units	Minimum 25th centile Median 75th centile Maximum	0 1 2 4 8	0 0 1 3 10	QF[f]	**0.001**
Apportionment	Community Hospital	50 32	761 196	1.00 2.48 (1.55, 3.98)	**0.003**
Admitted from	Own home Residential care Another hospital Other	57 18 7 1	825 69 36 31	1.00 3.78 (2.11, 6.77) 2.81 (1.20, 6.60) 0.47 (0.06, 3.48)	**<0.001**

n.e. = not estimable

[*]p-value for interaction

Bold p value - variable met criteria for inclusion in multivariable modelling or part of study hypothesis

QF Quadratic function, CF Cubic function

CF[a] CT Cycle Threshold counts. At least cubic: Linear: OR 7.82 (0.56, 108); Quadratic: OR 0.92 (0.83, 1.01); Cubic: OR 1.00 (1.00, 1.00)

CF[b] Serum haemoglobin g/L At least cubic: Linear: OR 1.38 (0.95, 2.01) Quadratic: OR 1.00 (0.99, 1.00) Cubic: OR 1.00 (1.00, 1.00)

QF[C] Total white cell count 10^9 / L. Linear: OR 1.15 (1.09, 1.21) Quadratic: OR 1.00 (1.00, 1.00)

QF[d] Serum creatinine mmol/L. Quadratic 2016/17:- Linear: 1.06 (1.02, 1.11), Quadratic: OR 1.00 (1.00, 1.00). 2017/18:- Linear: OR 1.02 (1.01, 1.04) , Quadratic: OR 1.00 (1.00, 1.00)

QF[e] Serum urea mmol/L. Quadratic 2016/17- Linear: OR 2.76 (1.59, 4.80) Quadratic: OR 0.96 (0.94, 0.99); 017/18:- Linear: OR 1.37 (1.21, 1.55), Quadratic: OR1.00 (1.00, 1.00)

QF[f] Non age adjusted Charlson Co-morbidity Index. Units. Linear: OR 2.03 (1.43, 2.89) Quadratic: OR 0.93 (0.89, 0.98)

F.4 Multivariable analysis of factors associated with mortality in patients with confirmed influenza, 2016-18

Original analysis performed by Neville Verlander, based on 852 observations.

Variable	Category	OR	95% CI	*P*-value
Age		1.05	1.02-1.07	<0.001
Apportionment	Community Hospital	1.00 2.93	 1.42-6.02	0.004
Admitted from	2016/17 Own home Residential care Another hospital Other 2017/18 Own home Residential care Another hospital Other	 1.00 0.68 0.39 0.00 1.00 8.28 4.70 6.94	 0.17-2.71 0.06-2.65 n.e. 2.83-24.2 0.85-25.8 0.71-68.1	0.008*
Non-invasive ventilation	Yes No	8.46 1.00	2.20-32.5	0.003
Admitted to critical care	Yes No	3.49 1.00	1.26-9.70	0.02
Radiological evidence of pneumonia	Yes No	2.43 1.00	1.25-4.72	0.009
Excessive alcohol use	Yes No	13.2 1.00	1.93-90.5	<0.001
Days between admission and first dose of oseltamivir		1.05	0.97-1.14	0.01
Haemoglobin		QF[a]		0.01*
Total white cell count		1.08	1.03-1.15	0.003
Urea		QF[b]		0.007*
Cycle Threshold (CT) value		CF[c]		0.06

*p-value for interaction n.e.= not estimable

QF Quadratic function, CF cubic function

QFa Haemoglobin g/L 2016/17:- Linear: OR 1.05 (0.83, 1.34) Quadratic: OR 1.00 (1.00, 1.00)

2017/18:-OR Linear: 0.84 (0.74, 0.96) Quadratic: OR 1.00 (1.00, 1.00)

QF[b] Serum urea mmol/L 2016/17:- Linear: OR 2.42 (1.16, 5.04) Quadratic: OR 0.96 (0.93, 1.00)

2017/18:-Linear: OR 1.18 (0.90, 1.55) OR Quadratic: 1.00 (0.99, 1.01)

CF[c] CT value Linear: OR 19.3 (0.54, 686) Quadratic: OR 0.89 (0.78, 1.01) Cubic: OR 1.00 (1.00, 1.00)

