

**Development of a military deployable
metagenomic next-generation sequencing
workflow for the diagnosis of infectious
disease**

Carl Matthew Halford

**Submitted for the degree of Doctor of Philosophy
University of East Anglia
Quadram Institute Bioscience**

September 2024

Abstract

The conditions that deployed military personnel are exposed to can lead to high incidences of infectious diseases, with a significant impact on personnel wellbeing and the operational effectiveness of the deployment. Therefore, it is crucial that military clinicians can rapidly diagnose infectious diseases so that the optimal treatments are administered. The diagnostic tests currently available in deployed environments are limited by the logistical requirements, the time to provide an actionable result and the targeted nature of the tests.

Clinical metagenomics (CMg) has the potential to provide rapid, untargeted identification of the causative pathogen of infection and additional clinically relevant information. However, there are several limitations associated with CMg, including the high ratio of human to pathogen genetic material, the complexity of the methods, the logistical requirements and the difficulty of interpreting the results to inform treatments.

In this study, we aimed to address these limitations to develop, optimise and evaluate CMg methods suitable for use in deployed medical treatment facilities (MTF). CMg workflows were developed for the diagnosis of wound, gastrointestinal (GI) and respiratory infections. These methods were applied to samples collected from patients with suspected or confirmed infections and the results were compared against routine diagnostic tests. This included CMg workflows that were operated by a military biomedical scientist (BMS) within a deployed MTF, highlighting the technical and logistical challenges associated with CMg and providing the basis for the future development of CMg for deployed infectious disease diagnosis.

This work has demonstrated that CMg has potential as a future deployed diagnostic test, however, further optimisation, including reduced method complexity through automation and a better understanding of the interpretation of CMg results, is required to enable the successful implementation of CMg within the deployed diagnostic pathway.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Contents

Abstract	2
Acknowledgements.....	9
Glossary	10
1. Introduction.....	13
1.1 Infectious diseases and the military	13
1.2 Defence Medical Services	15
1.3 Deployed infectious disease diagnosis	15
1.3.1 Role one	17
1.3.2 Role two	18
1.3.3 Role three.....	22
1.4 DNA sequencing for diagnostics	23
1.4.1 Nanopore sequencing.....	24
1.5 Clinical metagenomics.....	25
1.5.1 Amplicon metagenomic sequencing	27
1.5.2 Shotgun metagenomics sequencing	28
1.5.3 Advantages of clinical metagenomics.....	29
1.5.4 Limitations of clinical metagenomics.....	30
1.6 Study aims.....	31
2. Development of a clinical metagenomics workflow for the diagnosis of wound infections	33
2.1 Contribution statement.....	33
2.2 Introduction.....	33
2.2.1 Clinical metagenomics.....	33
2.2.2 Host DNA depletion	34
2.2.3 Microbial DNA extraction	35
2.2.4 Whole genome amplification.....	36
2.2.5 Nanopore sequencing library preparation	37
2.2.6 Nanopore sequencing and data analysis	39
2.2.7 Aims	40
2.3 Methods.....	40
2.3.1 Contrived sample generation	40
2.3.2 Preparation of DNA from bacterial cultures	41
2.3.3 Pre-sequencing sample preparation.	41
2.3.4 Comparison of DNA extraction platforms	42
2.3.5 DNA quantification.....	43
2.3.6 DNA purification.....	43
2.3.7 DNA fragment size analysis.....	44
2.3.8 Whole genome amplification.....	44

2.3.9	DNA de-branching	44
2.3.10	Nanopore DNA sequencing	45
2.3.11	Bioinformatic and statistical analysis.....	45
2.3.12	qPCR.....	46
2.4	Results	48
2.4.1	Host depletion and DNA extraction from contrived wound swabs	48
2.4.2	Storage of wound swab samples	50
2.4.3	Host depletion optimisation.....	50
2.4.4	Whole genome amplification incubation time	54
2.4.5	Wound swab sample full clinical metagenomics protocol proof of concept..	55
2.4.6	Evaluation of the amplification of species-specific DNA by whole genome amplification.....	59
2.4.7	Whole genome amplification method comparison.....	60
2.4.8	Genomiphi whole genome amplification optimisation.....	61
2.4.9	Whole genome amplification vs non-whole genome amplification comparison.....	63
2.4.10	DNA extraction platform comparison	64
2.4.11	Full optimised workflow testing	68
2.5	Discussion	70
2.5.1	Host depletion	71
2.5.2	Nucleic acid extraction.....	72
2.5.3	Whole genome amplification kit comparison	73
2.5.4	Future work and considerations for testing clinical samples.....	74
2.6	Conclusions.....	76
2.6.1	Recommended wound clinical metagenomics workflow	76
3	Evaluation of a clinical metagenomics workflow for the diagnosis of wound infections.	78
3.1	Contribution statement.....	78
3.2	Introduction.....	78
3.2.1	Wound infections	78
3.2.2	Impact of wound infections	79
3.2.3	Wound infection aetiology.....	80
3.2.4	Wound infection diagnosis	81
3.2.5	Wound infection treatment.....	85
3.2.6	Limitations of current diagnostics.....	86
3.2.7	Clinical metagenomics for wound infection diagnosis	88
3.2.8	Study aims.....	88
3.3	Methods.....	88
3.3.1	Ethical approval	88
3.3.2	Sample collection	88

3.3.3	Clinical metagenomics.....	89
3.3.4	Bioinformatic and statistical analysis.....	90
3.3.5	Routine clinical microbiological investigation	91
3.3.6	Interpretation of clinical metagenomics results	92
3.3.7	qPCR.....	92
3.4	Results	93
3.4.1	Prospective wound swab sample collection	93
3.4.2	Sample processing for clinical metagenomics	95
3.4.3	Routine microbiology	98
3.4.4	Clinical metagenomics species identification	99
3.4.5	qPCR.....	102
3.4.6	Detection of antimicrobial resistance genes.....	104
3.4.7	Species diversity.....	108
3.5	Discussion	109
3.6	Conclusions.....	113
4	Comparison of stool sample storage methods for deployed clinical metagenomics	114
4.1	Contribution statement.....	114
4.2	Introduction.....	114
4.2.1	Traveller's diarrhoea	114
4.2.2	Aetiology of traveller's diarrhoea.....	115
4.2.3	Diagnosis of traveller's diarrhoea.....	116
4.2.4	Clinical metagenomics for the diagnosis of traveller's diarrhoea	118
4.2.5	<i>Cryptosporidium</i>	118
4.2.6	<i>Cryptosporidium</i> outbreak.....	119
4.2.7	Aims	120
4.3	Methods.....	121
4.3.1	Ethical approval	121
4.3.2	Sample collection	121
4.3.3	DNA extraction	121
4.3.4	DNA quantification.....	122
4.3.5	DNA fragment analysis	122
4.3.6	DNA purification.....	122
4.3.7	Rapid Barcoding library preparation method.....	122
4.3.8	Optimised Rapid Barcoding library preparation method	123
4.3.9	Rapid PCR barcoding library preparation method.....	123
4.3.10	Bioinformatic analysis	124
4.3.11	qPCR.....	124
4.3.12	Statistical analysis	125
4.4	Results	126

4.4.1	Sample characteristics.....	126
4.4.2	Optimisation of the Rapid Barcoding library preparation method	126
4.4.3	Sample processing	128
4.4.4	Detection of <i>Cryptosporidium</i>	130
4.4.5	Comparison of the detection of additional species and antimicrobial resistance genes	138
4.4.6	Operational considerations	141
4.4.7	Species diversity and microbiome composition.....	141
4.5	Discussion	145
4.5.1	Stool sample storage method comparison	145
4.5.2	Library preparation method comparison	147
4.5.3	Microbiome storage	148
4.5.4	Detection of additional pathogens.....	148
4.6	Conclusions.....	150
4.6.1	Recommended stool sample clinical metagenomics workflow	150
5	Evaluation of clinical metagenomics in a deployed medical treatment facility.....	152
5.1	Contribution statement.....	152
5.2	Introduction.....	152
5.2.1	Lower respiratory tract infections	152
5.2.2	Deployed lower respiratory tract infections diagnostics.....	153
5.2.3	Deploying clinical metagenomics workflows to the RFA Argus	154
5.2.4	Aims	155
5.3	Methods.....	156
5.3.1	Exercise one: viral clinical metagenomics workflow	156
5.3.2	Exercise two: bacterial clinical metagenomics workflow.....	159
5.3.3	BioFire FilmArray pneumonia panel plus analysis.....	162
5.4	Results	162
5.4.1	Exercise one: Deployment of a viral pathogen clinical metagenomics workflow to the RFA Argus	162
5.4.2	Exercise two: deployment of a bacterial clinical metagenomics workflow to the RFA ARGUS.....	163
5.5	Discussion	168
5.5.1	Exercise one: viral clinical metagenomics workflow	168
5.5.2	Exercise two: bacterial clinical metagenomics workflow.....	170
5.5.3	The future of deployed diagnostics	173
5.6	Conclusions.....	175
6	Concluding remarks.....	176
7	Annex	178
8	References	190

Figures

Figure 1. Organisation of deployed MTFs with associated medical evacuation systems among NATO members and allied nations.....	17
Figure 2. A graphical representation of a CMg workflow for the diagnosis of infectious disease.	27
Figure 3. Schematic diagram of the wound CMg wet-laboratory workflow.....	33
Figure 4. Evaluation of the impact of the Amies charcoal agar on the host DNA depletion and bacteria DNA extraction methods.....	49
Figure 5. Evaluation of wound swab sample storage methods.	50
Figure 6. The effect of pre and post-depletion vortexing on the host depletion and <i>E. coli</i> DNA extraction methods.	51
Figure 7. The effect of bead-beating duration on the extraction of <i>S. aureus</i> DNA.	52
Figure 8. Comparison of the original and optimised DNA extraction methods.	53
Figure 9. Depletion of human genomic DNA by the optimised host depletion method.	54
Figure 10. Optimisation of the Repli-G WGA amplification step duration.	55
Figure 11. AMR gene coverage from host depleted sample 1.	58
Figure 12. Comparison of the amplification of purified bacteria DNA by Repli-G WGA.	60
Figure 13. Comparison of the amplification of bacteria DNA by Repli-G and Genomiphi WGA methods.	61
Figure 14. The concentration of purified bacteria DNA following WGA.	62
Figure 15. Evaluation of increasing the Genomiphi WGA amplification step duration.	63
Figure 16. Extraction of bacterial DNA from contrived wound swab sample using three DNA extraction methods.	65
Figure 17. Concentration of host DNA extracted from contrived wound swab samples using three DNA extraction methods.....	66
Figure 18. Proportion of spiked bacteria sequence reads from samples extracted by three DNA extraction methods.	67
Figure 19. Comparison of the gene coverage of spiked AMR genes from three DNA extraction methods.	68
Figure 20. Proportion of sequence reads assigned to spiked bacteria and the human genome from contrived wound swab samples.	69
Figure 21. Identification of AMR genes from contrived wound swab samples.	70
Figure 22. Routine microbiological investigation of wound swab samples carried out in NHS microbiology laboratories.....	84
Figure 23. Proportion of aerobic and obligate anaerobic bacteria species identified by CMg.....	102
Figure 24. Correlation of qPCR with CMg for the detection of four high priority pathogen species.	104
Figure 25. Comparison of wound swab sample species α -diversity.....	109
Figure 26. The life cycle of <i>C. hominis</i>	119
Figure 27. Comparison of the sequencing metrics from the RBK and RBK OP prepared libraries.	127
Figure 28. Comparison of the sequencing metrics from RBK OP and RPCR prepared libraries.....	129
Figure 29. <i>Cryptosporidium</i> sequence reads from the BioFire FilmArray GI panel positive samples.....	131
Figure 30. <i>C. hominis</i> Cq values from BioFire FilmArray GI panel positive samples stored using three methods.	132
Figure 31. Correlation of qPCR and CMg for the detection of <i>Cryptosporidium</i>	137
Figure 32. Frequency of AMR gene identifications.	140
Figure 33. Identification of AMR genes associated with antibiotic drug classes.....	141
Figure 34. Microbial species α -diversity from stool samples.....	142

Figure 35. Proportion of classified microbial reads assigned to the most abundant phyla.	143
Figure 36. Proportion of classified microbial reads assigned to the most abundant phyla per individual from samples stored in three storage methods and prepared using the RPCR method.	144
Figure 37. Image of the equipment and laboratory set-up used on board the RFA Argus to process samples for CMg.	156

Tables

Table 1. Primer and probe sequences for spiked bacteria species and host qPCR assays.	47
Table 2. Cq values of the spiked bacteria and human genomic DNA in the spiked depleted wound swab samples pre and post-WGA.	56
Table 3. Sequencing results from the full CMg workflow proof of concept experiment.	57
Table 4. CMg LOD sequencing results.	59
Table 5. WGA compared to non-WGA sequencing results.	64
Table 6. Primer and probe sequences for high priority pathogen qPCR assays	93
Table 7. Wound swab sample characteristics	94
Table 8. Results of CMg analysis of the wound swab samples.	95
Table 9. Results of routine microbiology testing and CMg analysis of wound swab samples.	99
Table 10. Sensitivity and specificity of CMg compared to qPCR and routine microbiology testing.	103
Table 11. Results of routine microbiology AST and AMR genes identified by CMg from wound swab samples.	105
Table 12. Primer and probe sequences for <i>C. hominis</i> qPCR assays.	125
Table 13. Comparison of the detection of <i>Cryptosporidium</i> by CMg using the RBK and RBK OP library preparation methods.	128
Table 14. Comparison of CMg and qPCR for the detection of <i>Cryptosporidium</i> .	133
Table 15 A-D. Sensitivity, NPV, specificity and PPV of CMg for the detection of <i>Cryptosporidium</i> .	135
Table 16. Detection of additional enteric pathogens by CMg.	139
Table 17. Results of the CMg analysis of nasopharyngeal swab samples processed on board the RFA Argus.	163
Table 18. Results of CMg from sputum samples processed on board the RFA Argus and additional routine microbiology and BioFire FilmArray pneumonia panel plus testing carried out in the UK.	164
Table 19. Results of AST and identification of AMR genes by BioFire FilmArray pneumonia plus panel and CMg.	166
Table A 1. Detection of high priority wound pathogen species by microbiology testing, qPCR and CMg.	178
Table A 2. Detection of AMR genes from DNA Shield RBK samples.	180
Table A 3. Detection of AMR genes from OMNIgene GUT RBK samples.	184
Table A 4. Detection of AMR genes from FTA RBK samples.	189

Acknowledgements

I would like to start by thanking my supervisory team for their support, guidance and time. To Dr. Simon Weller, thank you for believing in this project and for all your work to develop the wider diagnostics area at Dstl, without which my project would not have been possible. To Dr. Matthew Gilmour, thank you for stepping into the role of supervisor and providing a steady hand when you were only just at the beginning of your time at the Quadram Institute. I would also thank Dr. Justin O'Grady for helping to get the project started and for all the work conducted by his group that made the methods applied in this study possible.

I would like to acknowledge the UK Ministry of Defence, via the CBR programme at Dstl, for providing the funding and to the Quadram Institute Bioscience and the University of East Anglia for providing the academic partnership.

I would like to thank my colleagues at Dstl and all the external collaborators that have contributed to the project for their hard work, advice and guidance that has helped me to navigate through completing a PhD at Dstl. To Dr. Lluís Moragues-Solanas, thank you for taking the time to show me the ropes during my visits to the Quadram and good luck for the next phase of your career in Barcelona. I would also like to thank my mentor Prof. Sarah Harding and my colleague Dr. Sophie Smither for convincing me to aim for a PhD when my path forward was not so clear, and for going above and beyond to be helpful, both personally and professionally.

Finally, I would like to thank my family and friends. I am very lucky to have had you all with me. To my dad, your successful career in science has always been an inspiration. To my mum, who was here to see the start of the project, but could not be here for the end, I know that you would have been proud. To Georgie, thank you for providing an outlet when I needed it most. To my wife, Melissa. You have supported me throughout all the time it has taken to get here, which started long before this PhD. We are, and will always be, a team. Your love and support mean the world to me, and I would not have gotten here without you.

Glossary

AMR	Antimicrobial Resistance
API	Analytical Profile Index
ARMA	Antimicrobial Resistance Mapping Application
AST	Antimicrobial Susceptibility Testing
BATUK	British Army Training Unit Kenya
BMS	Biomedical Scientist
BNH	Basingstoke North Hospital
BP	Base Pairs
CARD	Comprehensive Antibiotic Resistance Database
cDNA	complementary DNA
CFU	Colony Forming Units
CLED	Cystine Lactose Electrolyte Deficient
CMg	Clinical Metagenomics
COTS	Commercial off the Shelf
COVID-19	Coronavirus disease 2019
Cq	Quantitation Cycle
CRP	C-Reactive Protein
CRU	Cryptosporidium Reference Unit
CTX-M	Cefotaximase-Munich
DCS	Damage Control Surgery
DLOD	Defence Lines of Development
DMS	Defence Medical Services
dsDNA HS	double-stranded DNA High-Sensitivity
DSTL	Defence Science and Technology Laboratory
EUCAST	European Committee on Antimicrobial Testing
FTA	Flinders Technical Associate
GI	Gastrointestinal
hbl	Hemolysin BL
HG	Heavy Growth
HLSAN	Heat Labile Salt Activated Nuclease
HRA	Health Research Authority
ICU	Intensive Care Unit

IRP	Immune Response Proteins
K ₂ EDTA	Dipotassium Ethylene Diamine Tetraacetic Acid
LFD	Lateral Flow Devices
LG	Light Growth
LOD	Limit of Detection
LRTI	Lower Respiratory Tract Infections
MALDI-TOF MS	Matrix-Assisted Laser Desorption/Ionisation - Time of Flight Mass Spectrometry
MDA	Multiple Displacement Amplification
MODREC	Ministry of Defence Research Ethics Committee
MRSA	Methicillin Resistant <i>Staphylococcus aureus</i>
MTF	Medical Treatment Facility
NATO	North Atlantic Treaty Organisation
NGS	Next-Generation Sequencing
NHS	National Health Service
NPV	Negative Predictive Values
NSSTI	Necrotising Skin and Soft-Tissue Infections
ompA	Outer Membrane Protein A
oprL	Outer-Membrane Protein L
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PPV	Positive Predictive Value
prlr	Prolactin Receptor
qPCR	Real-Time Polymerase Chain Reaction
RBK	Rapid Barcoding Library Preparation Kit V14
RBK OP	Rapid Barcoding Library Preparation Kit V14 Optimised
RFA	Royal Fleet Auxiliary
RPCR	Rapid PCR Barcoding Library Preparation Kit V14
rRNA	Ribosomal Ribonucleic Acid
RT	Reverse Transcription
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SBS	Sequencing-by-Synthesis
SDH	Salisbury District Hospital
SG	Scanty Growth

SPE	Streptococcal Pyrogenic Exotoxin
TD	Traveller's Diarrhoea
TEPIDOIL	Training Equipment Personnel Information Doctrine Organisation Infrastructure Logistics
UK SMI	UK Standards for Microbiology Investigations
URTC	Upper Respiratory Tract Commensal
vanA	Vancomycin A
WGA	Whole Genome Amplification
WGS	Whole Genome Sequencing
WIMP	What's In My Pot
WTA	Whole Transcriptome Amplification

1. Introduction

1.1 Infectious diseases and the military

Infectious diseases are defined as diseases caused by pathogenic microorganisms, including bacteria, viruses, fungi and protozoa (1). There are thousands of microbial species known to infect humans, with over 1500 bacterial pathogens having been identified alone (2) and novel pathogens emerge in human populations from a variety of sources, including transmission from an animal source to human populations (3). The variety and prevalence of pathogens affecting human populations means that infectious diseases are common, especially so in environments where the exposure of vulnerable populations is frequent. Therefore, infectious diseases have an enormous and profound global impact, both in terms of the human cost through loss of life and the social and economic costs (4).

Military personnel are at significant risk from infectious diseases (5) and there are several factors associated with operating in military environments that contribute to this risk. When undergoing training and operating in theatre, personnel are often housed in crowded communal quarters with relatively poor hygienic conditions, resulting in increased exposure to pathogens and disease transmission (6, 7). When operating abroad, personnel may encounter pathogens to which they have not been exposed to previously, either from environmental pathogens endemic to the deployment location (8) or pathogens transmitted from the local civilian populations (9). Military personnel may also be at higher risk of suffering traumatic injuries that allow microorganisms to contaminate the wound and lead to a wound infection (10). The threat of the intentional release of biological weapons to harm military personnel must also be considered. The combination of these factors leads to significant incidences of infectious diseases amongst military personnel in deployed environments.

Historically, diseases and injuries that did not directly result from combat were often responsible for more casualties in war than the military action itself and as a result, had a profound impact on the outcome of conflicts. In 1802, following the Haitian Revolution, Napoleon, the leader of the French Republic, dispatched a force of 30000 men to reconquer the island. Over the next two years, approximately 27000 men from Napoleon's forces were killed by an outbreak of yellow fever (11), a disease endemic to Haiti but to which Napoleon's forces had not been exposed to previously. As a result, Napoleon withdrew the French forces from Haiti completely and the survivors returned to focus on European conflicts (12). A century later, during the Anglo-Boer wars (1899-1902), 57684 British troops suffered from enteric fever caused by the bacterium *Salmonella enterica* serovar Typhi, resulting in 8225 deaths, greater than the 7582 killed through military action (13). Despite the significant advances made to the prevention and treatment of many infectious diseases over the past century through modern medicine, infectious

diseases continue to impact modern warfare. A trauma infectious disease outcomes study focusing on infectious complications in combat-injured US military personnel from the conflicts in Iraq and Afghanistan found that 34% of patients developed an infection during their initial hospitalisation and 38% developed a new related infection within the initial 88 days following injury, placing a significant burden on the military healthcare services and reducing the operational capability of the deployed forces (10).

The impact of infectious diseases on deployed forces means that the ability to rapidly diagnose pathogens of military relevance to limit their spread amongst deployed personnel has been a key area of research and in many cases has then transferred over to civilian medicine. One historical example is the Weil-Felix test, developed for the diagnosis of rickettsial infections in response to the outbreaks of typhus amongst German soldiers and civilian populations during World War I. The test utilised the patient's serum mixed with *Proteus* bacteria with known cross reactivity with *rickettsia* species. This resulted in a heterophile agglutination reaction that identified the presence of antibodies against the *Rickettsia* disease groups. This included *Rickettsia prowazekii*, the etiologic agent of epidemic typhus, spread to humans by the body louse (*Pediculus humanus corporis*) (14). Following the development of this assay, German forces implemented systematic testing of those suspected of typhus infection in an attempt to limit the spread of the disease. Despite this advance in the ability to diagnose the disease, epidemic typhus was responsible for the deaths of hundreds of thousands of civilians and soldiers during World War I (15).

As novel pathogens continue to develop, the need for the development of novel diagnostic tests continues into modern times. The outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of Coronavirus disease 2019 (COVID-19), at the end of 2019 provides a recent example of the continued need for the rapid development of diagnostic tests for military populations. As a highly contagious respiratory virus, SARS-CoV-2 presented a significant challenge to military populations globally. While the virus did not have the same impact in terms of deaths amongst military populations when compared to civilian populations, likely due to the age and fitness demographics of military populations (16), the ability to rapidly diagnose the disease to limit disruption to deployments was a significant challenge. Real-time polymerase chain reaction (qPCR) based diagnostic tests, targeting the RNA-dependent RNA polymerase gene with high sensitivity and specificity were published rapidly in response to the outbreak (17). Whilst qPCR based diagnostic tests were available, the absence of suitable equipment and logistical support to enable the large scale qPCR –based testing of UK deployed forces meant that testing was outsourced, both to the National Health Service (NHS) testing programme and to the Defence Science and Technology Laboratory (Dstl) (18). There were significant logistical challenges associated with sample collection and

transport, leading to turnaround times ranging from 24 hours to several days for the return of a result, depending on the operational requirement, causing delays to planned military deployments. Despite this, outbreaks of the SARS-CoV-2 virus occurred within deployed UK military populations, causing significant disruption (19) and demonstrating the continued importance of rapid and robust diagnostic tests for ensuring the continued safety and operational effectiveness of deployed military forces.

1.2 Defence Medical Services

The purpose of the Defence Medical Services (DMS) is to

“promote, protect and restore the health of the UK armed forces to ensure that they are ready and medically fit to deploy globally” (20).

More generally, this remit includes a wide range of tasks, from the provision of effective healthcare for all service personnel both at home and on deployment, to delivering clinical research related to defence medical care and providing advice on healthcare policy. A key aim of the DMS is the provision of healthcare to deployed personnel, known as medical force protection. The purpose of which is to prevent the loss of life of service personnel wherever they are situated in an operational context and by extension, enable the success of the mission. This includes four main tasks:

- i) Promotion of behaviour that promotes health through minimising risk to health and preventing disease.
- ii) Provide input to equipment programmes aimed at avoiding or reducing injury.
- iii) Provide input to protection systems for deployed forces.
- iv) Counter the effects of infectious disease, environmental extremes, industrial hazards and chemical, biological and radiological hazards (20).

While focused on deployed UK forces, the provision of healthcare may also be extended to the local civilian population if required, such as in the event of natural disasters.

It is the aim of countering the effects of infectious disease to which this study is focused, through the development of novel diagnostic methods, enabling the effective and rapid diagnosis of infectious disease and ultimately, reducing the impact of infectious diseases on deployed UK forces.

1.3 Deployed infectious disease diagnosis

The provision of deployed healthcare is organised into role one, two and three MTFs, as defined by the North Atlantic Treaty Organisation (NATO) (21). Role one is intended primarily for emergency care to prevent the immediate loss of life, while roles two and three enable more activities to be performed, including the diagnosis of infectious disease. The deployed diagnostics tests available for use on a deployment depend on a range of

factors and will vary according to the specific deployment. This includes the locations of the facilities (with each role from one to three likely to be positioned further from the deployed personnel positions), the presence of endemic infectious diseases, the total number of personnel and any medical intelligence that may indicate a requirement for a particular diagnostic test. The infrastructure at each role also determines the diagnostic tests available for use. The deployed environment includes many challenging aspects that can affect the use of diagnostic capabilities. Reduced logistical support and available infrastructure can restrict the laboratory equipment, reagents and suitably trained staff available for diagnostic testing. Therefore, the routine diagnostic tests that are available in hospitals or more permanent locations may not be suitable for use in deployed settings. The diagnostic tests may then be replenished according to the operational requirements and in response to the outbreak of an infectious disease during the deployment.

The diagnostic tests employed in deployed environments can be broadly separated into two categories: rapid molecular tests and microbiological culture-based tests. The main difference between these categories is that microbiological culture requires the growth and isolation of the microorganism to enable identification of the causative pathogen of infection. Dependent upon the pathogen species, the incubation step required for the growth of the organisms takes at least 18 hours, often increasing to 36 or 48 hours to allow for sufficient microbial growth of fastidious organisms and for additional tests such as antimicrobial susceptibility testing (AST) to be carried out. The incubation of the microorganisms also increases the logistical burden through the need for static incubators and the resupply of perishable reagents. Incubators are challenging to transport, require significant space within a fixed laboratory environment and constant power supply, which are logistically difficult to provide in deployed MTFs. Further, the reagents required for the growth of microorganisms, such as agar plates, require periodic resupply, which can be difficult to achieve. Rapid molecular tests, including polymerase chain reaction (PCR) and rapid immunoassays, target pathogen-specific nucleic acids or antigens, respectively, and do not require the prior isolation of the microorganism, with some exceptions. The fast and accurate diagnosis of an infection and identification of additional clinically relevant information such as the presence of antimicrobial resistance (AMR) genes can enable the provision of effective targeted treatments quickly, reducing the impact on the patient by reducing mortality and reducing the facilities required to care for the patients (22). Also, in contrast to microbiological culture-based tests, they do not require static incubators and if cold-chain storage is achievable, can be stored for extended periods of time without degradation, reducing the need for resupply.

1.3.1 Role one

Role one MTFs are flexible, modular healthcare units intended to provide emergency care and triage capabilities to injured military personnel. Therefore, role one MTFs will be located close to the deployed personnel to enable the rapid provision of emergency care and are likely to be very basic, both in terms of their structure and ability to provide medical care (21) (Figure 1).

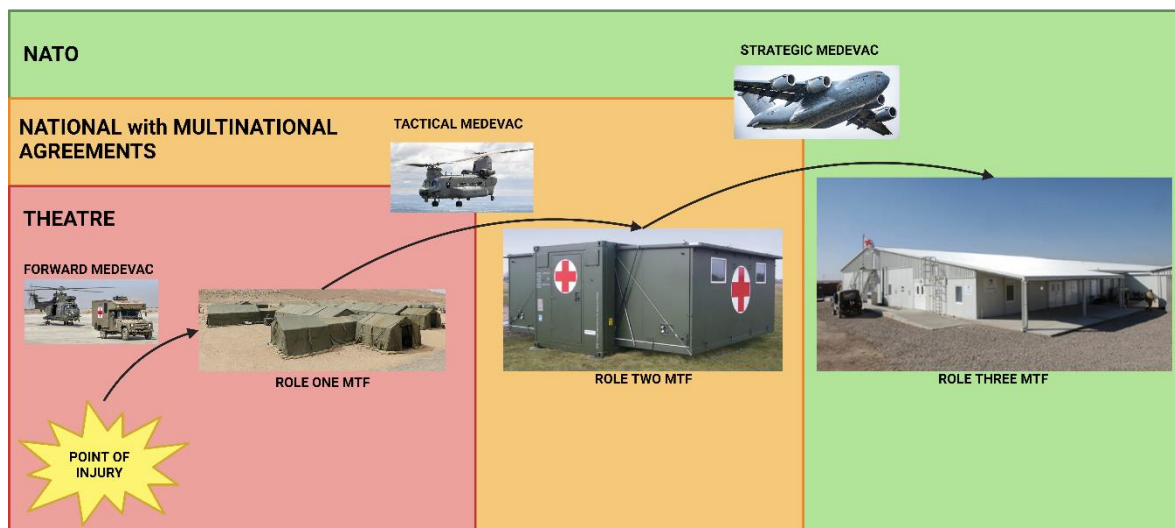


Figure 1. Organisation of deployed MTFs with associated medical evacuation systems among NATO members and allied nations.

Medevac = Medical evacuation. Images available under Open Government License. Figure created with Biorender.com.

Only basic first aid equipment is used, with antibiotic administration and the control of blood loss and shock being the majority of medical countermeasures provided. The infrastructure available at role one is influenced by the operating environment and location being positioned close to the deployed personnel, unique threats such as endemic diseases, accessibility to the site and suboptimal conditions such as extreme heat. In the event of a suspected infectious disease outbreak, isolation pods are used to contain individuals presenting with communicable diseases, such as sickness or diarrhoea, providing challenges to infection control if the pathogen is not identified. These factors affect the quality of medical intervention that can be delivered and ultimately results in the evacuation of the patient without the need or ability to further confirm injury through diagnostic assays.

Whilst role one environments are likely to be austere, limited diagnostic tests may be available, with the emphasis being on rapid time to result and low logistical burden. Diagnostic tests are generally limited to the use of target-specific lateral flow devices (LFD), with the tests selected for deployment according to the geographic location, endemic diseases and specific requirements of the deployment defined by the Centre for

Defence Pathology. These generally consist of LFD tests to which a whole sample can be added directly, without the need for extensive pre-treatment. When added to the test strip, the sample moves across the test strip and a visible band will form on the test strip in the presence of the target antigen in the sample. The antibody, conjugated to a coloured label present on the test strip, binds to the target antigen, indicating a positive test result. Negative results are often indicated by the lack of a visible band.

Due to the lack of sample pre-treatment and reagent preparation required, LFD tests can be performed easily and a result obtained within minutes. However, the result will only provide a binary positive or negative result for the target and will not provide any further information such as a quantitative measure of the pathogen in the sample. These include tests such as the BinaxNOW® malaria test, available for deployments to geographical locations where malaria is endemic, that detects the *P. falciparum* histidine-rich protein 2 and aldolase antigens with 84.2% sensitivity and 99.8% specificity (23), enabling the diagnosis of malaria without the need for microscopy, which cannot be performed in a role one environment.

During the global COVID-19 outbreak, rapid LFD tests were rolled out, in addition to the qPCR tests available in the UK, to enable the isolation of positive cases and reduce transmission in deployed environments. The tests were simple to use, enabling self-testing to be carried out and results to be provided within 30 minutes. However, the sensitivity of the LFDs was heterogeneous and depended on the test used, ranging from 37-99% for SARS-CoV-2 (24). This relatively poor sensitivity could lead to false-negative results and the continued spread of the pathogen amongst personnel. More complex molecular diagnostic tests that require additional equipment are not available at role one and require escalation to role two MTFs.

1.3.2 Role two

Role two MTFs are typically larger facilities than role one, dealing with a higher throughput of patients and providing more comprehensive patient care. Therefore, role two facilities are located further from the deployed personnel compared to role one but remain relatively mobile. Patient care includes deployed hospital care and damage control resuscitation by use of damage control surgery (DCS) providing blood transfusion and sedation ability. Evacuations due to limitations in capacity are commonplace, as well as the role two facility being bypassed for a role three facility for more severe cases and mass casualties. Role two facilities may not be deployed at all if the rapid evacuation of a patient to role three or four facilities can be routinely achieved, and this will be dependent upon the specific deployment. Within a role two environment, diagnostic equipment can include X-rays, magnetic resonance imaging, electrocardiograms, microscopy for pathogen morphology analysis and haematology to perform total blood counts. The DCS

capacity can include endotracheal intubation, laceration repairs and eye examinations. Infectious disease diagnostic capabilities, as with role one, rely primarily on rapid molecular tests. This includes the LFDs available at role one, as well as tests for additional targets, including Hepatitis B and C.

Crucially, the use of the BioFire® FilmArray® 2.0 platform (Biomerieux), which processes the BioFire FilmArray diagnostic testing panels, is also enabled. There are several factors that contribute to the usability of the BioFire FilmArray 2.0 platform in role two environments. The BioFire FilmArray 2.0 is a relatively small benchtop platform that performs a nucleic acid extraction followed by nested PCR analysis of a range of different clinical sample types, depending on the test pouch used. The test pouches used at role two include GI, respiratory and meningitis pathogen panels (25-27). Each pouch tests for a panel of 14-34 (dependent upon the test pouch used) common pathogens or relevant targets such as AMR genes, enabling the rapid identification of multiple pathogens from a single test. In addition, the dehydrated test pouch reagents are contained within a series of tubes that, prior to sample loading, are re-hydrated. The sample is loaded directly to the test pouch, which is then loaded into the BioFire FilmArray 2.0 platform, requiring less than five minutes to complete the sample loading process. The majority of the BioFire FilmArray test pouches do not require growth of the pathogens prior to loading the sample. This reduces the total time to result and means that the static incubator equipment is not required. This excludes the Blood Culture ID panel that targets pathogens associated with blood stream infections that does require blood culture samples to be added to the test pouch and therefore, can only be performed at role three.

Once the test has completed, a binary positive or negative result for each pathogen/gene target on the test panel is returned, with some test pouches, such as the *Pneumonia* panel plus, also providing quantification values for each target (28). The sensitivity and specificity of pathogen identification, although variable depending on the sample type and the specific target, are relatively high. The Respiratory Panel 2 achieves a sensitivity of 66.7% for *Bordetella pertussis* ranging up to 100% sensitivity for *Chlamydomphila pneumoniae*. The same panel achieved specificities of 96-100%, dependent on the target (26, 29). In addition, the BioFire FilmArray 2.0 platform is ruggedized for use in austere environments, reducing the potential for damage during transport, unloading and when in use.

Whilst the BioFire FilmArray test panels are simple to operate and can provide a result within 60 minutes, the key limitation of testing for a panel of known targets is that untargeted pathogens will not be identified by the test. For example, outbreaks of novel pathogens for which no test target has yet been developed cannot be identified using these targeted test panels. The test panel would need to be amended, re-validated and then redeployed to enable detection of novel pathogens, as was done with the BioFire

FilmArray Respiratory Panel 2.1 following the emergence of SARS-CoV-2 (30). In addition, infections caused by uncommon pathogens which are not included as targets within these panels will not be identified, requiring additional testing to confirm the diagnosis which may not be available in a role two environment. The genetic mutation of the target region of a rapid molecular test can also prevent the identification of the causative pathogen resulting in the need for further testing, slowing the time to result significantly. In addition, sample testing cannot be multiplexed to allow for the testing of multiple samples per run. In outbreak situations where it may be necessary to test entire cohorts of personnel within a limited timeframe, the relatively low throughput of testing one sample per hour may be prohibitive to the operational readiness of the deployment. However, the newer BioFire® FilmArray® Torch platform (Biomérieux), compatible with six of the syndromic BioFire FilmArray panels, does enable the simultaneous testing of multiple samples. Therefore, although this platform is not currently validated for use in deployed MTFs, the multiplex testing of samples may soon be possible.

In certain situations, an enhanced role two capability may be deployed. This will enable the deployment of all the capabilities used in role two, with some additional diagnostic tests included. Importantly, this includes the provision of microbiological culture, considered to be the gold standard for infectious disease diagnosis for many infection types, including bacteraemia (31), and respiratory infections (32). Microbiological culture-based methods are available at role two enhanced due to the increased availability of logistical support to enable the supply of equipment and reagents and the availability of the required infrastructure for solid and liquid growth media incubation. Microbiological culture involves collecting a sample from the patient and adding this sample to a liquid or solid growth medium. This growth medium is added to an incubator, a container heated to the optimal temperature for microbial growth, for 18 hours (or more if required) to grow the microorganism. For blood culture samples, samples are added to blood culture bottles and are incubated within automated continuous monitoring systems. Microbial growth is indicated by the measurement of volatile metabolites such as carbon dioxide. If microbial growth is detected, the sample is then sub-cultured on solid growth media to grow isolated colonies (33).

The advantage of microbiology-based methods is that they are well-established, and once the organism has been cultured there are several different tests available using relatively simple techniques. These include simple, rapid methods such as agglutination testing, in which target-specific antibodies are added to a sample of the isolated microorganism and a positive identification is indicated by clumping observed in the sample. At role two enhanced, this includes tests for *Salmonella* O/Vi antigens and *Escherichia coli* O157. In the case of *Salmonella*, this agglutination test enables the identification of the subspecies *S. enterica* serovar *Typhi*, the causative organism of typhoid fever, which expresses the Vi

antigen (34). Agglutination tests for *Staphylococcus aureus* and *Streptococcus pneumoniae*, common respiratory pathogens, are also included. In addition, the classical technique of performing a Gram stain, in which the purple crystal violet-iodine complex dye is added to the isolated microbes to determine whether the microorganism is Gram positive or Gram negative based upon the differential retention of the crystal violet precipitate due to the thicker cell wall in Gram positive species (35), is available. Analytical profile index (API) tests can also be performed from isolated microbes. This method differentiates between groups of bacteria and yeast, such as non-fastidious, non-enteric Gram negative rods (e.g. *Pseudomonas* spp. and *Acinetobacter* spp.), anaerobic bacteria and *Corynebacteria* spp. API tests are based on the different enzymatic activity exhibited by these groups, such as activity related to fermentation of carbohydrates. Despite being relatively simple to carry out, these methods provide information that informs the optimal antibiotic treatments.

For more detailed information on the optimal treatment, AST can also be performed on the isolated microbes. The isolated colonies are re-plated on a solid media in the presence of antimicrobial discs, known as the disk diffusion method (36). Organisms that are able to grow in the presence of an antimicrobial at concentrations above the defined breakpoint are considered resistant. This is key information required for treating an infection, as targeted use of antimicrobials reduces mortality and the faster the provision of targeted treatments the better the outcome for the patient (37). Finally, as blood culture is enabled, the BioFire FilmArray Blood Culture ID panel can be carried out for the identification of common causes of bacteraemia.

The key limitation of applying microbiological culture is that the time to result for these methods can range from 18 hours up to 72 hours plus, due to the need to grow the microorganisms in an incubator. This often results in the use of non-specific treatments, such as the administration of empiric antibiotics that may not be optimal for the treatment of the causative pathogen, prior to test results being obtained. Further to this, fastidious microorganisms may not be identified at all due to slow growth rates and many of the additional diagnostic tests are pathogen-specific, such that non-targeted pathogens will not be identified. However, the key disadvantage of microbiological culture is the logistical requirements of maintaining the capability within a deployed environment. This includes installation and maintenance of large plate incubators requiring significant power and space to operate. Further to this, the reagents, including plates and various agar types must accompany the capability and may require restocking periodically due to limited storage time of the reagents. Therefore, it is key that alternatives are developed that can provide results to the same or greater sensitivity and specificity without incurring the logistical costs that are currently associated with microbiology. For more extensive diagnostic testing capabilities, the role three environment is required.

In addition to microbiology, microscopy, considered to be the gold-standard diagnostic method for the identification of enteric parasites (38), will also be enabled at role two enhanced. Microscopy methods involve the preparation of a sample for observation under a microscope, usually through addition of the sample to a microscope slide and depending on the sample type, addition of a dye to aid in pathogen identification. A trained BMS will then examine the sample for the presence of the pathogen based on morphological characteristics. Deployed microscopy methods include the identification of enteric parasites, including *Cryptosporidium* spp. and *Giardia lamblia*. The parasites are concentrated in the sample using the Midi Parasep® faecal parasite concentrator, which applies the Ridley-Allen formol-ether sedimentation technique within an enclosed system (39). The concentrated sample is then examined using a light microscope at X10 and X40 objectives to identify the helminthic eggs or larvae and the protozoan cysts, oocysts or trophozoites (40). Microscopy based diagnosis can also be applied to the detection of the *Plasmodium* parasite species, the causative agents of the acute febrile illness malaria. Thick blood films are prepared and stained with a combination of methylene blue and eosin dyes to differentially stain the nucleus and the cytoplasm of the *Plasmodium* parasites replicating within the erythrocytes (41), and examined under oil immersion. Microscopy methods, when applied by trained personnel, are highly specific, with specificity of 98.3% achieved for the diagnosis of malaria from blood samples. However, sensitivity can be poor at just 39.3% (42). In addition, the processing of a sample can take more than 60 minutes and is labour intensive, requiring a trained BMS to process and examine each sample. Therefore, microscopy is not suitable for high-throughput diagnosis, which can delay the return of results, particularly during outbreak events in which increased sample numbers may be expected.

1.3.3 Role three

Role three facilities are field hospitals, evacuation hospitals (primary casualty receiving facility) and combat support hospitals, as well as advanced logistics support sites, with more permanent infrastructure than is possible at role one and two and include the highest level of medical care that can be provided with proximity to the battle zone. Role three MTFs should enable specialised surgery, care and additional services dependent upon the mission being supported. This increased capability means that role three facilities will be located further from deployed personnel and medical evacuation will be required. Role three facilities have surgical, preventative medicine and stress management teams. The higher capacity and holding level of patients enables more efficient diagnosis, treatment and recovery. Facilities are staffed by doctors, nurses and allied health professionals drawn from across Defence to enable more complex treatment to be administered. The facilities and equipment available at role three include resuscitation rooms, portable X-ray, ultrasound, CT scanners, haematology, biochemistry,

blood transfusion and microbiology. An extensively equipped operating theatre with capacity to perform three simultaneous surgeries, large capacity for over-night stays with 70 ward beds, 20 high dependency beds and ten intensive care beds will also be available as a minimum. For infectious disease diagnostics, microbiological culture-based tests and microscopy, as described for role two enhanced MTFs, is available and will always be included at role three.

As discussed, there are limitations associated with the current diagnostic methods available in deployed environments. In particular, the logistical burden of deploying the equipment and reagents required for microbiology and the time to result being a minimum of 18-36 hours, means that there is an opportunity to develop improved, novel methods of diagnosis. However, any new diagnostic technology must consider its implementation into the existing diagnostic pathway. All military capabilities are subject to Defence Lines of Development (DLOD) which provide a pan-defence taxonomy to enable the coherent, through life development and management of a defence capability. DLOD include training, equipment, personnel, information, doctrine, organisation, infrastructure and logistics (TEPIDOIL). Therefore, development of these new methods should consider the need for minimal logistical support, reduced time to result, and method complexity while enabling the identification of the causative pathogen in a target agnostic manner. The development of DNA sequencing and its application to infectious disease diagnostics has the potential to provide a solution to the current limitations of deployed diagnostics whilst taking the TEPIDOIL requirements into consideration.

1.4 DNA sequencing for diagnostics

DNA sequencing refers to the method of determining the sequence of nucleotides within the genetic content of an organism or a wider community of organisms in a sample. The method began with the development of the method known as Sanger sequencing in the 1970s, based on the use of chain-terminating DNA polymerase inhibitors to determine the sequence of nucleotides (43). Methods have since developed rapidly and arguably the DNA sequencing method that is now most widely used for a variety of applications is sequencing-by-synthesis (SBS) technology, developed for Illumina™ sequencing platforms. This method involves determining the nucleotide sequence of fragments of DNA up to 300 base pairs (bp) in length by incorporating fluorescent nucleotides into clusters of identical DNA fragments by stepwise primer elongation and capturing high resolution images in a cyclical reaction (44).

Crucially for diagnostic applications, the development of this method, amongst several other methods known collectively as next-generation sequencing (NGS), enabled a number of significant advances. This included the production of greater amounts of sequencing data per run from multiple samples at significantly lower cost compared to

previous technologies and within a shorter amount of time, with sequencing runs ranging between approximately 9.5-48 hours, depending on the platform. In addition, SBS is highly accurate, with error rates of $0.24 \pm 0.06\%$, depending upon the preparation methods (45). Further to this, the bioinformatics technology that processes, analyses and interprets the raw sequence data to provide a diagnosis or clinically relevant information has greatly improved over recent years. This includes a wide range of computational tasks, from data pre-processing, quality control (46), *de novo* assembly (47) and reference-based mapping so that the required information is elucidated from the sequence data within a clinically relevant timeframe.

The development of this technology has enabled DNA sequencing to become an increasingly important method for a range of uses across medicine, including infectious disease diagnosis, the identification of somatic genetic mutations and the high-throughput screening of gene variants for the diagnosis of cancers, amongst other applications (48-50). The use of whole genome sequencing (WGS) identified an outbreak of methicillin resistant *S. aureus* (MRSA) of a novel sequence type at a UK special care baby unit that had not been identified using conventional methods. Further to this, NGS demonstrated the transmission of the pathogen within the unit by a specific member of staff. Therefore, measures could be put in place to limit further spread (51).

However, while SBS sequencing continues to be highly useful in civilian medical environments, there are several limitations that prevent its use in deployed MTFs. The platforms are large and difficult to manoeuvre. The smallest Illumina sequencing platform, the iSeq, weighs 16 Kg and measures 30.5 cm × 33 cm × 42.5 cm. For the largest platform, the NovaSeq X, this increases to 587 Kg and measures 86.4 cm × 93.3 cm × 158.8 cm, making them unfavourable for the dynamic MTF environments. In addition, SBS platforms, due to the requirement for precise imagery, are sensitive to movement, vibration and extremes of temperature and humidity and are likely to require installation by a qualified engineer (52). Therefore, SBS NGS platforms are unlikely to be routinely operated within a deployed environment.

1.4.1 Nanopore sequencing

Another generation of nucleic acid sequencers, termed “third generation” sequencing technologies, has the potential to further the use of DNA sequencing for infectious disease diagnostics and enable DNA sequencing to be performed in deployed environments. Nanopore sequencing, developed by Oxford Nanopore Technologies Ltd., sequences DNA by measuring the alterations in the electrical current across a semi-permeable lipid membrane as a strand of DNA moves through a protein pore. Each individual nucleotide generates a distinct electrical signal analysed in real-time to determine the sequence of

nucleotides in the DNA strand (53). This is repeated across approximately 1500 pores across the membrane enabling high-throughput simultaneous sequencing of DNA strands. This sequencing technology has several advantages for use in deployed MTFs compared to SBS technologies. The technology is small and portable. The first nanopore sequencing device, the MinION, released in 2014, was just 105 mm x 23 mm x 33 mm and weighed 85 g. While this does require a connection to a suitable laptop with the required software, this set-up was significantly smaller and more lightweight compared to SBS-based technologies (54). Another nanopore sequencing device, the MinION Mk1c, further removes the need for a laptop connection, enabling nanopore sequencing runs to be set-up, run and the data base called on one integrated device, further reducing the space requirements.

Another significant advantage of nanopore sequencing is that the raw sequence data, consisting of the sequence of electrical signals generated as the DNA molecules move through the protein nanopores, are recorded in real-time. Therefore, the bioinformatics analysis can be started during the run, rather than once the run has finished, as is the case with SBS technologies, and workflows enabling real-time data analysis and interpretation of nanopore sequencing data are available (55). From an infectious disease diagnosis perspective, this has the potential to significantly reduce the time to result.

These advantages of nanopore technology mean that there are a number of examples of nanopore sequencing being applied to a range of scenarios in which SBS sequencing could not be used, including on board the International Space Station and in the field within the Arctic Circle (56, 57). Nanopore sequencing technology has also been applied to a range of more conventional purposes, including disease diagnostics, cancer genomics and plant breeding amongst a wide range of other uses, as reviewed by Zheng *P et al.* 2023 (58). The development of nanopore sequencing opens up the possibility of DNA sequencing based diagnostic tests being applied for the diagnosis of infectious disease close to deployed personnel in MTFs or even at the point of care.

1.5 Clinical metagenomics

CMg refers to the untargeted sequencing of the total genetic material, DNA and/or RNA, contained within a patient sample. While this can refer to both the microbial and host genetic material, for the purpose of this study, CMg refers to the sequencing of the microbial genetic content of a sample. The untargeted nature of CMg allows a wide range of information to be revealed and therefore, CMg has been applied to human health in several fields, as reviewed by Chiu and Miller, 2019 (59). Key examples include clinical microbiome analysis, used to inform disease prognosis, applying the untargeted nature of CMg to reveal the full diversity of species present in the microbiome of the patient. For example, CMg was applied to identify a significant reduction in gut microbiome species

richness in patients with cirrhosis of the liver compared to healthy subjects, associated with complications of cirrhosis and shown to predict 3-month survival (60). Analysis of the sputum microbiome revealed several bacterial biomarkers associated with the development of lung cancers, opening up the possibility of using CMg as a non-invasive method for monitoring the development and progression of the disease (61).

In the context of infectious disease diagnostics, the aim of CMg methods is to identify the organism or organisms that are causing the infection and provide additional clinically relevant information to aid the treatment of the infection. The methods vary depending on the sample type and the aim of the diagnostic test; however, the basic process includes the following steps and is summarised in Figure 2:

- i) Sample collection. CMg has been applied to a range of sample types, including respiratory, blood, stool and cerebrospinal fluid, amongst others (62-65).
- ii) Host DNA depletion. The downstream processing for each sample type differs to accommodate for variances in the composition of the sample such as the human and microbial content. Depending on the proportion of human cellular material in the sample, a human genetic material depletion step may be required to reduce the proportion of human DNA in the sequencing data so that microbial species are identified with greater sensitivity.
- iii) Nucleic acid extraction. The microbial genetic material is extracted and eluted into a suitable buffer. Depending on the aims, DNA, RNA or both are extracted. This process is widely automated, reducing the hands-on steps for the user.
- iv) Nucleic acid amplification. In low biomass samples, the nucleic acids may require amplification using whole genome amplification (WGA) and/or PCR to generate sufficient DNA for the sequencing step.
- v) Library preparation. The DNA is prepared for sequencing during a library preparation step. This step depends upon the sequencing method to be used, with protocols available for different sequencing platforms, workflow aims and DNA input.
- vi) DNA sequencing. The DNA will then be sequenced, and the data analysed to identify the species present in the sample. Additional information such as the presence of AMR genes can also be elucidated at this stage.

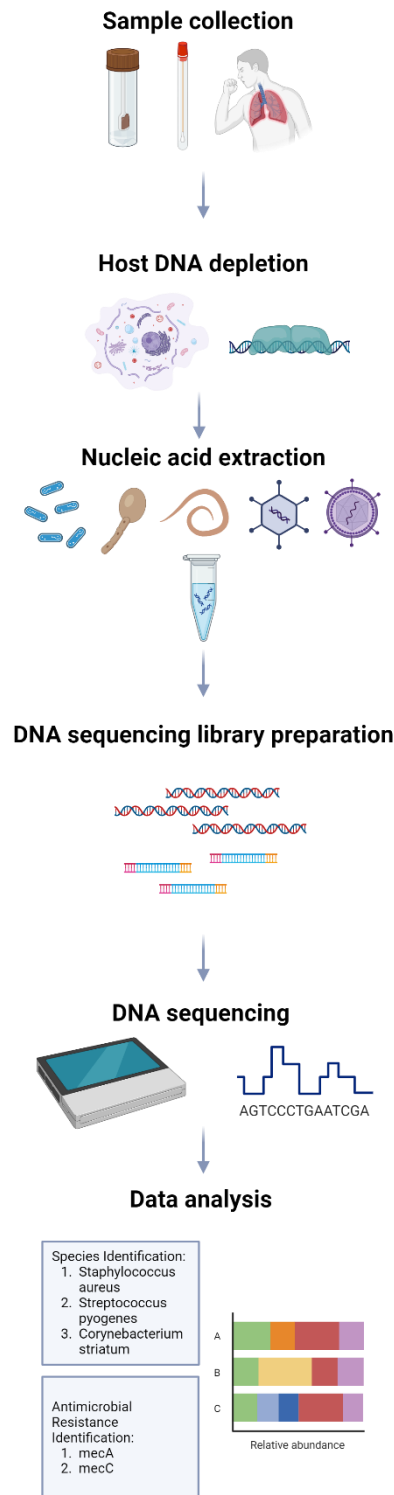


Figure 2. A graphical representation of a CMg workflow for the diagnosis of infectious disease.
Created with Biorender.com.

1.5.1 Amplicon metagenomic sequencing

CMg methods are broadly split into two approaches. Amplicon metagenomic sequencing refers to the amplification and sequencing of targeted regions of the genome to identify the pathogen species present in a sample. This method first amplifies by PCR universal hyper-variable target genes or regions that are flanked by conserved regions suitable for

oligonucleotide binding, such as the 16S ribosomal ribonucleic acid (rRNA) universal gene found in bacteria (66) or 18S rRNA gene found in fungi (67). The amplicons are then sequenced, and the resulting data analysed to compare the sequencing fragments present in the sample to a reference database. An example of this method being applied to the diagnosis of bacteraemia and fungaemia is the SepsiTTest (Molzym, Germany). Evaluation of this method demonstrated 86% concordance with blood culture results and a sensitivity and specificity of detection of 87% and 85.8% respectively (68). The PCR method can be completed within four hours and the sequencing requires at least an additional eight hours, dependent upon the technology used, a considerable time saving compared to a typical blood culture.

An additional advantage of the amplicon sequencing method is that it can identify any bacterial and fungal species present, if included in the relevant sequence database, and is not limited to a particular target or panel of targets, in contrast to PCR methods. However, it should be noted that viral pathogens will not be detected by 16S rRNA or 18S rRNA amplicon sequencing due to their lack of the required gene regions. In addition, inherent PCR biases, including differential primer mismatch and amplification efficiencies, can reduce the accuracy of microbial species detection using this method (69). Another key disadvantage of amplicon sequencing is the lack of additional information that can be gained from the sequencing data. While amplifying and sequencing the hyper-variable regions is effective for pathogen detection, it cannot be utilised to determine factors such as AMR and often cannot distinguish pathogens to a species or strain level. Therefore, additional testing would be required to gain this information, limiting the clinical utility.

1.5.2 Shotgun metagenomics sequencing

Shotgun metagenomics, in contrast to amplicon based sequencing, enables the untargeted sequencing of the total nucleic acid content of a clinical sample without the need for the prior amplification of targeted universal hyper-variable regions by PCR (70). During the library preparation stage, the DNA is fragmented, and specific adapters are attached to each end of the resulting DNA fragments. Fragmentation can be completed either through physical methods such as sonication or enzymatically. The ends of the DNA fragments are then repaired and ligated, with the enzymatic reaction differing depending on the sequencing technology used.

There are examples of shotgun metagenomics being applied in a clinical context. Grumaz *et al.* 2019 (71) compared CMg sequencing of cell-free DNA in plasma with routine microbiology methods for pathogen identification from patients at the onset of septic shock. CMg achieved a positivity rate of 72% at sepsis onset, compared to 33% for routine microbiology. Importantly, expert evaluation found that the treatment method would have been altered in 53% of cases based on the sequencing results. CMg for

respiratory infections is relatively advanced and has been trialled for routine use in diagnostic laboratories in the UK. A pilot service was established at a respiratory intensive care unit (ICU) at Guy's and St Thomas' NHS Foundation Trust, London. One hundred and twenty eight samples were processed using a previously validated CMg workflow, providing results within 6.7 hours and the identification of respiratory pathogens with 93% sensitivity and 81% specificity compared to standard of care testing (72). Importantly, 48% of the results provided by the pilot service informed changes to the patient's antimicrobial treatment, due to identification of the pathogen species or AMR genes.

1.5.3 Advantages of clinical metagenomics

A key advantage of the untargeted nature of shotgun CMg is the provision of additional clinically relevant information to inform the treatment of the infection, including the detection of AMR genes (73). The rise in the incidence of infections caused by pathogens resistant to one or more antimicrobial agents is of huge global concern, leading to increased rates of treatment failure, increased morbidity, mortality and treatment costs (74, 75). This has led to governments and global organisations including the World Health Organisation to take significant steps to try to control the development and spread of AMR and encourage the development of novel treatments (76), including for priority pathogens regarded as particular threats to the future of medical care due to AMR (77). From CMg data, if sufficient sequencing depth is achieved, the presence of genes associated with resistance to antimicrobials can be discerned. For example, the application of CMg to orthopaedic device infections enabled the identification of AMR genes, resulting in the correct prediction of phenotypic resistance in 87% of samples and the correct prediction of antibiotic phenotypic susceptibility of 100% (78).

It is not only genes associated with AMR that are of relevance to the treatment of an infection. The detection of virulence factors and toxins, which can affect the severity of an infection, provides useful information and guides treatment options. Shotgun CMg has been applied to successfully detect both virulence factors and toxin genes, however, particularly for complex sample types such as stool, more complex bioinformatics analysis was required to determine the presence of different pathotypes and the associated virulence genes (64). The prediction of virulence factors and bacterial toxins by CMg has also been demonstrated, both in simulated and case-control examples with high accuracy and specificity (79).

Whilst direct detection of the causative pathogen, AMR, toxin and virulence genes have obvious benefits, the range of additional data provided by CMg to inform on the progress of an infection and the appropriate treatment options extend beyond these. For sample types in which the sample contains commensal flora, the composition of this flora can be studied using CMg. Microbiome data, including species diversity, has been shown to differentiate patients with diarrhoea and healthy controls (80) and treatments developed to

increase microbiome diversity prevent re-infection (81). Further study of the composition of CMg data from a range of sample types in healthy controls compared to patients with a disease could reveal further data points that can inform treatments and the success of treatments over time. The untargeted sequencing of microbial DNA also means that any pathogen type can be detected and multiple pathogens can be detected from a single sequencing run, enabling the identification of the causative pathogen(s) even if the infection is caused by multiple microbial species (82). This advantage is particularly apparent in cases where the infection may be caused by a rare pathogen (83), during novel infectious disease outbreaks or the intentional release of undetermined pathogens where prior knowledge is not available.

1.5.4 Limitations of clinical metagenomics

While the potential for the use of CMg for the routine diagnosis of infectious disease in permanent medical facilities is clear and has been demonstrated, there are a number of limitations associated with current CMg methods that would prevent its routine use in deployed MTFs.

Firstly, CMg enables the sequencing of the total nucleic acid present in a sample, irrespective of its origin. The presence of high levels of host nucleic acid in the sample reduces the proportion of sequence reads assigned to the pathogen species and therefore reduces the sensitivity of pathogen detection (62, 84). This is particularly impactful for sample types that contain high proportions of host cellular material compared to pathogen genetic material, such as blood samples. Therefore, for these sample types, host genetic material must be depleted prior to sequencing or during the bioinformatics analysis to increase pathogen detection sensitivity.

Secondly, particularly for non-sterile sample types, CMg enables the detection of the nucleic acid of any microbial species present in the sample. Certain sample types, such as wound and respiratory samples, are likely to contain commensal microbial species irrespective of infection and this can impact the clinical utility of the sequencing data, masking the presence of pathogen genetic material. In some cases, commensal species are genetically similar to pathogen species, complicating the bioinformatics analysis and potentially leading to the misidentification of pathogen species as commensals and *vice versa*.

Similarly, AMR genes will be detected irrespective of their origin, including from commensal species unrelated to the aetiological agent of infection and mobile genetic elements that can be horizontally transferred between species, particularly genetically similar species present within ecological niches (85), complicating the interpretation of the output of CMg to identify the resistance profile of the infection. Further to this, even if an AMR gene is identified, this does not necessarily directly relate to the phenotypic

resistance of the pathogen. AMR can result from a number of complex pathways, including from the presence of specific promoters that increase the expression of an AMR gene (86) and the expression of multiple AMR genes that work in combination to result in antibiotic resistance (87), such that the detection of an AMR gene associated with resistance to a specific antibiotic does not necessarily translate to the phenotypic resistance of the pathogen. AMR genes can also be present within an isolate without conferring resistance, known as silent AMR genes (88).

In addition, exogenous contamination introduced during sample collection and processing may also be detected, potentially confounding the results and leading to false positives. The impact of contamination can be reduced by implementing thresholds to remove low level contamination (89), however, contamination is likely to be a significant problem associated with CMg for the foreseeable future. However, it should be noted that the colonising microbes and exogenous contamination issue is not unique to CMg and affects methods including culture and PCR.

Focusing on utilisation in deployed military diagnostic facilities, there are several limitations that prevent the use of CMg for infectious disease diagnosis using currently available methods. As discussed, the development of new military capabilities is subject to DLODs. Although all DLODs will apply to the development of a deployable CMg capability (some of which will only become detailed as a capability is developed), it is possible to examine some of the likely issues now. CMg sample processing, despite being culture-independent, can take between 12-48 hours to produce the data which then have to be analysed using complex bioinformatics software. In addition, the wet-lab processes required to prepare samples for sequencing are relatively complex and labour-intensive, significantly increasing the operator hands-on time and the total time to result. Subsequently, this complexity means that the training requirements to enable military personnel to produce and analyse sequencing data will be significant. The equipment required can have a significant footprint, energy requirements and can be disturbed by movement. Sensitive technologies of this nature may not be suitable for deployed laboratories in which space is at a premium. These limitations and the impact of the TEPIDOIL DLODs during the development and maintenance of an infectious disease diagnostics capability must be considered when developing a CMg diagnostic capability.

1.6 Study aims

Military personnel suffer from a range of infectious diseases, with wound, diarrhoeal and respiratory infections being of particular concern due to the high prevalence of these infection types and the challenges associated with diagnosis using current deployed methods (90). Whilst CMg is regarded as a potential solution to these limitations, the methods are not currently suitable for use in deployed MTFs due to method complexity,

time to result and the reagent and equipment requirements. In addition, technical limitations including poor pathogen detection sensitivity due to high proportions of host genetic material must be solved before CMg can be routinely applied for diagnostic purposes. The overall aim of this project was to develop, optimise and evaluate CMg methods for a range of relevant sample types that are routinely tested in deployed MTFs, providing a baseline for future development work towards deploying CMg as a standard diagnostic test.

Firstly, in Chapter Two, this included the development of a CMg workflow for the diagnosis of wound infections, aiming to enable the sensitive and specific identification of wound pathogens whilst efficiently depleting host DNA and minimising the complexity of the methods to enable the future use of the workflow in deployed MTFs.

In Chapter Three, the wound CMg workflow was then applied to wound swab samples prospectively collected from patients with suspected wound infections at two NHS healthcare facilities as part of a Ministry of Defence Research Ethics Committee (MODREC) approved study, with the aim of evaluating CMg for the diagnosis of wound infections and comparing the CMg results with current diagnostic tests. Further opportunities for method optimisation were identified for the improvement of the CMg workflow in future studies.

In Chapter Four, CMg was applied to stool samples, stored using three storage methods, collected from UK military personnel during an outbreak of traveller's diarrhoea (TD), caused by the enteric pathogen *Cryptosporidium hominis*, linked to the contamination of recreational water sources. This study aimed to compare the long-term storage methods for stool samples and optimise the stool sample CMg workflow for use in austere environments. Ultimately, this would enable military clinicians to store stool samples for extended periods at ambient temperatures, increasing the diagnostic testing options available during future outbreaks of TD.

Finally, Chapter Five aimed to evaluate the application of CMg workflows for the diagnosis of respiratory infections in a deployed MTF. Two CMg workflows, developed for the identification of viral and bacterial respiratory pathogens were deployed to a role three MTF and respiratory samples were processed by a military BMS. The aim of this study was to gain a better understanding of the technical and logistical challenges associated with deploying CMg workflows to enable the optimisation of CMg during future work.

2. Development of a clinical metagenomics workflow for the diagnosis of wound infections

2.1 Contribution statement

The wet-lab work for section “2.3.4 Comparison of DNA extraction platforms” was carried out by Katie Edge (Dstl) as part of a six-month graduate scheme project under the direct supervision of Carl Halford. Carl Halford was responsible for the experimental design and results interpretation, with assistance from Katie Edge. All other wet-lab work and results interpretation was carried out by Carl Halford. The bioinformatics analysis was carried out by Dr. Thanh Le-Viet (Quadram Institute Bioscience) in close collaboration with Carl Halford using a bioinformatics pipeline that Dr Thanh Le-Viet previously developed.

2.2 Introduction

2.2.1 Clinical metagenomics

CMg has the potential to provide a rapid, untargeted diagnostic test for a range of sample types relevant to deployed military personnel, including wound infections, using methods suitable for operation in austere environments. However, there are several limitations associated with CMg that prevent its routine application. The work carried out in this chapter aimed to develop a CMg workflow for the diagnosis of wound infections from wound swab samples. This focused on solving the limitations of CMg through the reduction of human DNA contamination whilst optimising for improved extraction of pathogen DNA and reducing the overall method complexity, time to result and equipment requirements. Figure 3 presents a schematic diagram of the wound swab CMg workflow.

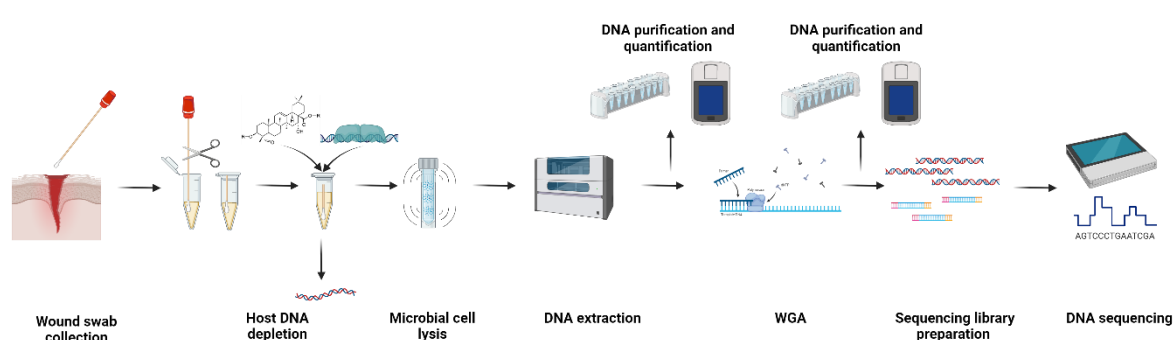


Figure 3. Schematic diagram of the wound CMg wet-laboratory workflow.

Created with Biorender.com.

2.2.2 Host DNA depletion

The collection of wound swab samples in clinical environments involves the identification of a small area of clean viable tissue at the base of the wound and rotating the swab on this area for five seconds, avoiding necrotic tissue and the wound edges (91). Therefore, wound swabs are likely to consist of a complex matrix of human epidermal and immunological cells, commensal microbial flora and, if present, pathogen species. The proportion of human and microbial cells collected on the wound swab is likely to vary due to the amount of genetic material in microbial species, particularly bacteria, being approximately 1000-fold lower than human cells. The presence of even a small number of human cells could interfere with the identification of microbial nucleic acids. Therefore, the first step of the process for preparing wound swab samples for metagenomics sequencing is the depletion of the human DNA content of the sample.

Various commercially available methods of human DNA depletion are available, including the Molzym Ultra-Deep Microbiome Prep (Molzym GmbH & Co.), NEBNext Microbiome DNA Enrichment kit (New England Biolabs), QIAamp DNA Microbiome kit (Qiagen) and Zymo HostZERO microbial DNA kit (Zymo Research). A previous evaluation of these methods in human intestinal samples found that the QIAamp DNA Microbiome kit efficiently reduced the proportion of host DNA contamination, increasing bacterial DNA from 1% to 28% of sequence reads (92). In contrast, evaluation of the MoLYsis kit in cerebrospinal fluid and nasopharyngeal aspirate samples spiked with known pathogen species did not improve the ratio of human to pathogen DNA (93). However, treatment with 0.025% final concentration of saponin, a non-ionic surfactant, resulted in a >95% reduction in human DNA without affecting the pathogen DNA; resulting in a 20-to 650-fold increase in the ratio of pathogen reads to human reads in the sequencing data. Incubation of human cells with saponin resulted in increased permeability of the lipid bilayer to macromolecules (94), but importantly, saponin did not have the same effect on bacterial cells, leaving them intact (95). The addition of a DNase enzyme, which enters the permeabilised human cells and digests the human DNA whilst the microbial DNA remains protected by the cell wall, effectively removed human DNA from the sample.

A CMg workflow developed for the rapid identification of respiratory pathogens applied a 2.5% final concentration saponin-based depletion of human DNA from lower respiratory tract infection samples (96). This resulted in an average of 10^3 – fold depletion in human DNA in depleted samples compared to non-depleted controls. When comparing the identification of an antimicrobial resistant pathogen, MRSA in this example, after two hours of DNA sequencing, the depleted sample resulted in a $47.9 \times$ mean genome coverage compared to $3.9 \times$ mean coverage in the non-depleted control (96). This increase in pathogen genome coverage was also demonstrated in additional samples, highlighting the importance of human DNA depletion for pathogen identification. Following

further optimisation of the saponin-based human DNA depletion method to reduce the time and number of hands-on steps in spiked whole blood samples, application of the saponin-based host depletion method increased the proportion of bacteria DNA from 0.86% to 68.8% (63). The effective depletion of human DNA when treated with saponin followed by a suitable DNase enzyme led to the testing and evaluation of this method for the depletion of human DNA from wound swab samples as part of this study.

2.2.3 Microbial DNA extraction

Following the host depletion step, the next step in the sample processing workflow is the isolation and extraction of the microbial DNA. The efficiency of this step is critical to ensuring that the microorganisms present in the initial sample are identified by downstream processes. Further, the selection of the extraction methods can lead to the introduction of bias to the results, leading to the incorrect conclusions to be drawn from the resulting test (97).

DNA was first extracted in 1869, having been crudely purified by the addition of acid to leukocytes during experiments conducted on leukocyte proteins (98, 99), almost a century prior to the discovery of the double-helical structure of DNA by Franklin (100), Watson and Crick (101). The study of the sequence of nucleotides within DNA has since become a cornerstone of molecular biology and the methods for the extraction and isolation of DNA have progressed significantly. The main steps of modern DNA extractions start with the disruption of the cytoplasm and nuclear membranes, followed by the isolation of the DNA and removal of other cellular structures, including proteins and lipids, before purification of the DNA and elution into a final buffer. The efficient disruption of the cytoplasmic membranes is of particular importance to CMg workflows, due to the potential for the presence of a wide variety of microorganisms, including Gram positive bacterial and fungal species. In general, it is more difficult to extract DNA from Gram positive bacteria due to the outer cell wall of Gram positive bacteria comprising a thicker peptidoglycan layer. The outer cell wall consists of polysaccharides including teichoic acids and lipoteichoic acids attached and inserted into the membrane, respectively (102). DNA extraction from fungal cells is also made more difficult by the presence of fibrous carbohydrate polymers with a range of proteins and additional components that combine to make very strong cell walls (103).

The disruption of the cytoplasmic membrane is usually achieved through mechanical, chemical or enzymatic methods, or a combination of these. Mechanical disruption relies on the physical breaking of the membranes through the addition of beads of various sizes and subjection to high intensity vibration. Factors including the bead material, bead size and the duration and intensity of bead beating can be tailored to the specific sample type to improve the DNA yield (104) and the inclusion of a bead-beating step increases the

diversity of microbial species revealed from gut microbiome samples (105). Chemical lysis methods include the addition of the sample to buffers containing agents that disrupt the outer membrane or cell wall. This includes guanidine thiocyanate, a chaotropic agent that destabilises macromolecules including proteins, disrupting cell and organelle membranes (106) and preventing the degradation of nucleic acids by nucleases (107). Enzymatic lysis methods utilise lytic enzymes to disrupt the cell membranes. One example, lysozyme, acts by lysing the cell wall peptidoglycan through the activation of autolytic enzymes in the bacterial cell wall (108). Both mechanical and enzymatic methods have been shown to be effective in different contexts, however, bead-beating has been shown to be more effective for the extraction of DNA from samples including Gram positive bacteria (109).

The two most utilised methods for the isolation of DNA following cell membrane disruption rely on silica-based matrices or magnetic beads to bind the DNA. Silica matrices have high-affinity for DNA due to the presence of positively charged ions which attract the negatively charged DNA while the cell lysate can be removed (110). Similarly, magnetic beads consisting of magnetic nanoparticles coated with a DNA binding molecule, bind DNA and allow for the separation of the DNA from the cell lysate through the use of a magnet (111). In this project, initial experiments were carried out using the MaxWell 16 LEV platform (Promega), an automated DNA extraction device available at Dstl that enabled the reproducible and relatively high-throughput extraction of microbial DNA from host depleted samples. Following the host depletion and prior to addition of the sample to the device, the sample is added to a lysis buffer with Proteinase K and incubated at 65°C for five minutes. The lysis buffer contains Guanidine Thiocyanate and Triton X-100 detergent, which further lyse the cell membranes and nuclei of the cells. The Proteinase K degrades various proteins, and the histones bound to the DNA which enables the release of the genomic DNA ready for isolation. The MaxWell 16 LEV device then uses a sequential series of wells in a removable cartridge. The sample is moved from one well to the next, starting with a well containing paramagnetic particles that bind the DNA. The DNA is separated and washed with ethanol to purify the DNA, before being eluted into the elution buffer, with the process taking approximately 45 minutes.

2.2.4 Whole genome amplification

The inclusion of a WGA step is necessary for CMg workflows when working with low biomass samples, such as host depleted samples. In addition, nanopore based sequencing often requires higher DNA input in comparison to Illumina SBS based library prep methods. Therefore, a WGA step was included in the CMg workflow for this study. WGA refers to the non-selective amplification of the total DNA content of a sample, resulting in a significant increase in total DNA concentration reflecting the entire genetic content of the sample. This enables the study of DNA content from very low biomass samples including trace DNA and even single cells (112).

WGA can be categorised into two main methods of amplification: PCR and isothermal amplification. PCR based methods, such as primer extension pre-amplification PCR, uses random four-nucleotide primers that anneal to many sites along the DNA sample. The reaction is heated and cooled in a cyclical manner with a DNA polymerase enzyme to amplify the DNA. This results in the amplification of 75% of the genome after 50 cycles, requiring approximately three hours of amplification time (113). Isothermal amplification includes methods such as multiple displacement amplification (MDA), the most applied WGA method. MDA uses a DNA polymerase with high extension activity and random primers to amplify the genome and, in contrast to PCR-based methods, uses isothermal conditions without the need for cyclical heating and cooling. Advantages of this method include the uniform amplification of the whole genome, enabling fragments of 10-100 kilobases in length to be amplified with high fidelity (114). Further to this, the method is generally simple to carry out, does not require pre-purification of the DNA requiring minimal steps and is available through a range of commercial off the shelf (COTS) products. In addition, advances in microfluidic technologies may enable the transfer of MDA WGA methods onto automated microfluidic devices, further reducing the user hands-on time, power requirements, potential for contamination and human error (112). Therefore, this method was selected for evaluation as part of this project.

Despite being required for the generation of sufficient DNA from low biomass samples, there are disadvantages to applying WGA methods. Firstly, the amplification stage adds two to three hours to the total time to result, varying according to kit selection, DNA input and desired DNA yield. Also, whilst amplification is uniform across the genome, WGA does introduce bias to the identification of species by sequencing according to factors such as GC content and DNA conformation of the genome that can lead to the preferential amplification of some DNA sequences compared to others (115). Therefore, it may be necessary to better understand and account for the impacts of these biases on the detection of pathogen species. Finally, the amplification of DNA by MDA forms branched intermediate DNA structures and chimeras (116) that can block the protein nanopores during sequencing, reducing the total number of sequence reads. Therefore, a DNA de-branching step, involving the incubation of the DNA with an endonuclease to break up the branched structures, is required. This reduces the average read length to an average of two to three kilobases (117) and further adds to the overall method complexity and time to result.

2.2.5 Nanopore sequencing library preparation

As discussed in Chapter One, there are several types of DNA sequencing methods available, with Illumina SBS technology currently being the most widely used. However, the nanopore based DNA sequencing technology developed by Oxford Nanopore Technologies is currently the most suitable technology for use in austere deployed

environments. This is due to the availability of rapid and simple library preparation methods, the small size and portability of the sequencing device and the ability for real-time data analysis. These factors combine for reduced time to result and improved suitability for use in deployed MTFs compared to the alternatives.

Part of the reduced time to result for nanopore sequencing is due to the availability of relatively simple library preparation methods. One example is the Rapid PCR Barcoding Kit (Oxford Nanopore Technologies) (SQK-RPB004). This method involves the enzymatic fragmentation of the DNA by an enzymatic complex termed the “transposome” that both cleaves the DNA template whilst simultaneously adding a DNA tag that contains PCR primer binding sites. Due to the use of 12 primers that each contain a unique barcode, up to 12 samples can be multiplexed onto one run, increasing the sample throughput. The tagged DNA fragments are amplified by PCR, generating amplicons suitable for DNA sequencing. This amplification step, whilst adding to the total time to result, enables significantly more reads to be generated from DNA inputs as low as 1-5 ng. Sequencing adapters which facilitate the translocation of the DNA strands through the protein nanopores during the sequencing process are then added. This library preparation method requires approximately 15 minutes of hands-on time with additional time required for the PCR step which varies depending on the required number of amplification cycles. There are examples of this method being applied to low biomass samples to enable metagenome sequencing to be carried out for diagnostic purposes. When an optimised version of the Rapid PCR barcoding method was applied to host depleted urine samples for the detection of *Neisseria gonorrhoeae* from patients with symptomatic urethral gonorrhoea, researchers were able to achieve $\geq 92.8\%$ coverage of the *N. gonorrhoeae* reference genome at ≥ 10 -fold depth in 70% of samples (118). This was achieved directly from the samples and without the need for culture.

An alternative library preparation method, the Rapid Barcoding Kit (SQK-RBK004), employs the same transposome enzymatic tagmentation reaction followed by the addition of DNA sequencing adapters but omits the PCR reaction. Therefore, the total time to result is reduced by approximately two hours compared to the PCR-based method. However, the kit requires an increased DNA input of 400 ng per sample, far higher than the five-ng required for the PCR based method. The ease of use of this method has enabled it to be employed in field-based exercises in deployed environments. A multicentre evaluation of the effect of different analytical approaches for the identification of bacteria from a range characterised DNA samples showed that the Rapid Barcoding kit was suitable for use, even in the austere field based environment (119). In addition, the sequencing data produced from the Rapid Barcoding kit more closely matched the composition of the input DNA compared to the PCR-based method. However, in this example, sufficient DNA was produced from the samples to enable the use of the Rapid

Barcoding kit without prior amplification. As discussed, low biomass samples that require pre-amplification to generate sufficient DNA for sequencing can potentially introduce distortion of the DNA composition compared to the sample input.

2.2.6 Nanopore sequencing and data analysis

Once prepared for sequencing, the multiplexed sample is sequenced on a nanopore sequencing device. A key advantage of nanopore based sequencing is the ability to analyse the data in real-time. The analysis of metagenomics data, due to the amount and complexity of the data, is difficult. Further, depending on factors such as the aim of the workflow and the sample type, bespoke analysis pipelines are required, needing input from expert bioinformaticians to establish. The selection of the optimal bioinformatics pipeline is critical for metagenomics studies and significantly influences the sensitivity of pathogen detection, particularly when the pathogens are present at low abundance (120).

As the DNA strands move through the protein nanopore, it applies a different voltage across the membrane. This raw data is stored in Fast5 format by the MinKNOW software (Oxford Nanopore Technologies) in real-time and the first step in the bioinformatics pipeline is converting this raw data into a DNA sequence. The MinKNOW software has in-built base calling software that can convert the Fast5 data to Fastq read data. Alternatively, base calling software can be applied offline using adjustable features that can be modified according to the pipeline aims.

Following base calling, a series of quality control steps are often implemented. These usually remove shorter DNA reads, sequencing adapters, duplicate reads and reads that have not achieved a specified quality score (70). Non-target DNA and known contaminants, such as human DNA, can also be removed at this stage by aligning the sequences to a reference genome. Alignment software such as minimap2 are used for this (121) and is often one of the most time consuming steps (122).

For the identification of microbial species, the next step is usually to compare the remaining reads to a suitable reference database. The choice of database will depend upon the aims of the workflow, with comprehensive bacterial, viral and fungal species databases available. The rise in the application of metagenomics has led to an increase in the number of databases available, with more compiled every year (123). There are different strategies for aligning sequences to reference databases. Microbial reads or microbial reads assembled into contigs can be aligned to the database. Alternatively, to reduce the time and computational power required, analysis of *k-mers* can be carried out, although this reduces the sensitivity of the output. However, this can be improved through the use of shorter *k-mers* (124). One example is the WIMP pipeline, developed by Oxford Nanopore Technologies, that utilises Centrifuge, a *k-mer* based identification with Burrow-

Wheeler's Transform and the Ferragina-Manzini index to classify reads into species using the RefSeq database in real-time (125).

Metagenomic data can also be interrogated to identify AMR genes. This is commonly achieved by aligning single reads to a reference AMR gene database, using tools such as AMRFinder which utilises the Bacterial Antimicrobial Resistance Reference Gene Database (126). The more challenging aspect is to then link the AMR gene to a specific organism and infer the phenotypic resistance to inform optimal treatments. Using nanopore sequencing of long reads can enable the host species to be identified by aligning the AMR gene flanking regions. This method was able to attribute tetracycline resistance genes to *N. gonorrhoeae* from metagenomics data, as opposed to other contaminating species (127). The identification of certain AMR genes and the host organism can be used to discern the phenotypic resistance with confidence as they have been extensively studied and are well understood. The AMR gene Cefotaximase-Munich (CTX-M) has been characterised as an extended-spectrum β -lactamase present in *Enterobacteriaceae* species for many years (128). Similarly, the link between the *mecA* gene in *S. aureus* and resistance to methicillin is well understood due to its global significance and prevalence in healthcare environments (129).