Appendix G Detailed transmission networks

G.1 SARS-CoV-2 transmission networks in CUH

No.	Size	Acquisition				Dates		Epidemiological links			Notes
		CAIs	Possible HAIs*	HAIs	HCWs	First Case	Last Case	Strong	Possible	None	
1	15	4	4	5	2	20/3/20	16/4/20	14	0	1	Eight patients swabbed during a suspected outbreak on ward 1. The first patient to be swabbed was recently discharged from ward 2 and then re-admitted to ward 3. The two HCWs worked on ward 2; one had direct contact with the first patient to be swabbed in this cluster. Two patients, distinct from the larger outbreak, co-habit (spouses). The final patient has no identifiable association with any other case in this cluster.
2	15	10	1	0	4	27/3/20	20/4/20	9	5	1	8 patients were resident in the care home A. One patient works as a carer in the same home. Another patient also works as a carer in an unspecified home. Two of the HCWs were paramedics. The other two HCWs work in different clinical areas, but both live with paramedics. One patient had no identifiable contact with the other cases in this cluster.
3	10	1	2	3	4	1/4/20	20/4/20	7	1	2	Four patients swabbed during a suspected outbreak on ward 4 (ward B); a fifth patient was discharged from ward 4 and re-admitted with COVID within 2 weeks of discharge. Two HCWs work on ward 4 and a third works in the same department as one of these HCWs. There are no identified associations between the remaining patient or the fourth HCW.

No.	Size	Acquisition				Dates		Epidemiological links			Notes
		CAIs	Possible HAIs*	HAIs	HCWs	First Case	Last Case	Strong	Possible	None	
4	8	7	0	0	1	20/3/20	17/4/20	1	0	7	No identified epidemiological link between the seven patients. The HCW had direct contact with two of these patients.
5	5	0	3	2	0	27/3/20	11/4/20	5	0	0	Five patients all swabbed on ward 5 (ward A)
6	5	4	0	1	0	27/3/20	19/4/20	0	0	5	No identified epidemiological link between the 5 patients.
7	5	0	4	0	1	20/3/20	16/4/20	4	1	0	Three patients swabbed during a suspected outbreak on ward 2. HCW works on ward 2. The first patient in this cluster was recently discharged from ward 8 and readmitted to ward 8. Wards 2 and 8 are in the same department and share some staff
8	5	2	0	0	3	7/4/20	17/4/20	5	0	0	Two patients were resident in care home B; one of these patients admitted to ward 6. Two HCWs work on ward 6, including one with direct contact with this patient. The third HCW works on ward 8.
9	4	2	0	0	2	4/4/20	16/4/20	4	0	0	One patient resident in care home B, admitted to ward 7. Second patient is a carer in care home B. One HCW works on wards 7 and 8. The second HCW works on ward 8. The isolates from clusters 8 and 9 diverge by 1 SNP, but share epidemiological links of both patients and HCWs.
10	6	0	6	0	0	1/4/20	20/4/20	6	0	0	All six patients receive dialysis at the same unit
11	3	2	1	0	0	9/4/20	17/4/20	0	0	3	Two patients live in separate care homes. The third patient is a community carer but has no other identified association with these patients
12	3	3	0	0	0	11/4/20	20/4/20	0	2	1	One patient is a resident in a specialist dementia care home. A second patient works in a specialist dementia care home, but it is unclear whether this is the same home as the first patient. The third patient has no other identified associations with the other patients.
13	3	0	0	1	2	31/3/20	17/4/20	0	0	3	The first patient acquired infection in hospital. The HCWs have no known association with each other or the patient.
14	3	0	0	2	1	12/4/20	20/4/20	3	0	0	The two patients were co-located on ward 10. The HCW worked on the same ward, but there is no documented direct contact with either patient.

No.	Size	Acquisition				Dates		Epidemiological links			Notes
		CAIs	Possible HAIs*	HAIs	HCWs	First Case	Last Case	Strong	Possible	None	
15	2	1	0	0	1	2/4/20	13/4/20	2	0	0	The patient was admitted to ward 6 with documented direct contact with the HCW.
16	2	1	1	0	0	4/4/20	5/4/20	2	0	0	Two patients co-located on ward 9.
17	2	0	0	0	2	16/4/20	17/4/20	0	2	0	Two HCWs, working in separate wards in the same department.
18	2	1	0	0	1	15/4/20	18/4/20	0	0	2	The HCW lives with a care assistant in an unspecified care home. The patient was admitted from a care home. No other identified association
19	2	1	0	0	1	19/4/20	20/4/20	0	0	2	The patient works in a local community hospital. The HCW works in a rehabilitation unit. There are no other known epidemiological associations.
20	2	0	0	1	1	9/4/20	12/4/20	0	0	2	The HCW and patient have moved between multiple ward areas. No direct contact documented.
21	2	1	0	1	0	3/4/20	3/4/20	0	2	0	Two patients briefly co-located within 24 hrs of testing, likely insufficient duration for transmission. One patient was resident with community acquired infection in a care home; the second patient was on a rehabilitation ward prior to swabbing
22	2	1	0	1	0	22/3/20	27/3/20	0	0	2	No identified associations between the two patients
23	2	1	1	0	0	30/3/20	19/4/20	0	2	0	No direct associations between two patients, but they were co-located on neighbouring wards (wards 1, 4 and 10), which share some HCWs
24	2	2	0	0	0	29/3/20	7/4/20	0	0	2	No identified associations between the two patients
25	2	0	0	2	0	29/3/20	7/4/20	2	0	0	Two patients co-located on ward 11
26	2	2	0	0	0	1/4/20	2/4/20	2	0	0	Two patients co-habiting (siblings)
Total	**114**		**23**	**19**	**25**			**66**	**15**	**33**	

G.2 SARS-CoV-2 transmission networks in the University of Cambridge

(a) Summary of the 16 genomic clusters of two or more university cases, generated using the clustering tool CIVET (UoC = University of Cambridge; HCW = healthcare workers).

Cluster	Lineage	Number of cases					Epiweek in which cases identified						Notes
							Whole cohort		UoC		Community		
		Total	UoC	HCW	Patients	Community	First Case	Last Case	First Case	Last Case	First Case	Last Case	
1	B.1.160.7	354	337	1	1	15	41	49	41	49	43	48	See supplementary table (b) below
2	B.1.36	32	30	0	0	2	41	43	42	43	41	41	Of the 30 university members: 24 were students living in the same accommodation block in College A; 4 were students living in different accommodation in the same College, 3 of whom were living in the same household; 2 students in different colleges had no identified associations with other cases in this cluster.
3	B.1.177.16	35	20	0	0	15	41	50	42	47	41	50	Of the 20 university members: 5 belong to 2 households on neighbouring staircases in College B; 4 belong to 2 neighbouring households in a different accommodation block in College B; 1 further student is resident in a different accommodation block in College B; 2 further cases from College C, and one from College D share the same course and year of study as a student from College B; the 2 students in College C are named contacts of each other in university contact tracing and share a common exposure with an individual at College B in national contact tracing; a student from College E lives in the same household as a student whose isolate did not sequence, but who is on the same course as the students from Colleges C and D and is named in university contact tracing by the student in College D; 2 further students from College A live in the same household but, as with the remaining two students, have no identified epidemiological associations with any other student in this cluster. No further growth of the cluster is seen amongst students after week 3, but 2 infections are noted in week 7, both in university staff members who share a household.
4	B.1.177	201	25	38	30	108	40	51	41	49	40	51	Of the 25 university members: 2 share the same household in College G; 1 student shares the same course and year of