Linking AMR to the presence of specific genes is not always this simple. Resistance to antimicrobials is dependent upon a range of complex factors, including AMR gene presence/absence and gene expression that may involve multiple genes. Increased expression of efflux pumps and/or reduced expression of porins cause resistance to antibiotics by reducing the concentration of antibiotics in the cell. Increased AMR gene copy number, which is difficult to infer from metagenomics data alone, can also confer resistance (130). To solve this, machine learning approaches are being developed to generate confidence scores for the phenotypic resistance of an organism in the presence of an AMR gene. This has been shown to be accurate when confidence scores were high, but struggled with identifying resistance conferred by gene mutations (131).

2.2.7 Aims

The aim of the work described in this chapter was to develop and optimise a CMg workflow for the rapid diagnosis of wound infections from wound swab samples suitable for use in austere environments, specifically deployed MTFs, prior to the prospective collection of samples from patients with suspected wound infections.

2.3 Methods

2.3.1 Contrived sample generation

Contrived wound swab samples were generated by spiking blank Transwab® Amies Charcoal (Medical Wire & Equipment) swabs with 50 μ L whole human blood in

dipotassium Ethylene Diamine Tetraacetic Acid (K₂EDTA) collected from healthy donors (Cambridge Bioscience) or defibrinated *Ovis aries* blood (ThermoFisher), depending upon access to the suitable laboratory facilities. Whole human blood was mixed with serial dilutions of clinically relevant bacteria strains ranging from 5.0×10^4 to 5.0×10^2 colony forming units (CFU) per swab, with these concentrations selected based on the previous quantification of bacteria collected from infected wound sites by wound swab sampling (132). Spiked strains included *Escherichia coli* 10- β (New England Biolabs), *Bacillus thuringiensis* (ATCC 10792), *bla*_{CTX-M} positive *E. coli* (NCTC 13441), *mecA* positive *S. aureus* (NCTC 14245), *vanA* positive *Enterococcus faecium* (NCTC 13633) and *Acinetobacter baumannii* (NCTC 13486). Cultures were grown in nutrient broth at 37°C at 180 rpm for 16 h and quantified by plating in triplicate onto nutrient agar (*E. coli*, *S. aureus*, *B. thuringiensis*) or Sheep Blood agar (*E. faecium* and *A. baumannii*).

2.3.2 Preparation of DNA from bacterial cultures

To generate highly concentrated bacterial DNA from each spiked species, DNA was extracted from overnight cultures using the Qiagen UCP Pathogen Mini DNA extraction kit (Qiagen), including the pretreatment of microbial DNA from cultures steps, according to the manufacturer's instructions. One millilitre of overnight culture was added to the pathogen lysis tube and centrifuged for 5 min at $14000 \times g$. The supernatant was removed, and the pellet was resuspended in 500 μ L Buffer ATL. The pathogen lysis tube was placed in a FastPrep-24™ 5G (MP Biomedicals) bead-beater and bead beaten for 60 s at 6.0 m/s. The tube was then centrifuged for 5 s at $8000 \times g$. Four hundred microliters of the supernatant were transferred to a 2 mL microcentrifuge tube. Forty microliters of Proteinase K were added, and the sample incubated at 56°C for 10 min. Two hundred microliters of Buffer APL2 was added, and the sample was incubated at 70°C for 10 min. Three hundred microliters of ethanol were added to the sample and 600 μ L of the sample was added to the QIAamp UCP Mini spin column. The spin column was centrifuged at $6000 \times g$ for 1 min and the filtrate was discarded. This step was repeated until the whole sample had been filtered through the spin column and then 600 μ L Buffer APW1 was added and the spin column and centrifuged at $6000 \times g$ for 1 min. The filtrate was discarded, and this step was repeated with 750 μ L Buffer APW2 and the spin column centrifuged at $20000 \times g$ for 3 min. The spin column was then dried at 56°C for 3 min and 50 μ L Buffer AVE was applied to the spin column. The spin column was incubated at room temperature for 1 min and then centrifuged at $20000 \times g$ to elute the DNA.

2.3.3 Pre-sequencing sample preparation.

2.3.3.1 Original host depletion and DNA extraction method

The spiked swab sample was added to the charcoal media swab tube. The swab tip was added to 200 μ L phosphate-buffered saline (PBS) and vortexed for 20 s at full speed. To

facilitate host DNA depletion, 40 μ L 30% saponin (in PBS) was added to a final concentration of 2.5% with 200 μ L heat labile salt activated nuclease (HLSAN) buffer (5.5 M NaCl and 100 mM MgCl_2 in molecular-grade water) and 10 μ L HLSAN DNase (ArcticZymes). The sample was incubated at 37°C for 10 min at 1000 rpm. The sample was transferred to a new 2 mL tube and the swab tip disposed of. Nine hundred microliters of PBS were added and the sample was centrifuged at $12000 \times g$ for 3 min. The supernatant was discarded, and the pellet resuspended in 600 μ L PBS. The resuspended pellet was transferred to a Matrix lysing E tube (MP Biomedicals) and bead beaten for 30 s at 6.0 m/s on a FastPrep-24 5G bead beater and centrifuged at $17000 \times g$ for 60 s. Four hundred microliters of the clear supernatant was transferred to a new tube with 300 μ L lysis buffer and 30 μ L Proteinase K (Promega). The sample was incubated at 65°C for 5 min at 1000 rpm and added directly to a MaxWell 16 LEV Blood cartridge (Promega). The DNA extraction was carried out by the MaxWell 16 LEV system Blood DNA protocol and eluted into 50 μ L elution buffer.

2.3.3.2 Optimised host depletion

The optimised host depletion method aimed to improve the yield of bacteria DNA. The vortex step was switched to after the saponin and HLSAN DNase incubation step and the PBS washed pellet was resuspended in 550 μ L PBS and bead beaten for 60 s at 6.0 m/s.

2.3.4 Comparison of DNA extraction platforms

Comparison of three DNA extraction methods was carried out. This included the MaxWell 16 LEV platform, as previously described and the following methods:

PDQeX (MicroGEM): an automated DNA extraction platform that utilises a cyclical heating process and thermophilic enzymes within a closed tube filtration system to lyse bacterial cells and isolate the DNA, resulting in the elution of purified bacterial DNA. Once the sample has been prepared and added to the platform filter tube, the method does not require further hands-on steps and takes approximately 30 min to complete. The extraction mix was prepared using the following reagents: 15 μ L 10X Green buffer, 2 μ L Lysozyme, 15 μ L Enhancer, 2 μ L *prep*GEM and 116 μ L sample. The extraction mix was then added to a PDQeX filter column and loaded into the platform column rack. The extraction run was then started using the following programme: 37°C 15 min, 75°C 15 min and 95°C 2 min.

ZYMO Quick-DNA Fungal/Bacterial Miniprep Kit (ZYMO Research), is a manual DNA extraction kit that utilises a silica column-based DNA extraction method. Bacterial cells are lysed, and the DNA bound to the column matrix and the debris and contaminants are removed. The DNA is then released from the silica matrix into the elution buffer. This manual process takes approximately 30 min but requires a high number of hands-on steps. Firstly, 200 μ L of sample was added to a labelled ZR BashingBead™ Lysis Tube

(0.1 mm & 0.5 mm). Seven hundred and fifty microliters of BashingBead™ Buffer was then added and the tubes were bead beaten for 30 s at 6.0 m/s on a FastPrep-24 5G bead beater and centrifuged at $17000 \times g$ for 60 s. The tubes were then centrifuged at 12200 rpm for 1 min and 400 μ L of supernatant was transferred to a Zymo-Spin™ III-F Filter unit placed in a collection tube and centrifuged at 11000 rpm for 1 min. The filter unit was removed and 1200 μ L of Genomic Lysis Buffer was added to the filtrate. Eight hundred microliters of the mixture were added to a new Zymo-Spin™ IICR Column placed in a new collection tube and centrifuged at 12200 rpm for 1 min. The flow through was discarded and the step repeated. The Zymo-Spin™ IICR Column filter unit was placed into a new filter unit and 200 μ L DNA Pre-Wash Buffer was added to the column and was centrifuged at 12200 rpm for 1 min. Five hundred microliters of g-DNA Wash Buffer were added to the column and centrifuged at 12200 rpm for 1 min. The column filter unit was added to a 1.5 mL microcentrifuge tube and 50 μ L DNA Elution Buffer was added directly to the column matrix. The column was centrifuged at 12200 rpm for 30 s to elute the DNA. The 50 μ L DNA extract was then either stored at -20°C or progressed directly to library preparation.

2.3.5 DNA quantification

Total DNA concentration was quantified using the double-stranded DNA high-sensitivity (dsDNA HS) assay kit on the Qubit 4™ fluorometer (Invitrogen) according to the manufacturer's instructions. One hundred and ninety-nine microliters of dsDNA HS buffer was mixed with 1 μ L of dsDNA HS dye to generate the working solution. One hundred and ninety-eight microliters of working solution was mixed with 2 μ L of DNA sample in a Qubit Assay Tube and incubated for 2 min at room temperature. For instrument calibration, 190 μ L of working solution was added to Qubit Assay Tubes. Ten microliters of dsDNA HS Standard #1 was added to tube one and 10 μ L dsDNA HS Standard #2 was added to tube two and incubated for 2 min at room temperature. The standards and DNA samples were quantified using the dsDNA HS programme (Invitrogen).

2.3.6 DNA purification

To increase the concentration of the DNA and reduce the sample volume prior to WGA, the DNA extract was purified and concentrated by AMPure XP bead (Beckman Coulter) purification according to the manufacturer's instructions. AMPure XP beads were removed from the refrigerator and kept at room temperature for 15 min prior to use. AMPure XP beads were then mixed with the DNA sample in a 1.5 mL Lo-Bind tube (Eppendorf) at a ratio of $1.8 \times$ beads to DNA sample. The samples were then stored for 5 min at room temperature on a HulaMixer (Invitrogen) using the following settings: rotation speed 15 and timer 10, reciprocal tilting turning angle 45° & timer 5, vibration turning angle 5° and timer 5. The samples were pulse-spun on a micro centrifuge and added to a magnetic

tube rack for 2 min until the supernatant was clear. The clear supernatant was then removed, and the beads washed with 200 μ L of 70% ethanol for 30 s and the ethanol removed and the wash step repeated. The tubes were pulse-spun on a micro centrifuge and placed back on the magnetic rack. Residual ethanol was removed and the beads air dried for 90 s. The beads were then resuspended in 10 μ L nuclease-free water and incubated at room temperature for 2 min before being placed back on the magnetic rack for 2 min. The clear supernatant was then transferred to a new 1.5 mL Lo-Bind tube.

2.3.7 DNA fragment size analysis

DNA fragment size was quantified using the TapeStation 4200 system with the Genomic DNA Screentape assay (Agilent). Ten microliters of the Genomic DNA assay sample buffer and 1 μ L of the DNA sample or DNA ladder were added to the TapeStation tubes. The tubes were briefly vortexed and then pulse centrifuged. The tubes were added to the TapeStation 4200 system along with the Genomic DNA screentape reagent. The analysis was then carried out and visualised using the TapeStation analysis software (Agilent).

2.3.8 Whole genome amplification

Qiagen Repli-G Single Cell kit WGA (Qiagen) (herein referred to as Repli-G): 2 μ L Buffer D2 was added to the DNA sample and the reaction made up to 15 μ L with nuclease-free water in a 0.2 mL PCR tube and incubated at 65°C for 10 min. Three microliters of stop solution were then added at room temperature. The polymerase master mix was prepared by adding 29 μ L of single cell reaction buffer to 2 μ L of the REPLI-g single-cell DNA polymerase per sample and mixed by pipetting. Thirty-one microliters of polymerase master mix was added to the sample and the sample incubated at 37°C for 90 min followed by heat inactivation at 60°C for 3 min. For the WGA incubation time experiments, 5 pg of purified bacteria DNA was amplified according to the described protocol. Five microliter aliquots of each reaction were removed at 1.5, 2.0 and 2.5 h time points, heat inactivated at 60°C for 3 min and stored on ice until further processing.

Illustra™ GenomiPhi™ V3 Ready-to-Go™ WGA (Cytiva) (herein referred to as Genomiphi): 10 μ L of purified DNA was added to 10 μ L of the 2 $\times$ denaturation buffer in a 0.2 mL PCR tube and heated to 95°C for 3 min and cooled to 4°C on ice. The 20 μ L reaction was added directly to the freeze-dried reagent cake and incubated at 37°C for 90 min followed by a heating step at 65°C for 10 min. The sample was cooled to 4°C on ice.

2.3.9 DNA de-branching

Two microliters of the NEBuffer and 1 μ L of the T7 Endonuclease (New England Biolabs) were added to a maximum of 1 μ g amplified DNA in 17 μ L total volume, adjusted with nuclease-free water. The sample was then incubated at 37°C for 15 min.

2.3.10 Nanopore DNA sequencing

2.3.10.1 Library preparation and sequencing

Rapid PCR barcoding (SQK-RPB004) library preparation kit (Oxford Nanopore Technologies): libraries were prepared according to the manufacturer's instructions with the following modifications: i) up to 10 ng DNA used as the template in a total volume of 7.5 μ L ii) 2.5 μ L fragmentation mix added per sample and the PCR reaction volume was doubled. This consisted of 2 μ L barcode, 50 μ L LongAmp Taq 2 $\times$ Master Mix (New England Biolabs) and 38 μ L nuclease-free water iii) the PCR programme used was 3 min at 95°C followed by 25 cycles of 15 s at 95°C, 15 s at 56°C, 4 min at 65°C and a final extension of 4 min at 65°C.

Rapid Barcoding (SQK-RBK004) library preparation kit (Oxford Nanopore Technologies): libraries were prepared according to the manufacturer's instructions. Up to 400 ng of the DNA in a total volume of 7.5 μ L was added to a 0.2 mL PCR tube. Two point five microliters of the fragmentation mix RB01-12 was added, and the sample incubated at 30°C for 1 min and 80°C for 1 min and the sample was cooled briefly on ice. Libraries were pooled into a 1.5 mL LoBind tube (Eppendorf) and mixed with an equal volume of AMPure XP beads. The pooled libraries underwent AMPure XP bead purification as previously described and eluted into 10 μ L of 10 mM Tris-HCl pH 7.5-8.0 with 50 mM NaCl. One microliter of rapid adapter was added to 10 μ L of the purified pooled libraries and the sample incubated at room temperature for 5 min.

2.3.10.2 Flow cell priming

An R9.4.1 Flowcell (Oxford Nanopore Technologies) was inserted into the MK1c Sequencing device (Oxford Nanopore Technologies). The flow cell was then checked using the Flow Cell Check programme to ensure >800 pores were available for sequencing. The priming port was opened, and any air bubbles removed using a pipette. The priming mix was prepared by adding 30 μ L flush tether directly to the tube of flush buffer and mixed by pipetting at room temperature. Eight hundred microliters of the priming mix were added to the flow cell via the priming port and the flow cell incubated at room temperature for 5 min. The SpotOn port was opened and the remaining 200 μ L of priming mix was added to the flow cell via the priming port. The DNA libraries were then added to the flow cell via the SpotOn port. The SpotOn port and priming ports were then closed and the Mk1c lid was closed. A new experiment was then set up using the MinKnow software on the Mk1c device. The prepared libraries were sequenced for up to 16 h and the data stored in Fast5 format.

2.3.11 Bioinformatic and statistical analysis.

ONT Taxonomic Profiling, a custom bioinformatics pipeline for the taxonomic profiling of microbial species and for the identification of AMR genes, previously developed for the

analysis of metagenomics datasets for infectious disease diagnostics, was applied (63). Fast5 files were base-called using the Super Accuracy model using the guppy_basecaller and de-multiplexed with the guppy_barcode of the Guppy software v6.2.x. The minimum barcoding score was set to 65. Human reads were then removed from the data by mapping against the *Homo sapiens* reference genome CHM13v2 (133) using minimap2 (version 2.17) (121). The remaining reads were then mapped to the RefSeq database using minimap2 v2.17. The reads were then filtered based on a score defined as a harmonic mean of identity and coverage of the mapped reads using the following equation:

$$(2 \times \frac{[identity + coverage]}{[identity \times coverage]})$$

Reads with a taxa score ≥ 85 were included for taxonomic profiling, which identified the taxa and their relative abundance. The pipeline also included applying the KMA (version 1.4.9) tool (134) to screen for AMR genes with the amrfinderplus (v3.10) database. Microbial species identification was presented as the total number of reads and the percentage of the total classified microbial reads aligned to the species. Species were concluded to be present if ≥ 5 reads, minus the number of reads of the same species identified in the negative control sample, were assigned to that species. AMR genes were concluded to be present if their template identity and coverage were $\geq 90\%$ and their depth of coverage reported by KMA were $\geq 1\times$. Data are expressed as means + standard deviation and were analysed using GraphPad Prism 10 (version 10.0.2) software. When multiple comparison analyses were performed, *p*-values were obtained by performing an ordinary one-way or two-way ANOVA test or mixed effects analysis when the *n* number varied, with post hoc Tukey's test. When two samples were individually compared, results were obtained using a two-tail unpaired t-test.

2.3.12 qPCR

Taqman probe-based qPCR assays were performed to quantify the spiked bacteria and human DNA in the DNA extracts. All assays were performed on the QuantStudio™ 7 Flex (ThermoFisher) platform and quantitation cycle (Cq) values determined by the QuantStudio™ software (ThermoFisher). Cq values below 38.0 were considered positive. Positive control reactions of purified bacterial DNA at a concentration of 1.0 ng/μL and negative control reactions of nuclease-free water only were run on each plate. The details of each assay are described in Table 1. The TaqMan Fast Advanced Master Mix (Applied Biosystems) was used for all assays. The following conditions were applied for each assay:

E. coli 16S rRNA gene: 25 μL reactions using 12.5 μL master mix, final concentrations of 900 nM forward primer, 900 nM reverse primer, 200 nM of probe, and 1 μL of template.

Thermal cycling conditions were 1 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 30 s at 60 °C.

E. coli *bla*_{CTX-M-15}, *S. aureus* *mecA*, *A.baumannii* outer membrane protein A (*ompA*), *E. faecium* vancomycin A (*vanA*), *O. aries* prolactin receptor (*prlr*): 25 µL reactions using 12.5 µL master mix, final concentrations of 400 nM forward primer, 400 nM reverse primer, 100 nM of probe, and 1 µL of template. Thermal cycling conditions were 1 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 30 s at 60 °C.

S. aureus protein coding sequence (NCBI Protein accession number: YP_500811.1): 10 µL reactions using 5 µL master mix, 1 µM forward primer, 1 µM reverse primer, 200 nM probe, and 1 µL template. Thermal cycling conditions were 1 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 30 s at 57 °C.

Human β -actin gene: 25 µL reactions using 12.5 µL master mix, 500 nM forward and reverse primers and 200 nM probe and 1.0 µL of template DNA. The reaction was then made up to a final volume of 25 µL with nuclease-free water. The reaction conditions were 95°C for 1 min, followed by 40 cycles of 95°C for 15s and 60°C for 30s.

B. thuringiensis Hemolysin BL (*hbl*) C gene: 25 µL reactions using 12.5 µL master mix, 400 nM forward and reverse primers and 200 nM probe and 1.0 µL of template DNA. The reaction was then made up to a final volume of 25 µL with nuclease-free water. The reaction conditions were 95°C for 2 min, followed by 45 cycles of 95°C for 15 s and 58°C for 1 min.

Table 1. Primer and probe sequences for spiked bacteria species and host qPCR assays.

Assay target	Forward primer (5'-3')	Reverse primer (5'-3')	Probe sequence (5'-3')	Reference
<i>E. coli</i> 16S rRNA	CATGCCGCGTGT ATGAAGAA	CGGGTAACGTCAA TGAGCAAA	TTCAGCGGGGAGG AAGGGAGTAAAGT TAATA	(135)
<i>E. coli</i> <i>bla</i> _{CTX-M-15}	CGGGCRATGGC GCAAC	TGCRCCGGTSGTA TTGCC	CCARCGGGCGCAG YTGGTGAC	(136)
<i>S. aureus</i> <i>mecA</i>	GCTCAAATTTCAA ACAAAAATTTAGA TAATG	TGAAAGGATCTGT ACTGGGTTAATCA GT	AGCTGATTCAGGT TACGGACAAGGTG A	(137)
<i>A. baumannii</i> <i>ompA</i>	TCTTGGTGGTCA CTTGAAGC	ACTCTTGTGGTTGT GGAGCA	AAGTTGCTCCAGTT GAACCAACTCCA	(138)
<i>E. faecium</i> <i>vanA</i>	GCCGGA AAAAGG CTCTGAA	TTTTTGGCGTTTC CTGTATCC	CGCAGTTATAACC GTTCCCGCAGACC	(139)
<i>S. aureus</i> protein coding sequence (YP_500811.1)	AACTACTAGGGG AGCCTAATRAT	GGTACTAACCAAAT CAGGTCATAA	TGGCTGAGATGAA YTGTTCAGACCC	(140)

<i>O. aries prlr</i>	CCAACATGCCTT TAAACCCTCAA	GGAAGTGTAGCCT TCTGACTCG	TGCCTTTCCTTCCC CGCCAGTCTC	(141)
Human β-actin	CCACACTGTGCC CATCTACG	AGGATCTTCATGA GGTAGTCAGTCAG	ATGCCCTCCCCCA TGCCATCCTGCGT	(142)
<i>B. thuringiensis</i> hblC	AAGATAGAGAAG GGTTTTCGGATA GA	TGCCGTTGCGTAT TTTCTTG	AGTACTTCAAGAAA TGGCTATG	[in-house assay]

2.4 Results

2.4.1 Host depletion and DNA extraction from contrived wound swabs

The workflow applied to CMg for wound swab samples was split into five sequential steps: host DNA depletion, DNA extraction, WGA, library preparation and DNA sequencing. Being the first steps in the process, it was essential to determine whether the wound swab samples collected with the Transwab Amies Charcoal swab, commonly used for the collection of wound swab samples from patients in the NHS, were compatible with the saponin-based host DNA depletion and DNA extraction steps. Contrived samples were generated by spiking blank wound swabs with 5×10^3 CFU of *E. coli* 10- β mixed with defibrinated *O. aries* blood and added to the charcoal media tube before undergoing the one-pot 2.5% saponin host depletion protocol. The saponin depletion reduced the host DNA concentration compared to non-depleted control samples, with no detection of the *O. aries prlr* gene by qPCR in the depleted samples. Non-depleted control samples returned mean Cq values of 31.71 in the charcoal swab samples (Figure 4B). There was no reduction in the efficiency of host DNA depletion between samples added to the charcoal media and control samples (Figure 4B). There was no detection of the *prlr* gene in any depleted sample, including charcoal and control swabs. In addition, there was no reduction in the concentration of *E. coli* DNA extracted from wound swab samples added to the charcoal media compared to the control samples (Figure 4A). These results demonstrated that the host DNA depletion and DNA extraction methods were not impacted by the presence of the Amies Charcoal Agar and were therefore suitable for the processing of wound swab samples for CMg analysis.

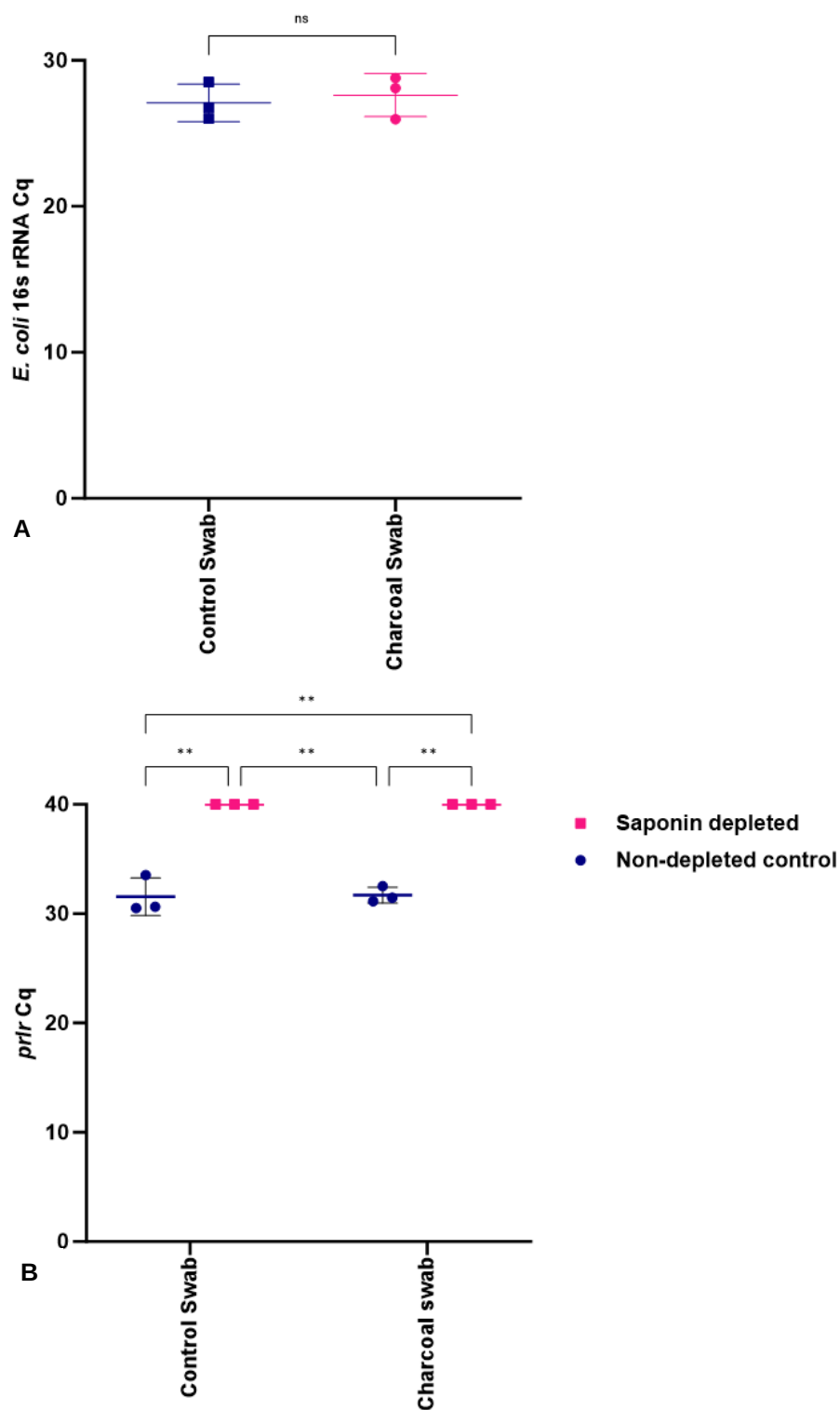


Figure 4. Evaluation of the impact of the Amies charcoal agar on the host DNA depletion and bacteria DNA extraction methods.

Comparison of the *E. coli* 16s rRNA (A) and *O. aries prlr* (B) Cq values obtained from spiked swab samples added to a Transwab Amies Charcoal swab compared to control swabs. Addition to the charcoal agar did not impact the concentration of *E. coli* DNA or the efficiency of host DNA depletion compared to control swab samples.

±SD. n=3. ns $p > 0.05$ (not shown) ** $p < 0.01$.

2.4.2 Storage of wound swab samples

To evaluate whether the short-term storage of the wound swab samples, likely to be required due to the sporadic nature of wound swab sample collections, impacted the detection of spiked bacteria, experiments were carried out to quantify the concentration of bacteria DNA following storage for seven days. There was no change in the concentration of *E. coli* DNA, as measured by qPCR Cq values, extracted from the swab samples stored at 4°C for seven days compared to control samples processed immediately following swab spiking. However, the DNA extracted from the samples stored at room temperature for seven days contained an increased *E. coli* DNA concentration, with a ΔCq of 6.63 lower compared to the control samples (Figure 5).

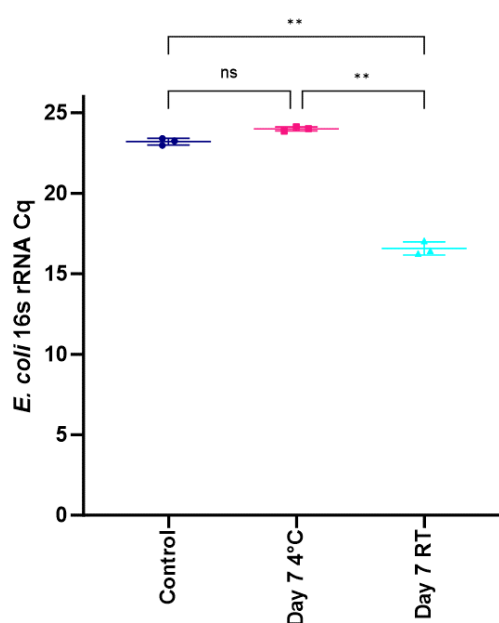


Figure 5. Evaluation of wound swab sample storage methods.

Comparison of the *E. coli* 16s rRNA Cq values obtained from spiked swab samples stored for 7 days at 4 °C and at room temperature (RT) compared to control swabs processed immediately. Storage at 4°C for seven days did not change the concentration of *E. coli* DNA compared to the control swabs processed immediately. Storage at room temperature resulted in increased *E. coli* DNA concentration.

±SD. $n=3$. ns $p>0.05$ ** $p<0.01$.

2.4.3 Host depletion optimisation

A preliminary workflow was designed for the processing of wound swab samples for CMg. Several of the steps were evaluated for optimisation to improve the extraction of microbial DNA. Firstly, the inclusion of a vortexing step was deemed necessary to remove the material from the swab tip into solution to enable the subsequent extraction of microbial DNA. However, previous work carried out by group members and shared in a conversation by J. Latham (April 2021) showed that the vortexing of bacterial cells

resulted in bacteria cell damage and DNA loss. It was hypothesised that damaging the bacteria cells by vortexing prior to the host depletion step would result in a loss of bacterial DNA by enabling the DNase enzyme to access the bacteria DNA. To evaluate this, experiments were carried out to compare the efficiency of host depletion and microbial DNA extraction when vortexing was carried out prior to and following the host depletion step. No decrease in host depletion efficiency, determined by *O. aries prlr* gene Cq value, was identified between pre and post-depletion vortexing (Figure 6A). However, carrying out vortexing, at all three tested durations of 20, 60 and 120 s, pre-depletion was shown to significantly increase the *E. coli* 16s rRNA Cq value, representing a decrease in the concentration of *E. coli* DNA, obtained from spiked swab samples compared to post-depletion vortexing (Figure 6B). Therefore, post-depletion vortexing was applied moving forward.

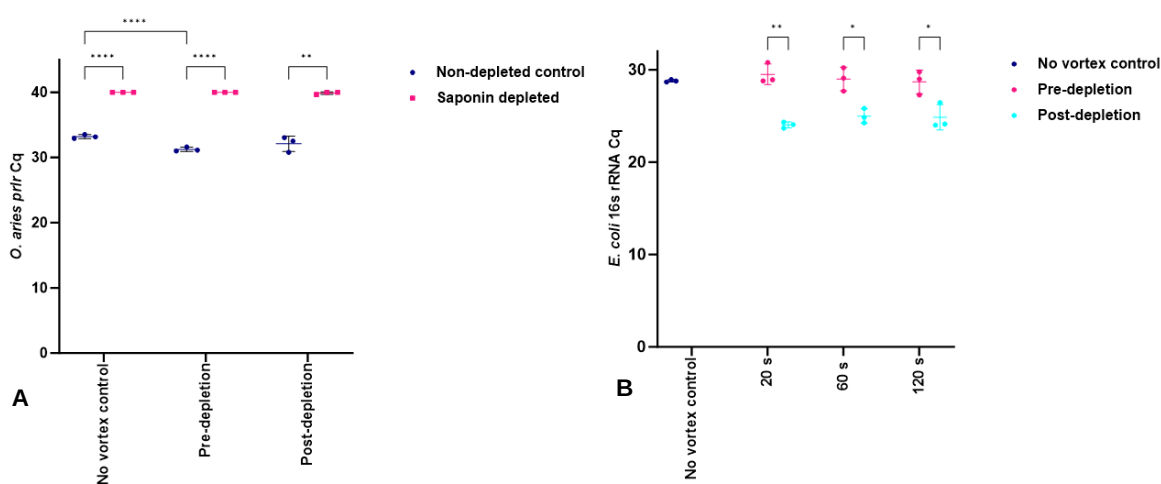


Figure 6. The effect of pre and post-depletion vortexing on the host depletion and *E. coli* DNA extraction methods.

Comparison of the host *prlr* (A) and *E. coli* (B) Cq values obtained from spiked swab samples that underwent vortexing pre and post-host depletion step. There was no difference in *O. aries* Cq values, representing no impact on the host deletion efficiency. Vortexing post-depletion resulted in a lower *E. coli* Cq value, representing an increase in *E. coli* DNA concentration compared to pre-depletion vortexing.

±SD. n=3. ns $p>0.05$ (not shown) * $p<0.05$ ** $p<0.001$ **** $p<0.00001$

The bead-beating step was also evaluated. Data obtained from group members and shared in a conversation with L. Moragues-Solanas (May 2021) indicated that the efficiency of DNA extraction from Gram positive bacteria, including *S. aureus*, was poor. Therefore, the duration of the bead-beating step was evaluated. An increased duration of bead beating from the initial 30 s at 6.0 m/s to 60 s at 6.0 m/s was compared for the extraction of DNA from the representative Gram positive bacteria *S. aureus*. There was no significant change in the Cq values obtained from each bead-beating duration for either the *S. aureus* genomic DNA target assay or the *mecA* AMR gene target assay (Figure 7).

Despite there being no significant increase in *S. aureus* DNA concentration by bead-beating for 60 s, to ensure that the extraction method was suitable for microorganisms including fungal pathogens, subsequent experiments included the 60 s bead-beating duration.

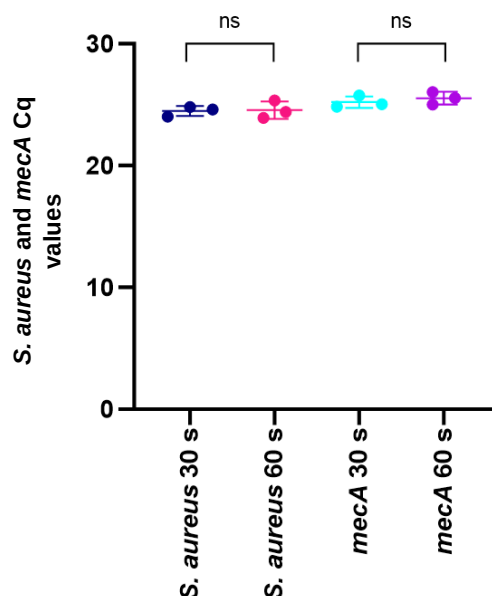


Figure 7. The effect of bead-beating duration on the extraction of *S. aureus* DNA.

Comparison of 30 s and 60 s bead beating steps on the concentration of *S. aureus* DNA extracted from spiked contrived wound swab samples as determined by qPCR. There was no change in the *S. aureus* and *mecA* AMR gene Cq values from DNA extracted with a 60 s bead beating step compared to 30 s bead beating.

±SD $n=3$ ns $p>0.05$.

Additional steps were optimised to improve the ease of operation, and the optimised method was compared to the original for the extraction of *E. coli* and *S. aureus* DNA from spiked contrived wound swab samples. Blank wound swabs were spiked with one clinically relevant bacteria species including *E. coli*, *S. aureus*, *E. faecium* and *A. baumannii* mixed with human blood in K₂ EDTA at approximately 5.0×10^3 CFU per swab. There was no significant change in the *E. coli* Cq values from the original and optimised depletion protocols. However, in *S. aureus*, *E. faecium* and *A. baumannii* the Cq values reduced significantly, from 35.01 to 29.48, 28.84 to 28.04 and 35.04 to 32.11 respectively, representing an increase in DNA concentration (Figure 8).

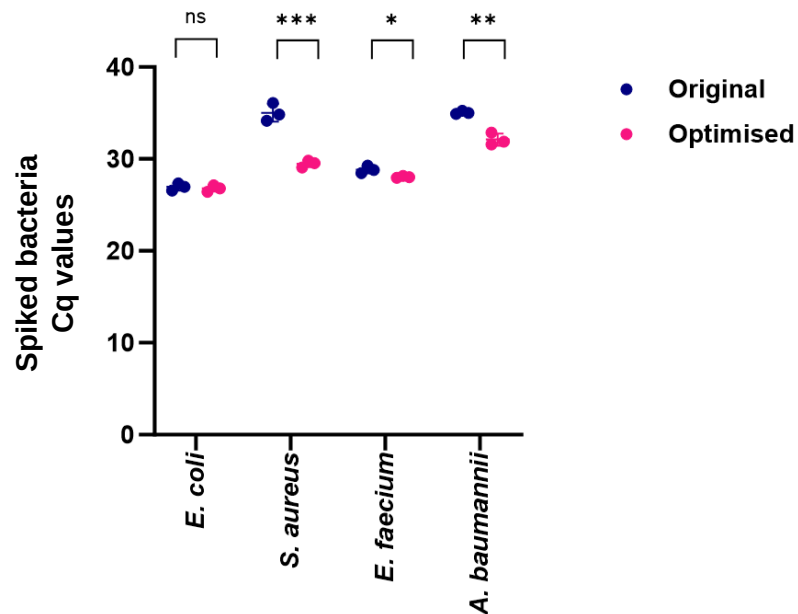


Figure 8. Comparison of the original and optimised DNA extraction methods.

Comparison of the spiked bacteria Cq values obtained from samples extracted using the original and optimised DNA extraction methods. Contrived wound swab samples were spiked with bacteria and extracted using the original or the optimised host depletion and DNA extraction methods. The optimised DNA extraction method resulted in lower Cq values for samples spiked with *S. aureus*, *E. faecium* and *A. baumannii*. There was no change in the Cq values of *E. coli* DNA extracts.

±SD n=3 ns $p > 0.05$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

To determine the efficiency of the host DNA depletion step using the optimised method, blank wound swabs were spiked with human blood in K₂ EDTA and processed according to the optimised protocol. The concentration of human genomic DNA in depleted and non-depleted control samples was evaluated by qPCR. Results showed a ΔCq value of 10.57 between the depleted and non-depleted control samples, corresponding to a 1520-fold ($2^{10.57}$) depletion in human genomic DNA (Figure 9).

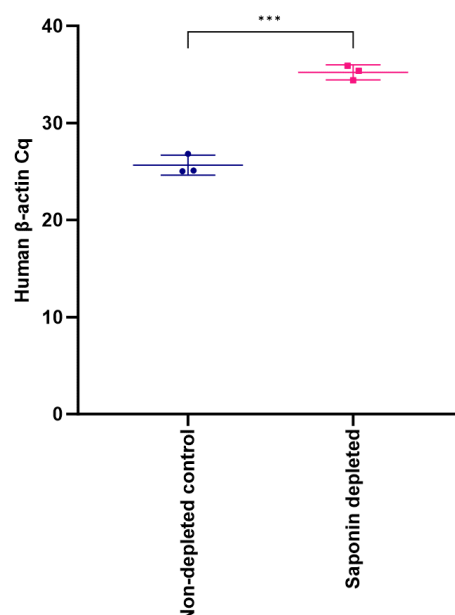


Figure 9. Depletion of human genomic DNA by the optimised host depletion method.

Comparison of the human β -actin Cq values obtained from depleted and non-depleted control samples using the optimised host depletion and DNA extraction method. Contrived wound swab samples spiked with whole human blood were processed using the optimised host depletion and DNA extraction methods and the concentration of human genomic DNA was measured by qPCR. The depletion resulted in a Δ Cq of 10.57, a 1520-fold depletion of human genomic DNA.

$\pm$ SD. $n=3$. *** $p<0.001$

2.4.4 Whole genome amplification incubation time

Previous work had demonstrated that the inclusion of a WGA step was required to generate sufficient concentrations of DNA for nanopore based DNA sequencing and a previously optimised Repli-G single cell kit protocol was identified as the most efficient WGA method from low DNA input samples (63). This protocol was further investigated to identify the optimal amplification step incubation time. There was a reduction in the Cq values of *E. coli* samples amplified for 1.5 and 2.0 h, but no significant increase in total DNA concentration in these samples. Conversely, there was an increase in total DNA concentration in the *S. aureus* DNA samples incubated for 2.5 h (Figure 10A) and a reduction in *S. aureus* Cq values when incubated for 2.0 and 2.5 h compared to 1.5 h. However, the total DNA concentration of the WGA NTC sample increased when incubated for 2.5 h (85.00 ng/ μ L) compared to 1.5 h (1.59 ng/ μ L). Therefore, the amplification time was maintained at 1.5 h to minimise the amplification of contaminating DNA.

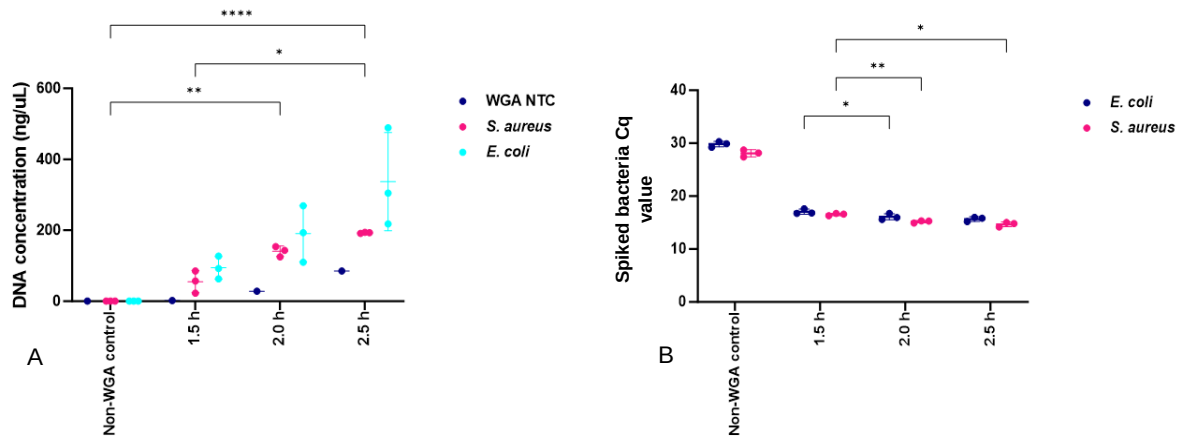


Figure 10. Optimisation of the Repli-G WGA amplification step duration.