Cluster	Lineage	Number of cases					Epiweek in which cases identified						Notes
							Whole cohort		UoC		Community		
		Total	UoC	HCW	Patients	Community	First Case	Last Case	First Case	Last Case	First Case	Last Case	
													study as one of the students from College G; 2 students in separate colleges are in the same year and course as each other; 2 staff members work in the same college (no students are identified from this college) and live very close to one another; 4 students are clinical medical students in the same block of accommodation in College H; a 5th clinical medical student is from College H but lives in a different household and is in a different year of study; 7 are clinical medical students in neighbouring households in College I; 2 further clinical students in different colleges are named contacts of the index case in College I; 2 students and 2 staff members have no obvious association with anyone else in this cluster
5	B.1.177	7	4	0	0	3	44	48	44	45	45	48	All 4 of the university members are students living in the same household
6	B.1.177	4	3	0	0	1	46	46	46	46	46	46	Of the 3 university members: 2 are students in the same college, but different households and courses; there is no evident association with the 3rd case, a member of staff
7	B.1.177	4	2	0	0	2	43	45	45	45	43	45	The 2 university members are students who share the same household and course
8	B.1.177	3	3	0	0	0	45	45	45	45			Of the 3 university members: 2 are students in the same household and course; there is no evident association with the 3rd student
9	B.1.177.17	41	6	2	2	31	40	50	43	48	40	50	Of the 6 university members: 2 are members of staff in the same college; 1 of these lives in the same household as another individual in this cluster; there is no evident association with the other 3 members
10	B.1.177	10	9	0	0	1	42	45	42	44	45	45	Of the 9 university members: 2 share the same course and year of study in College J; 1 further student in College J shares the same course, but is in a different year of study; 3 further students in College K share the same course as the 2 students in College J, with 2 being in the same year of study; 1 further student in College J shares the same year of study with other students, but is on a different course; 1 further student in College K shares the same year of study, but a different course; one student in College C is a named contact of a student from College J; 1 student has no identified association with any other students.

Cluster	Lineage	Number of cases					Epiweek in which cases identified						Notes
							Whole cohort		UoC		Community		
		Total	UoC	HCW	Patients	Community	First Case	Last Case	First Case	Last Case	First Case	Last Case	
11	B.1.177.4	19	7	0	1	11	44	47	45	47	44	47	Of the 7 university members, 4 are staff and 3 are students; 2 staff work in the same 'additional personnel' department. There is no identified association between the remaining members of this cluster
12	B.1.1.315	21	6	3	1	11	39	47	45	47	39	46	Of the 6 university members: 2 students share the same household and are both PhD students in the same department; a 3rd student is also a PhD student in this department but living in a different household and college; 2 individuals share the same household, but like the remaining individual have no association with the other cases in this cluster
13	B.1.258	12	3	0	0	9	42	46	42	42	44	46	Of the 3 university members: 2 students share the same postgraduate course and work in the same department; there is no identified association with the third student
14	B.1	5	3	0	0	2	45	48	46	46	45	48	Of the 3 university members: 2 staff members share the same household; there is no identified association with the third member, a student
15	B.1.177	27	2	2	1	22	41	49	47	48	41	49	Both university cases are staff members with no identified associations
16	B.1.1.153	4	2	0	0	2	41	44	42	42	41	44	Both university cases are students in the same academic year with no other identified associations

(b) Summary of the 19 genomic clusters of two or more university cases, using a SNP difference threshold of 0 SNPs, based on isolates from cluster 1 identified by CIVET in table (a) above (UoC = University of Cambridge; HCW = healthcare workers).

Cluster	Lineage	Number of cases					Epiweek in which cases identified						Notes
							Whole cohort		UoC		Community		
		Total	UoC	HCW	Patients	Community	First Case	Last Case	First Case	Last Case	First Case	Last Case	
A	B.1.160.7	182	176	0	0	6	44	48	44	47	46	48	There are a large number of cases that emerge in the same week, making further analysis challenging. Of the 176 university members: 113 are students sharing a household with at least one other individual in this cluster; the largest household cluster is 11 students living on the same staircase; 155 students share a course and year of study with at least one other individual in the cluster, with some overlap with college household structure; the largest cluster sharing course, year of study and college is 7 students.
B	B.1.160.7	6	6	0	0	0	46	48	46	48			Of the 6 university members: all 6 are students living in shared or neighbouring households in the same college; 3 of these students are in the same course and year of study
C	B.1.160.7	2	2	0	0	0	46	47	46	47			The 2 university members are students that share the same year and college, and are identified contacts in university contact tracing
D	B.1.160.7	2	2	0	0	0	46	46	46	46			Both cases are students, but have no identified association.
E	B.1.160.7	62	60	0	1	1	41	46	41	46	44	46	Of the 62 university members: 21 students share a household with at least one other individual in this cluster; the largest household cluster is 4 students living on the same staircase; 29 students share a course and year of study with at least one other individual in the cluster, with some overlap with household structure; the largest cluster sharing course and year of study is 5 students; of note, two students in the first week that this cluster was identified report attending Venue A on the same day in the first week of term.
F	B.1.160.7	3	3	0	0	0	44	45	44	45			Of the 3 university members: 2 students share the course and year of study; there is no identified association with the 3rd student.
G	B.1.160.7	2	2	0	0	0	42	42	42	42			Both cases are students, but have no identified association.
H	B.1.160.7	3	3	0	0	0	46	46	46	46			All 3 university cases are students, but have no identified association.

Cluster	Lineage	Number of cases					Epiweek in which cases identified						Notes
							Whole cohort		UoC		Community		
		Total	UoC	HCW	Patients	Community	First Case	Last Case	First Case	Last Case	First Case	Last Case	
I	B.1.160.7	2	2	0	0	0	46	46	46	46			Both university cases are students, and share the same household and course
J	B.1.160.7	4	4	0	0	0	44	45	44	45			Of the 4 university members: 3 are students from the same college (2 in the same household) but different courses; 1 member of staff has no known associations with the students
K	B.1.160.7	2	2	0	0	0	44	44	44	44			Both university cases are students, and share the same course and year of study
L	B.1.160.7	2	2	0	0	0	43	44	43	44			Both cases are students, but have no identified association.
M	B.1.160.7	2	2	0	0	0	44	44	44	44			Both cases are students, but have no identified association.
P	B.1.160.7	9	9	0	0	0	45	48	45	48			Of the 9 university members: 4 live in the same/neighbouring households in College L; 2 live in neighbouring households from College J; a further student is also at College J; 2 the remaining 2 students have no identified association with the rest of the cluster
Q	B.1.160.7	13	12	0	0	1	44	46	44	46	46	46	Of the 13 university members: 3 are students on the same course and year of study (1 is named as a contact of the first; another is in the same college as the first); 6 live in the same block of accommodation in a different college; the 4 remaining members have no identified association.
R	B.1.160.7	4	4	0	0	0	45	47	45	47			Of the 4 university members: 2 students live in the same household; a 3rd student lives in the same college on a different course and year of study; 1 of the students has no identified association with other members of this cluster; one of the students lives in the same household and shares a course and year of study with a student in cluster Q, which is one SNP different from R
S	B.1.160.7	18	15	0	0	3	43	45	43	45	43	45	Of the 15 university members: 5 are students sharing the same course and household at College L; 2 live in a different household at College L; 4 live in the same accommodation block in College M; 4 live in College N, of which 1 is a named contact of another in university contact tracing; 1 of the students in College L and another at College M share the same course and year of study
T	B.1.160.7	2	2	0	0	0	41	41	41	41			These are the earliest 2 isolates from this large cluster during the study. There is no identified association between these 2 students; 1 student has an isolate that is 1 SNP different from a household contact from cluster E