The total DNA concentration (A) and *E. coli* and *S. aureus* Cq values (B) of purified bacteria DNA amplified by the Repli-G WGA kit for 1.5, 2.0 and 2.5 h. *E. coli* and *S. aureus* DNA was extracted from overnight cultures. 5 pg of purified DNA underwent WGA by Repli-G WGA kit and aliquots of the reaction were taken at 1.5, 2.0 and 2.5 h of the amplification step. The total DNA and spiked bacteria DNA concentrations were quantified by Qubit dsDNA HS assay and qPCR, respectively. Increasing the incubation step to 2.5 h increased the total DNA concentrations from the *S. aureus* DNA samples and the NTCs. The *E. coli* and *S. aureus* Cq values were lower, representing increased DNA concentrations, when the DNA was amplified for 2.0 and 2.5 h respectively.

±SD $n=3$ ns $p>0.05$ (not shown) * $p<0.05$ ** $p<0.01$ **** $p<0.0001$

2.4.5 Wound swab sample full clinical metagenomics protocol proof of concept

A full workflow proof of concept experiment was carried out to highlight any technical issues prior to progressing to the processing of patient samples. Blank wound swabs were spiked with *E. coli* (NCTC 13441), positive for the *bla*_{CTX-M-15} AMR gene and *S. aureus* (NCTC 14245) positive for the *mecA* AMR gene mixed in human blood in K₂EDTA at 5.0×10^3 CFU per species. qPCR, carried out pre and post-WGA, demonstrated the amplification of spiked bacteria DNA in depleted samples, resulting in a ΔCq of 6.99 and 8.55 pre and post-WGA for *S. aureus mecA* and *E. coli bla*_{CTX-M-15} qPCR assays respectively (Table 2). Conversely, the Cq value of human β -actin gene increased by 1.02 Cq following WGA in the depleted spiked sample. This contrasted with the non-depleted control sample in which the human β -actin Cq value decreased by 4.7 Cq post-WGA. Therefore, the ΔCq for post-WGA human β -actin in depleted and non-depleted samples was 13.24 Cq, a 9674-fold ($2^{13.24}$) depletion in human genomic DNA.

Table 2. Cq values of the spiked bacteria and human genomic DNA in the spiked depleted wound swab samples pre and post-WGA.

qPCR assay target	Pre-WGA Cq	Post-WGA Cq	ΔCq
<i>S. aureus mecA</i>	28.10	21.11	-6.99
<i>E. coli bla</i> _{CTX-M-15}	24.36	15.81	-8.55
Human β -actin	34.64	35.66	+1.02

Spiked wound swab samples were prepared for CMg. Following the host depletion and DNA extraction steps, the DNA underwent WGA. The spiked bacteria and human genomic DNA was quantified by qPCR pre and post-WGA. Values are the mean of biological replicates ($n=3$).

The saponin depletion of human genomic DNA reduced the proportion of sequencing reads assigned to the human genome and resulted in a higher proportion of reads aligning to the spiked bacteria species compared to the non-depleted control. In the saponin depleted samples, an average of 80.34% of the total sequencing reads were identified as *E. coli* and 0.26% of reads were identified as human. In comparison, in the non-depleted control, an average of 96.2% of the total reads were identified as human and the proportion of *E. coli* reads was reduced to 2.5% (Table 3). The efficient removal of human genomic DNA also resulted in improved genome coverage of the *E. coli bla*_{CTX-M-15} AMR gene, with an average of $68.23 \times$ gene coverage achieved in the saponin depleted samples. This was reduced to $1.13 \times$ coverage in the non-depleted control.

In contrast to the significant proportion of reads assigned to *E. coli*, the proportion of reads assigned to *S. aureus* in the saponin depleted samples was an average of 1.69%. This is reflected in the reduced coverage of the *mecA* AMR gene, with $0.79 \times$ average coverage in the depleted spiked samples. The coverage across the *bla*_{CTX-M-15} and *mecA* AMR genes obtained from depleted sample one are displayed in Figure 11 for comparison. Therefore, further evaluation of the amplification of *S. aureus* DNA was required to determine the cause of the low number of sequencing reads.

Table 3. Sequencing results from the full CMg workflow proof of concept experiment.

Sample	Total reads	<i>E. coli</i> reads (proportion of total reads)	<i>bla</i> _{CTX-M-15} alignments (gene coverage ×)	<i>S. aureus</i> reads (proportion of total reads)	<i>mecA</i> alignments (gene coverage ×)	Human reads (proportion of total reads)
Negative control	346,912	18,127 (5.23)	0 (0)	0 (0)	0 (0)	29929 (8.63)
Depleted sample one	529,729	434,917 (82.10)	100 (61.3)	4,104 (0.77)	2 (0.63)	1090 (0.21)
Depleted sample two	524,246	423,224 (80.73)	225 (128.51)	3,016 (0.58)	3 (0.18)	1036 (0.20)
Depleted sample three	123,887	96883 (78.20)	28 (14.87)	4612 (3.72)	3 (1.56)	466 (0.38)
Non-depleted control	125,472	3,486 (2.78)	4 (1.13)	51 (0.04)	0 (0)	120934 (96.38)

Contrived wound swab samples were spiked with *E. coli* and *S. aureus* mixed with whole human blood at known concentrations and processed for CMg. The depleted samples resulted in an average of 80.34% *E. coli* reads, 1.69% *S. aureus* reads and 0.26% human reads. The non-depleted sample resulted in 96.38% human reads.

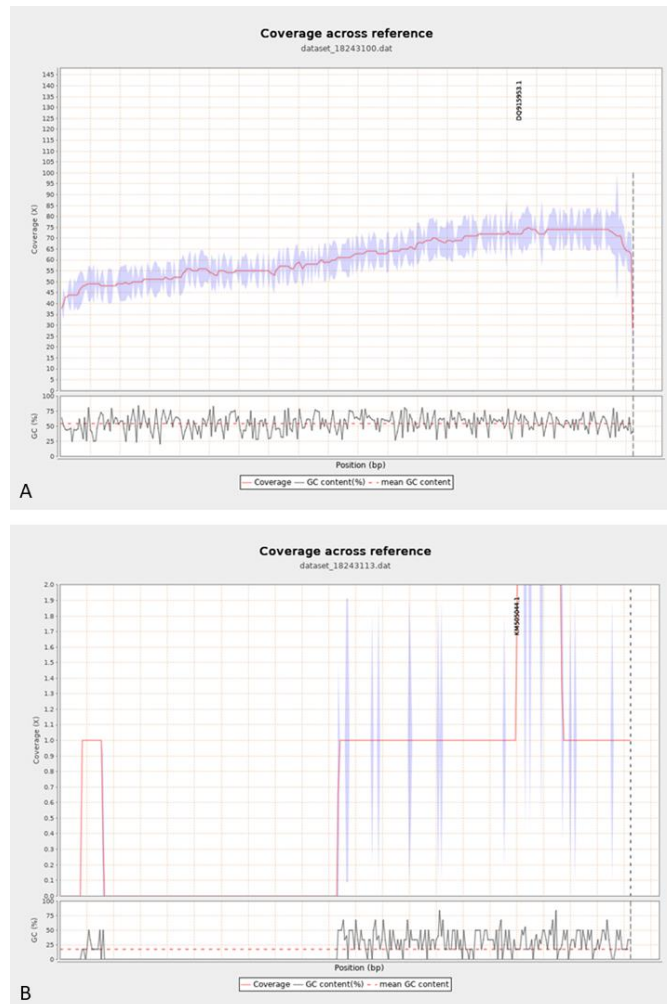


Figure 11. AMR gene coverage from host depleted sample 1.

Comparison of the average coverage across the *bla*_{CTX-M-15} (A) and the *mecA* AMR genes (B) from saponin depleted sample 1. Contrived wound swab samples were spiked with *E. coli* and *S. aureus* mixed with whole human blood and processed for CMg. Average coverage of $61.3 \times$ was achieved across the *bla*_{CTX-M-15} gene, compared to $0.63 \times$ across *mecA*.

To investigate the feasibility of utilising the Rapid Barcoding library preparation method, which had a reduced time to result, and hands-on time compared to the previously used rapid PCR barcoding method and to determine the limit of detection (LOD) of the full workflow, an LOD experiment was carried out on contrived samples spiked with four clinically relevant species. The samples were then prepared for sequencing with the Rapid Barcoding library preparation method. Blank wound swab samples were spiked with four serially diluted clinically relevant bacteria species ranging from 50000 to 50 CFU per species per swab. The total classified reads generated from these runs reduced as the bacteria CFU input was reduced, ranging from 123,002 reads at the 50000 CFU input to 4182 reads from the 50 CFU input. All four bacteria species were detected at the 50000 CFU input, ranging from 1.12% of total classified reads for *S. aureus* up to 43.44% for *E. coli*. As with previous experiments, a high proportion of the negative control sample reads were identified as *E. coli* contamination present in the negative control. Only *bla*_{CTX-M-15}

and *vanA* AMR genes were detected above 1× gene coverage at the 50000 CFU input, achieving 22 and 31 × gene coverage respectively and *mecA* was identified at 0.4 × gene coverage. At the 5000 CFU input, both *S. aureus* and *E. faecium* made up 0.93 and 0.75% of the total reads respectively. *vanA* was identified at the highest coverage at 7.2 × gene coverage, despite the low number of reads aligned to *E. faecium*. At the 50 CFU input, all three spiked species were identified at low proportions of total reads and none of the spiked AMR genes were identified (Table 4).

Table 4. CMg LOD sequencing results.

Sample	Total classified reads	<i>S. aureus</i> reads (proportion of total classified reads)	<i>mecA</i> reads (gene coverage ×)	<i>E. coli</i> reads (proportion of total classified reads)	<i>bla</i> _{CTX-M-15} reads (gene coverage ×)	<i>E. faecium</i> reads (proportion of total classified reads)	<i>VanA</i> (gene coverage ×)	<i>A. baumannii</i> reads (proportion of total classified reads)	Human reads (proportion of total classified reads)
Negative control	7326	3	0 (0)	1114	0 (0)	1	0 (0)	37	1719
50000 CFU input	123,002	1383 (1.12)	4 (0.4×)	53,428 (43.44)	18 (22×)	2046 (1.66)	32 (31×)	26,255 (21.35)	6175 (5.02)
5000 CFU input	71,937	667 (0.93)	1 (0.1×)	25,650 (35.66)	2 (1.8×)	542 (0.75)	6 (7.2×)	13,534 (18.81)	11,981 (16.65)
500 CFU input	35,820	70 (0.20)	0 (0×)	7845 (21.90)	0 (0×)	70 (0.20)	1 (0.13×)	3454 (9.64)	11,885 (33.18)
50 CFU input	4182	2 (0.05)	0 (0×)	667 (15.95)	0 (0×)	5 (0.12)	0 (0×)	37 (0.88)	502 (12.00)

Contrived wound swab samples were spiked with serial dilutions of bacteria and processed for CMg. *E. coli* was the only species identified >1% of total classified reads at the 50 CFU input. *S. aureus* and *E. faecium* were not identified at >1% of total classified reads at any input <50,000 CFU.

2.4.6 Evaluation of the amplification of species-specific DNA by whole genome amplification

The initial proof of concept experiments indicated that *S. aureus* and *E. faecium* reads were under-represented in the sequencing data in comparison to *E. coli* and *A. baumannii*. To investigate the impact of the Repli-G WGA step on the representation of bacteria species in the sequencing data, 250 pg of purified *S. aureus* and *E. coli* DNA was amplified by WGA, both separately and when combined in the same reaction and the resulting change in total DNA concentration and fold change in bacteria DNA concentration pre and post-WGA was determined by qubit and qPCR. This was repeated with *A. baumannii* and *E. faecium*. The average total post-WGA DNA concentration of *E. coli* DNA was not significantly higher compared to *S. aureus* when amplified in isolation (260.7 ng/μL vs. 124.0 ng/μL) (Figure 12A). However, the fold-change in *E. coli* DNA concentration was higher compared to *S. aureus* when amplified in a combined reaction (127.60 vs. 54.11) (Figure 12B). There was no significant difference in the total DNA concentration of the amplified *A. baumannii* and *E. faecium* DNA when amplified in isolation (94.18 vs 26.19 ng/μL) (Figure 12C). However, the fold-change in the concentration of *E. faecium* DNA was significantly higher when amplified in isolation

compared to *A. baumannii* in isolation (65.60 vs. 1.54). Notably, when amplified in a combined reaction, the fold change in concentration of *E. faecium* DNA was reduced to 3.55-fold (Figure 12D).

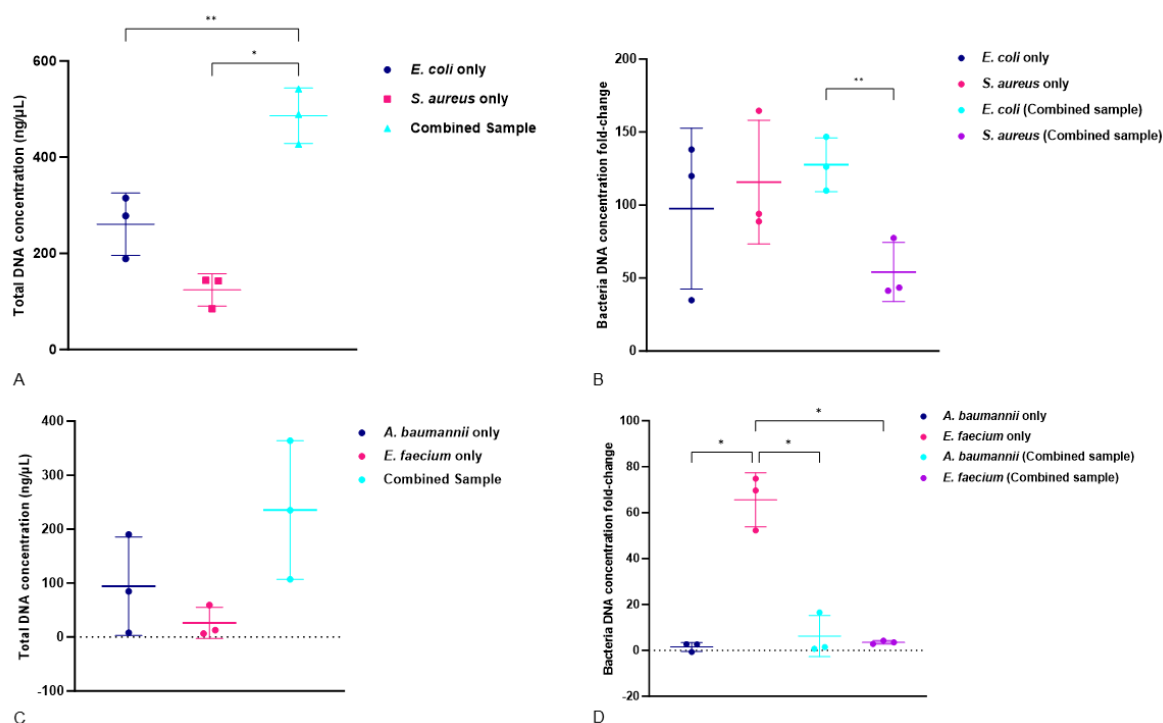


Figure 12. Comparison of the amplification of purified bacteria DNA by Repli-G WGA.

Comparison of the total DNA concentration (A,C) and pre and post-WGA fold-change in bacteria DNA concentration as determined by qPCR Cq values in single and combined reactions (B,D) by Repli-G WGA. Purified bacteria DNA was amplified by WGA when in isolation and when combined with the DNA of an additional species. The fold-change in *E. coli* DNA concentration was higher compared to *S. aureus* when amplified in a combined reaction. The fold-change in *E. faecium* DNA concentration pre and post-WGA was higher when amplified in isolation compared to both *A. baumannii* in isolation and *E. faecium* when amplified in the same reaction as *A. baumannii*.

±SD n=3. ns $p > 0.05$ (not shown) * $p < 0.05$ ** $p < 0.01$

2.4.7 Whole genome amplification method comparison

Previous experiments demonstrated that the amplification of bacteria DNA by the Repli-G WGA method was resulting in the over-representation of *E. coli* DNA in comparison to other spiked bacteria species in contrived wound swab samples. Therefore, an alternative WGA kit, the Genomiphi WGA kit was evaluated and compared to the Repli-G WGA kit. Two hundred and fifty pg of purified DNA from four bacteria species was amplified using each WGA method and the total bacteria DNA concentrations were compared. The Genomiphi kit resulted in lower total DNA concentrations compared to Repli-G amplified DNA samples for both sample types (see Figure 13A). However, the Cq values for bacteria species DNA concentrations, measured by qPCR, were lower in Genomiphi

amplified samples compared to Repli-G single cell in all species, excluding *E. coli* (Figure 13B). For *E. faecium* and *A. baumannii*, amplification by Repli-G single cell resulted in increased Cq values post-WGA, indicating poor DNA amplification. This was consistent with previous results that the Repli-G WGA kit resulted in increased amplification of *E. coli* DNA compared to other spiked species.

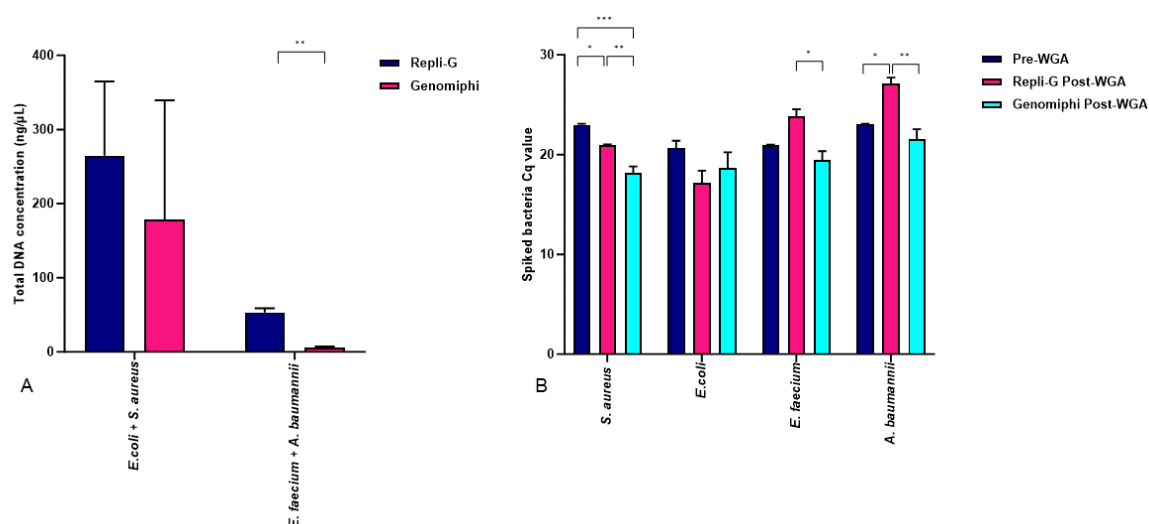


Figure 13. Comparison of the amplification of bacteria DNA by Repli-G and Genomiphi WGA methods.

Total DNA concentrations obtained from purified bacteria DNA amplified by Repli-G and Genomiphi WGA kits (A) and pre and post-WGA Cq values from bacteria DNA amplified with the RepliG and Genomiphi kits (B). Amplification by the Genomiphi kit resulted in lower total DNA concentrations compared to Repli-G. However, the Genomiphi kit resulted in lower Cq values for spiked bacteria compared to Repli-G, excluding *E. coli* in which Repli-G resulted in lower Cq values, representing higher DNA concentrations.

+SD $n=3$ ns $p>0.05$ (not shown) * $p<0.05$ ** $p<0.01$ *** $p<0.001$

2.4.8 Genomiphi whole genome amplification optimisation

The ease of use, ambient temperature storage requirements and improved amplification of bacteria DNA led to the selection of the Genomiphi WGA kit for ongoing sample processing. However, the reduced total DNA concentration required optimisation to ensure that low input samples could be processed and sequenced. Firstly, an AMPure XP bead purification step was added following the DNA extraction step and pre-WGA to concentrate the DNA sample. Purified DNA was then amplified by the Genomiphi WGA kit and compared to the amplification of an unpurified DNA sample. Pre-WGA AMPure XP bead purification resulted in an increased average total DNA yield of 4.89 ng/μL, compared to 1.30 ng/μL in the unpurified sample, although this was not significant (Figure 14A). There was a reduction in the Cq values of *E. coli* and *S. aureus*, from 25.07 to 24.85 and 20.20 to 18.79 respectively, although not significant. The human genomic DNA Cq

values increased when the sample was purified prior to the WGA step from 36.33 to 38.59, although this was not significant (Figure 14B).

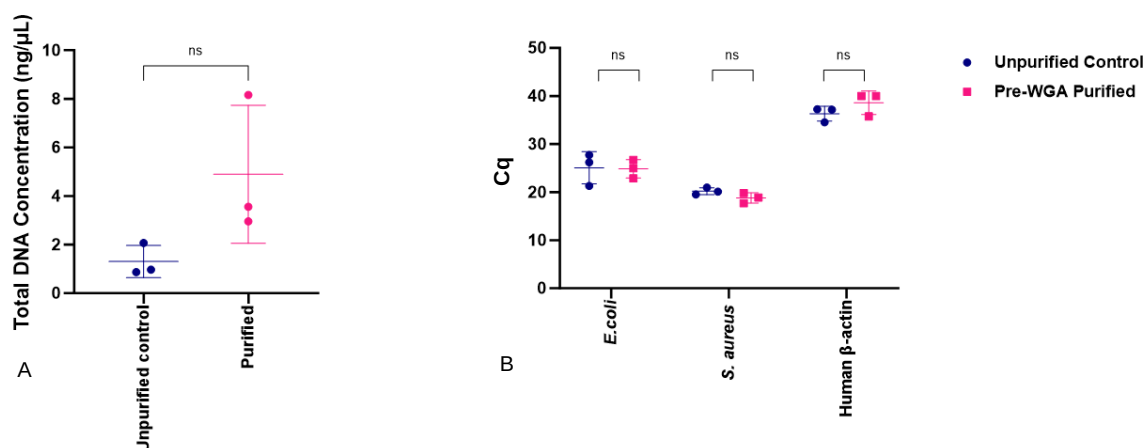


Figure 14. The concentration of purified bacteria DNA following WGA.

Total DNA concentration (A) and bacteria DNA Cq values (B) from pre-WGA purified DNA samples compared to unpurified controls. The DNA extracted from contrived spiked samples was purified prior to Genomiphi WGA and the total and species-specific DNA concentrations were quantified by Qubit dsDNA HS assay and qPCR, respectively. There was no significant increase in total DNA concentration or Cq values between purified and non-purified samples.

±SD $n=3$ ns $p>0.05$

Increasing the amplification step from 1.5 h, as recommended by the manufacturer, to 2.0 h was also evaluated. DNA samples obtained from spiked swab samples were amplified for 1.5 and 2.0 h. The total DNA concentration was determined by Qubit analysis and species-specific DNA concentrations by qPCR. The average total DNA concentration increased significantly in the 2.0 h samples compared to 1.5 h (3.22 vs 12.46 ng/μL) (Figure 15A). Average Cq values were reduced for all bacteria species, excluding *E. faecium*, although not significantly. There was also no significant difference in human genomic DNA Cq values (Figure 15B). Further experiments utilised a 2 h incubation step to increase the total DNA concentration to a level suitable for nanopore sequencing.

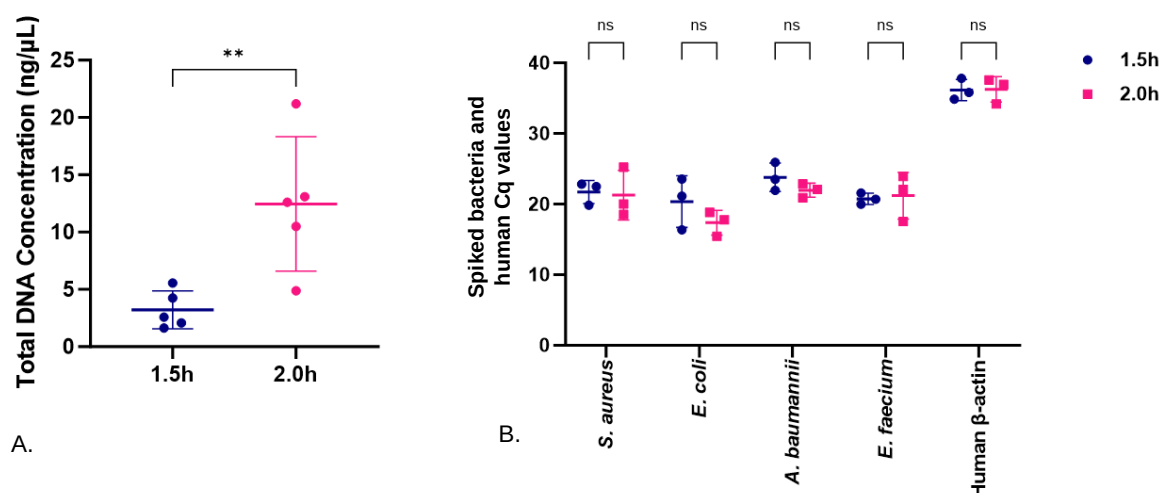


Figure 15. Evaluation of increasing the Genomiphi WGA amplification step duration.

Evaluation of the effect of increasing the Genomiphi WGA amplification step incubation time from 1.5 to 2 h on the total DNA concentration (A) and bacteria Cq values (B). DNA extracts from contrived spiked wound swab samples were amplified by Genomiphi WGA with 1.5 and 2 h incubation times and the total and species-specific DNA concentrations were quantified by Qubit dsDNA HS assay and qPCR, respectively. The total DNA concentration increased when samples were amplified for 2 h compared to 1.5 h, although there was no change in the spiked bacteria Cq values.

±SD Total DNA concentration: $n=5$ Cq values: $n=3$ ns $p>0.05$ ** $p<0.01$

2.4.9 Whole genome amplification vs non-whole genome amplification comparison

Following the optimisation of the Genomiphi WGA method, the amplification of bacteria DNA from contrived spiked samples was evaluated and compared to non-WGA samples to determine any significant differences in the proportion of bacteria reads. Amplification of the DNA by the Genomiphi WGA kit resulted in an increase in the average proportion of *E. coli* reads, compared to the non-WGA control samples, increasing from 21.43% to 51.59% of reads. Conversely, the mean proportion of *S. aureus* reads was reduced from 57.23% of the total reads in the non-WGA samples to 28.76% in the WGA samples. However, despite the increased proportion of *E. coli* reads, the average number of reads and gene coverage of the *bla*_{CTX-M-15} gene was lower in the WGA samples compared to the non-WGA sample (31.33 vs 14.67 and 16.29× vs 8.27×). Conversely, the average number of reads and gene coverage of the *mecA* gene increased in the WGA sample, despite *S. aureus* comprising a lower proportion of the total reads (17.67 vs 68.67 and 9.88× vs 33.66×) (Table 5).

Table 5. WGA compared to non-WGA sequencing results.

Sample	Total classified reads	<i>E. coli</i> reads (proportion of total classified reads)	<i>bla</i> _{CTX-M-15} reads (mean gene coverage ×)	<i>S. aureus</i> reads (proportion of total classified reads)	<i>mecA</i> reads (mean gene coverage ×)	Human reads (proportion of total classified reads)
Spiked sample non-WGA 1	6593	1283 (19.46)	8 (2.37)	3792 (57.51)	3 (1.98)	384 (5.82)
Spiked sample non-WGA 2	17831	3514 (19.71)	23 (13.8)	11232 (62.99)	13 (6.96)	440 (2.47)
Spiked sample non-WGA 3	35379	8883 (25.11)	63 (32.7)	18112 (51.19)	37 (20.69)	730 (2.06)
Spiked sample WGA 1	74396	52151 (70.10)	13 (7.54)	6778 (9.11)	108 (51.57)	1439 (1.93)
Spiked sample WGA 2	70041	15543 (22.19)	21 (11.44)	43289 (61.81)	80 (39.3)	667 (0.95)
Spiked sample WGA 3	11611	7255 (62.48)	10 (5.84)	1785 (15.37)	18 (10.1)	565 (4.87)

Contrived spiked wound swab samples were processed for CMg, including the Genomiphi WGA step and compared to samples that omitted this step. Inclusion of the WGA step increased the average total number of classified reads from 19934 to 52016. The proportion of *E. coli* reads increased from 21.43% in the non-WGA sample to 51.59% in the WGA sample. Conversely, the proportion of *S. aureus* reads reduced from 57.23% in the non-WGA samples to 28.76% in the WGA samples.

2.4.10 DNA extraction platform comparison

The extraction of microbial DNA from the host-depleted sample is of critical importance to the success of the workflow, both in terms of the identification of the causative pathogen and for the ease of use for the operator. Therefore, further work was carried out to compare two automated DNA extraction platforms and a silica-column based manual DNA extraction kit using contrived wound swab samples spiked with clinically relevant bacteria species. The identification of the spiked bacteria species and the gene coverage of the associated AMR genes and the suitability of each platform for use in deployed MTF environments was compared.

Blank wound swabs were spiked with *E. coli* and *B. thuringiensis* mixed with defibrinated *O. aries* blood at a known concentration, to generate contrived wound swab samples. Negative controls and non-depleted controls were processed concurrently. Samples then underwent the optimised host depletion method as previously described followed by one of the three DNA extraction methods. The bacteria and host DNA were quantified by qPCR before the DNA was then prepared for DNA sequencing.

The PDQeX platform required approximately 30 min to complete with just five minutes of those requiring operator hands-on steps, compared to 44 min for the MaxWell 16 LEV method, of which seven minutes required operator hands-on steps. The manual Zymo method required 32 min to complete, however, all 32 min required operator input.

Following the host depletion and DNA extraction steps, the MaxWell platform DNA extracts had the highest concentrations of *E. coli* DNA, with an average of 1917 fg/μL. This was compared to the Zymo kit average of 1334 fg/μL and the PDQeX average of 217 fg/μL (Figure 16A). For *B. thuringiensis*, the Zymo kit extracted an average of 4811 fg/μL DNA compared to the MaxWell device which averaged 4390 fg/μL DNA. The PDQeX averaged 957 fg/μL (Figure 16B).

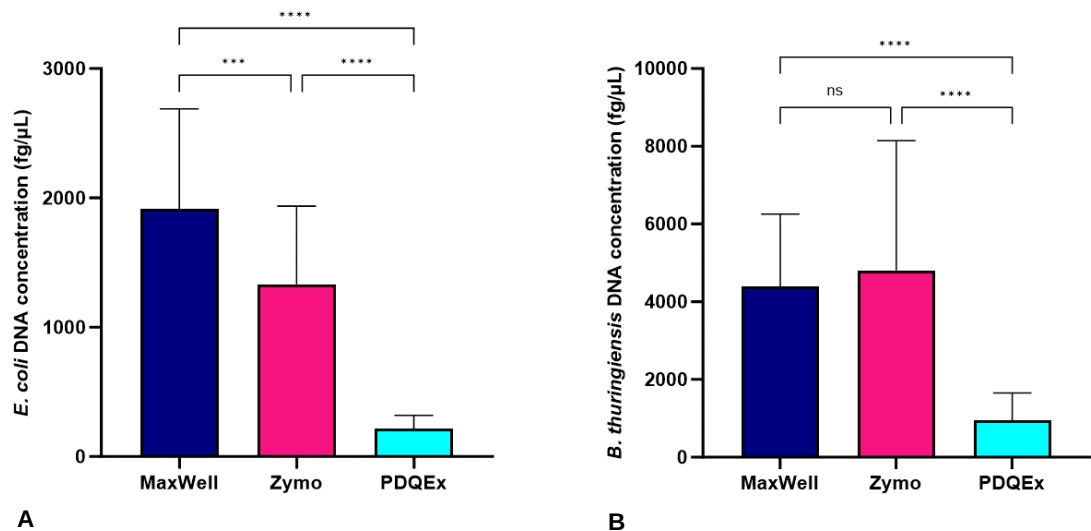


Figure 16. Extraction of bacterial DNA from contrived wound swab sample using three DNA extraction methods.

E. coli (A) and *B. thuringiensis* (B) DNA concentrations extracted by MaxWell, Zymo and PDQeX DNA extraction methods. Contrived wound swab samples were spiked with *E. coli* and *B. thuringiensis* and underwent the host depletion method before undergoing DNA extraction using three DNA extraction methods. The bacteria DNA concentrations were then determined by qPCR absolute quantification. The MaxWell platform extracted the highest concentration of *E. coli* DNA, whilst both the MaxWell and Zymo methods extracted higher concentrations of *B. thuringiensis* DNA compared to the PDQeX platform.

+SD $n=18$. ns $p>0.05$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$ **** $p<0.0001$.

The concentration of host DNA in each DNA extract was quantified by qPCR assay targeting the *O. aries prlr* gene. The MaxWell platform DNA extracts contained the highest concentration of host DNA (0.01 ng/μL) compared to the Zymo (0.00 ng/μL) and PDQeX (0.00 ng/μL) methods (Figure 17).

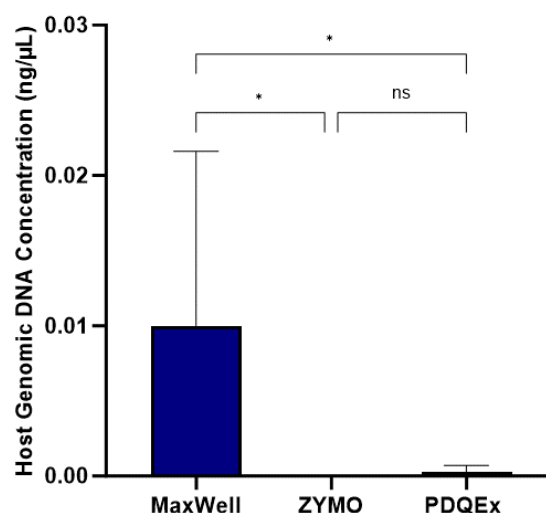


Figure 17. Concentration of host DNA extracted from contrived wound swab samples using three DNA extraction methods.

Host genomic DNA concentrations extracted by MaxWell, Zymo and PDQEx DNA extraction methods. Contrived wound swab samples were spiked with bacteria mixed in defibrinated *O. aries* blood and underwent host depletion and DNA extraction using three DNA extraction methods. The *O. aries* DNA concentrations were then determined by qPCR absolute quantification by standard curve. The MaxWell platform extracted the highest concentration of host DNA of 0.01 ng/μL compared to the Zymo and PDQEx methods at 0.00 ng/μL.

+SD $n=12$ ns $p>0.05$ * $p<0.05$.

DNA extracts were then progressed through to DNA sequencing using the optimised workflow, as previously described, to determine the proportion of sequencing reads assigned to the spiked bacteria (Figure 18). The MaxWell platform DNA extracts had the highest proportion of *E. coli* reads, averaging 13.67% of the total sequence reads, significantly higher compared to Zymo at 8.45% and higher than PDQEx at 8.12%, although not significantly. There was no significant difference in the proportion of *B. thuringiensis* reads between the MaxWell and PDQEx platforms, however, the Zymo method had the highest percentage of *B. thuringiensis* reads averaging 68.78%, higher than the MaxWell averaging 59.52% and the PDQEx averaging 58.91%. The percentage of total sequence reads assigned to the host genome was low in all depleted samples, averaging 0.78% in MaxWell samples, 0.09% for Zymo and 0.16% for PDQEx samples.

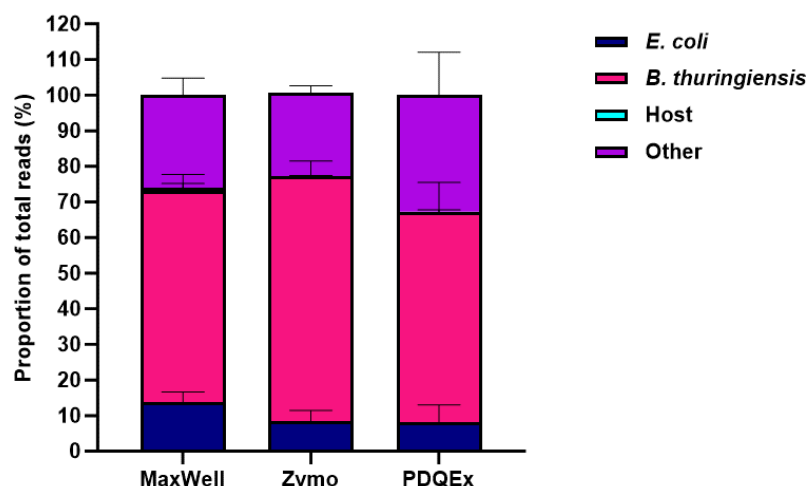


Figure 18. Proportion of spiked bacteria sequence reads from samples extracted by three DNA extraction methods.

Contrived wound swab samples were spiked with bacteria mixed in defibrinated *O. aries* blood and underwent the full CMg workflow including DNA extraction by one of three DNA extraction methods and the proportion of sequence reads assigned to the spiked bacteria were compared. The MaxWell platform DNA extracts contained the highest proportion of *E. coli* reads of 13.67% compared to the Zymo method at 8.45% and the PDQeX platform at 8.12. The Zymo method DNA extracts contained the highest percentage of *B. thuringiensis* reads at 68.78% compared to the MaxWell platform at 59.52% and the PDQeX at 58.91%.

+SD $n=9$

The DNA extraction platforms were then compared for the identification of spiked AMR genes from extracted DNA from contrived wound swab samples. The Zymo method DNA extracts contained the highest average coverage for all three AMR genes, generating 21× coverage for the *S. aureus mecA* gene, 20.8× coverage for the *E. coli bla_{CTX-M-15}* gene and 26.5× for the *E. faecium vanA* gene. This was compared to 15.52×, 12.49× and 20.13× for *MecA*, *bla_{CTX-M-15}* and *VanA* respectively for the MaxWell platform and 7.84×, 7.90× and 15.80× for the PDQeX platform, although this increase in average gene coverage was not significant (Figure 19).

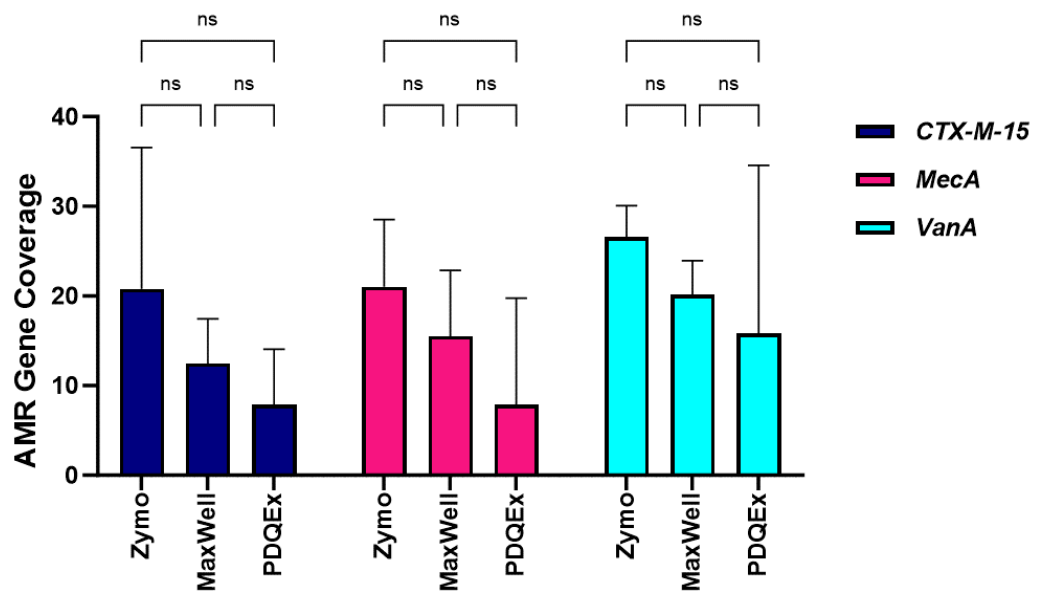


Figure 19. Comparison of the gene coverage of spiked AMR genes from three DNA extraction methods.

Three bacteria species containing known AMR genes were mixed with defibrinated *O. aries* blood and spiked onto blank swab samples at a known concentration and the swabs processed for CMg using each DNA extraction method. The Zymo method DNA extracts contained the highest AMR gene coverage for all three AMR genes, generating 21×, 20.8× and 26.5× gene coverage for the *mecA*, *bla*_{CTX-M-15} and *vanA* genes, respectively.

+SD. *n*=3. ns *p*>0.05.

2.4.11 Full optimised workflow testing

Following the workflow optimisation work using contrived samples, the full workflow was carried out in a final proof of concept experiment. For the depletion comparison, blank wound swabs were spiked with *E. coli* or *S. aureus* mixed with whole human blood at 5.0×10^4 CFU per swab. To determine the detection of spiked bacteria at lower concentrations, overnight cultures were serially diluted and mixed with whole human blood and spiked onto blank wound swabs at concentrations of 5.0×10^4 to 500 CFU per swab. The spiked samples then underwent the optimised host depletion and DNA extraction workflow, including the post-host depletion vortexing and increased bead-beating duration steps. The samples were then prepared for sequencing using the Genomiphi V3 WGA kit and the Rapid Barcoding library preparation kit and the prepared DNA was sequenced on an MK1c nanopore sequencing device to determine the proportion of reads assigned to the spiked bacteria and human genome.