Appendix H Supplementary COVID-19 at CUH data

H.1 Characteristics of COVID-19 cases requiring critical care at CUH

NB only laboratory-confirmed cases at CUH included

	Cohort 1 (n=95)	**Cohort 2 (n=21)**	**Cohort 3 (n=186)**	**Total (n=302)**
Age, mean (range) median (IQR)	56 (0-82) 61 (48-69)	59 (3-83) 60 (56-65)	57 (1-87) 59 (49-67)	57 (0-87) 60 (49-67)
Paediatric (age <16 yrs)	4 (4%)	1 (5%)	4 (2%)	9 (3%)
Male gender (%)	66 (69%)	16 (76%)	118 (63%)	200 (66%)
Female gender (%)	29 (31%)	5 (24%)	68 (37%)	102 (34%)
Ethnicity – White	63 (66%)	14 (67%)	93 (50%)	170 (56%)
Ethnicity – Black, Asian and minority ethnic	15 (16%)	4 (19%)	37 (20%)	56 (19%)
Ethnicity – missing/not stated	17 (18%)	3 (14%)	56 (30%)	76 (25%)
Charlson co-morbidity index, median (IQR)	1 (0-2)	1 (0-3)	1 (0-2)	1 (0-2)
Length of stay (days)	**Cohort 1**	**Cohort 2**	**Cohort 3**	**Total**
Pre-ICU* – no. of pts mean (range) median (IQR)	79 1.7 (0.0-34.0) 0.3 (0.1-1.7)	15 3.0 (0.1-23.3) 0.9 (0.2-3.9)	121 1.8 (0.1-16.4) 1.0 (0.3-2.3)	215 1.9 (0.0-34.0) 0.6 (0.3-2.3)
On ICU** – no. of pts mean (range) median (IQR)	95 17.0 (0.3-63.6) 13.0 (5.3-24.5)	21 11.1 (2.4-32.3) 7.7 (4.9-13.8)	181 12.8 (0.2-52.6.8) 8.1 (3.5-19.5)	297 14.0 (0.2-68.6) 9.3 (4.6-20.3)
Post-ICU*** – no. of pts mean (range) median (IQR)	53 16.7 (0-103.8) 7.8 (4.1-14.1)	12 10.1 (0-58.1) 5.3 (3.1-8.1)	97 9.7 (0.3-51.0) 6.1 (4.1-12.0)	162 12.0 (0-103.8) 6.1 (4.0-13.0)

Length of stay (min 30 days post first positive result)	**Cohort 1**	**Cohort 2**	**Cohort 3**	**Total**
On ICU** – number of pts mean (range) median (IQR)	95 17.0 (0.3-63.6) 13.0 (5.3-24.5)	21 11.1 (2.4-32.3) 7.7 (4.9-13.8)	176 12.6 (0.2-51.6) 8.2 (3.5-19.5)	292 14.0 (0.2-68.6) 9.3 (4.6-20.2)
Post-ICU*** – no. of pts mean (range) median (IQR)	53 16.7 (0-103.8) 7.8 (4.1-14.1)	12 10.1 (0-58.1) 5.3 (3.1-8.1)	96 9.7 (0.3-51.0) 6.1 (4.1-12.0)	161 12.0 (0-103.8) 6.1 (4.0-13.0)
Outcome (min 30 days post first positive result)	**Cohort 1 (n=95)**	**Cohort 2 (n=21)**	**Cohort 3 (n=179)**	**Total (n=170)**
Discharged to ward	0	0	4 (2%)	4 (1%)
Current Inpatient on ICU	0	0	3 (1%)	3 (1%)
Transferred	12 (13%)	0	20 (12%)	32 (11%)
Discharged from hospital	55 (58%)	12 (57%)	96 (54%)	163 (55%)
Died in hospital or within 30 days of positive swab	28 (30%)	9 (43%)	56 (32%)	93 (32%)

*community acquired cases, excluding transfers from other hospitals; **cases discharged or transferred from critical care, or deceased; ***cases discharged from hospital

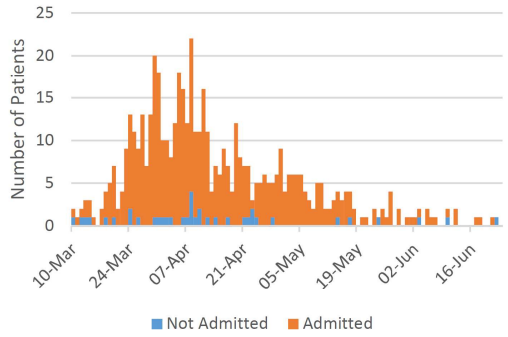
H.2 CUH COVID report 10: 23rd June 2020

Example report describing the burden of infection in CUH, alongside the demographic and clinical characteristics of the COVID-19 patient cohort. These reports were produced weekly during the first wave and disseminated to clinical teams, senior management and, via internal communications, to all staff at CUH.