When *E. coli* was spiked alongside human blood on a charcoal swab, the saponin-based host depletion protocol reduced the mean proportion of human DNA from 99.63% to 0.59% of total classified reads and increased the proportion of *E. coli* reads from 0.09% to 97.67% in the resulting sample metagenomes (Figure 20A). In contrived wound swabs

with *S. aureus*, host depletion reduced the mean proportion of human DNA from 99.73% to 2.87% and increased the mean proportion of *S. aureus* reads from 0.03% to 96.57% (Figure 20C). In samples containing serial dilutions of *E. coli* from 5.0×10^4 CFU to 500 CFU per swab, the mean proportion of *E. coli* reads remained high, decreasing from 98.41% to 91.27% (Figure 20B). The mean proportion of *S. aureus* reads reduced from 97.56% in the 5.0×10^4 CFU input to 40.77% in the 500 CFU input samples (Figure 20D).

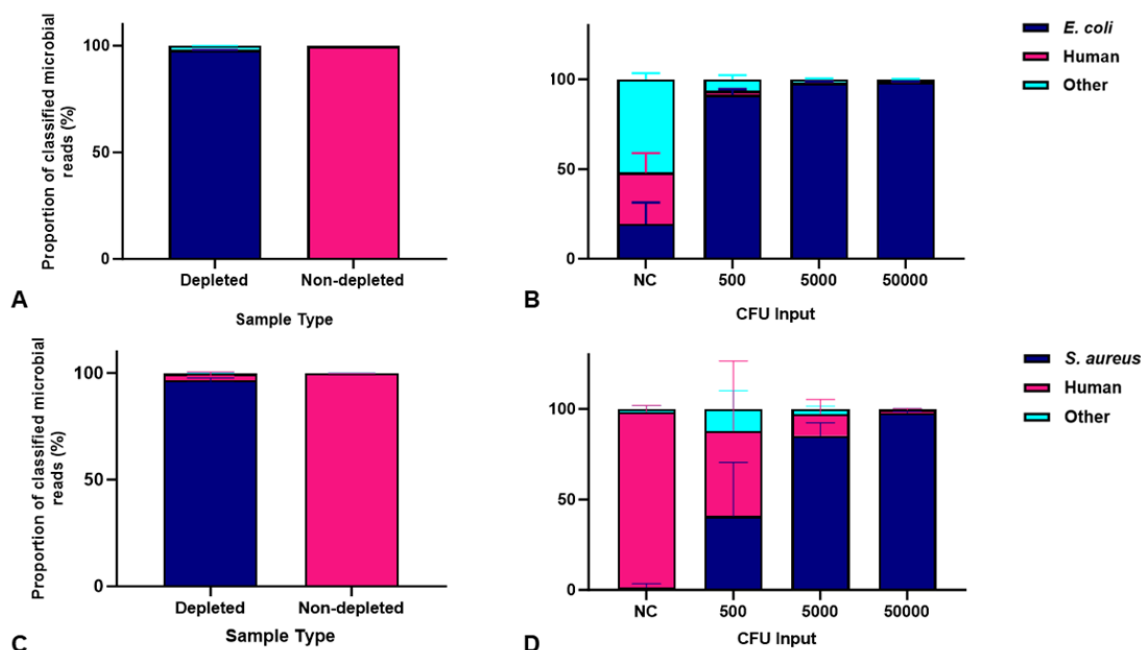


Figure 20. Proportion of sequence reads assigned to spiked bacteria and the human genome from contrived wound swab samples.

Blank wound swabs were spiked with *E. coli* (A) and *S. aureus* (C) and whole human blood and underwent the full CMg workflow. The proportion of total classified sequence reads assigned to the spiked bacteria and the human genome were compared to non-depleted control samples. Contrived samples were spiked with serial dilutions of *E. coli* (B) and *S. aureus* (D) (500 - 5.0×10^4 CFU) and whole human blood and processed for CMg and the proportions of the total sequence reads were compared. The saponin-based host depletion protocol reduced the mean proportion of human DNA from 99.63% to 0.59% of total classified reads and increased the proportion of *E. coli* reads from 0.09% to 97.67% (A). The host depletion reduced the proportion of human DNA from 99.73% to 2.87% and increased the proportion of *S. aureus* reads from 0.03% to 96.57% (C). Samples spiked with 5.0×10^4 CFU to 500 CFU of *E. coli* per swab contained 98.41% to 91.27% *E. coli* reads (B). The proportion of *S. aureus* reads reduced from 97.56% in the 5.0×10^4 CFU input to 40.77% in the 500 CFU input samples (D).

+SD NC= Negative control $n=3$.

The identification of AMR genes in the sample metagenomes was also compared. Host depletion of samples containing *E. coli* resulted in a significant increase in *bla*_{CTX-M} coverage depth to 34.61× in depleted samples from 0× in the non-depleted samples (Figure 21A). There was also a significant increase in coverage of the *mecA* gene to 1.84× in the depleted sample from 0× in the non-depleted samples (Figure 21C). From the

serial dilutions, *bla*_{CTX-M} was identified at all spiked concentrations, ranging from 9.76× coverage in the 500 CFU input to 82.5 in the 5.0×10^4 CFU input (Figure 21B). In samples containing *S. aureus*, *mecA* was identified at the highest CFU input (5.0×10^4 CFU) at 1.78× depth of coverage but was not identified at any other input (Figure 21D).

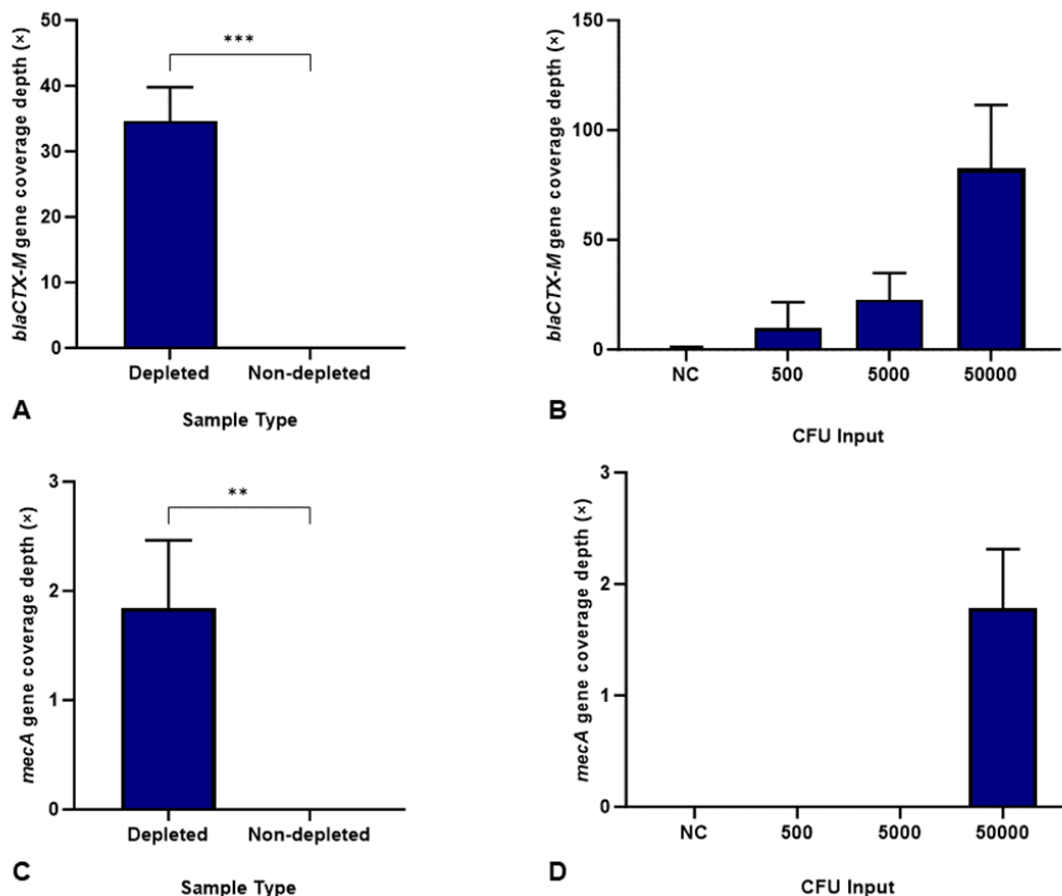


Figure 21. Identification of AMR genes from contrived wound swab samples.

The depth of AMR gene coverage from *E. coli* (A) and *S. aureus* (C) spiked samples was compared to non-depleted control samples and from contrived samples spiked with serial dilutions of *E. coli* (B) and *S. aureus* (D). Host depletion of samples spiked with *E. coli* resulted in increased in *bla*_{CTX-M} coverage depth from from 0× in the non-depleted samples to 34.61× in depleted samples (A). Coverage of the *mecA* gene increased to 1.84× in the depleted samples from 0× in the non-depleted samples (C). *bla*_{CTX-M} was identified at all spiked serial dilutions, ranging from 9.76× coverage in the 500 CFU input to 82.5× in the 5.0×10^4 CFU input (B). *mecA* was identified at the highest CFU input at 1.78× depth of coverage but was not identified at any other dilution (D).

+ SD. NC= Negative control. *n*=3. ***p* < 0.01. ****p* < 0.001

2.5 Discussion

The overall aim of this thesis was the development of CMg workflows for sample types relevant to deployed military personnel and to optimise the methods towards enabling their routine utilisation in deployed MTFs to improve the treatment of infectious disease.

Therefore, the key focus of the work described in this chapter was to develop a CMg workflow, including depletion of the host genomic DNA, extraction of the microbial DNA, WGA, DNA sequencing library preparation and nanopore based DNA sequencing steps, suitable for wound swab samples obtained from patients with suspected wound infections. Using contrived wound swab samples consisting of blank swabs spiked with host cellular material and bacterial pathogens, optimisation of the methods focused on reducing the hands-on steps, the total time taken to reach the sequencing step and equipment requirements. The quantification of host and pathogen genetic material by qubit, qPCR and comparison of the reads assigned to the spiked bacteria species was applied to identify the optimal methods. Ultimately, an optimised workflow was established in preparation for the collection and processing of patient wound swab samples as part of a proof of concept study.

2.5.1 Host depletion

Prior to the collection of wound swab samples from patients, the development of a robust set of methods to minimise the risk of method failure was required. Importantly, the ethical approval for the collection and processing of wound swab samples for CMg as part of a proof of concept study was contingent on previous testing having been carried out to provide evidence of the suitability of the sample type for CMg, to reduce the risk of study participant samples being wasted. The addition of host cellular material was achieved through the use of whole human and defibrinated *O. aries* blood, with defibrinated *O. aries* blood used in instances where access to the suitable containment level two laboratory was not available due to extended building shutdowns and COVID-19 lockdowns. The host cellular material was spiked with known concentrations of key bacterial pathogen species and the mixture added directly onto blank wound swabs before being added to the charcoal matrix to generate replicable contrived samples for method development.

Preparation of the model sample type demonstrated that the application of the optimised saponin-based host depletion method resulted in the efficient depletion of human DNA and significantly increased the proportion of spiked bacterial sequence reads. Comparison of host depleted samples to non-depleted controls showed that the depletion reduced the proportion of host DNA from >99% down to <3% of total reads. Reduction of host DNA sequences from the sequence data is key to improving the sensitivity of pathogen detection by increasing the proportion of reads assigned to the microbial species present, rather than human genetic material (143). Whilst the effectiveness of a saponin-based host DNA depletion has been demonstrated before (62), these results provided evidence that the host depletion methods were compatible with the wound swab sample type, including the charcoal agar containing known inhibitors of downstream processing of DNA samples (144). This work enabled the ethical approval for the evaluation of CMg for the diagnosis of wound infections using samples collected from NHS patients to be obtained.

2.5.2 Nucleic acid extraction

Once it had been demonstrated that the initial host depletion step was compatible with the sample type, further work was carried out to optimise the extraction of microbial DNA from the depleted sample. This included increasing the bead beating step and evaluating pre and post-depletion vortex steps. The importance of bead-beating for the extraction of Gram positive bacteria is well documented (145) due to the increased resistance of these bacteria to mechanical lysis. However, somewhat surprisingly, in this study, increasing the duration of the bead beating step did not result in a significant increase in the DNA yield of the representative Gram positive organism *S. aureus*. It may be that the intensity and duration of bead beating was already sufficient for the lysis of *S. aureus* cell walls and an increased duration was not required. However, this study did not include the evaluation of the bead beating for fungal species, which have cell walls that are tough and difficult to lyse and require a bead beating step (146). Therefore, although the improved DNA extraction yield was not observed for Gram positive bacteria, increased bead beating was included for further work. It may be necessary for future optimisation work to evaluate the methods for a larger range of microbial species, including fungal species, to ensure that nucleic acids are effectively extracted from all microbial species.

The selection of a suitable nucleic acid extraction method can have a significant impact on nucleic acid yield and quality and therefore, the success of the workflow (147, 148). Comparisons of nucleic acid extraction methods for wound sample metagenomics studies are rare, particularly so for workflows that utilise nanopore based sequencing with the express aim of being applied in austere environments. The comparison carried out in this chapter demonstrated that the automated PDQeX platform generated significantly lower bacteria DNA concentrations compared to the MaxWell and Zymo methods. This was reflected by the reduced AMR gene coverage obtained from the PDQeX samples sequencing data. However, the PDQeX platform had the smallest footprint, weight, fewest hands-on steps and the lowest time to completion, making it the most suitable for use in austere environments. This agrees with previous evaluations of the PDQeX platform that found that, although suitable for use in resource-limited settings, the platform yielded reduced DNA concentrations and eluted proteins that would inhibit downstream processing (149). Therefore, to ensure that sufficient sensitivity of pathogen and AMR gene detection was achieved whilst reducing the method complexity, the MaxWell platform was selected for future work and used for the processing of patient wound swab samples in the evaluation of CMg for wound swab samples, carried out following this work. These results highlighted the potential trade-off between the ease of use and DNA yield when working in deployed MTFs. As the field of microfluidics progresses and more automated nucleic acid solutions are developed, these should continue to be monitored

for the development of novel methods that increase the DNA yield whilst maintaining ease of use suitable for austere environments.

2.5.3 Whole genome amplification kit comparison

The inclusion of the WGA step was based on previous examples of CMg pipelines in which the sample type did not yield sufficient DNA to directly progress to the library preparation step required for nanopore based sequencing, that requires higher DNA input (400 ng) compared to Illumina SBS sequencing methods (500 pg) (150). The original workflow employed the Repli-G WGA kit, previously optimised for low DNA input (63). This method generated significant increases in total DNA concentration from low biomass samples, hence its previous application for sample types such as host depleted blood samples. However, when the full workflow was first carried out as a proof of concept experiment, *E. coli* sequence reads were over represented compared to the other spiked species.

It was hypothesised that the cause of this phenomenon was two-fold. Firstly, that *E. coli* was a contaminating species in the laboratory reagents, leading to its amplification during the WGA step compared to other species. *E. coli* is a known contaminant of molecular biology reagents, including the reagents used for the generation of metagenomic sequencing data (151) and specifically the WGA step has been shown to amplify contaminants, particularly in low biomass samples (152). This was further corroborated by the presence of a significant proportion of *E. coli* reads present in the negative control samples. Therefore, it may be necessary to incorporate steps for the reduction of contamination into the workflow and for the identification of possible contaminants in the sequencing results, to reduce the impact on the identification of the causative pathogen of infection. In addition, it was hypothesised that *E. coli* DNA was preferentially amplified by the WGA step compared to the DNA from other spiked bacterial species. Factors such as GC content and DNA conformation of the genome can lead to the preferential amplification of DNA sequences compared to others (115). The high proportion of *E. coli* reads in the sequence data may have been as a result of one of these factors.

Importantly, whilst the bacteria DNA was shown to be amplified by the Repli-G WGA step, the human DNA, following the depletion step, was not amplified and was lower in concentration post-WGA. The absence of amplification of human genomic DNA by WGA in the depleted samples may have been due to the saponin and DNase digestion step causing fragmentation of the human genomic DNA so that it could not be effectively amplified. The degree of DNA fragmentation impacts amplification by WGA (153). The efficient depletion of human genomic DNA and preferential amplification of non-fragmented bacteria DNA is an additional benefit to the use of a saponin-based host depletion step and will increase the sensitivity of bacteria detection.

As a result of the potential for the introduction of bias by the Repli-G WGA kit, an alternative WGA kit was evaluated. The Genomiphi WGA kit has several advantages over alternative methods, particularly for use in austere environments. Firstly, the reagents require ambient storage temperatures, reducing the requirement for cold chain transport. In addition, the number of hands-on steps is reduced compared to the Repli-G kit, reducing the burden on the operator. However, studies that had compared a previous version of the Genomiphi kit, Genomiphi V2, against the Repli-G kit for low biomass samples found that the Repli-G WGA kit was superior for allele recovery (154). Another study comparing these kits for infectious disease diagnosis found that the Repli-G kit provided the highest proportion of pathogen reads and the Genomiphi V2 kit identified the highest amount of background contaminant DNA (155). When compared to the Repli-G kit for the amplification of purified bacteria DNA in this study, the Genomiphi kit resulted in reduced total DNA yield from spiked samples, but, in contrast to previous studies, produced lower concentrations of contaminant DNA in non-template control samples. The Genomiphi kit reduced the Cq values for the spiked bacteria DNA in all species excluding *E. coli*, indicating improved amplification of spiked bacteria DNA. Therefore, the Genomiphi kit was selected for further evaluation.

A limitation of the evaluation of the automated DNA extraction platforms and WGA kits was the use of qPCR for the quantification of bacterial and human DNA. The qPCR assays targeted a single gene per species and therefore, the quantification of each species may have been impacted by factors such as the efficiency of the particular qPCR assay, rather than differences in the recovery of DNA by the extraction method or the amplification of the DNA by WGA. In addition, there may have been differences in the recovery and amplification of the DNA across the genome of each species, which would not have been reflected by the use of a single qPCR assay per species. Future work should consider the use of multiple qPCR assays per species to reduce the impact of these variables.

2.5.4 Future work and considerations for testing clinical samples

Although the generation of contrived wound swab samples did enable the comparison of methods to optimise each step of the CMg workflow, there were several limitations of this approach. Wound swab samples present several challenges for the application of CMg. The composition, in terms of the ratio of human and microbial DNA, of patient wound swab samples varies from wound to wound. Wound swab samples can be collected from a wide range of wound types, wound locations and wounds at different stages of infection ranging from contaminated wounds through to spreading infections and pre and post antimicrobial treatment. For example, although experiments carried out in this chapter demonstrated that use of the Genomiphi WGA kit was optimal compared to Repli-G using contrived samples, it was not known whether WGA would be required for real patient

samples. Alternatively, and more complex, it may be that some samples contain sufficient DNA to warrant omission of the WGA step whilst others require amplification. Future work processing patient samples for CMg in a proof of concept study was required to determine whether the methods developed would be suitable and highlight further opportunities for method optimisation.

Further to this, a wide range of microbial species have been obtained from wound swab samples, including bacterial and fungal species. However, only bacterial species were included in the method development experiments in this study. Future work should focus on including fungal pathogens in the method development process to ensure that they are identified by CMg analysis. In addition, although seemingly rarer compared to bacterial and fungal wound infections, viral pathogens have also been implicated in wound infections (156). The methods developed in this study have not included steps to enable the identification of viral species that have RNA based genomes and therefore, these species will not be identified. In addition, whilst DNA based viruses may be identified, it is unlikely, due to the comparatively small genome of viral species, that these will be detected with the sensitivity required in a clinical context without the addition of viral enrichment steps, of which there are several options including sequence-independent, single-primer-amplification (157) and hybridisation probes (158). A key focus of future work should be to develop an all-encompassing workflow that can identify all possible pathogen species, including RNA and DNA viral species. This may involve the development of a separate viral pathogen arm to the workflow, to allow for the preparation and enrichment of viral nucleic acids before pooling the sample for sequencing.

Alongside the human and pathogen genetic material, patient wound swab samples will also have commensal flora species present. While the use of a defined contrived sample type did allow for the optimisation of methods, the effect of the commensal flora on the detection of pathogen species was not quantified. The impact of this could have been further quantified through the generation of a mock wound microbiome, consisting of a mixture of known wound commensal species at known concentrations spiked into the human cellular matrix alongside pathogen species. Previous studies have used reagents containing a larger number of defined pathogen species at a known concentration to replicate human microbiomes to improve metagenomics method development (159). Alternatively, collection of patient wound swab samples and spiking known pathogens onto these samples would also enable the impact of the presence of commensal species to be quantified and to provide a more accurate measurement of pathogen detection sensitivity.

The experiments carried out in this chapter did demonstrate that the proportions of species DNA in the sequencing data was altered between the spiking of the bacteria and DNA sequencing. It was hypothesised that the inclusion of a WGA step was resulting in

amplification bias and therefore, an alternative WGA kit was selected for use. However, despite this, the results from the final full workflow run through still demonstrated reduced proportions of sequencing reads assigned to *S. aureus* compared to *E. coli* at the lowest CFU input and lower reads assigned to the *mecA* AMR gene compared to *bla*_{CTX-M-15}. This may be due to inherent bias introduced during the CMg workflow, including the DNA extraction and WGA steps (153, 160). The reduced depth of coverage of the *mecA* gene may also be attributed to the chromosomal location of the staphylococcal cassette chromosome containing the *mecA* gene (161) compared to the *bla*_{CTX-M-15} gene, located on a plasmid (162), potentially impacting the number of AMR gene copies added at each CFU input. Future work should aim to understand further the impact of the reduced proportion of reads assigned to Gram positive species on the sensitivity of detection and if required, identify the cause and optimise the methods to reduce this bias. Further work to understand fully the cause of this affect was not carried out and therefore, cannot be accounted for in the sequencing data obtained from real patient samples. The sensitivity and specificity of *E. coli* and *S. aureus* detection should be compared from patient samples to determine whether the alterations in sensitivity observed in spiked samples are also observed in wound swab samples.

2.6 Conclusions

The work carried out in this chapter developed and evaluated a CMg workflow for wound swab samples and demonstrated the effective depletion of host DNA and the detection of spiked bacteria species and the associated AMR genes. This work provided the basis for the planning of a prospective wound swab sample CMg study and enabled the project to gain the required ethical approval to collect samples from patients with suspected wound infections. This was integral to the progression of CMg methods for wound swab samples and enabled the future work to identify opportunities for method optimisation. This will ultimately contribute to the further understanding of the clinical interpretation of CMg results to improve the treatment of deployed personnel with wound infections.

2.6.1 Recommended wound clinical metagenomics workflow

The work carried out in this chapter resulted in the establishment of an optimised wound swab CMg workflow. The recommended workflow consisted of the following steps:

- Sample storage: unless immediate sample processing is available, the wound swab sample should be refrigerated at 4-8°C to minimise alterations to the composition of the microbial species present on the swab. The sample should not be frozen to avoid lysing the microbial cells and the subsequent loss of genetic material during the saponin-based host depletion step.
- Host depletion: the swab sample is added to PBS and undergoes a saponin-based host DNA depletion step. Saponin is added to a final concentration of 2.5%

alongside the HLSAN DNase and HLSAN buffer and incubated for 10 minutes. The swab is then vortexed post-depletion to ensure that the microbial cellular material is removed from the the swab matrix for downstream processing without lysing the microbial cells prior to the addition of the DNase enzyme, reducing the loss of microbial genetic material.

- DNA extraction: the sample is then bead-beaten for 60 seconds at 6.0 m/s to ensure the lysis of microbial cells, including Gram positive bacteria and fungal species. The sample then undergoes a further chemical lysis step and DNA extraction using the automated MaxWell 16 LEV blood DNA method. This platform was shown to combine the sufficient recovery of microbial DNA with reduced user hands-on time compared to the manual extraction method. As new platforms are developed, these should be evaluated for both the recovery of microbial DNA and their ease-of-use in austere environments.
- WGA: the extracted DNA is then amplified using the Genomiphi V3 WGA kit. This kit uses lyophilised reagents, reducing the need for cold-chain storage and results in the more uniform amplification of microbial species DNA, resulting in sequencing results that better reflect the sample input.
- Library preparation: the amplified DNA is prepared for nanopore-based sequencing using the Rapid Barcoding library preparation kit. This method uses minimal hands-on steps and reduces the total turnaround time compared to alternative nanopore library preparation methods.
- DNA sequencing: the prepared DNA is sequenced using nanopore-based sequencing on an MK1c sequencing device for up to eight hours. Although not utilised in this study, nanopore sequencing enables real-time data analysis, potentially reducing the time required to provide a result. In addition, the equipment is small and portable, making nanopore sequencing suitable for use in deployed MTFs.

3 Evaluation of a clinical metagenomics workflow for the diagnosis of wound infections.

3.1 Contribution statement

The work detailed in this chapter was carried out as a collaborative study with partners from the NHS and the Quadram Institute Bioscience, which has been submitted to a peer-reviewed journal for publication, for which Carl Halford was the first author. Wound swab samples were collected by Dr. Paul Russell, Fiona Trim (Salisbury District Hospital (SDH)) and Dr. Nathan Moore (Basingstoke North Hospital (BNH)). Routine microbiology was carried out by the NHS microbiology laboratories at each NHS site. Dr. Paul Russell provided the methodology for the routine microbiology investigations and the identification of reportable wound pathogens from the CMg results. The bioinformatics analysis was carried out by Dr. Thanh Le-Viet (Quadram Institute Bioscience) in close collaboration with Carl Halford using a bioinformatics pipeline that Dr Thanh Le-Viet previously developed. The MODREC and Health Research Authority (HRA) study protocol designs, the wet-lab work for the processing of the wound swab samples for CMg and the results interpretation were carried out by Carl Halford.

3.2 Introduction

3.2.1 Wound infections

Wounds are defined as injuries that break the integrity of the skin barrier such as burns, lacerations, punctures, blast injuries and surgical sites. The skin is a barrier that prevents microbes from coming into contact with the underlying tissues, providing protection against infection. A break in the skin barrier enables microbes to colonise the wound site, which provides a suitable environment for microbial growth (163). In some cases, the colonising microorganisms become pathogenic, a process dependent upon the location and severity of the wound, the microbial burden, the species and virulence of the microbes present on the wound and the host immune response (164).

The development of a wound infection has been defined into five distinct stages based on the presence and proliferation of microbes and the associated clinical symptoms, by the International Wound Infection Institute, known as the wound infection continuum (163). The majority of wounds are contaminated by microbes as a matter of course. The contaminating microbes can originate from a range of sources including the body's own flora from sites such as mucosal barriers of the oral cavity, the gut and the skin. Environmental contaminants introduced during the injury contaminate wounds, particularly if foreign material, such as soil, is deposited onto the wound at the point of injury. Nosocomial transmission of microbes from healthcare facilities has also been identified as a route of microbial contamination of wounds, including from healthcare equipment and

personnel (165). Contaminating microorganisms are non-proliferating and at this stage and do not evoke a host immune response. However, wounds sites can provide a suitable environment for microbial proliferation. This stage of the continuum is defined as colonisation. This does not illicit a host response and the healing of the wound is not impaired. However, wound colonisation can progress to local infection, where the microbes proliferate at a rate that does cause a host reaction. Clinical symptoms are present, including delayed wound healing and increased pain but are limited to the wound site or structure. This can then progress to spreading and then systemic infections. In these cases, the microbes present on the wound reach critical colonisation and invade the surrounding tissue. The microbes spread via the vascular or lymphatic systems resulting in an immune response, inflammation and tissue damage. This can have serious consequences for the patient, including systemic inflammatory response and sepsis.

Some of the known host risk factors associated with the development of a wound infection include but are not limited to poorly managed diabetes, conditions associated with hypoxia and immune system disorders. Risk factors associated with the wound itself include the presence of soil or plant matter and other foreign materials, necrotic wound tissue, the involvement of deep tissue and the injury being classed as traumatic (163). Hospitalisation can also lead to nosocomial wound infections, with wound infections being the most common of all nosocomial infection types and infection by AMR isolates being of particular concern (166, 167).

3.2.2 Impact of wound infections

Wound infections can have a significant and negative impact on the patients and the wider healthcare systems. These outcomes are dependent upon the type of wound, wound location and severity, the comorbidities of the patient and the access to an appropriate healthcare setting and treatments. In any setting, the development of a wound infection is a serious issue in the patient's recovery and is associated with delayed wound healing (168) and increased mortality (169). Mortality commonly results from secondary sepsis, as the pathogens colonise the wound and then invade the surrounding tissue via the bloodstream. Fifty four percent of burn-related deaths are due to septic shock and subsequent organ dysfunction following invasive infection (170). The incidence of post-operative surgical site infections, despite the wound being clean at the time of the surgery, has been shown to be as high as 31% (171) and can incur significantly increased costs, mainly due to the increased hospital stay resulting from the infection (172).

Due to the risks of incurring an injury, such as burns, lacerations, punctures, blast injuries and surgical sites, associated with military operations, military personnel are at risk of suffering from a wound infection. In the first 6.5 years of Operation Iraqi Freedom, over 30000 US military personnel were wounded in action (173). Recent conflicts have been

characterised by the use of improvised explosive devices that result in complex tissue damage, primarily of the limbs (174, 175). Improvements in body armour and the immediate care of personnel who have suffered a traumatic injury have led to reduced mortality in the immediate aftermath of the event (176). However, the significant and complex injuries resulting from explosive blasts and the microbial contamination from the environment mean patients are particularly susceptible to invasive bacterial and fungal infections (177). Therefore, the primary complication resulting from a combat-related extremity wound is infection. The risk of infection differs dependent upon the wound type, 43% of US personnel that suffered a combat-related extremity wound resulting in lower extremity amputation developed a wound infection, rising to 68% of those requiring above-knee amputations (178).

3.2.3 Wound infection aetiology

Infection represents the greatest risk to life in the combat casualty that survives the initial two hours post-injury, and wound infections are the most common cause for military hospital infectious disease consultations (179). Infection rates in the Iraq and Afghanistan conflicts among injured personnel ranged from 28-34% (177) and were caused by a range of pathogenic species, including *Acinetobacter* spp., *E. coli*, *Pseudomonas* spp. and *Enterococcus* spp. as the predominant organisms (180).

The causative pathogen of wound infection depends on a range of factors. Seasonality has been shown to impact on the cause of wound infections. Analysis of the data collected as part of a retrospective observational study of infectious complications among military personnel wounded during deployment in Afghanistan found that Gram negative species, including multi-drug resistant isolates, were more commonly isolated from wound cultures in the summer and autumn seasons compared to winter and spring. *E. coli* and *Enterobacter* spp. were the most commonly identified Gram negative species whilst infecting Gram positive species did not vary by season (181).

In addition to seasonality, the location of injury was associated with differences in the causative pathogen of infection. The exposure of combat-related extremity wounds to soil and debris material in Afghanistan's southern region, characterised by warmer temperatures and lower elevation, resulted in the increased incidence of invasive fungal wound infections, associated with increased morbidity and mortality. Such was the proportion of wounded personnel developing an invasive fungal wound infection during operations in Afghanistan, the US Department of Defence subsequently conducted an outbreak investigation (182). From the Ukraine-Russia war, *Pseudomonas aeruginosa* (22.5%), *Proteus mirabilis* (21.3%), *S. aureus* (23.8%), *Streptococcus pyogenes* (18.8%), and *E. coli* (13.8%) are most commonly identified (183).

The pathogens identified at a wound site are also related to the type of wound. Organisms commonly identified at burn wounds include *S. aureus*, β -haemolytic *streptococci*, *pseudomonads*, as well as fungal species including *Candida albicans* (184-186). While bite wounds are likely to be infected by species such as *Pasteurella multocida*, *S. aureus*, α -haemolytic *Streptococci* and the *Streptococcus angiosus* group (187). In civilian healthcare environments, *P. aeruginosa* and *S. aureus* are commonly identified in community and hospital-acquired wound infections, with elderly patients receiving wound care at home or in long-term care facilities particularly susceptible to infection by *S. pyogenes* (188-191).

It is not only the causative species that has an impact on the severity of a wound infection. The presence of AMR is also vitally important to treating the infection. Infections caused by pathogens with resistance to antimicrobials, such as MRSA, are associated with increased mortality and healthcare costs (192). AMR is found in a number of species commonly identified in wound infection samples. Eighty eight percent of isolates of the bacteria *A. baumannii* obtained from wound infection samples are resistant to trimethoprim and 80% of *P. mirabilis* isolates are resistant to piperacillin (188). Therefore, determination of the resistance profile of an infection is key to enabling the targeting of antimicrobial therapies to prevent the administering of ineffective treatments. Treatment is further complicated by the development of biofilms, composed of bacterial communities situated within a polysaccharide matrix. Biofilms are more resistant to disinfection, the action of the host immune system, antibiotics and more likely to result in a persistent wound infection (193, 194).

Therefore, the ability to determine whether a wound is infected, identify the causative pathogen and the AMR profile so that the optimal treatment can be provided is key to reducing mortality. The availability of rapid diagnostic methods to provide this information in deployed environments and in hospital diagnostic laboratories are of vital importance.

3.2.4 Wound infection diagnosis

3.2.4.1 Current wound infection diagnostic tests

The identification of a suspected wound infection is based upon the observation of clinical signs and symptoms, with diagnostic tests unlikely to be ordered until a spreading infection is suspected (163). Indicators of infection include hypergranulation, increased exudate and delayed wound healing. If the infection continues to progress to a local infection, this is signified by erythema, increased temperature, swelling, purulent discharge, increasing pain and an increase in wound size. In the absence of a sufficient immune response and effective treatment, this progresses to a spreading infection associated with swelling of the lymph glands, crepitus, breakdown of the wound site and spreading inflammation away from the wound site. Further progression to a systemic

infection will result in lethargy, general patient deterioration, fever, severe sepsis and septic shock with associated organ failure. The observation of these symptoms as well as evaluation of the patient's risk factors will be incorporated to make a clinical decision as to whether a wound has become infected (163).

Clinical guidance may vary between medical facilities, however, NHS guidance states that diagnostic tests are not ordered until a spreading infection is suspected. Once requested, diagnostic tests include haematological, radiological and microbiological investigations. Haematological investigations include measuring for increased white blood cell count or C-reactive protein (CRP) and erythrocyte sedimentation rate. Radiological investigations may be employed to identify the presence of an abscess associated with a wound infection, including an X-ray or magnetic resonance imaging (163). These methods indicate the presence of an infection, but do not provide information on the causative pathogen or inform on suitable treatment methods.

Microbiological testing, based on the isolation and growth of the microorganisms present on the wound site, is used to confirm the diagnosis of a wound infection, identify the causative pathogen and the optimal treatment option based on the AMR of the isolated microbes. The sample types collected for wound culture investigations include deep-tissue biopsy, wound aspirate and wound swabs. Tissue biopsy refers to obtaining a tissue sample from the wound site and a quantifying the number of microorganisms per Gram of tissue by microscopic evaluation. Biopsies are known to yield more optimal results, particularly for the accurate identification of the invading pathogens and associated antibiotic resistance and are considered the gold-standard method (195). However, due to the invasive, painful nature of the sample collection and specialised equipment requirements, tissue biopsies are rarely collected. Wound aspirates involve the collection of the fluid within an abscess or a puncture wound to determine the concentration of microorganisms within the abscess. Similarly to the collection of a tissue biopsy, this method is relatively invasive and painful and therefore, is not often collected.

In contrast, the quick, simple and non-invasive nature of wound swab collection means that this sample type is the most frequently collected (196). This method relies on the collection of microorganisms onto a swab which is then added to a growth media to grow the microorganisms for subsequent identification. The wound is first cleansed with a saline solution prior to swab collection to remove colonising organisms from the surface of the wound. The swab is pressed onto and moved across the surface of the wound to collect microbes onto the swab. The culture media selected will depend on the features of the wound, the suspected pathogen and the clinical circumstances of the suspected infection. Figure 22 details the culture media, conditions and the target organisms for the microbiological investigation of wound swabs, as detailed by the UK Standards for Microbiology Investigations (UK SMI): Investigation of swabs from skin and superficial soft

tissue infections (184). All specimens are likely to be cultured on blood agar at 35-37°C in a 5-10% CO₂ atmosphere for 40-48 hours, as well as Cystine–lactose–electrolyte-deficient (CLED)/MacConkey agar at 35-37°C in air for 18-24 hours. While more specific clinical circumstances may result in the use of specific growth conditions. For example, wounds resulting from human bites are also cultured on chocolate agar at 35-37°C at 5-10% atmospheric CO₂ for 40-48 hours to enable the growth of fastidious organisms. Therefore, the downstream analysis of isolated cultures will vary depending on the clinical features of the wound. For example, the suspected isolation of *S. aureus* from blood agar is indicated by colonial appearance, appearing as glistening, smooth, raised and translucent colonies with a golden pigment (197). *S. aureus* identification is confirmed using a number of different tests, the use of which will depend on the test availability at the laboratory location. Matrix-assisted laser desorption/ionisation - time of flight mass spectrometry (MALDI-TOF-MS) is an increasingly commonly utilised method in the UK that analyses the protein composition of an isolated colony enabling rapid species level identification using relatively simple preparation methods (198). PCR tests can also be employed. This method relies upon the amplification and detection of species-specific or gene -specific DNA sequences. Therefore, the presence of genes associated with resistance to antimicrobials can also be identified and commercial tests are available for this (199).

There are a range of additional tests that may be used to identify an isolated colony. If required, methods including WGS, pulsed field gel electrophoresis and multiple-locus variable number tandem repeat analysis can be applied to identify the bacteria down to the strain level. Depending on the clinical features of the wound and the microbial species isolated, AST is carried out on an isolated colony. In the UK, the antimicrobial test panels will depend on the species and the methods based upon the European Committee on Antimicrobial testing (EUCAST) recommendations. The minimum inhibitory concentration of the isolated species will be determined by disk diffusion and this will be used to inform the antimicrobial treatment of the infection (36).

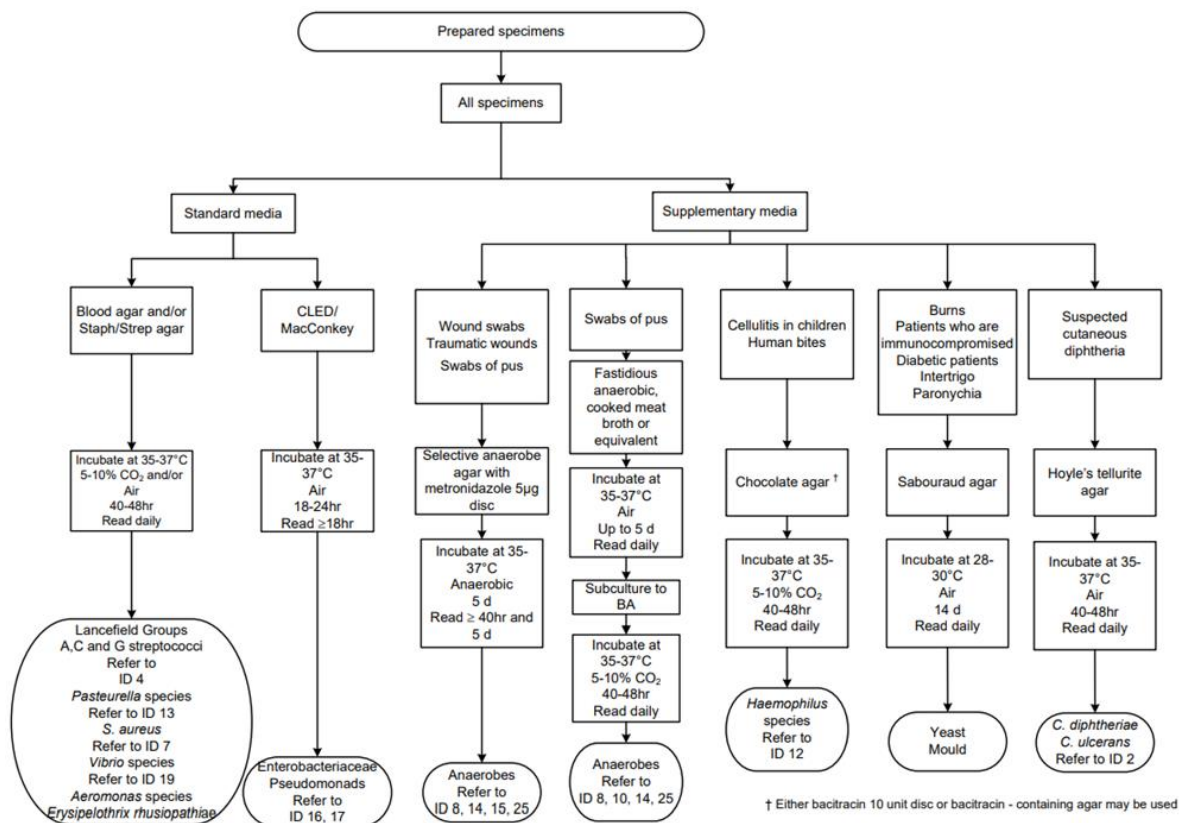


Figure 22. Routine microbiological investigation of wound swab samples carried out in NHS microbiology laboratories.

Flow chart detailing the culture media and conditions used for the investigation of wound swabs by microbiology according to the clinical context of the wound. Figure published in the UK SMI: Investigation of swabs from skin and superficial soft tissue infections (184).