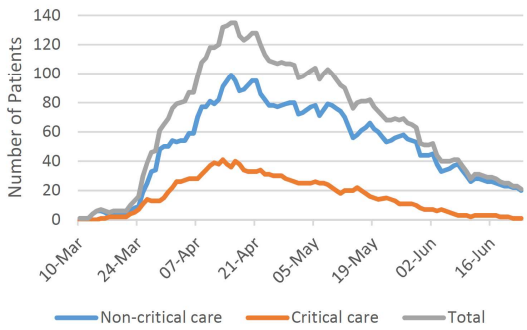
CUH COVID Report 10 – 23rd June 2020

Report author – Ben Warne, ben.warne@addenbrookes.nhs.uk. Data correct as of 6AM 22nd June 2020

Confirmed Cases (by Swab Date)



Bed Occupancy of All Cases



Confirmed COVID Cases Summary:

Current Status of Patients	
Current inpatients on medical wards	28 (5%)
Current inpatients on critical care	4 (1%)
Discharged from inpatient stay	312 (59%)
Swabbed in ED/N2/outpatients only	36 (7%)
Deceased	124 (24%)
Inpatient transfer	21 (4%)
Total	**525**

Patients discharged from ED/N2 (n=52)*

Age, mean (range) median (IQR)	47 (22-88) 45 (32-57)
Male gender (%)	32 (62%)
Female gender (%)	20 (38%)
Readmissions	19 (37%)

*Further 8 patients swabbed only in outpatient settings

All Admissions (n=489)

Age, mean (range) median (IQR)	66 (0-98) 70 (55-82)
Paediatric (age <16 yrs)	11 (2%)
Male gender (%)	290 (59%)
Female gender (%)	199 (41%)
Ethnicity – White	356 (73%)
Ethnicity – Black, Asian and minority ethnic	39 (8%)
Ethnicity – missing/not stated	94 (19%)
LOS* (days), mean (range) median (IQR)	10.8 (0.3-73.1) 7.7 (2.4-15.0)

*(n=213 discharged pts, not including readmissions, with community acquired infections); LOS = Length of stay, IQR = interquartile range

Data underlying this report generated from tools prepared by Peter Driscoll and Stuart Wills.

Admissions requiring critical care (n=96)

Age (yrs)	Mean, (range) median (IQR)	56 (0-82) 61 (48-69)
Gender	Male	68 (71%)
	Female	28 (29%)
Ethnicity	White	54 (56%)
	Black, Asian and minority ethnic	14 (15%)
	missing/not stated	28 (29%)
Inpatient transfers		13 (14%)
LOS (days)	Pre-ICU (mean, range) median (IQR)*	1.7 (0.0-34.0) 0.3 (0.1-1.7)
	On ICU (mean, range) median (IQR)**	18.8 (0.3-84.6) 13.3 (5.3-26.6)
	Post-ICU (mean, range) median (IQR)***	7.7 (0-31.8) 4.9 (3.7-10.1)
Current Status	Discharged to ward	11 (11%)
	Deceased	28 (29%)
	Inpatient on ICU	4 (4%)
	Transferred	11 (11%)
	Discharged from hospital	42 (44%)

*(n=78 community acquired cases, excluding transfers from other hospitals); **(n=91 cases discharged or transferred from critical care, or deceased); ***(n=41 cases discharged from hospital)

Fatalities (n=124 since positive COVID swab taken)*

Age (mean, range) (median, IQR)	78 (0-97) 80 (73-87)
Gender (male)	76 (61%)
Gender (female)	48 (39%)
Ethnicity – White	95 (77%)
Ethnicity – Black, Asian, minority ethnic	4 (3%)
Ethnicity – missing/not stated	25 (20%)
LOS** (days), mean (range) median (IQR)	9.7 (0.2-29.7) 8.7 (4.3-13.5)
Critical care admission	28 (23%)

*123 within 30 days of swab **(n=78 community acquired cases)

Appendix I Supplementary COVID-19 data at the University of Cambridge

I.1 Tests conducted

Week	Dates	Eligible students	Participating students	Asymptomatic screening programme								University symptomatic testing	
				Screening tests	Valid results	Scale*	Mean swab count per pool	Students screened (estimated)	Individual confirmatory tests	Valid results	Students screened per total tests	Individual tests	Valid results
1	5 Oct – 11 Oct	15,479	11,638	1,867	1,837	2 students per pool	1.87	3,435	36	34	1.81	102	96
2	12 Oct – 18 Oct	15,511	12,100	1,890	1,866	2 students per pool	1.92	3,583	56	56	1.84	221	219
3	19 Oct – 25 Oct	15,488	12,195	1,913	1,886	Half-pool	2.47	4,658	105	104	2.31	195	194
4	26 Oct – 1 Nov	15,440	12,383	1,923	1,900	Half-pool	2.92	5,548	109	109	2.73	109	109
5	2 Nov – 8 Nov	15,385	12,372	1,873	1,865	Half-pool	2.49	4,644	79	77	2.38	112	110
6	9 Nov – 15 Nov	15,323	12,350	1,864	1,851	Half-pool	2.89	5,349	228	226	2.56	265	263
7	16 Nov – 22 Nov	15,307	12,424	1,743	1,727	Half-pool	2.42	4,179	102	102	2.27	87	87
8	23 Nov – 29 Nov	15,309	12,498	1,919	1,889	Whole pool	5.02	9,483	45	44	4.83	28	28
9	30 Nov – 6 Dec	15,310	12,544	1,953	1,938	Whole pool	4.90	9,496	52	52	4.74	27	27
Weeks 1-2	5 Oct – 18 Oct			3,757	3,703	2 students per pool		7,018	92	90	1.82	323	315
Weeks 3-7	19 Oct – 22 Nov			9,316	9,229	Half-pool		24,379	623	618	2.45	768	763
Weeks 8-9	23 Nov – 6 Dec			3,872	3,827	Whole pool		18,979	97	96	4.78	55	55
All weeks	**5 Oct – 6 Dec**	**15,561**	**12,979**	**16,945**	**16,759**			**50,376**	**812**	**804**	**2.84**	**1,146**	**1,133**

*Phase 1 (weeks 1-2): two students per testing pool screened each week; phase 2 (weeks 3-7): half of each testing pool screened on alternating week; phase 3 (weeks 8-9): whole testing pool screened each week.

I.2 Case ascertainment

Week	Dates	Asymptomatic screening programme							University symptomatic testing	Other*	Total positive individual tests	% ascertained through screening
		Positive pooled screening tests	Confirmed positive pools (at least one positive individual test)	Positive individual tests	Positive individual tests per positive pool (mean, range)		Positive individual tests per positive household (mean, range)		Positive tests			
1	5 Oct – 11 Oct	19	11	12	1.1	1 to 2	1.1	1 to 2	6	2	20	60.0%
2	12 Oct – 18 Oct	28	27	35	1.3	1 to 2	1.2	1 to 2	79	8	122	28.7%
3	19 Oct – 25 Oct	38	36	38	1.1	1 to 2	1.2	1 to 3	50	12	100	38.0%
4	26 Oct – 1 Nov	35	30	38	1.3	1 to 3	1.3	1 to 3	21	7	66	57.6%
5	2 Nov – 8 Nov	26	22	23	1.1	1 to 2	1.1	1 to 2	37	17	76	28.9%
6	9 Nov – 15 Nov	67	59	80	1.4	1 to 4	1.4	1 to 4	122	26	228	35.1%
7	16 Nov – 22 Nov	22	17	27	1.6	1 to 4	1.6	1 to 4	15	6	48	56.3%
8	23 Nov – 29 Nov	7	3	3	1.0	1	1.0	1	1	0	4	75.0%
9	30 Nov – 6 Dec	10	0	0	-	-	-	-	4	2	6	0.0%
Weeks 1-2	5 Oct – 18 Oct	47	38	47	1.2	1 to 2	1.2	1 to 2	85	10	142	33.1%
Weeks 3-7	19 Oct – 22 Nov	188	164	206	1.3	1 to 4	1.3	1 to 4	245	68	519	39.7%
Weeks 8-9	23 Nov – 6 Dec	17	3	3	1.0	1	1.0	1	5	2	10	30.0%
All weeks	**5 Oct – 6 Dec**	**252**	**205**	**256**	**1.3**	**1 to 4**	**1.3**	**1 to 4**	**335**	**80**	**671**	**38.2%**

*Other includes: positive results obtained through other testing pathways, such as NHS testing facilities, and reported to the university by Public Health England; and 7 asymptomatic positive cases identified through a distinct programme of screening undertaken by the university in conjunction with local public health teams during an outbreak investigation in a single accommodation block during term week 3. An additional 2 cases reported to the university by Public Health England during the study period were excluded, because their test dates could not be confirmed