3.2.4.2 Future diagnosis of wound infections

Compared to the laboratory-based methods previously described, the use of point-of-care diagnostic methods reduce the time to result and the need for empirical antibiotic treatments prescribed in the absence of conclusive identification of the organism. The measurement of the host response to the onset of an infection can be utilised to detect early infection. In the immediate aftermath of an injury, there are a plethora of immune response proteins (IRP) involved in the initial response to achieve homeostasis, prevent infection and facilitate wound healing (200). The levels of specific IRPs can change dramatically over the course of an infection presenting a multitude of potential diagnostic test targets. CRP, as previously mentioned, is a well-established target for the detection of inflammation at the onset of an infection (201). Levels of CRP increase in human serum from approximately 1.4 mg/µL to >100 mg/µL at the onset of infection (202, 203) such that measuring levels of CRP can determine whether a wound is infected (203).

To enable point-of-care CRP testing, electrochemical sensing devices to detect the levels of CRP have been developed. Wang *et al.* 2017 developed a CRP fibre optic surface plasmon resonance biosensor with a detection range of 10-20000 ng mL⁻¹. The wide

detection range of this technology ensured that the onset of an infection could be detected as well as subsequent reductions in CRP as the infection responded to treatments. Critically, this technology was easily manufactured at low cost and was disposable, enabling use of the sensor at the point-of-care for rapid CRP detection in response to a suspected wound infection (204).

An alternative method is to quantify the bacterial load in the wound tissue. The MolecuLight i:X (MolecuLight, Inc.) device detects the fluorophores emitted by bacterial species including *Staphylococcus*, *Proteus*, *Klebsiella*, and *Pseudomonas* spp. and can determine when the bacterial load has reached the critical bioburden at which a wound is considered to be infected (205). Another method that has potential for measuring the bioburden of a wound is through detection of the volatile organic compounds released by pathogens by gas chromatography ion mobility spectrometry (206). However, these technologies currently cannot identify specific species and the ability to identify fungal pathogens has not been demonstrated.

Smart dressings refer to wound dressings with incorporated technology to monitor wound healing progress via one or several chemical stimuli associated with an infection and in some cases, actively stimulate wound healing. Examples of this technology include a hydrogel based smart dressing that quantifies bacterial load via the wound pH and detecting alterations in pH associated with the bacterial environment (207). The detection of a bacterial environment triggered the release of an antibacterial treatment to facilitate wound healing. The advantages of smart dressings are that the wound can be continuously monitored without the need for a healthcare professional to repeat measurements for infection and antibiotic treatments are applied immediately following detection of a bacterial environment. However, this technology in its current form would be unable to identify the causative pathogens and the antimicrobial susceptibility and tailor the antimicrobial treatment accordingly. Therefore, additional testing would still be required in the event of an infection.

3.2.5 Wound infection treatment

The treatment for a wound infection depends upon the stage of the infection and the clinical judgement of the care team that will factor in any comorbidities, such as diabetes mellitus or immunosuppression. Wounds that have been colonised by microbial species but have not progressed to a localised infection are debrided and treated with an appropriate wound dressing. No antimicrobial dressings or antibiotics will be used at this stage to prevent the development of an antibiotic-resistant infection. If the infection has progressed to a localised infection, antimicrobial irrigation solutions and topical wound dressings are applied to disrupt the formation of a biofilm and prevent the spread of the infection. Antimicrobial dressings can include the use of honey, iodine and silver, all of

which have antimicrobial properties (208). However, if the wound progresses to a spreading infection, the care team are then likely to carry out more extensive interventions to control the infection. As described previously, a wound swab will be collected to identify the causative organism, and systemic antibiotics will be prescribed. Until identification of the organism is provided, flucloxacillin is often prescribed as the first line antibiotic, with metronidazole added if the wound is located in the abdomen or pelvis. Doxycycline is prescribed if the patient is allergic to penicillin, or the infection does not respond to flucloxacillin. Specific wounds, such as animal or human bites, are prescribed with co-amoxiclav (209). Once the causative organism is identified, the antimicrobial treatment is targeted. For example, wounds positive for MRSA will be treated with doxycycline, which has better coverage against MRSA infections compared to alternative antibiotics. Fungal wound infections are treated with amphotericin B or azole antifungal agents, depending on the specific causative organism. If the infection has further progressed to a systemic infection, the patient will be monitored closely for signs of sepsis. Systemic antibiotics and antimicrobial wound dressings will be continued, and blood cultures may be carried out to monitor for systemic infection.

3.2.6 Limitations of current diagnostics

Despite being the key to identifying a wound infection, the use of clinical symptoms is highly fallible to subjectivity and the variability of the clinical signs, due to factors such as the specific strain of the pathogen causing the infection and development of biofilms. Therefore, it has been demonstrated that clinical symptoms are not a suitable indicator of bacterial load or of wound infection when used in isolation (210).

Further to this, while wound swab culture is the most common method for wound infection diagnosis, the method has several key limitations. Firstly, wound swab culture methods have a relatively poor pooled test sensitivity and specificity rate of 49% and 62% respectively (211). This is caused by a number of factors, including the presence of biofilm formation preventing access to the causative pathogen during sample collection and the presence of fastidious species including anaerobic bacteria that are not identified by culture (212). For example, *Actinomyces* spp. may require up to 7 days of growth to be detected. Often, unless there is a specific clinical indication that extended growth periods are required, cultures will not be incubated for this length of time, preventing pathogen identification. In addition, the previous use of empiric antibiotics immediately upon the suspicion of a wound infection may mean that the bacteria present on the wound do not grow in culture, leading to a negative result. Finally, the wound swab cannot physically access the invasive microbes present within the tissue rather than on the surface of the wound and therefore, these organisms will not be identified.

In the context of a deployed military MTF, there are a number of critical limitations associated with the use of wound culture methods. Firstly, the time to result for culture methods is significant. Bacterial growth generally requires 18-36 hours and increased times may be required for particularly fastidious or slow growing bacteria and AST requires further incubation time. This delay in achieving a result means that initial antimicrobial therapies, in the absence of any information relating to the causative pathogen and the antimicrobial susceptibility, are empirical and broad range, to provide the best chance of targeting the infection. However, the use of broad-range antimicrobials can cause serious side effects for the patient and has been implicated in the further development of antimicrobial resistance amongst bacterial populations.

Further to this, culture methods have been shown to underestimate the number of bacteria and the various types of bacteria present in a wound sample. This is an issue particularly for fastidious bacteria that may not grow efficiently in the culture medium, such as obligate anaerobes. Zhou-Fleming *et al.* 2019 used 16S rRNA metagenomic sequencing to study necrotising skin and soft-tissue infections (NSSTI) and compared the results against standard culture testing. The results of the sequencing analysis found that anaerobic species, including *Prevotella*, *Bacteroides*, *Anaerococcus*, *Peptoniphilus*, and *Porphyromonas* species were frequently missed by culture testing. It was also found that anaerobes made up the majority of bacterial species identified by 16S rRNA sequencing in NSSTI non-survivors compared to patients that did survive the infection, demonstrating the importance of detecting these species (213). Culture methods are also unable to detect the full diversity of fungal infections unless additional specific testing is carried out. This can severely impair the progress of wound infection diagnosis. A case study that followed the progress of a combat-related wound infection demonstrated the importance of fungal species in wound infections. In this case, the presence of the fungal species was initially identified by culture, however, post-mortem PCR testing revealed further diversity of fungal species that had not been previously identified (214).

Finally, the facilities required for microbiology testing, including large incubators, require a central diagnostics laboratory. Forward deployed settings at role one and role two do not have access to such infrastructure and therefore no information on the infection can be gained until the patient has been evacuated to a role two enhanced, role three or role four environment, with role four often being stationed in the patient's home country or that of an allied nation. By this stage, the infection may have progressed significantly.

Despite the issues associated with traditional microbiological culture-based diagnosis of wound infections, the use of molecular-based diagnostics is not widespread within the NHS and is not operated in deployed military MTFs for wound infection diagnosis. Therefore, alternative methods for the rapid diagnosis of wound infections are required.

3.2.7 Clinical metagenomics for wound infection diagnosis

Studies involving the use of CMg for wound infection diagnosis are rare, particularly so for combat-related wounds. However, metagenomics has been applied to the diagnosis of soft skin and tissue infections in a civilian hospital setting. Wang *et al.* 2020 (215) demonstrated an increase in sensitivity of pathogen detection using metagenomics, particularly so for anaerobic bacteria and viruses, compared to standard culture-based methods (67.7% vs 35.4%). Metagenomics also increased the detection of poly-microbial infections. This example demonstrated a key advantage of metagenomics, where fastidious pathogens not detected by culture-based methods are identified by CMg, enabling more targeted treatments to be administered. This increase in sensitivity meant that targeted antibiotic therapies could be utilised in a significantly higher number of cases when metagenomics was used.

3.2.8 Study aims

This study aimed to evaluate the CMg workflow previously developed for wound swab samples and compare the results against routine microbiology testing carried out by the NHS laboratories. Importantly, this would provide a better understanding of the clinical interpretation of the CMg results for the identification of microbial species and additional clinically relevant information including the presence of AMR genes. In addition, additional opportunities for the future optimisation of the workflow would be identified to enable the future deployment of the workflow to deployed MTFs, including reducing the equipment requirements and the time to result.

3.3 Methods

3.3.1 Ethical approval

The study protocol for the prospective collection of 40 wound swab samples, the provision of routine microbiology results and previous antibiotic treatments from NHS patients with a suspected wound infection was approved by the Dstl Scientific Advisory Committee and the MODREC (Reference: 2149/MODREC/22). Approval for the collection of wound swab samples at participating NHS sites was then provided by the HRA (Project ID: 307383) and the study was incorporated into the National Institute for Health Research clinical research network portfolio.

3.3.2 Sample collection

Forty participants with suspected wound infections from open wounds, defined as cuts, lacerations, abrasions, avulsions, punctures, bites, grades 2-4 pressure ulcers, arterial, venous or diabetic ulcers and burns, were prospectively recruited from SDH and BNH. Written informed consent was obtained prior to sample collection. One additional wound swab sample was collected according to best practice (184) immediately following the

collection of a standard of care wound swab. The additional swab sample was given an anonymised unique sample identifier and stored at 4 - 8°C. Samples were processed for CMg analysis following the protocol described below.

3.3.3 Clinical metagenomics

3.3.3.1 Host depletion

The swab tip was added to 200 μ L PBS in a 2.0 mL tube. Forty microliters of 30% saponin (in PBS) was added with 200 μ L HLSAN Buffer (5.5M NaCl and 100mM MgCl₂ in molecular-grade water) and 10 μ L HLSAN DNase. The sample was incubated at 37°C for 10 min at 1000 rpm on a thermomixer. The sample was vortexed for 20 s at full speed and the sample was transferred to a new 2 mL tube and the swab tip disposed of. Nine hundred microliters of PBS were added and the sample was centrifuged at 12000 $\times g$ for 3 min. The supernatant was discarded, and the pellet resuspended in 600 μ L PBS. Negative controls, consisting of a blank swab processed alongside the participant swab samples were included prior to the host-depletion step.

3.3.3.2 DNA extraction

The resuspended pellet was transferred to a Matrix lysing E tube (MP Biomedicals) and bead beaten for 60 s at 6.0 m/s on a FastPrep-24 5G bead beater and centrifuged at 17000 $\times g$ for 60 s. Four hundred microliters of clear supernatant was transferred to a new tube with 300 μ L lysis buffer and 30 μ L Proteinase K. The sample was incubated at 65°C for 5 min at 1000 rpm and added directly to a MaxWell 16 LEV Blood cartridge. The DNA extraction was carried out by the MaxWell 16 LEV platform using the Blood DNA protocol and eluted into 50 μ L elution buffer.

3.3.3.3 DNA quantification

The total DNA concentration was quantified using the dsDNA HS assay kit on the Qubit 4 fluorometer according to the manufacturer's instructions, as previously described in Chapter Two section 2.3.5.

3.3.3.4 DNA purification

The DNA extract was purified by AMPure XP bead purification according to the manufacturer's instructions. AMPure XP beads were added to the DNA extract at a ratio of 1.8 $\times$ beads to DNA extract and processed according to the methods described in Chapter Two section 2.3.6.

3.3.3.5 Whole genome amplification

The purified DNA underwent WGA using the Genomiphi WGA kit, as previously described in Chapter Two section 2.3.8. Ten microliters of purified wound swab DNA was added to 10 μ L of the 2 $\times$ denaturation buffer in a 0.2 mL PCR tube and heated to 95°C for 3 min

and cooled to 4°C on ice. The 20 µL reaction was added directly to the freeze-dried reagent cake and incubated at 37°C for 2 h followed by a heating step at 65°C for 10 min. The sample was cooled to 4°C on ice. The resulting amplified DNA was then re-quantified by Qubit dsDNA HS assay, as previously described.

3.3.3.6 DNA de-branching

The amplified DNA was then de-branched as described in Chapter Two section 2.3.9. The de-branched DNA was quantified by Qubit dsDNA HS assay as previously described and purified using a 0.6 × ratio of AMPure XP beads to DNA extract, as previously described.

3.3.3.7 Library preparation and sequencing

The de-branched DNA was prepared for sequencing using the Rapid Barcoding library preparation kit according to the manufacturer's instructions, as previously described in Chapter Two section 2.3.10.1. Five to seven samples, plus a negative control, were run per multiplexed sequencing run. An R9.4.1 flowcell was inserted into the MK1c Sequencing device and the flow cell was then checked and primed for sequencing and the prepared libraries loaded into the flowcell according to the manufacturer's instructions. A new experiment was then set up using the MK1c MinKnow user interface. The library preparation kit was selected, and the sequencing run set for 16-h duration. The sequencing run was started, and the data stored in Fast5 format.

3.3.4 Bioinformatic and statistical analysis

ONT Taxonomic Profiling, the custom bioinformatics pipeline, previously described in Chapter Two section 2.3.11, was applied. Microbial species identification was presented as the total number of reads and the percentage of the total classified microbial reads aligned to the species. Species were concluded to be present if ≥ 5 reads, minus the number of reads of the same species identified in the negative control sample, matched that species. Alternative thresholds for considering a positive identification were applied to compare the sensitivity when these were applied, including ≥ 50 reads assigned to a species and the total reads assigned to a species comprised $\geq 0.1\%$ of the total classified microbial reads. Sequencing depth was calculated by dividing the total bp assigned to a species by the total genome size of the species. The pipeline included applying the KMA (version 1.4.9) tool to screen for AMR genes with the amrfinderplus (v3.10) database. AMR genes were concluded to be present if their template identity and coverage were $\geq 90\%$ and their depth of coverage reported by KMA were $\geq 1\times$. AMR genes were linked to the resistances reported by AST through the Comprehensive Antibiotic Resistance Database (CARD), which lists the species associated with specific AMR genes and the antibiotic drug classes to which they are associated (216). Bacterial species were grouped into aerobes, facultative anaerobes and obligate anaerobes according to the BacDive database (217). Data are expressed as means and + standard

deviation and were analysed using GraphPad Prism 10 (version 10.0.2) software. When multiple comparison analyses were performed, *p*-values were obtained by performing an ordinary one-way or two-way ANOVA test or mixed effects analysis when the *n* number varied, with post hoc Tukey's test. When two samples were individually compared, results were obtained using a two-tail unpaired t-test. Sensitivity and specificity values were calculated according to the following calculations as previously described (218):

$$\text{Sensitivity} = [(True\ Positives)/(True\ Positives + False\ Negatives)] \times 100$$

$$\text{Specificity} = [(True\ Negatives)/(False\ Positives + True\ Negatives)] \times 100$$

True positives = species identified by CMg and routine microbiology and/or qPCR.

False Negative = species identified by routine microbiology and/or qPCR but not identified by CMg.

True Negative = species not identified by CMg or routine microbiology and/or qPCR.

False positive = species identified by CMg but not by routine microbiology and/or qPCR.

Species α -diversity was calculated using the Shannon diversity index. The Shannon diversity index is an estimator for species richness and evenness (219). The formula is as follows:

$$\text{Shannon diversity index} = - \sum [(p_i) \times \ln(p_i)]$$

$$p_i = \frac{n}{N}$$

n = number of reads assigned to a microbial species in a sample

N = total reads assigned to microbial species in a sample

3.3.5 Routine clinical microbiological investigation

Microbiology testing of the routine wound swab sample was carried out by NHS staff at each participating NHS trust according to the standard protocols. Wound swabs were collected using a Transwab charcoal swab. Within 48 h of collection, wound swabs were streaked on to semi-solid agar media including columbia blood and chocolate agar, incubated at 35-37°C in 5% CO₂; columbia selective agar and CLED agar, incubated at 35-37°C under aerobic conditions and fastidious anaerobe agar with neomycin incubated at 35-37°C under anaerobic conditions. Additional media, such as sabouraud agar for the growth of fungal and yeast species, were used for wounds swabs collected from specific anatomical sites, or where particular clinical information or specific clinician instruction determined their use based on the UK SMI (184). Plates were examined at 18-24 h and

again at 48 h post-inoculation for growth with extended incubation if directed by the consultant microbiologist. Potentially significant isolates were selected and identified according to colony morphology, growth characteristics on particular media (e.g. haemolysis, growth on selective agar), Gram stain and rapid biochemical tests. Confirmation of colony identity was carried out by MALDI-TOF-MS or by extended biochemical tests including API. The level of identity reported was in accordance with the UK SMI (184) based on the anatomical site from where the swab was taken, the clinical presentation, the isolate and following any specific clinical instructions.

AST was carried out by disc diffusion assay. Pure isolates were streaked on to Mueller-Hinton agar against antibiotic panels according to the isolate. Conduct of antibiotic susceptibility testing and interpretation of zones of inhibition was done according to EUCAST guidelines (36).

3.3.6 Interpretation of clinical metagenomics results

The interpretation of the sequencing results to identify the species considered to be reportable in a clinical context was carried out by a consultant microbiologist in the context of the UK SMI (184). Reportable species, including species that would only be reported in specific circumstances including when identified in isolation, were reported.

3.3.7 qPCR

Taqman probe-based qPCR assays were performed in duplicate for four high priority pathogen species. All assays were performed on the QuantStudio 7 Flex platform, and Cq values were determined by the QuantStudio software. Cq values below 38.0 were considered positive. Positive control reactions of purified bacterial DNA at a concentration of 1.0 ng/μL and negative control reactions of nuclease-free water only were run on each plate. The primer and TaqMan probe sequences of each assay are described in Table 6. The TaqMan Fast Advanced Master Mix (Applied Biosystems) was used for all assays. The following conditions were applied for each qPCR assay:

E. coli 16S rRNA: 25 μL reactions using 12.5 μL master mix, final concentrations of 900 nM forward primer, 900 nM reverse primer, 200 nM of probe, and 1 μL of template. Thermal cycling conditions were 1 min at 95°C followed by 40 cycles of 15 s at 95°C and 30 s at 60°C.

S. aureus protein coding sequence (NCBI Protein accession number: YP_500811.1): 10 μL reactions using 5 μL master mix, 1 μM forward primer, 1 μM reverse primer, 200 nM probe, and 1 μL template. Thermal cycling conditions were 1 min at 95°C followed by 40 cycles of 15 s at 95°C and 30 s at 57°C.

P. aeruginosa outer-membrane protein (opr) L: 10 μL reactions using 5 μL master mix, 300 nM forward primer, 300 nM reverse primer, 200 nM TaqMan probe, and 1 μL

template. Thermal cycling conditions were 1 min at 95°C followed by 40 cycles of 15 s at 95°C and 30 s at 60°C.

S. pyogenes streptococcal pyrogenic exotoxin (spe) B: 25 µL reactions using 12.5 µL master mix, final concentrations of 100 nM forward and reverse primer, 150 nM probe and 1 µL template. Thermal cycling conditions were 1 min at 95°C followed by 40 cycles of 20 s at 95°C and 20 s at 60°C.

Table 6. Primer and probe sequences for high priority pathogen qPCR assays

Assay target	Forward primer (5'-3')	Reverse primer (5'-3')	Probe sequence (5'-3')	Source
<i>E. coli</i> 16S rRNA gene	CATGCCGCGTGAT GAAGAA	CGGGTAACGTCAAT GAGCAAA	TTCAGCGGGGAGGA AGGGAGTAAAGTTA ATA	(135)
<i>P. aeruginosa</i> oprL gene	CAGGTCGGAGCTGT CGTACTC	ACCCGAACGCAGGC TATG	AGAAGGTGGTGATC GCACGCAGA	(220)
<i>S. aureus</i> protein coding sequence (NCBI Protein accession number: YP_500811.1)	AACTACTAGGGGAG CCTAATRAT	GGTACTAACCAAATC AGGTCATAA	TGGCTGAGATGAAY TGTTTCAGACCC	(140)
<i>S. pyogenes</i> speB gene	CTAAACCCTTCAGCT CTTGGTACTG	TTGATGCCTACAACA GCACTTTG	CGGCGCAGGCGGCT TCAAC	(221)

3.4 Results

3.4.1 Prospective wound swab sample collection

A total of 40 wound swab samples, prospectively collected from study participants with suspected wound infections, were processed for CMg and the results were compared to routine microbiology results obtained from a swab collected immediately prior. Samples were collected from wound types including 29 (72.50%) ulcers, three (7.50%) animal bites, three (7.50%) unspecified non-healing wounds, two (5.00%) pressure wounds and one (2.50%) each of burn, amputation and laceration wounds. Twenty-one participants (52.50%) had previously been prescribed antibiotics and 15 participants (37.5%) were not prescribed antibiotics prior to sample collection, with antibiotic history unknown for four (10%) participants. Sample characteristics are presented in Table 7.

Table 7. Wound swab sample characteristics

Sample Number	Wound Type	Wound Location	Antibiotics prescribed prior to sample collection
01	Laceration	Palmar aspect of metacarpophalangeal joint of index finger	No antibiotics at time of sampling.
02	Animal bite	Leg	No antibiotics at time of sampling
03	Unspecified non-healing	Ankle	No antibiotics at time of sampling
04	Animal bite	Finger	No antibiotics at time of sampling
05	Ulcer	Lateral upper aspect of foot	No antibiotics at time of sampling
06	Burn	Wound location not specified	Flucloxacillin
07	Ulcer	Upper aspect of toe	Flucloxacillin
08	Ulcer	Left medial malleolus	No antibiotics at time of sampling
09	Ulcer	Lower leg	Flucloxacillin and clindamycin
10	Ulcer	Lower leg	Flucloxacillin
11	Non-healing amputation	Toe	No antibiotics at time of sampling
12	Unspecified non-healing	Plantar aspect of toe	Flucloxacillin
13	Ulcer	Heel	No antibiotics at time of sampling
14	Ulcer	Medial aspect of foot over metatarsal head	No antibiotics at time of sampling
15	Ulcer	Toe apex	No antibiotics at time of sampling
16	Ulcer	Hallux	Co-amoxiclav, gentamicin
17	Ulcer	Foot	Clindamycin
18	Pressure	Sacrum	Clindamycin
19	Ulcer	Ankle	Amoxicillin, meropenem
20	Ulcer	Lower leg	Flucloxacillin, gentamycin
21	Ulcer	Lower leg	Clindamycin
22	Ulcer	Heel	Co-amoxiclav, flucloxacillin
23	Ulcer	Lower leg	Flucloxacillin
24	Ulcer	Heel	Metronidazole, ciprofloxacin, cotrimoxazole
25	Ulcer	Lower leg	Linezolid, meropenem, clindamycin
26	Animal bite	Lower leg	Doxycycline, ceftriaxone
27	Ulcer	Heel	Co-amoxiclav
28	Ulcer	Foot	No antibiotics at time of sampling
29	Ulcer	Heel	Co-amoxiclav, flucloxacillin
30	Ulcer	Lower leg	No antibiotics at time of sampling
31	Ulcer	Lower leg	No antibiotics at time of sampling
32	Pressure	Heel	Doxycycline
33	Ulcer	Ankle	No antibiotics at time of sampling
34	Ulcer	Lateral foot over metatarsal head	No antibiotics at time of sampling
35	Unspecified non-healing	Plantar aspect of foot	Antibiotic history unknown
36	Ulcer	Plantar aspect of foot	Antibiotic history unknown
37	Ulcer	Lateral aspect of foot	Antibiotic history unknown
38	Ulcer	Plantar aspect right foot	Antibiotic history unknown
39	Ulcer	Toe	Co-amoxiclav
40	Ulcer	Toe	Co-amoxiclav

3.4.2 Sample processing for clinical metagenomics

The mean total DNA concentration following the host-depletion and DNA extraction steps was 3.53 ng/μL, ranging from below the Qubit dsDNA HS assay LOD (0.05 ng/μL) to 21.60 ng/μL. The mean total number of sequencing reads per sample was 105688, with a mean number of classified microbial reads of 32760 and mean number of human reads of 24657 and a mean read length of 1249 bp. The host depletion, DNA extraction, WGA and library preparation steps were completed within approximately four hours and fifteen minutes per batch of five to seven samples. Full sample processing for CMg results are provided in Table 8.

Table 8. Results of CMg analysis of the wound swab samples.

Sample Number	Post-depletion DNA Extract Concentration (ng/μL)	Total Reads	Mean Read Length (bp)	Classified Microbial Reads	Human Reads	Organisms identified by metagenomics sequencing (≥ 1% classified microbial reads)	Species reads	% classified microbial reads
01	2.31	217440	1505	157878	32790	<i>Enterobacter hormaechei</i> <i>Enterobacter</i> sp. DSM 30060 <i>Enterobacter roggenkampii</i> <i>Enterobacter cloacae</i> complex spp. ECL414 <i>Klebsiella pneumoniae</i>	111837 21481 7014 6398 1707	70.84 13.61 4.44 4.05 1.08
02	0.313	308679	1247	8161	293523	<i>Mycobacteroides chelonae</i> <i>Toxoplasma gondii</i> <i>Mycobacteroides abscessus</i> <i>Mycobacterium stephanolepidis</i> <i>Mycobacteroides salmoniphilum</i>	6326 497 318 96 93	77.52 6.09 3.90 1.18 1.14
03	<0.05	196621	917	2764	190054	<i>S. aureus</i> <i>T. gondii</i> <i>Finegoldia magna</i> <i>Lactocaseibacillus paracasei</i> <i>Lentilactobacillus buchneri</i>	1550 245 231 35 28	64.84 10.22 9.63 1.46 1.17
04	<0.05	294741	927	8666	171392	<i>P. multocida</i> <i>Tannerella forsythia</i> <i>Fusobacterium nucleatum</i> <i>Porphyromonas gingivalis</i> <i>T. gondii</i> <i>Campylobacter</i> sp. CCUG 57310 <i>Fusobacterium pseudoperiodonticum</i> <i>Fusobacterium hwasookii</i> <i>Campylobacter</i> sp. RM16192 <i>Fusobacterium canifelinum</i>	2708 2390 771 480 451 362 170 152 113 105	31.25 27.58 8.90 5.54 5.20 4.18 1.96 1.75 1.30 1.21
05	11.7	288444	1461	247600	1298	<i>Enterococcus faecalis</i> <i>Klebsiella aerogenes</i>	232570 8879	93.93 3.59
06	0.135	39997	797	112.0	38457	<i>Staphylococcus epidermis</i> <i>Aspergillus luchuensis</i> <i>Corynebacterium glucuronolyticum</i>	60 15 13	53.57 13.39 11.61

07	5.94	63333	1294	15474	19	<i>Streptococcus oralis</i>	3480	22.49
						<i>Parvimonas micra</i>	3020	19.52
						<i>Veillonella parvula</i>	2310	14.93
						<i>Streptococcus sp. oral taxon 064</i>	1380	8.92
						<i>Streptococcus constellatus</i>	1030	6.66
						<i>Streptococcus sp. NSJ-72</i>	770	4.98
						<i>Streptococcus milleri</i>	716	4.63
						<i>Prevotella buccalis</i>	417	2.69
						<i>Anaerococcus obesiensis</i>	307	1.98
						<i>Anaerococcus vaginalis</i>	234	1.51
						<i>Streptococcus anginosus</i>	166	1.07
						<i>Peptoniphilus harei</i>	163	1.05
08	7.88	109061	1377	45571	3948	<i>E. coli</i>	20400	44.77
						<i>E. hormaechei</i>	9940	21.81
						<i>K. pneumoniae</i>	4420	9.70
						<i>E. faecalis</i>	1840	4.04
						<i>E. cloacae</i>	1810	3.97
						<i>Corynebacterium jeikeium</i>	1780	3.91
						<i>Citrobacter portucalensis</i>	945	2.07
						<i>F. magna</i>	943	2.07
						<i>Corynebacterium coyleae</i>	457	1.00
09	0.097	5636	944	642	31	<i>Rothia dentocariosa</i>	405	63.08
						<i>Rothia aeria</i>	44	6.85
						<i>Cutibacterium acnes</i>	34	5.30
						<i>Corynebacterium tuberculostearicum</i>	33	5.14
						<i>Staphylococcus epidermidis</i>	23	3.58
						<i>Staphylococcus haemolyticus</i>	16	2.49
						<i>Staphylococcus warneri</i>	12	1.87
						<i>Staphylococcus pasteurii</i>	10	1.56
						<i>Malassezia restricta</i>	9	1.40
10	0.170	44840	1151	2735	34271	<i>S. pyogenes</i>	1540	56.31
						<i>S. aureus</i>	1130	41.32
11	<0.05	2112	365	51	299	<i>Dermabacter jinjuensis</i>	6	11.76
12	0.931	64218	1461	31411.0	12773	<i>Citrobacter koseri</i>	29995	95.49
13	7.76	34286	1194	23145	48	<i>B. fragilis</i>	15096	65.22
						<i>Streptococcus dysgalactiae</i>	2830	12.23
						<i>A. obesiensis</i>	1010	4.36
						<i>Bacteroides ovatus</i>	790	3.41
						<i>A. vaginalis</i>	594	2.57
						<i>F. magna</i>	561	2.42
						<i>P. harei</i>	463	2.00
						<i>Helcococcus kunzii</i>	293	1.27
14	4.98	43397	1084	23118	7243	<i>S. aureus</i>	23086	99.86
15	5.54	33311	1227	15395	8541	<i>S. aureus</i>	15369	99.83
16	0.156	14647	876	5787	2468	<i>S. aureus</i>	1937	33.47
						<i>F. magna</i>	1880	32.49
						<i>Corynebacterium striatum</i>	1380	23.85
						<i>Acinetobacter johnsonii</i>	138	2.38
17	0.559	20929	963	3525	856	<i>Pseudomonas sp. NIBR-H-19</i>	3127	88.71
						<i>C. striatum</i>	86	2.44
						<i>Klebsiella oxytoca</i>	68	1.93
18	0.781	151500	1474	53009	303	<i>E. coli</i>	35004	66.03
						<i>Lactocaseibacillus rhamnosus</i>	8859	16.71
						<i>C. albicans</i>	3273	6.17
						<i>C. striatum</i>	2587	4.88
						<i>Lactobacillus gasseri</i>	1803	3.40
						<i>S. epidermidis</i>	537	1.01
19	0.186	43525	1212	1498	32139	<i>C. striatum</i>	1396	93.19
						<i>Corynebacterium simulans</i>	24	1.60
20	2.77	161465	1555	30104	19182	<i>Corynebacterium urealyticum</i>	16700	55.47
						<i>Corynebacterium propinquum</i>	7830	26.01
						<i>Corynebacterium aurimucosum</i>	1580	5.25
						<i>C. tuberculostearicum</i>	469	1.56
						<i>C. jeikeium</i>	338	1.12
21	1.08	191440	1168	67266	9304	<i>C. koseri</i>	59255	88.09
						<i>Citrobacter sp. TBCP-5362</i>	3843	5.71
						<i>Mammaliococcus lentus</i>	1279	1.90

22	9.47	205000	1554	59660	82	<i>Morganella morganii</i>	45270	75.88
						<i>Streptococcus gordonii</i>	5522	9.26
						<i>A. obesiensis</i>	1256	2.11
						<i>F. magna</i>	914	1.53
						<i>A. vaginalis</i>	899	1.51
23	1.22	150218	1433	38028	22082	<i>C. striatum</i>	36283	95.41
24	1.71	43689	1347	8437	1350	<i>C. albicans</i>	3288	38.97
						<i>E. faecalis</i>	2495	29.57
						<i>C. striatum</i>	2259	26.77
						<i>S. haemolyticus</i>	217	2.57
25	0.989	70854	1309	18975	581	<i>Staphylococcus pseudintermedius</i>	17986	94.79
26	<0.05	43218	988	8430	6111	<i>S. pyogenes</i>	6994	82.97
						<i>S. aureus</i>	544	6.45
						<i>S. epidermidis</i>	538	6.38
27	0.337	75430	1172	10952	14475	<i>C. striatum</i>	5646	51.55
						<i>F. magna</i>	1156	10.56
						<i>A. vaginalis</i>	1140	10.41
						<i>A. obesiensis</i>	1100	10.04
						<i>S. aureus</i>	868	7.93
						<i>S. aureus</i>	868	7.93
						<i>C. simulans</i>	130	1.19
28	1.83	395134	1737	42295	13010	<i>C. striatum</i>	30433	71.95
						<i>Rhodococcus erythropolis</i>	1541	3.64
						<i>F. magna</i>	1356	3.21
						<i>Rhodococcus sp. P-2</i>	1334	3.15
						<i>P. aeruginosa</i>	1310	3.10
						<i>Anaerococcus sp. Marseille-Q7828</i>	1173	2.77
						<i>P. harei</i>	1006	2.38
						<i>Anaerococcus mediterraneensis</i>	943	2.23
						<i>Schaalia radingae</i>	607	1.44
29	0.368	34373	779	95	31336	<i>H. kunzii</i>	592	1.40
						<i>S. aureus</i>	62	65.26
30	21.6	292064	2046	122166	23705	<i>T. gondii</i>	21	22.11
						<i>P. aeruginosa</i>	101165	82.81
31	<0.05	1993	651	218	756	<i>Pseudomonas putida</i>	18103	14.82
						<i>Staphylococcus simulans</i>	175	80.28
32	0.091	42928	1220	16523	2544	<i>S. epidermidis</i>	15	6.88
						<i>C. striatum</i>	9621	58.23
						<i>S. dysgalactiae</i>	4783	28.95
						<i>D. jinjuensis</i>	424	2.57
						<i>F. magna</i>	320	1.94
						<i>E. faecalis</i>	217	1.31
						<i>C. simulans</i>	184	1.11
33	10.9	200437	1765	95559	433	<i>B. fragilis</i>	70466	73.74
						<i>M. morganii</i>	10540	11.03
						<i>Peptoniphilus sp. SAHP1</i>	3211	3.36
						<i>C. striatum</i>	2297	2.40
						<i>Bacteroides uniformis</i>	2221	2.32
						<i>P. mirabilis</i>	1938	2.03
34	1.19	22931	1474	13068	2269	<i>C. striatum</i>	10922	83.58
						<i>P. aeruginosa</i>	1224	9.37
						<i>C. simulans</i>	153	1.17
35	1.27	40228	1346	18067	85	<i>Staphylococcus pettenkoferi</i>	9809	54.29
						<i>C. striatum</i>	3799	21.03
						<i>S. dysgalactiae</i>	1239	6.86
						<i>E. faecalis</i>	1198	6.63
						<i>S. simulans</i>	510	2.82
						<i>H. kunzii</i>	260	1.44
						<i>S. aureus</i>	252	1.39
						<i>S. epidermidis</i>	182	1.01
36	20.2	38948	1285	23440	393.00	<i>S. dysgalactiae</i>	18000	76.79
						<i>S. aureus</i>	1377	5.87
						<i>C. striatum</i>	1267	5.41
						<i>S. pyogenes</i>	631	2.69
						<i>Streptococcus agalactiae</i>	547	2.33
						<i>P. aeruginosa</i>	314	1.34
37	7.23	44806	1119	27416	27	<i>S. dysgalactiae</i>	20847	76.04
						<i>Prevotella corporis</i>	2089	7.62
						<i>B. fragilis</i>	1180	4.30

						<i>S. pyogenes</i>	722	2.63
						<i>S. agalactiae</i>	452	1.65
						<i>Peptoniphilus</i> sp. SAHP1	312	1.14
						<i>F. magna</i>	278	1.01
38	5.66	56748	1594	32422	13.00	<i>Klebsiella michiganensis</i>	12535	38.66
						<i>S. anginosus</i>	3261	10.06
						<i>K. oxytoca</i>	1667	5.14
						<i>C. striatum</i>	1475	4.55
						<i>K. pneumoniae</i>	1390	4.29
						<i>Citrobacter freundii</i>	1269	3.91
						<i>E. faecalis</i>	1182	3.65
						<i>A. obesiensis</i>	890	2.75
						<i>B. fragilis</i>	757	2.33
						<i>Klebsiella grimontii</i>	731	2.25
						<i>Peptoniphilus</i> sp. SAHP1	713	2.20
						<i>P. micra</i>	658	2.03
						<i>A. vaginalis</i>	617	1.90
						<i>E. coli</i>	576	1.78
						<i>P. aeruginosa</i>	463	1.43
<i>E. hormaechei</i>	442	1.36						
<i>M. morgani</i>	329	1.01						
39	<0.05	36541	1253	7929	6088	<i>F. magna</i>	2946	37.15
						<i>S. simulans</i>	830	10.47
						<i>A. obesiensis</i>	625	7.88
						<i>A. vaginalis</i>	589	7.43
						<i>P. harei</i>	359	4.53
						<i>H. kunzii</i>	347	4.38
						<i>A. mediterraneensis</i>	313	3.95
						<i>S. pettenkoferi</i>	129	1.63
						<i>S. pyogenes</i>	103	1.30
<i>S. agalactiae</i>	97	1.22						
40	3.52	102369	1626	22819	1986	<i>Alcaligenes faecalis</i>	4218	18.48
						<i>F. magna</i>	4110	18.01
						<i>H. kunzii</i>	2203	9.65
						<i>Anaerococcus</i> sp. Marseille-Q7828	983	4.31
						<i>C. aurimucosum</i>	903	3.96
						<i>A. obesiensis</i>	787	3.45
						<i>P. harei</i>	747	3.27
						<i>Corynebacterium amycolatum</i>	737	3.23
						<i>A. vaginalis</i>	664	2.91
						<i>A. mediterraneensis</i>	587	2.57
						<i>P. aeruginosa</i>	518	2.27
						<i>C. jeikeium</i>	408	1.79
						<i>Anaerococcus prevotii</i>	395	1.73
						<i>Corynebacterium</i> sp. ATCC 6931	347	1.52
						<i>Fastidiosipila sanguinis</i>	233	1.02

Additional wound swabs collected from study participants were processed for CMg. DNA extracts were quantified prior to WGA and library preparation. Human reads were quantified by aligning sequences to a reference human genome prior to further analysis. Microbial species identified $\geq 1\%$ classified microbial reads are reported. % classified microbial reads = % of species reads over the total number of taxonomically classified microbial reads obtained.

3.4.3 Routine microbiology

Routine microbiology testing, carried out by the NHS microbiology laboratories, reported the identification of a microorganism according to species, genus or phenotypic characteristic in 32 (80.00%) samples. Seventy-one microorganisms were reported in total, with multiple microorganisms reported in 21 positive samples (Table 9). The remaining eight (20.00%) samples reported no microbial growth or mixed colonising organisms (MCO) only without a microorganism species, genus or phenotypic characteristic reported.

3.4.4 Clinical metagenomics species identification

CMg identified a corresponding organism matching 62 of the 71 microorganism species, genus or phenotypic characteristics reported by routine microbiology testing at ≥ 5 reads, an overall sensitivity of 87% (Table 9). Additional organisms that would be considered reportable, as defined by consultant microbiologists reviewing the CMg results, were identified in 8/8 samples that reported no growth or MCO only by routine microbiology (≥ 5 CMg reads). Further, the reportable organisms included those identified at a high proportion of microbial reads ($\geq 1\%$ classified microbial reads) in 4/8 of these samples. From the 32 microbiology positive samples, CMg identified additional reportable pathogens at a high abundance ($\geq 1\%$ classified microbial reads) in 13 samples. This included high priority pathogen species *E. coli* in sample 38, *P. aeruginosa* in samples 30, 36, 38 and 40 and *S. pyogenes* in samples 10, 36, 37 and 39. CMg also identified the fungal pathogen *C. albicans* in sample 18.

Table 9. Results of routine microbiology testing and CMg analysis of wound swab samples.

Sample Number	Routine microbiology report	Corresponding species identified by CMg	Reads (% classified microbial reads)	Sequencing Depth (x)	Additional reportable wound pathogens identified by CMg ($\geq 1\%$ classified microbial reads)	Reads (% classified microbial reads)	Sequencing Depth (x)
01	<i>Enterobacter cloacae</i> ^a	<i>Enterobacter hormaechei</i>	111837 (70.84)	29.76	<i>Enterobacter</i> sp. DSM 30060 <i>E. roggenkampii</i> <i>E. cloacae</i> complex sp. ECL414 <i>K. pneumoniae</i>	21481 (13.61) 7014 (4.44) 6398 (4.05) 1707 (1.08)	7.16 1.04 0.22 0.33
02	No growth	-	-	-	-	-	-
03	<i>S. aureus</i>	<i>S. aureus</i>	1555 (64.85)	0.39	<i>F. magna</i> ^c	231 (9.63)	0.06
04	<i>P. multocida</i>	<i>P. multocida</i>	2708 (31.25)	0.62	-	-	-
05	MCO <i>Enterococci</i> Coliform ^b	N/A <i>E. faecalis</i> <i>K. aerogenes</i>	N/A 232570 (93.93) 8879 (3.59)	93.48 2.46	-	-	-
06	MCO ^b	-	-	-	-	-	-
07	<i>S. oralis</i> <i>E. cloacae</i>	<i>S. oralis</i> Not identified	3482 (22.49) -	1.81 -	<i>S. constellatus</i> ^c <i>S. milleri</i> ^c <i>S. anginosus</i> ^c	1030 (6.66) 716 (4.63) 166 (1.07)	0.45 0.54 0.02
08	<i>E. coli</i> <i>K. pneumoniae</i> <i>Enterococci</i>	<i>E. coli</i> <i>K. pneumoniae</i> <i>E. faecalis</i>	20441 (44.77) 4423 (9.70) 1839 (4.04)	5.19 1.13 0.67	<i>E. hormaechei</i> <i>E. cloacae</i> <i>C. portucalensis</i> <i>F. magna</i> ^c	9940 (21.81) 1810 (3.97) 954 (2.07) 953 (2.07)	2.69 0.77 495.75 0.44
09	No growth	-	-	-	-	-	-
10	<i>S. aureus</i>	<i>S. aureus</i>	1126 (41.17)	0.39	<i>S. pyogenes</i>	1542 (56.31)	0.89
11	<i>S. aureus</i> Coagulase negative <i>Staphylococcus</i> Group D <i>Streptococcus</i>	Not identified Not identified Not identified	-	-	-	-	-
12	Coliform ^b	<i>C. koseri</i>	29984 (95.49)	6.89	-	-	-

13	Anaerobes ^b	<i>B. fragilis</i>	15092 (65.22)	2.29			
	Group A	<i>S. pyogenes</i>	138 (0.60)	0.05	<i>S. dysgalactiae</i>	2830 (12.23)	0.08
	<i>Streptococci</i>						
	<i>Enterococci</i>	<i>E. faecalis</i>	110 (0.48)	0.03	<i>F. magna</i> ^c	561 (2.42)	0.11
	Coliform ^b	<i>C. koseri</i>	50 (0.22)	0.01			
14	<i>S. aureus</i>	<i>S. aureus</i>	23082 (99.84)	7.11	-	-	-
15	<i>S. aureus</i>	<i>S. aureus</i>	15365 (99.81)	4.92	-	-	-
16	<i>S. aureus</i>	<i>S. aureus</i>	1933 (33.40)	0.39	<i>F. magna</i> ^c	1880 (32.49)	0.46
17	Coliform ^b	<i>K. oxytoca</i>	68 (1.93)	0.01			
	<i>E. faecalis</i>	Not identified	-	-	<i>Pseudomonas</i> sp.	3127 (88.71)	0.76
	<i>Proteus</i> overgrowth	Not identified	-	-	NIBR-H-19		
	Diphtheroids ^b	<i>C. striatum</i>	86 (2.44)	0.04			
18	<i>E. coli</i>	<i>E. coli</i>	35004 (66.03)	8.00			
	Anaerobes ^b	<i>L. rhamnosus</i>	8859 (16.71)	1.68			
	Coagulase negative	<i>S. epidermidis</i>	537 (1.01)	0.18	<i>C. albicans</i>	3273 (6.17)	2.25
	<i>Staphylococcus</i>						
	Diphtheroids ^b	<i>C. striatum</i>	2587 (4.88)	1.27			
19	Scanty growth (SG)	Not identified	-	-			
	coagulase negative						
	<i>Staphylococcus</i>				-	-	-
	SG Diphtheroids ^b	<i>C. striatum</i>	1396 (93.19)	0.65			
20	<i>S. aureus</i>	<i>S. aureus</i>	186 (0.62)	0.05			
	<i>Enterococcus</i>	<i>E. faecium</i>	161 (0.53)	0.05			
	Non-lactose fermenting coliform ^b	<i>E. coli</i>	23 (0.08)	0.00	-	-	-
21	Group B	<i>S. agalactiae</i>	323 (0.48)	0.14	<i>C. koseri</i>	59255 (88.09)	12.93
	<i>Streptococcus</i>				<i>Citrobacter</i> sp.		
	<i>S. aureus</i>	<i>S. aureus</i>	94 (0.14)	0.02	TBCP-5362	3843 (5.71)	1.03
22	Alpha-haemolytic	<i>S. gordonii</i>	5522 (9.26)	3.43			
	<i>Streptococcus</i>						
	<i>Enterococci</i>	<i>E. faecalis</i>	359 (0.60)	0.14	<i>M. morganii</i>	45270 (75.88)	0.34
	Coliform ^b	<i>E. coli</i>	73 (0.12)	0.02	<i>F. magna</i> ^c	914 (1.53)	0.18
	Oxidase negative non-lactose fermenter ^b	<i>P. mirabilis</i>	18 (0.03)	0.00			
	Diphtheroids ^b	<i>C. striatum</i>	94 (0.16)	0.05			
23	SG Group C	<i>S. dysgalactiae</i>	280 (0.74)	0.02			
	<i>Streptococcus</i>				-	-	-
	SG <i>S. aureus</i>	<i>S. aureus</i>	212 (0.56)	0.06			
	<i>Proteus</i>	<i>P. mirabilis</i>	28 (0.07)	0.01			
24	Yeasts ^b						
	Coagulase negative	<i>C. albicans</i>	3288 (38.97)	1.90	<i>E. faecalis</i>	2495 (29.57)	0.64
	<i>Staphylococcus</i>	<i>S. haemolyticus</i>	217 (2.57)	0.06			
25	Coagulase negative	<i>S. epidermidis</i>	85 (0.45)	0.02	-	-	-
	<i>Staphylococcus</i>						
26	Group A	<i>S. pyogenes</i>	6994 (82.97)	3.10			
	<i>Streptococcus</i>				-	-	-
	<i>S. aureus</i>	<i>S. aureus</i>	544 (6.45)	0.16			
27	<i>S. aureus</i>	<i>S. aureus</i>	868 (7.93)	0.17			
	Group G	<i>S. dysgalactiae</i>	6 (0.05)	0.00	<i>F. magna</i> ^c	1156 (10.56)	0.30
	<i>Streptococcus</i>						
	Diphtheroids ^b	<i>C. striatum</i>	5646 (51.55)	2.53			
28	<i>Pseudomonas</i>	<i>P. aeruginosa</i>	1294 (3.06)	0.20			
	SG coagulase negative	<i>Staphylococcus equorum</i>	14 (0.03)	0.00	<i>F. magna</i> ^c	1356 (3.21)	0.64
	<i>Staphylococcus</i>						
	<i>Proteus</i>	<i>P. mirabilis</i>	11 (0.03)	0.00			
	Diphtheroids ^b	<i>C. striatum</i>	30433 (71.95)	15.63			
29	<i>S. aureus</i>	<i>S. aureus</i>	62 (65.26)	0.00	-	-	-

30	<i>E. faecalis</i>	<i>E. faecalis</i>	561 (0.46)	0.08	<i>P. aeruginosa</i>	101154 (82.80)	16.36
	<i>Strenotrophomonas maltophilia</i>	<i>S. maltophilia</i>	221 (0.18)	0.06		18103 (14.82)	3.87
	<i>Diphtheroids^b</i>	Not identified	-	-	<i>P. putida</i>		
31	MCO ^b	-	-	-	-	-	
32	Group G	<i>S. dysgalactiae</i>	4783 (28.95)	0.27			
	<i>Streptococcus</i>				<i>F. magna^c</i>	320 (1.94)	0.14
	<i>Enterococcus</i>	<i>E. faecalis</i>	217 (1.31)	0.08			
33	Coagulase negative						
	<i>Staphylococcus</i>	Not identified	-	-			
					<i>B. fragilis^c</i>	70466 (73.74)	18.52
34					<i>M. morgani</i>	10540 (11.03)	3.92
					<i>Peptoniphilus sp. SAHP1^c</i>	3211 (3.36)	2.55
					<i>B. uniformis^c</i>	2221 (2.32)	0.73
35	<i>Pseudomonas spp.</i>	<i>P. aeruginosa</i>	1224 (9.37)	0.17	-	-	0.69
36	Group C	<i>S. dysgalactiae</i>	1239 (6.86)	0.05			
	<i>Streptococcus</i>				<i>E. faecalis</i>	1198 (6.63)	0.30
	<i>S. aureus</i>	<i>S. aureus</i>	252 (1.39)	0.05			
37	Group C	<i>S. dysgalactiae</i>	17990 (76.75)	0.69	<i>S. pyogenes</i>	631 (2.69)	0.10
	<i>Streptococcus</i>				<i>S. agalactiae</i>	547 (2.33)	0.10
	<i>S. aureus</i>	<i>S. aureus</i>	1377 (5.87)	0.32	<i>P. aeruginosa</i>	314 (1.34)	0.05
38	Group C	<i>S. dysgalactiae</i>	20847 (76.04)	0.79	<i>P. corporis^c</i>	2089 (7.62)	1.15
	<i>Streptococcus</i>				<i>B. fragilis^c</i>	1180 (4.30)	0.17
	<i>S. aureus</i>	<i>S. aureus</i>	23 (0.08)	0.01	<i>S. pyogenes</i>	722 (2.63)	0.12
39					<i>S. agalactiae</i>	452 (1.65)	0.08
					<i>F. magna^c</i>	278 (1.01)	0.07
40					<i>K. michiganensis</i>	12527 (38.66)	2.31
					<i>S. anginosus^c</i>	3261 (10.06)	1.02
					<i>K. oxytoca</i>	1667 (5.14)	0.14
41					<i>K. pneumoniae</i>	1390 (4.29)	0.24
					<i>C. freundii</i>	1269 (3.91)	0.27
					<i>E. faecalis.</i>	1182 (3.65)	0.30
42					<i>B. fragilis^c</i>	757 (2.33)	0.13
					<i>K. grimontii</i>	731 (2.25)	1.63
					<i>E. coli</i>	576 (1.78)	0.14
43					<i>P. aeruginosa</i>	463 (1.43)	0.07
					<i>E. hormaechei</i>	442 (1.36)	0.03
					<i>M. morgani</i>	329 (1.01)	0.08
44					<i>F. magna^c</i>	2946 (37.15)	0.54
					<i>S. pettenkoferi^c</i>	129 (1.63)	0.04
					<i>S. pyogenes</i>	103 (1.30)	0.03
45					<i>S. agalactiae^c</i>	97 (1.22)	0.03
46					<i>F. magna^c</i>	4110 (18.01)	0.97
					<i>P. aeruginosa</i>	518 (2.27)	0.07

Species identified by routine microbiology and the corresponding species identified by CMg at the highest classified microbial read proportion are reported. Additional reportable pathogens identified by CMg, as determined by a consultant microbiologist, identified $\geq 1\%$ classified microbial reads are listed. % classified microbial reads = % of species reads over the total number of taxonomically classified microbial reads obtained. Sequencing depth was calculated by dividing the total bp assigned to a species by the total genome size.

a= routine microbiology test reporting refers to a group of *Enterobacter* species as *E. cloacae* for routine reporting purposes This group includes *Enterobacter hormaechei*, listed as the corresponding organism in sample 01.

b=microbial species or genus identification not provided by routine microbiology report.

c= organism reported by routine microbiology if identified in pure culture only.

CMg reported the presence of obligate anaerobic species at a high proportion of microbial reads ($\geq 1\%$) in 17 of the 40 samples (Figure 23). The presence of anaerobic species was specifically reported by routine microbiology in two samples, 13 and 18. CMg reported the presence of anaerobic species in both samples. CMg reported high proportions of anaerobic species in several samples including 04, 07, 33, 39 and 40 in which routine microbiology did not specifically report the presence of anaerobes.

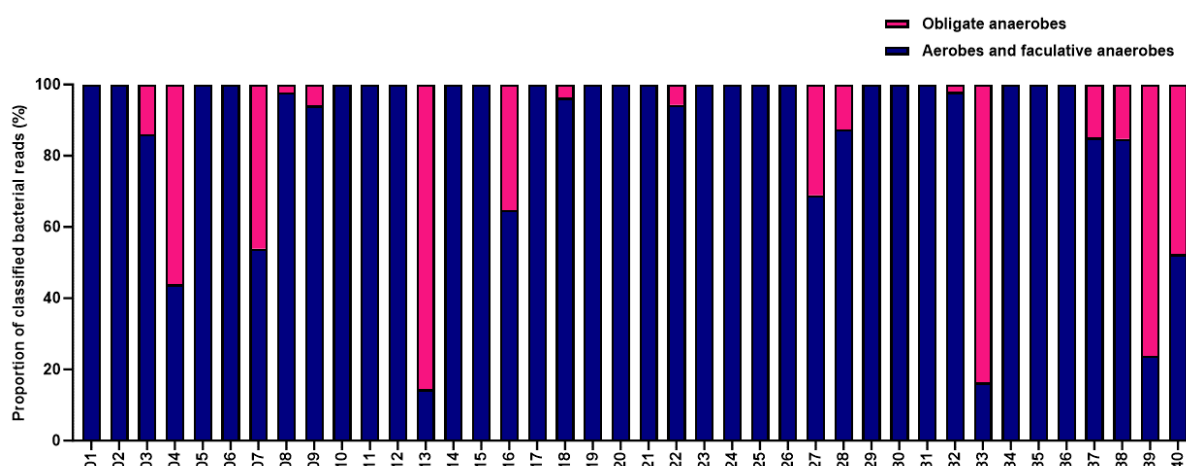


Figure 23. Proportion of aerobic and obligate anaerobic bacteria species identified by CMg.

All bacterial species identified at $\geq 1\%$ of classified microbial reads were included and grouped as either obligate anaerobes or aerobes and facultative anaerobes and presented as the proportion of total bacterial reads. The presence of anaerobic species was reported in samples 13 and 18 in agreement with routine microbiology. Anaerobic species were reported by CMg in samples 04, 07, 33, 39 and 40 in which routine microbiology did not report the presence of anaerobes.

3.4.5 qPCR

To provide additional corroboration of the CMg species identifications compared to microbiology testing, four high priority pathogens were selected for confirmatory testing by qPCR: *E. coli*, *S. aureus*, *P. aeruginosa* and *S. pyogenes*. The full list of qPCR Cq values and target species sequence reads and read lengths are provided in Annex A Table A1. Applying the threshold of ≥ 5 reads assigned to a species for positive identification; CMg achieved a mean sensitivity of 83.82% and specificity of 66.64% across the four species (Table 10). Increasing the threshold to ≥ 50 reads reduced the mean sensitivity to 71.04% but increased the specificity to 90.66%. Applying a threshold of 0.1% of the total classified microbial reads achieved a sensitivity of 76.94% and a specificity of 82.85%. Modification of the thresholds had differing impacts on the sensitivity and specificity according to species. For example, the application of the 0.1% classified microbial read threshold or the ≥ 50 read threshold did not change the sensitivity of 80.00% for *S. aureus*. In contrast, this resulted in an increase in sensitivity for the detection of *P. aeruginosa* from 66.67% to 77.78%. Of the 24 samples that identified a high priority pathogen at a high concentration

by qPCR (<30 Cq), the pathogen was identified by CMg at ≥5 reads in 24 (100%) samples. This reduced to 21 (87.50%) and 22 (91.67%) when applying the ≥50 reads and 0.1% classified microbial read thresholds respectively. The mean sequence read lengths of reads assigned to *S. aureus* and *S. pyogenes*, at 916 and 905 bp, respectively, were lower than both *E. coli* and *P. aeruginosa* at 1236 and 1474 bp, respectively.

Table 10. Sensitivity and specificity of CMg compared to qPCR and routine microbiology testing.

Target species	Sensitivity (%)			Specificity (%)		
	≥5 reads	≥50 reads	0.1% classified microbial reads	≥5 reads	≥50 reads	0.1% classified microbial reads
<i>E. coli</i>	62.50	37.50	50.00	78.13	87.50	84.38
<i>P. aeruginosa</i>	77.78	66.67	77.78	87.10	96.77	96.77
<i>S. aureus</i>	95.00	80.00	80.00	50.00	100.00	80.00
<i>S. pyogenes</i>	100.00	100.00	100.00	51.35	78.38	70.27

Sensitivity and specificity of CMg compared to qPCR and microbiology testing for four high priority pathogen species according to three reporting thresholds; ≥5 reads assigned to the target species, ≥50 reads assigned to the species and 0.1% of the total classified microbial reads assigned to the target species. % classified microbial reads = % of species reads over the total number of taxonomically classified microbial reads obtained.

To determine whether the number of sequencing reads assigned to a high priority pathogen represented the concentration of pathogen DNA in the DNA extract, the correlation of qPCR Cq values to the sequencing reads was calculated. There was a negative correlation between the Cq values and sequence reads for all four species, indicating that the number of reads assigned to the species was representative of the concentration of pathogen DNA in the sample (Figure 24).

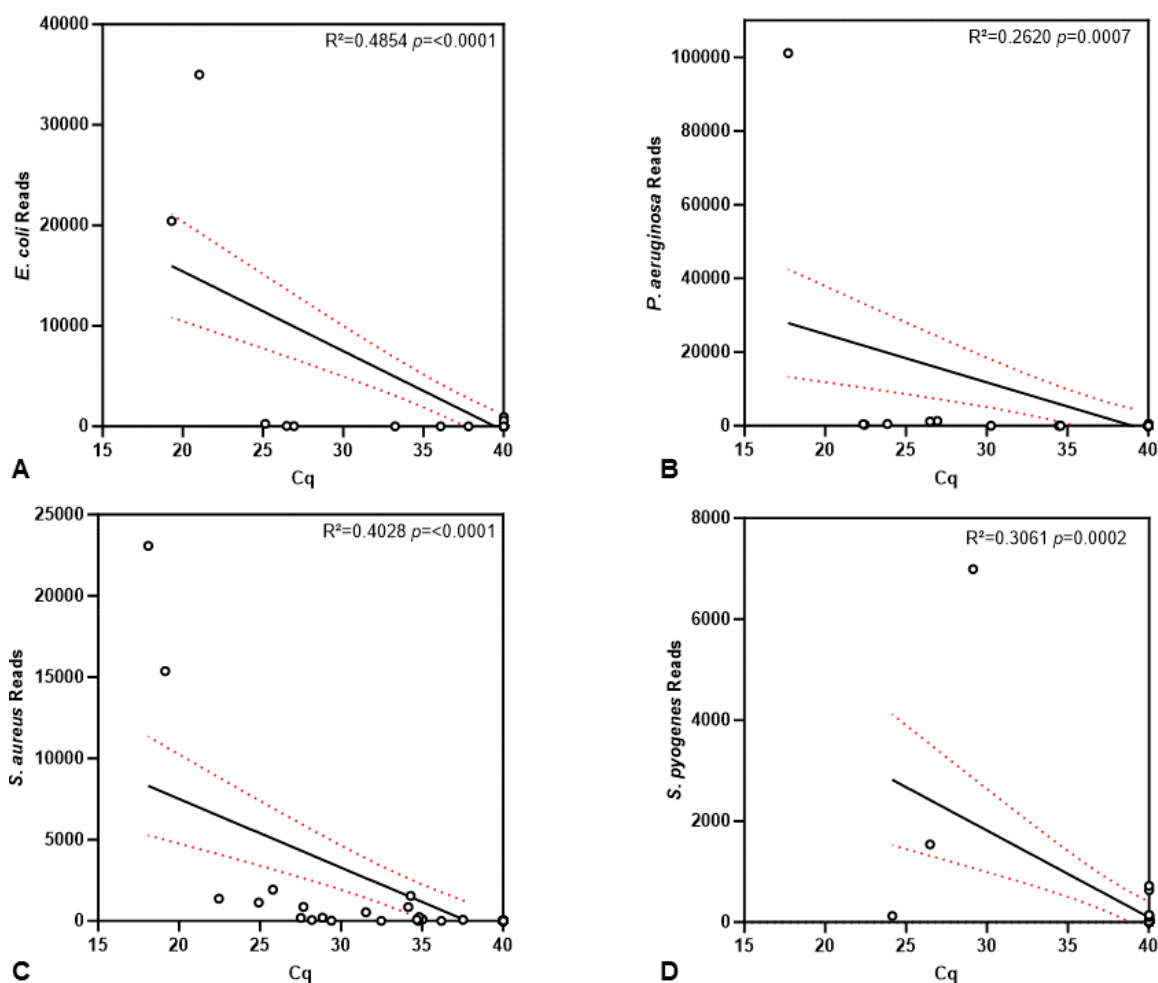


Figure 24. Correlation of qPCR with CMg for the detection of four high priority pathogen species.

The qPCR Cq values and the number of reads assigned to *E. coli* (A) *P. aeruginosa* (B) *S. aureus* (C) and *S. pyogenes* (D) by CMg were compared, showing correlation for all four species. The linear regression line and 95% confidence interval are presented.

3.4.6 Detection of antimicrobial resistance genes

AST was carried out on 34 colonies isolated from 24 samples. Of these 34 colonies, 20 reported resistance to one or more antibiotics and 14 were sensitive to all tested antibiotics. CMg reported the identification of ≥ 1 AMR genes corresponding to the organism and resistance profile identified by AST for 10 (50%) of the 20 resistant colonies. For example, *tetK*, associated with resistance to tetracycline and previously identified in *S. aureus* (222) was identified in sample 23, in which AST had identified *S. aureus* resistant to doxycycline. Across all samples, CMg identified ≥ 1 AMR genes in 32/40 samples. AMR genes associated with resistance to macrolide antibiotics were most identified with 48 AMR gene identifications, followed by streptogramins and lincosamides, with 47 and 44 AMR gene identifications, respectively (Table 11).

Table 11. Results of routine microbiology AST and AMR genes identified by CMg from wound swab samples.

Sample Number	AST Report	AMR genes identified by CMg sequencing	Template Coverage	Depth of coverage (x)	Associated antibiotic drug class resistance
01	<i>E. cloacae</i> : Amoxicillin	<i>fosA</i> <i>bla</i> _{ACT-23}	100 100	23.75 32.11	Phosphonic acid Penam, cephamycin, cephalosporin, carbapenem
02	No growth: AST not carried out.	-	-	-	-
03	<i>S. aureus</i> : Sensitive (Levofloxacin intermediate)	-	-	-	-
04	<i>P. multocida</i> : Sensitive <i>S. aureus</i> : Sensitive (Levofloxacin intermediate)	- -	- -	- -	-
05	MCO: AST not carried out.	<i>bla</i> _{CMY-2} <i>vgaA</i> <i>emrD</i> <i>qacR</i> <i>dfcS1</i> <i>bla</i> _z <i>aac(6)</i> / <i>le</i> <i>bla</i> _z <i>lsaA</i> <i>vgaB</i>	100 100.25 103.04 105.47 100 104.26 104.1 102.36 100 98.07	1 2.05 1.47 1.58 1 1.52 1.82 1.84 59.89 1.64	Penam, cephamycin, cephalosporin, carbapenem Streptogramin, lincosamide, pleuromutilin Phenicol Fluoroquinolone Diaminopyrimidines Penam Aminoglycoside Penam Streptogramin, lincosamide, pleuromutilin Streptogramin, pleuromutilin
06	MCO: AST not carried out.	-	-	-	-
07	<i>S. oralis</i> : Sensitive <i>E. cloacae</i> : Amoxicillin	<i>mefA</i> <i>msrD</i> <i>tetM</i> <i>tetL</i> <i>ermF</i> <i>bla</i> _{CFXA-3}	100.08 101.5 100.16 99.71 100 100.21	5.33 2.6 1.79 3.84 5.19 2.81	Streptogramin, macrolide Streptogramin, macrolide Tetracycline Tetracycline Streptogramin, macrolide, lincosamide Cepharmycin
08	<i>E. coli</i> : Sensitive <i>K. pneumoniae</i> : Amoxicillin <i>Enterococcus</i> species: AST not carried out.	<i>ermX</i> <i>tetD</i> <i>kdeA</i> <i>acrF</i> <i>ermA</i> <i>aph(3')</i> -IIIa <i>ermB</i> <i>oqxA</i> <i>bla</i> _{CFXA} <i>emrE</i> <i>bla</i> _{EC} <i>fosA2</i> <i>bla</i> _{ACT} <i>emrD</i>	100.23 101.35 99.59 100 102.73 100.25 100.95 100.43 95.34 100 100.09 101.41 100.17 99.16	3.68 1.8 2.36 8.96 1.81 2.72 2.01 2.88 1.6 9.96 5.63 2.01 3.03 2.55	Streptogramin, macrolide, lincosamide Tetracycline Multidrug resistance Fluoroquinolone, penam, cephamycin, cephalosporin Streptogramin, macrolide, lincosamide Aminoglycoside Streptogramin, macrolide, lincosamide * Oxacephem, cephamycin Multidrug resistance Cephalosporin Phosphonic acid Carbapenem, cephalosporin, cephamycin, penam Phenicol
09	MCO: AST not carried out.	-	-	-	-
10	<i>S. aureus</i> : Sensitive	<i>tet38</i> <i>bla</i> _z	100.23 101.35	3.68 1.8	Tetracycline Penam
11	<i>S. aureus</i> : Sensitive Coagulase negative <i>Staphylococcus</i> : AST not carried out Group D <i>Streptococcus</i> : AST not carried out	-	-	-	-
12	Coliform: Amoxicillin	<i>bla</i> _{CKO}	100	8.46	Penam
13	Group A <i>Streptococci</i> : Tetracycline	<i>tetM</i> <i>bla</i> _{CepA}	100 100	3.73 7.61	Tetracycline Cephalosporin

14	<i>S. aureus</i> : Levofloxacin	<i>mepA</i> <i>bla_Z</i> <i>bla_{R-1}</i> <i>tet38</i> <i>lmrS</i>	100.07 100 100 100 100	4.69 25.67 31.91 5.47 55.75	Tetracycline, glycylcycline Penam Penam Tetracycline **
15	<i>S. aureus</i> : Levofloxacin	<i>mepA</i> <i>bla_Z</i> <i>bla_{R-1}</i> <i>tet38</i> <i>lmrS</i>	100 100 100 100 103.19	9.77 29.72 28.4 8.48 3.82	Tetracycline, glycylcycline Penam Penam Tetracycline **
16	<i>S. aureus</i> : Levofloxacin	<i>fosB</i>	100.95	2.01	Phosphonic acid
17	AST not carried out.	-	-	-	-
18	<i>E. coli</i> : Amoxicillin, CoTrimoxazole, Ceftriaxone, CoAmoxiclav, Cefpodoxime, Gentamicin	<i>dfrA17</i> <i>Sul1</i> <i>vgaA</i> <i>acrF</i> <i>mdtM</i> <i>bla_{TEM}</i> <i>mphA</i> <i>aph(6)-Id</i> <i>aadA5</i> <i>aac(3)-IId</i> <i>tetW</i> <i>tetA</i> <i>sul2</i> <i>bla_{CTX-M}</i> <i>emrE</i> <i>ermX</i> <i>aph(3'')-Ib</i> <i>bla_{EC}</i> <i>emrD</i>	100 100 100.25 99.97 98.95 100 100 100 100 100.21 100 100 100 100 100 100 99.88 100 100	24.01 14.14 1.45 18.44 14.4 14.55 25.32 26.15 16.61 30.91 4.53 21.1 26.14 5.88 27.67 2.48 26.74 16.08 17.66	Diaminopyrimidine Sulfonamide Streptogramin, lincosamide, pleuromutilins Fluoroquinolone, penam , cephamycin, cephalosporin Nucleoside, fluoroquinolone, phenicol, lincosamide Penam , monobactam, cephalosporin , penem Macrolide Aminoglycoside Aminoglycoside Aminoglycoside Tetracycline Tetracycline Sulfonamide Cephalosporin Multidrug resistance Streptogramin, macrolide, lincosamide Aminoglycoside Cephalosporin Phenicol
19	AST not carried out.	<i>sul1</i> <i>ermX</i>	103.5 100	2.03 3	Sulfonamide Streptogramin, macrolide, lincosamide
20	<i>S. aureus</i> : Erythromycin, Clindamycin	<i>ermX</i> <i>tetW</i> <i>qacR</i> <i>dfrS1</i> <i>sul1</i> <i>ant(2'')-Ia</i> <i>aph(6)-Id</i> <i>tetK</i> <i>ermC</i> <i>tet33</i> <i>ermX</i> <i>ermX</i>	99.88 100 101.59 103.7 100 100 101.91 100 100.14 100 100 100.13	6.46 114.6 2.01 1.96 71.52 133.1 2.02 16.92 3.62 128.19 25.8 256.3	Streptogramin, macrolide, lincosamide Tetracycline Fluoroquinolone Diaminopyrimidines Sulfonamide Aminoglycoside Aminoglycoside Tetracycline Streptogramin, macrolide , lincosamide Tetracycline Streptogramin, macrolide, lincosamide Streptogramin, macrolide, lincosamide
21	Group B <i>Streptococcus</i> : Sensitive <i>S. aureus</i> : Sensitive	<i>tetM</i> <i>bla_{MAL-1}</i> <i>tetK</i> <i>ermC</i> <i>spd</i> <i>fosA7.3</i>	99.69 100 100 100 100 100	1.19 34.79 22.67 35.48 52.6 13.21	Tetracycline Carbapenem Tetracycline Streptogramin, macrolide, lincosamide Aminoglycoside Phosphonic acid
22	AST not carried out.	<i>bla_{DHA-5}</i> <i>mefA</i> <i>ermA</i> <i>catA2</i> <i>ant(6)-Ia</i>	100 101.07 100 99.84 101.87	27.76 2.01 4.45 24.59 1.94	Cephamycin, cephalosporin Macrolide Streptogramin, macrolide, lincosamide Phenicol Aminoglycoside
23	<i>S. aureus</i> : Doxycycline, Erythromycin, Clindamycin, Fusidic acid, Norfloxacin Group G <i>Streptococcus</i> : Doxycycline	<i>tetK</i> <i>tetW</i> <i>ermA</i> <i>ermX</i> <i>aac(3)-XI</i>	100 100 107.1 100.13 100	7.64 67.76 2.06 16.3 6	Tetracycline Tetracycline Streptogramin, macrolide , lincosamide Streptogramin, macrolide, lincosamide Aminoglycoside
24	AST not carried out.	<i>dfrC</i> <i>sul1</i> <i>tet(W)</i> <i>ermC</i> <i>aac(6')-Ie</i> <i>ermB</i>	100 100.71 100 100 100.07 100.41	9.28 2.01 4.74 5.04 6.17 2.79	Diaminopyrimidine Sulfonamide Tetracycline Streptogramin, macrolide, lincosamide Aminoglycoside Streptogramin, macrolide, lincosamide

25	AST not carried out.	<i>dfrC</i>	100	2	Diaminopyrimidine
		<i>sat4</i>	100	8.83	Nucleoside
		<i>bla_Z</i>	100	11.2	Penam
		<i>ermB</i>	100	4.8	Streptogramin, macrolide, lincosamide
		<i>bla_{R-1}</i>	100.06	6.36	Penam
		<i>aph(3')-IIIa</i>	100	4.29	Aminoglycoside
		<i>ant(6)-Ia</i>	100	6.18	Aminoglycoside
26	<i>S. aureus</i> : Norfloxacin Group A <i>Streptococcus</i> : Doxycycline	<i>bla_{TEM}</i>	100	3	Penam, monobactam, cephalosporin, penem
		<i>tetM</i>	100	6.12	Tetracycline
27	<i>S. aureus</i> : Doxycycline, Erythromycin , Clindamycin Group G <i>Streptococcus</i> : Doxycycline	<i>ermX</i>	100	11.97	Streptogramin, macrolide, lincosamide
		<i>bla_{TEM}</i>	102.67	1.84	Penam, monobactam, cephalosporin, penem
		<i>ermA</i>	100	5.73	Streptogramin, macrolide , lincosamide
		<i>ermX</i>	100.13	109.45	Streptogramin, macrolide, lincosamide
		<i>tetW</i>	100	46.8	Tetracycline
28	AST not carried out.	<i>ermX</i>	100	7	Streptogramin, macrolide, lincosamide
		<i>iri</i>	100	2	Rifamycin
		<i>tetM</i>	100.16	4.14	Tetracycline
		<i>sat4</i>	100.18	6.01	Nucleoside
		<i>tetW</i>	100	33.01	Tetracycline
		<i>ermA</i>	100	25.91	Streptogramin, macrolide, lincosamide
		<i>aph(3')-IIIa</i>	100	5.81	Aminoglycoside
		<i>ant(6)-Ia</i>	100	2.89	Aminoglycoside
		<i>cmx</i>	102.47	2.34	Phenicol
		<i>ermX</i>	100.13	15.15	Streptogramin, macrolide, lincosamide
		<i>ermX</i>	98.83	1.71	Streptogramin, macrolide, lincosamide
		<i>crpP</i>	100.51	2.01	Fluoroquinolone
		<i>aac(3)-XI</i>	100	6.78	Aminoglycoside
29	<i>S. aureus</i> : Sensitive	-	-	-	-
30	AST not carried out.	<i>bla_{PDC}</i>	100	2.28	Monobactam, carbapenem, cephalosporin
		<i>bla_{OXA-50}</i>	100	27.1	Penam, cephalosporin
		<i>mexA</i>	100	46.28	***
		<i>fosA</i>	100	12.8	Phosphonic acid
		<i>catB7</i>	100	19.85	Phenicol
		<i>aph(3')-IIb</i>	100.12	3.9	Aminoglycoside
		<i>mexX</i>	100	6.54	****
		<i>mexE</i>	100	7.04	Diaminopyrimidine, phenicol, fluoroquinolone
31	MCO: AST not carried out.	<i>ermC</i>	105.03	1.94	Streptogramin, macrolide, lincosamide
32	Group G <i>Streptococcus</i> : Doxycycline	<i>tetM</i>	99.69	2.64	Tetracycline
		<i>tetW</i>	100	12.42	Tetracycline
		<i>ermX</i>	100.13	8.34	Streptogramin, macrolide, lincosamide
33	MCO: AST not carried out.	<i>tetD</i>	100	6.72	Tetracycline
		<i>bla_{DHA}</i>	100	7.49	Cephalosporin, cephamycin
		<i>sat4</i>	100.37	2	Nucleoside
		<i>tetW</i>	100	12.67	Tetracycline
		<i>tetM</i>	100.16	3.63	Tetracycline
		<i>ermA</i>	100.82	2.01	Streptogramin, macrolide, lincosamide
		<i>aph(3')-IIIa</i>	100.5	2.01	Aminoglycoside
		<i>tetQ</i>	100	14.38	Tetracycline
		<i>tetJ</i>	100.33	1.96	Tetracycline
		<i>bla_{CFXA-3}</i>	100	405.12	Cephamycin
		<i>bla_{CepA}</i>	100	27.96	Cephalosporin
		<i>ermX</i>	100.13	12.35	Streptogramin, macrolide, lincosamide
		<i>aac(3)-XI</i>	100	1	Aminoglycoside
34	<i>Pseudomonas</i> spp.: Sensitive (ciprofloxacin, ceftazidime, piperacillin-tazobactam, gentamicin intermediate)	<i>tetW</i>	100	15.68	Tetracycline
		<i>ermX</i>	100.13	6.92	Streptogramin, macrolide, lincosamide
35	Group C <i>Streptococcus</i> : erythromycin, clindamycin. <i>S. aureus</i> :	<i>qacB</i>	101.94	1.41	Fluoroquinolone
		<i>dfrS1</i>	100	5.51	Diaminopyrimidine
		<i>tetW</i>	100	5.95	Tetracycline
		<i>bla_Z</i>	100	8.44	Penam
		<i>bla_{R-1}</i>	101.37	2.67	Penam
		<i>qacC</i>	100	11.83	Fluoroquinolone
		<i>bla_Z</i>	100	8.12	Penam
		<i>bcrC</i>	100	3	Peptide

	Sensitive (levofloxacin intermediate)	<i>bcrB</i> <i>ermX</i> <i>aac(3)-XI</i>	100 100.13 100	5 10.37 3	Peptide Streptogramin, macrolide, lincosamide Aminoglycoside
36	Group C <i>Streptococcus</i> : erythromycin, clindamycin. <i>S. aureus</i> : Sensitive (levofloxacin intermediate)	<i>ermX</i> <i>ermA</i> <i>fosB</i> <i>catB7</i> <i>mepA</i>	100.23 101.23 100.71 100.31 101.18	1.66 1.01 2 1.84 2.33	Streptogramin, macrolide, lincosamide Streptogramin, macrolide, lincosamide Phosphonic acid Aminoglycoside Tetracycline, glycylcycline
37	Group C <i>Streptococcus</i> : erythromycin, clindamycin. <i>S. aureus</i> : Sensitive (levofloxacin intermediate)	<i>ermA</i>	100.27	1.84	Streptogramin, macrolide, lincosamide
38	MCO: AST not carried out.	<i>tetM</i> <i>bla_{oxy-1-7}</i> <i>tetW</i> <i>aph(3')-IIIa</i> <i>ant(6)-Ia</i> <i>aad9</i> <i>ermF</i>	101.09 101.35 99.96 100.68 101.64 100.99 100	3.99 9.84 2.02 1.8 2.01 15.95 16.62	Tetracycline Penam, cephalosporin, monobactam Tetracycline Aminoglycoside Aminoglycoside Aminoglycoside Streptogramin, macrolide, lincosamide
39	MCO: AST not carried out.	<i>ermA</i> <i>ermC</i> <i>qacC</i> <i>ermX</i> <i>qacG</i> <i>tetM</i>	100 100 100 100.7 100 100.52	5.74 14.53 41.27 2 1 2.75	Streptogramin, macrolide, lincosamide Streptogramin, macrolide, lincosamide Fluoroquinolone Streptogramin, macrolide, lincosamide Fluoroquinolone Tetracycline
40	MCO: AST not carried out.	<i>ermX</i> <i>tetM</i> <i>ermA</i> <i>ermC</i> <i>qacC</i> <i>tet33</i> <i>qacG</i>	100 100 100 100 100 97.55 100	20.74 5.49 14.48 17.01 17.6 1.17 2.98	Streptogramin, macrolide, lincosamide Tetracycline Streptogramin, macrolide, lincosamide Streptogramin, macrolide, lincosamide Fluoroquinolone Tetracycline Fluoroquinolone

Results of routine microbiology AST from each routine wound swab and the AMR genes identified from the corresponding additional wound swab sample are reported. The antibiotic class to which each AMR gene is associated, according to the CARD is reported. AMR genes associated with resistance to the antibiotic drug classes identified by AST in the same sample are highlighted in bold.

* Tetracycline, diaminopyrimidine, glycylcycline, nitrofurantoin, fluoroquinolone

** Aminoglycoside, macrolide, phenicol, diaminopyrimidine, oxazolidinone

*** Peptide, sulfonamide, Diaminopyrimidine, phenicol, aminocoumarin, tetracycline, penam, cephalosporin, carbapenem, monobactam, fluoroquinolone, macrolide, penem, cephamycin

**** Aminoglycoside, phenicol, tetracycline, carbapenem, fluoroquinolone, macrolide, penam, cephamycin, cephalosporin

3.4.7 Species diversity

The species α -diversity in CMg data sets from each sample was determined by calculating the Shannon diversity index. There was a reduction in the α -diversity from samples in which routine microbiology testing reported the presence of a colony according to species, genus or phenotypic characteristic compared to samples that reported no growth or MCOs only (Figure 25A). There was also a reduction in the species α -diversity associated with the sampling location, with the species diversity lower in samples collected at BNH compared to SDH (Figure 25C). However, there was no change in α -diversity observed between patients that had undergone antibiotic treatment prior to sample collection (Figure 25B).

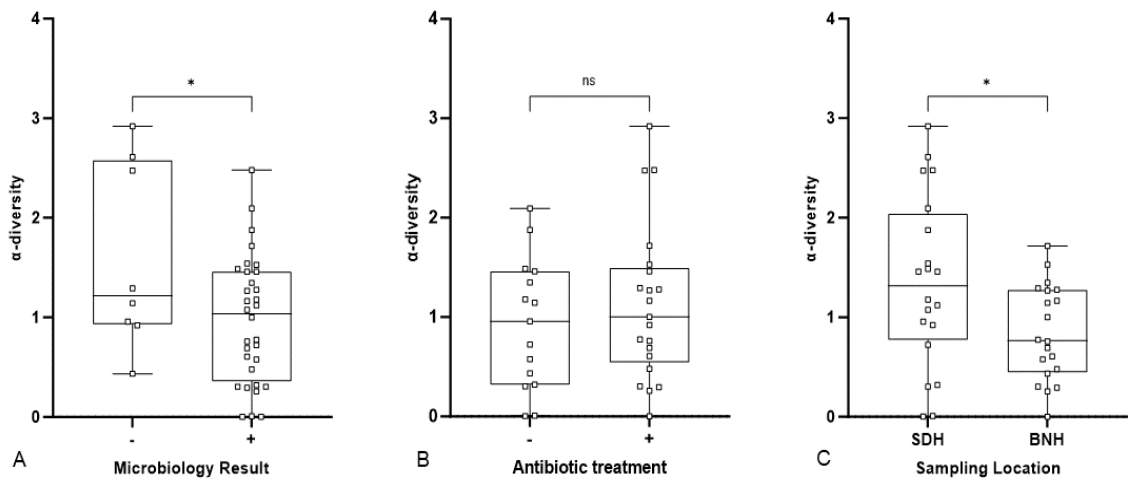


Figure 25. Comparison of wound swab sample species α -diversity.

The Shannon diversity index was calculated for each wound swab sample and the results compared according to microbiology test results (A), antibiotic treatment prior to sample collection (B) and sampling location (C). The antibiotic history of four participants was unknown and these samples were omitted from the antibiotic treatment comparison. Increased α -diversity was observed in samples that reported no growth or colonising microorganisms only by routine microbiology and samples collected from SDH.

Error bars represent the minimum and maximum values. ns $p > 0.05$. * $p < 0.05$. Microbiology and sampling location $n = 40$. Antibiotic treatment $n = 36$.

3.5 Discussion

Microbiology based diagnostic tests are well established for use in fixed laboratories and enable a plethora of tests to identify the causative pathogen and inform treatment. Therefore, these methods continue to be the gold standard for wound infection diagnosis. However, the equipment, time and BMS expertise required to grow and identify the microorganisms, particularly for fastidious and fungal species, mean that there are opportunities to develop additional improved methods. This is critical for austere environments such as deployed military MTFs in which time and physical space are at a premium. CMg offers a potential alternative to microbiology-based testing and has been applied to numerous sample types, including respiratory, blood and cerebrospinal fluid (223-225).

This study presents a CMg workflow that identifies reportable wound pathogen species with good concordance to routine microbiology testing and qPCR. Wound swab samples were processed directly without the need for a growth step, removing the requirement for incubator equipment, reducing the time to result, reaching the DNA sequencing step in approximately four hours, and enabling untargeted pathogen detection. The identification of AMR genes, key to informing rapid antimicrobial treatment, was demonstrated and additional data provided by CMg, including the microbial species α -diversity, could be

applied to improve the clinical interpretation of CMg outputs. Additional studies applying CMg to wound swab samples to inform clinical interpretation and method optimisation focused on increased automation are now required to enable the routine use of CMg for wound infection diagnosis in austere environments.

Regarding the sample processing, there was variation observed in the total DNA yield obtained from each sample. This complicated the processing of the samples in batches and may require the use of a different library preparation method depending on the DNA obtained from the sample. Future optimisation of the methods may require the use of specific DNA yield thresholds to determine the library preparation method used. In addition, the proportion of reads assigned to the human genome varied between samples. This may be due to reduced effectiveness of the host-depletion or a reduced yield of bacteria DNA. For example, wound swabs obtained from patients with a low bacterial load may result in higher proportions of human DNA due to the absence of bacteria in the original sample. It was observed that the samples with the highest DNA yields resulted in sequence data with the lowest proportion of human reads and vice versa, providing further evidence for this.

Consistent with previous studies (226, 227), CMg identified a greater diversity of species than was reported by routine microbiology, including additional reportable pathogen species in microbiology negative samples, fungal species and anaerobic species. Routine microbiology is selective in nature, with the selection of the growth media, evaluation of plates by the BMS and reports to clinicians all focused on the pathogen species relevant to the clinical context of the patient (184). For example, anaerobic species were identified in several samples by CMg in which they were not reported by routine microbiology. Anaerobic species are only tested for according to the sample type, including wound swabs of traumatic wounds or swabs of pus. Therefore, the additional identification of anaerobic species by CMg may not reflect the detection of additional anaerobic species, but that routine microbiology did not test for these species at all. CMg therefore represents an opportunity to better understand the presence of anaerobic species in wound infections and the impact of their presence on patient outcomes. Further to this, pathogen species may be ignored by microbiology testing if not considered relevant or missed completely if mixed in with morphologically similar colonising species. In-keeping with the UK SMI recommendations, microbiology reporting may only specify the genus or phenotypic characteristics of a microorganism, further complicating the interpretation of CMg results.

To understand the CMg outputs in relation to microbiology testing, four high priority pathogen species were selected for species-specific qPCR testing, resulting in non-concordant results between culture, qPCR and CMg. Species misidentifications are a common issue for metagenomics studies, especially when pathogen species may share similar genomic regions to commensal species that are also present in high abundance.

To minimise false-positive identifications, previous CMg studies have applied reporting thresholds, such as the total number of reads or the proportion of microbial reads assigned to a species. The reporting threshold of ≥ 5 reads assigned to a species applied in this study resulted in an average sensitivity of 83.82% and specificity of 67.19% across the four high priority species. Application of the reporting thresholds of ≥ 50 reads and 0.1% of classified microbial reads increased the specificity for all tested species. However, changing the reporting thresholds had different impacts on each tested high priority species due to the higher proportion of false-positive samples for *S. aureus* and *S. pyogenes* with ≥ 50 reads, resulting in increased specificity observed in these species when this threshold was applied, as an example. An improved understanding of the interpretation of the identification of wound pathogens, including the relationship between the proportions/total number of sequence reads assigned to a specific pathogen and the cause of the wound infection, is required. This should include the development of a comprehensive list of wound pathogens and statistically robust reporting thresholds, developed by the scientific and clinical community, for optimal sensitivity and specificity. Development of these reporting thresholds would aid in the interpretation of CMg to inform infection treatment. The output of CMg workflows can include a myriad of microbial species of variable relevance to the treatment of the infection. Improved understanding of the consequences of the identification of each species and whether the treatment should be adjusted accordingly will be key to the implementation of CMg for wound infections.

Invasive fungal wound infections have been reported amongst military personnel (182). The identification of fungal species by routine microbiology requires culturing on additional growth media, using additional time and resources and may only be reported generally as yeasts, as with sample 24 in this study. The CMg workflow enabled the identification of fungal pathogens at the species level, including in sample 18 where routine microbiology did not report the presence of a fungal species as the sample was not tested for fungal species. The rapid identification of fungal species without the need for additional testing could enable clinicians to target fungal pathogens faster, reducing the spread of invasive fungal wound infections.

The untargeted nature of CMg enables the provision of additional clinically relevant information. One metric that could be applied in a clinical context is the diversity of microbial species present at a wound. There was a significant reduction in the species α -diversity between samples that tested positive by routine microbiology testing compared to negative samples. Previous studies have linked bacterial diversity, as a component of wound bioburden, to poor patient outcomes including delayed wound healing (228). The reduced α -diversity observed in microbiology positive samples may be related to the increased proportion of reads assigned to the pathogen species as the infection progresses. Therefore, α -diversity could be applied to provide clinicians with additional

information on the progress of a wound infection. There was also a significant difference in the α -diversity from samples collected at each sampling location. This may reflect differences in the environmental contaminants that initially colonise the wound site at each location. Therefore, interpretation of the α -diversity when evaluating a wound infection may need to be location specific. Surprisingly, previous antibiotic treatment did not result in changes to species diversity. It may be expected that antibiotic treatment would reduce species diversity through the removal of sensitive bacteria species, as observed at other sites, such as the gut microbiome (229). However, it may be that removal of sensitive species at the wound site by antibiotics results in increased colonisation by resistant bacteria and fungal species, without impacting the overall species diversity. Further work should be carried out to understand the relationship between species α -diversity, wound infection severity, antibiotic treatment and delayed healing.

In addition to identifying the causative pathogen, determining the antibiotic resistance is key to enabling optimal treatment of a wound infection. These results have demonstrated that AMR genes associated with the resistance observed by AST are identified from CMg sequencing data. However, further understanding of the relationship between the presence of an AMR gene and the phenotypic resistance profile of a pathogen is required. AMR genes were also identified in samples that did not report the presence of a wound pathogen and in samples with colonies determined to be sensitive by AST, potentially originating from the commensal species present in the sample. Further work should be carried out to better understand the consequence of AMR gene identification and identify a comprehensive list of reportable AMR genes associated with specific pathogen species for wound infections, as has been done for other sample types (230). Targeting a specific list of AMR gene targets would reduce the complexity of the bioinformatics analysis and the reporting of the CMg results.

A key aim of this study was to ensure that the methods developed minimised the method complexity so that the workflow can be adapted for use in a deployed military environment. Adapting methods for use in austere environments could also open up the application of CMg-based diagnostics in additional environments with limited access to permanent healthcare facilities, such as for remote expeditions. The caveat was that the sensitivity and specificity of pathogen detection must remain as high as possible to inform patient treatment. Several steps, ordinarily included in CMg workflows that involved large, bulky equipment, were omitted. Specific reagent kits developed for ease-of-use, including the Genomiphi WGA kit and the nanopore Rapid Barcoding library preparation kit were selected. The use of nanopore-based sequencing itself was employed due to the reduced size of the equipment, robustness against movement and reducing the total time to result through enabling real-time data analysis, having previously been applied for point-of-care and in-field diagnostics for a range of sample types (231-233). Further optimisation of the

workflow should now be carried out. In a routine clinical context, the use of a spike in control would be a key requirement and future iterations of the workflow should incorporate a spike-in control allowing the identification of inter-run variation and to differentiate true negative and false negative results. Method optimisation to reduce the hands-on steps and method complexity via increased method automation will reduce the burden on deployed BMSs. The workflow in this study required approximately four hours to start sequencing and the samples were sequenced for 16 hours to ensure that sufficient reads were generated. The use of a real-time analysis pipeline would further reduce the time to result, allowing sequencing runs to continue until a reportable pathogen is identified at the reporting threshold. With this, it is feasible that the methods developed in this study could be adapted for the rapid diagnosis of wound infections in deployed environments.

3.6 Conclusions

The work reported in this chapter includes the development and proof of concept evaluation of a CMg workflow for the diagnosis of wound infections. The workflow required approximately four hours to reach the start of DNA sequencing using methods and equipment suitable for use in austere environments. The results showed good concordance with routine microbiology and qPCR testing and enabled the identification of AMR genes. CMg identified additional reportable wound pathogens including fungal pathogens, demonstrating the advantages of applying untargeted, non-selective diagnostic methods. Whilst this study shows the potential for CMg for deployed diagnostics, increased method automation, understanding of the interpretation of CMg results including a threshold for species identification and a list of reportable AMR genes is required before CMg can be routinely implemented in a deployed healthcare setting.

4 Comparison of stool sample storage methods for deployed clinical metagenomics

4.1 Contribution statement

The work detailed in this chapter was carried out as a collaborative study with partners from the Liverpool School of Tropical Medicine, the Defence Centre of Pathology and the Quadram Institute Bioscience. Stool samples were previously collected from UK military personnel as part of routine testing, including by the BioFire FilmArray GI panel, by Sergeant Romeo Toriro. The stool samples were then provided to Dstl for CMg analysis under a Material Transfer Agreement by the Liverpool School of Tropical Medicine. Some of the CMg wet-lab work was carried out by Emily Rowlands, Dstl, as part of a six-month graduate scheme project under the direct supervision of Carl Halford. Carl Halford was responsible for the experimental design and results interpretation, with assistance from Emily Rowlands. The bioinformatics analysis was carried out by Dr. Thanh Le-Viet (Quadram Institute Bioscience) in close collaboration with Carl Halford using a bioinformatics pipeline that Dr Thanh Le-Viet previously developed.

4.2 Introduction

4.2.1 Traveller's diarrhoea

TD is defined as the passage of ≥ 3 loose stools within a 24 hour period, alongside additional symptoms including one or a combination of vomiting, abdominal cramps, nausea, fever, blood/mucus in the stools and faecal urgency, experienced whilst travelling or within ten days of travel (234). Annually, approximately 100-500 million people travel to low- or middle-income countries for business or touristic purposes and between 10-40% of these traveller's develop TD (235, 236). The destination of travel greatly affects the incidence of TD. Regions including South-East Asia, North and West Africa and South and Central America are high-risk regions. For example, 15.2% of a group of traveller's to Mexico, monitored for TD as part of a study, reported experiencing all-cause diarrhoea during travel (237). Activities whilst at a destination also impact the risk of developing TD, with eating at all-inclusive buffets considered high-risk (238). Although often self-limiting, GI infections, including TD, can have significant morbidity and mortality rates amongst vulnerable populations, including young children, the elderly and the immunocompromised (239).

Diarrhoeal diseases are a common medical complaint amongst deployed military personnel and have been throughout human history, often having a significant impact on the outcome of conflicts. During the American civil war, 44,558 Union soldiers were recorded as having died from diarrhoea or dysentery, compared to 110,070 combat-related deaths (240). The impact of diarrheal diseases have continued to affect deployed

personnel into modern times, affecting 76.8% and 54.4% of deployed U.S troops in Iraq and Afghanistan, respectively, between 2003 and 2004, with 80% of patients requiring medical care (241).

Advancements in the understanding of the microbiology and pathophysiology of diarrheal diseases have led to more effective methods of prevention and treatment. Prevention of diarrheal disease focuses on improving the levels of sanitation and hygiene, including measures such as water chlorination to prevent pathogen transmission, shown to significantly reduce the incidence of diseases such as cholera (242). More recently, modulation of the gut microbiota through the pre-emptive use of pre and probiotics as a preventative measure for diarrheal disease have been investigated, although meta-analyses have concluded that the evidence for the reduction of diarrheal disease using these supplements is weak (243). Treatments include oral rehydration solutions with added electrolytes, of which an imbalance is a significant cause of mortality during acute diarrhoea, particularly amongst vulnerable patients in resource-poor settings (244). Anti-diarrheal medications, including loperamide, a piperidine derivative that slows intestinal motility through the activation of μ -opiate receptors, have also been developed to manage the symptoms (245). Antibiotics can also be administered in severe cases of bacterial diarrhoea, although this carries the risk of inducing further imbalances in the gut microbiome and causing additional complications, including infection with *Clostridioides difficile* (246).

As a result of these advancements, the incidence and severity of diarrheal disease, particularly in economically developed nations, have significantly reduced. However, whilst military populations do not include those most vulnerable to TD due to the selection criteria, the relatively poor hygiene in deployed military environments means that TD is still common, resulting in reduced operational capability, lost ability to work and in some cases additional GI disorders including irritable bowel syndrome, constipation and dyspepsia, increasing pressures on the provision of deployed healthcare (247-249). Therefore, TD is a key concern for deployed military healthcare.

4.2.2 Aetiology of traveller's diarrhoea

GI infections, including TD, are caused by bacteria, viruses or parasites that infect the gut (usually the stomach or intestines) (250), in some cases releasing toxins, resulting in inflammation. The majority of cases of TD in military populations are linked to bacterial pathogens, varying according to region (251), with diarrheagenic *E. coli* being the most prevalent (248). *E. coli* is a common component of the human gut microbiome (252), however, some *E. coli* strains have acquired virulence characteristics that enable them to cause GI infection. The type and combination of virulence characteristics are used to define the strains of pathogenic *E. coli* into different pathotypes. Each pathotype has

different disease presentation, including the mode of transmission, symptoms, host preferences and global prevalence, as reviewed by Gomes *et al.* 2016 (253) and Jesser and Levy, 2020 (254).

Whilst diarrheagenic *E. coli* is the most common cause of TD amongst military personnel and civilian travellers globally, there are variations observed by region. *Campylobacter* spp. are a common cause of TD in South-East Asia, accounting for 15-63% of cases (249) and the incidence is increasing over time (255). By comparison, in Sub-Saharan Africa, West Africa, East Africa, Middle Africa and Southern Africa, *Campylobacter* spp. account for between 0-9% of cases (249). Viruses are also known to cause TD, although less common than bacterial pathogens. However, viral TD is associated with more severe symptoms and misuse of antibiotics (256). These include pathogens such as norovirus, characterized by acute nausea, abdominal cramps, vomiting and diarrhoea, with outbreaks often spreading rapidly, particularly in healthcare environments, due to its low infectious dose (257). Parasites also cause TD, although less commonly than both bacterial and viral pathogens, and are associated with persistent symptoms (258). Examination of persistent TD identified the parasite *G. lamblia* as the most common cause, with other parasites including *Blastocystis hominis*, *Cyclospora* and *Entamoeba histolytica* also implicated (259).

The causative pathogen of TD is often not identified. As the disease is self-limiting, it is often unreported or medical advice is not sought (260). Even when medical treatment is required, factors including poor stool culture sensitivity and the complexity of differentiating pathotypes may prevent the pathogen from being accurately identified and an appropriate treatment pathway being determined.

4.2.3 Diagnosis of traveller's diarrhoea

The diagnosis of TD is challenging. Often, the symptoms associated with different aetiological agents are very similar, but the optimal treatments are different, such that it is important that the pathogen is correctly identified to aid in treatment. In cases of bacterial TD, primary diagnostics includes culturing the causative pathogen from a stool sample using selective media. Media is selected to enable the identification of common bacterial causes of TD including *Salmonella*, *Shigella*, *Campylobacter* and Shiga-toxin producing *E. coli*, with more specific media selected based on the context of the patient, including travel history and known exposure events (261). Faecal samples are diluted and spread across the plates and incubated in different conditions (dependent upon the target pathogen) to identify individual colonies. As an example, xylose lysine deoxycholate agar is used, and cultured at 37°C for 24 hours in air for *Salmonella* and *Shigella* species. The recommended process for UK laboratories include 13 different combinations of growth media dependent upon the suspected causative pathogen (261). Once cultured, an

isolated colony can undergo additional testing including AST to determine the presence of resistance to specific treatments. Testing for the enteric parasites *G. lamblia* and *Cryptosporidium* is recommended alongside the common bacterial pathogens during primary testing. This can be carried out using nucleic acid amplification tests including PCR and/or antigen-based tests, with the methods varying depending upon the laboratory (262). However, despite being less sensitive and requiring time and expertise to carry out (263), microscopy is more commonly applied for the identification of parasites. For *Cryptosporidium*, a faecal sample is prepared by staining with auramine phenol and observing the stained sample under a microscope. Under observation the parasites fluoresce compared to the dark background of the sample and additional confirmation is achieved by Ziehl–Neelsen staining (264).

In deployed military environments, diagnosis of TD relies upon multiplex PCR panels, including the BioFire FilmArray GI panel (265). The stool sample is mixed with the sample buffer and added directly to the sample pouch and the platform carries out the DNA extraction and multiplex PCR without further input from the operator. The panel includes 22 targets, including common bacterial pathogens such as *C. difficile* and *Salmonella*. Different diarrheagenic *E. coli* pathotypes, including enteroaggregative and enterotoxigenic *E. coli* are differentiated through the detection of virulence determinant genes. The panel also enables the detection of parasites and viral pathogens. A multicentre evaluation found the panel to be highly sensitive and specific, achieving 94.5–100% sensitivity and >97.1% specificity, depending on the target (25). However, whilst highly sensitive and specific, this platform only allows for the detection of targeted pathogens and does not include the provision of additional clinically relevant information, such as the presence of AMR genes and quantification of the targeted pathogens. Pathogens that have been genetically modified through natural mutation or intentional genetic manipulation might also be missed. Further, each platform has the capacity to run one sample pouch per hour, reducing testing throughput, which is critical during outbreaks in which many samples may require testing within a short time frame.

In addition to PCR, microscopic identification of parasites can also be carried out. Parasitic ova, larvae, cysts and oocysts are concentrated from stool samples using devices such as the MIDI PARASEP® before the sample is stained and examined under the microscope. However, this method achieves moderate sensitivity of 70–80% and requires a trained BMS to carry out the relatively complex sample preparation and use of sensitive equipment that is unsuitable for high throughput diagnostics (263). Therefore, additional rapid, untargeted diagnostic methods would be beneficial for the diagnosis of TD amongst deployed military personnel.

4.2.4 Clinical metagenomics for the diagnosis of traveller's diarrhoea

As previously discussed, CMg, the untargeted sequencing of the genetic content of a sample without the prior isolation of an organism (59) has the potential to provide an additional option for deployed TD diagnosis.

Due to its importance to human health across a range of areas, the gut microbiome is well-studied. As the field of NGS has progressed, metagenomic DNA sequencing has been applied to stool samples for a variety of purposes, including determining the composition of the gut microbiome. These have included the changes to the gut microbiome related to international travel and incidences of diarrhoea (266, 267), the altered gut microbes associated with a range of common intestinal diseases (268), cancer diagnosis (269) and diarrhoeal pathogen identification (270). Shotgun metagenomic sequencing identified *C. difficile* with a sensitivity of 90.9% and 86.3%, respectively whilst enabling the co-detection of additional pathogens and antibiotic resistance genes. Metagenomics has also been applied for the identification of enteric *E. coli* infections, enabling the reliable detection of the cultured isolate with a sensitivity of 97%. However, the presence of the virulence genes associated with enterotoxigenic *E. coli* on plasmids prevented their inclusion in the metagenome-assembled genomes, demonstrating a limitation of metagenomics for the differentiation of *E. coli* pathotypes (64). Metagenomic DNA sequencing has also been applied for the detection of enteric parasites, for outbreak investigations and the identification of parasites in wastewater sources (271, 272).

The challenges associated with the detection of enteric pathogens, particularly closely related strains and pathotypes, may be one reason that CMg is not currently routinely applied for the diagnosis of TD, in addition to the limitations associated with complex wet-laboratory procedures and equipment requirements, particularly for deployed MTFs.

4.2.5 *Cryptosporidium*

Species belonging to the *Cryptosporidium* genus, the causative agents of cryptosporidiosis, are protozoan enteric obligate parasites with a global distribution and are a significant cause of TD with limited effective therapeutic options available (273, 274). First identified as a human pathogen in the 1970s (275), there are at least 44 *Cryptosporidium* species described, with 19 identified in humans. Two species, *C. hominis* and *Cryptosporidium parvum* are responsible for ~95% of infections in humans, with *C. hominis* being the most common (276). Common symptoms include watery diarrhoea, which can be severe and prolonged, abdominal pain, myalgia, weakness and headaches (275). Transmission is via the faecal-oral route through water, food or direct contact with human or animals contaminated with the oocyte life cycle stage. The life cycle initiates when the oocyst is ingested and undergoes excystation before entering the intestinal epithelial cells and undergoing asexual reproduction and subsequent sexual reproduction

life stages. The oocysts are shed in faeces so that the life cycle can repeat (Figure 26) (277).

Cryptosporidiosis is associated with significant morbidity and mortality amongst vulnerable populations, including children under five years of age and the immunocompromised. It was estimated to have caused over one million deaths globally between 2005 and 2015, with close to 50% being children under five (278). Although *Cryptosporidium* spp. have a global distribution, the majority of cases occur in low-income countries, with prevalence greatest in countries including Mexico, Nigeria and Bangladesh (279).

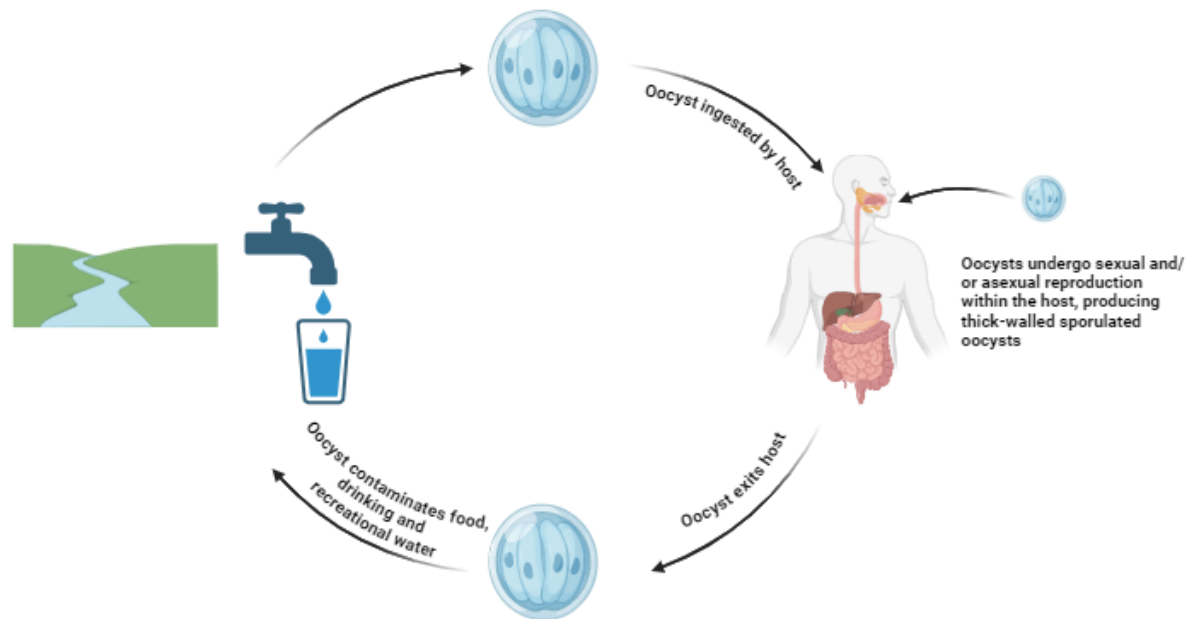


Figure 26. The life cycle of *C. hominis*.
Figure created with Biorender.com.

4.2.6 *Cryptosporidium* outbreak

In February 2022, there was an increase in the reports of profuse watery diarrhoea amongst British Army personnel stationed at the British Army Training Unit Kenya (BATUK), in Nanyuki, Kenya, approximately 200 Km north of Nairobi. Over 69 days, 172 of the 1200 personnel displayed diarrhoeal symptoms (280). An outbreak investigation was conducted, including the on-site testing of 106 primary stool samples by the BioFire FilmArray GI panel and molecular characterization of repatriated samples by the *Cryptosporidium* Reference Unit (CRU; Swansea, UK).

Of the 106 samples tested, 63 were found to be positive for *Cryptosporidium*. This included 38 samples that tested positive for *Cryptosporidium* solely and 25 that tested positive for *Cryptosporidium* and an additional pathogen. Twenty-seven tested positive for non-*Cryptosporidium* pathogens and 16 were negative. Of the 63 *Cryptosporidium* positive samples, 58 were subsequently genotyped as *C. hominis*, with the remaining not

genotypable. Of the *C. hominis* positive samples, 27 were sub-typed; 26 as the unusual ImA13G1 subtype and one as the more common *gp60* subtype IaA11G3T3 (281).

Epidemiological investigations were also carried out to determine the source of the infection. This included the analysis of individual sleeping arrangements, activities and ablution locations and exposure to the suspected point source or close contact with anyone who had visited the location. It was concluded that the rare *C. hominis* ImA13G1 subtype responsible for the outbreak originated from faecal contaminated recreational swimming water. Public health interventions, including restricted access to the outbreak source, PCR testing of food handlers stool samples, isolation of symptomatic cases and encouraging improved personal hygiene were then implemented to limit the outbreak (280).

During the sample collection process, primary stool samples were prepared for long-term storage to allow for sample repatriation to the UK for further testing and for the comparison of three stool sample storage methods: i) Flinders Technical Associate (FTA) (Whatman™) cards, ii) OMNIgene® GUT tubes (DNAgenotek®) and iii) DNA Shield™ Faecal Collection Tubes (Zymo Research). Studies have shown that the method used for stool sample storage can have a significant impact on the species composition (282-284) and is a key challenge for enabling the deployed application of CMg for TD diagnosis and for wider deployed diagnostic methods where immediate testing is not available. This may be during outbreaks of GI infections or for the repatriation of samples to home countries for more extensive laboratory investigations when standard deployed diagnostic testing is not sufficient. Cold-chain storage at -80°C, the preferred method of stool sample preservation (285), is not always feasible due to limited equipment, power and space availability and room temperature alternatives are required. The development of sample preparation methods with minimal equipment, complexity and time to result are also required to enable exploitation in austere deployed environments in which space, time and personnel availability are at a premium.

4.2.7 Aims

The work detailed in this chapter aimed to develop and optimise CMg methods for stool samples for use in austere environments. In addition, three long-term stool sample storage methods for the detection of *C. hominis* from a retrospective cohort of stool samples collected from British military personnel during a significant outbreak event were compared. Long-term stool sample storage would allow military clinicians to store and transport stool samples at ambient temperatures for extended periods of time without the need for cold-chain logistics, allowing for the storage of samples prior to testing during periods when the availability of diagnostic tests is limited or overwhelmed and the repatriation of samples to home countries for more extensive testing.

4.3 Methods

4.3.1 Ethical approval

Ethical approval for the collection and additional testing of the primary stool samples was provided by the MODREC (2076/MODREC/21). All samples were anonymised prior to transfer to Dstl for CMg analysis.

4.3.2 Sample collection

Stool samples were collected from British Army personnel displaying symptoms of TD whilst undergoing training at the BATUK, situated near to Nanyuki, Kenya, in February 2022 following an increase in the prevalence of TD, as previously described (280). Primary stool samples collected from personnel displaying symptoms of TD were evaluated using the BioFire FilmArray GI multiplex PCR panel. Stool samples were then split and transferred to three long-term storage media: FTA cards ($n=35$), OMNIgene GUT tubes ($n=39$) and DNA Shield tubes ($n=39$). Samples were stored at room temperature until the CMg analysis was carried out.

4.3.3 DNA extraction

DNA was extracted from 250 μ L of OMNIgene GUT and DNA Shield samples using the QIAamp® PowerFecal® Pro DNA extraction kit (Qiagen) according to the manufacturer's instructions, with minor modifications, in batches of up to ten samples. The 250 μ L of stool sample was added to the Matrix Lysing E tubes with 800 μ L of Solution CD1. The tubes were bead beaten for 60 s at 6.0 m/s on a FastPrep-24 5G and the tubes were centrifuged at $15000 \times g$ for 60 s. The supernatant was transferred to a 2 mL microcentrifuge tube and 200 μ L of Solution CD2 was added. The microcentrifuge tube was centrifuged at $15000 \times g$ for 60 s and 700 μ L of the supernatant was transferred to a new 2 mL microcentrifuge tube. Six hundred microliters of Solution CD3 was added and 650 μ L of the lysate was loaded onto an MB Spin Column and centrifuged at $15000 \times g$ for 60 s. The flow through was discarded and the step repeated until all the lysate had been filtered through the spin column. The spin column was transferred to a new 2 mL collection tube and 500 μ L of Solution EA was added to the spin column. The spin column was centrifuged at $15000 \times g$ for 60 s and the flow through was discarded. Five hundred microliters of Solution C5 was added to the spin column and centrifuged at $15000 \times g$ for 60 s and the flow through was discarded. The spin column was dried by centrifuging at $16000 \times g$ for 2 min and placed into a new 1.5 mL elution tube. Fifty microliters of Solution C6 was added directly to the spin column and incubated at room temperature for 30 s to elute the DNA. The spin column was centrifuged at $15000 \times g$ for 60 s and the eluted DNA stored at 4-8°C.

DNA was eluted directly from the FTA cards according to a previously described method (286). Three 5 $\times$ 5 mm punches per sample were collected from the FTA cards and added

directly to a Monarch[®] DNA clean-up column (New England Biolabs) with 175 μ L Elution Buffer (Qiagen). The columns were incubated at 37°C at 1000 rpm for 1 h and centrifuged for 10 min at 13000 $\times g$. The FTA card DNA eluent then underwent an AMPure XP bead purification step. AMPure XP beads were added to the 175 μ L Elution Buffer at a ratio of 1.2 $\times$ volume of beads to sample in a 1.5 mL LoBind tube. The sample was then incubated at room temperature for 5 min on a shaking HulaMixer using the following settings: rotation speed 15 and timer 10, reciprocal tilting turning angle 45° & timer 5, vibration turning angle 5° and timer 5. The samples then underwent the AMPure XP bead purification according to the manufacturer's instructions, as previously described, and were eluted into 50 μ L nuclease-free water. Negative controls consisting of PBS only and positive controls consisting of 30 μ L microbial community standard II (Log distribution) (Zymo Research) made up to a total volume of 250 μ L with PBS were processed alongside each batch of samples.

4.3.4 DNA quantification

The total DNA concentration was quantified using the dsDNA HS assay kit on the Qubit 4 fluorometer according to the manufacturer's instructions, as described in Chapter Two section 2.3.5.

4.3.5 DNA fragment analysis

The DNA fragment size of extracted DNA was measured using the TapeStation 4200 system with the Genomic DNA Screentape assay according to the manufacturer's instructions, as described in Chapter Two section 2.3.7.

4.3.6 DNA purification

The 50 μ L DNA extract was purified by AMPure XP bead purification according to the manufacturer's instructions at a ratio of 1.8 $\times$ volume of beads to DNA extract in a 1.5 mL LoBind tube and eluted into 10 μ L nuclease-free water, as described in Chapter Two section 2.3.6.

4.3.7 Rapid Barcoding library preparation method

DNA was prepared for sequencing using the Rapid Barcoding library preparation kit V14 (SQK-RBK114.24) (Oxford Nanopore Technologies) (RBK) according to the manufacturer's instructions. Up to 12 libraries, including a negative and positive control, were multiplexed per sequencing run. The purified DNA samples were quantified using the Qubit dsDNA HS assay, as previously described. Up to 50 ng of the DNA libraries in a total volume of 10 μ L per sample was added to a 0.2 mL PCR tube. One microliter of Rapid Barcode (RB01-24) was added, and the samples were incubated at 30°C for 2 min, 80°C for 2 min and then cooled briefly on ice. The libraries were pooled into a 1.5 mL LoBind tube and mixed with an equal volume of AMPure XP beads, and the library DNA

purified as previously described, except the concentration of ethanol was increased from 70% to 80%. The DNA libraries were eluted into 12 µL of Elution Buffer. The pooled libraries were again quantified by Qubit dsDNA HS assay as previously described and up to 800 ng of the pooled libraries were added to a 0.2 mL PCR tube. One microliter of the diluted Rapid Adapter was added, and the reaction was incubated at room temperature for 5 min. The pooled libraries were loaded onto a Nanopore R10.4.1 flowcell (Oxford nanopore Technologies) according to the manufacturer's instructions and sequenced for a total of 24 h. The data was stored in POD5 format.

4.3.8 Optimised Rapid Barcoding library preparation method

The RBK library preparation method was optimised (RBK OP) to increase the yield of total sequencing reads. This included the addition of TapeStation Genomic DNA Screentape analysis of the purified DNA samples prior to the library preparation. The DNA fragment size and DNA concentration were then used to calculate the library molarity according to the following formula:

$$\text{Library concentration (nM)} = \frac{\text{concentration } \left(\frac{\text{ng}}{\mu\text{L}}\right)}{660 \left(\frac{\text{g}}{\text{mol}}\right) \times \text{library size (bp)}} \times 10^6$$

Up to 14 fmol of the DNA libraries in a total volume of 10 µL per sample was then added to a 0.2 mL PCR tube, increased from 50 ng per sample. Then, up to 1600 ng of pooled libraries were then taken forward for sequencing, increased from 800 ng. The libraries were then sequenced as previously described.

4.3.9 Rapid PCR barcoding library preparation method

To determine the performance of another library preparation method, DNA was also prepared for sequencing using the Rapid PCR barcoding library preparation kit V14 (SQK-RPB114.24) (Oxford Nanopore Technologies) (RPCR) according to the manufacturer's instructions without amendment. Up to 5 ng of the DNA libraries in a total volume of 3 µL per sample was added to a 0.2 mL PCR tube. One microliter of the Fragmentation Mix was added, and the samples incubated at 30°C for 2 min, 80°C for 2 min and then cooled briefly on ice. Libraries then underwent PCR amplification using the following reaction mix: 4 µL of the tagmented DNA was added to 1 µL of the Rapid barcodes (RLB 01-24) and 25 µL of the LongAmp Taq 2 × master mix (New England Biosciences) and the reaction was made up to 50 µL total volume with nuclease-free water. The PCR reaction underwent the following cycling conditions; initial denaturation of 95°C for 3 min followed by 14 cycles of denaturation at 95°C for 15 s, annealing at 56°C for 15 s and extension at 65°C for 6 min with a final extension at 65°C for 6 min. Once the PCR was completed, 4 µL of EDTA was added to each sample and incubated at room temperature for 5 min. Libraries were pooled in equimolar ratios to a final combined concentration of up to 400 fmol per library.

Pooled libraries were further purified by AMPure XP bead purification at a ratio of 0.6 × beads to library volume, as previously described. Libraries were eluted into 15 µL Buffer EB (Qiagen) and the DNA concentration and fragment length quantified, as previously described. One microliter of the diluted Rapid Adapter was added to 10-50 fmol of the pooled libraries in a total volume of 11 µL and incubated for 5 min at room temperature. An R10.4.1 flow cell (Oxford Nanopore Technologies) was then primed for sequencing according to the manufacturer's instructions. Prepared libraries were sequenced for 24 h and the data stored in POD5 format.

4.3.10 Bioinformatic analysis

ONT Taxonomic Profiling, a custom bioinformatics pipeline for the taxonomic profiling of microbial species and for the identification of AMR genes, previously developed for the analysis of metagenomics datasets for infectious disease diagnostics, was applied (63). Fast5 files were base-called and de-multiplexed using the guppy_basecaller Super Accuracy model and the guppy_barcode of the Guppy software v6.2.x, respectively. The minimum barcoding score was set to 65. Human reads were then removed from the data by mapping against the *Homo sapiens* reference genome CHM13v2 and the remaining reads were mapped to the RefSeq database using minimap2 v2.17. The reads were then filtered based on the harmonic mean of identity and coverage of the mapped reads using the following equation:

$$(2 \times \frac{[identity + coverage]}{[identity \times coverage]})$$

Reads with a taxa score ≥ 85 were included for taxonomic profiling. The pipeline also included applying the KMA (version 1.4.9) tool to screen for AMR genes with the amrfinderplus (v3.10) database. Microbial species identification was presented as the total number of reads and the percentage of the total classified microbial reads aligned to the species. Species were concluded to be present if ≥5 reads, minus the number of reads of the same species identified in the negative control sample, were assigned to that species. AMR genes were concluded to be present if their template identity and coverage were ≥ 90% and their depth of coverage reported by KMA were ≥ 1×.

4.3.11 qPCR

To provide additional information from a comparison diagnostic method, a *C. hominis* species-specific Taqman probe-based qPCR assay was performed in triplicate on all samples. All assays were performed on the QuantStudio 7 Flex platform and the Cq values were determined by the QuantStudio software. Positive control reactions of synthetic *C. hominis* DNA (PRA-3011SD™, American Type Culture Collection) at a concentration of 0.5 × 10⁵ genome copies per µL and negative control reactions of nuclease-free water only were run on each plate. The primer and probe sequences are

described in Table 12. The TaqMan Fast Advanced Master Mix (Applied Biosystems) was used for all assays. The qPCR assay was run in 25 µL total volume using 12.5 µL of master mix, final concentrations of 500 nM of forward primer, 500 nM of reverse primer, 250 nM of probe, and 2.0 µL of template. Thermal cycling conditions were 2 m at 95 °C followed by 45 cycles of 10 s at 95 °C and 1 min at 60 °C.

Table 12. Primer and probe sequences for *C. hominis* qPCR assays.

Assay target	Forward primer (5'-3')	Reverse primer (5'-3')	Probe sequence (5'-3')
<i>C. hominis</i> Lib13 (287)	TCCTTGAAATGAATA TTTGTGACTCG	AAATGTGGTAGTTGC GGTTGAAA	CTTACTTCGTGGCG GCGT

4.3.12 Statistical analysis

Data are expressed as means and standard deviation and were analysed using GraphPad Prism 10 (version 10.0.2) software. When multiple comparison analyses were performed, *p*-values were obtained by performing an ordinary one-way or two-way ANOVA test or mixed effects analysis when the *n* number varied, with post hoc Tukey's test. When two samples were individually compared, results were obtained using a two-tail unpaired *t*-test. Sensitivity, specificity, positive predictive values (PPV) and negative predictive values (NPV) were calculated according to the following calculations (218):

$$\text{Sensitivity} = [(True\ Positives)/(True\ Positives + False\ Negatives)] \times 100$$

$$\text{Specificity} = [(True\ Negatives)/(False\ Positives + True\ Negatives)] \times 100$$

$$\text{PPV} = [True\ Positive/(True\ positive + False\ negative)] \times 100$$

$$\text{NPV} = [True\ Negative/(False\ negative + True\ Negative)] \times 100$$

Species α-diversity was calculated using the Shannon diversity index. The Shannon diversity index is an estimator for species richness and evenness (219). The formula is as follows:

$$\text{Shannon diversity index} = - \sum [(p_i) \times \ln(p_i)]$$

$$p_i = \frac{n}{N}$$

n = number of reads assigned to a microbial species in a sample

N = total reads assigned to microbial species in a sample

4.4 Results

4.4.1 Sample characteristics

Thirty-nine samples, stored using three storage methods, were selected from the samples collected from UK military personnel displaying symptoms of TD during an outbreak of *C. hominis* and retrospectively analysed using CMg. From the BioFire FilmArray GI panel results obtained from the primary samples at the time of the outbreak, 16 (41.0%) samples contained *Cryptosporidium* only and eight samples (20.5%) contained *Cryptosporidium* and ≥ 1 additional pathogen(s). The remaining samples included eight that were positive for bacterial pathogens, two non-*Cryptosporidium* parasitic infections and one polymicrobial infection containing *C. difficile* and norovirus. The remaining four (10.3%) tested negative. Of the 39 samples included for CMg analysis, four *Cryptosporidium* positive samples included in the study had not been stored on FTA cards and were not available for CMg analysis.

4.4.2 Optimisation of the Rapid Barcoding library preparation method

Ten samples were selected for initial method evaluation using the RBK library preparation method according to the manufacturer's instructions. The mean total reads per sample were 43880, 117570 and 4557 reads from DNA Shield, OMNIgene GUT and FTA samples, respectively (Figure 27). In addition, the number of sequencing reads aligned to the *Cryptosporidium* genus in samples that tested positive for *Cryptosporidium* by BioFire FilmArray GI panel testing was low, with no identification above the threshold of five reads from DNA Shield and FTA samples and a maximum of 18 reads from OMNIgene GUT samples (Table 13).

The RBK library preparation method was then optimised. This included using the library molarity rather than just the DNA concentration, with libraries normalised up to 14 fmol per library. The optimisation steps resulted in increased total reads per sample, to 119000 for DNA Shield samples, 169000 for OMNIgene GUT samples and 17000 for the FTA samples (Figure 27A). In addition, the average read length increased significantly in the DNA Shield and OMNIgene GUT samples, from 1439 bp to 1839 bp and 1486 bp to 1968 bp respectively (Figure 27B), although there was no significant change in the FTA samples. The number of *Cryptosporidium* reads also increased, with a maximum of 9, 26 and 12 *Cryptosporidium* reads identified in DNA Shield, OMNIgene GUT and FTA samples respectively (Table 13). The RBK OP method was therefore taken forward for further evaluation.

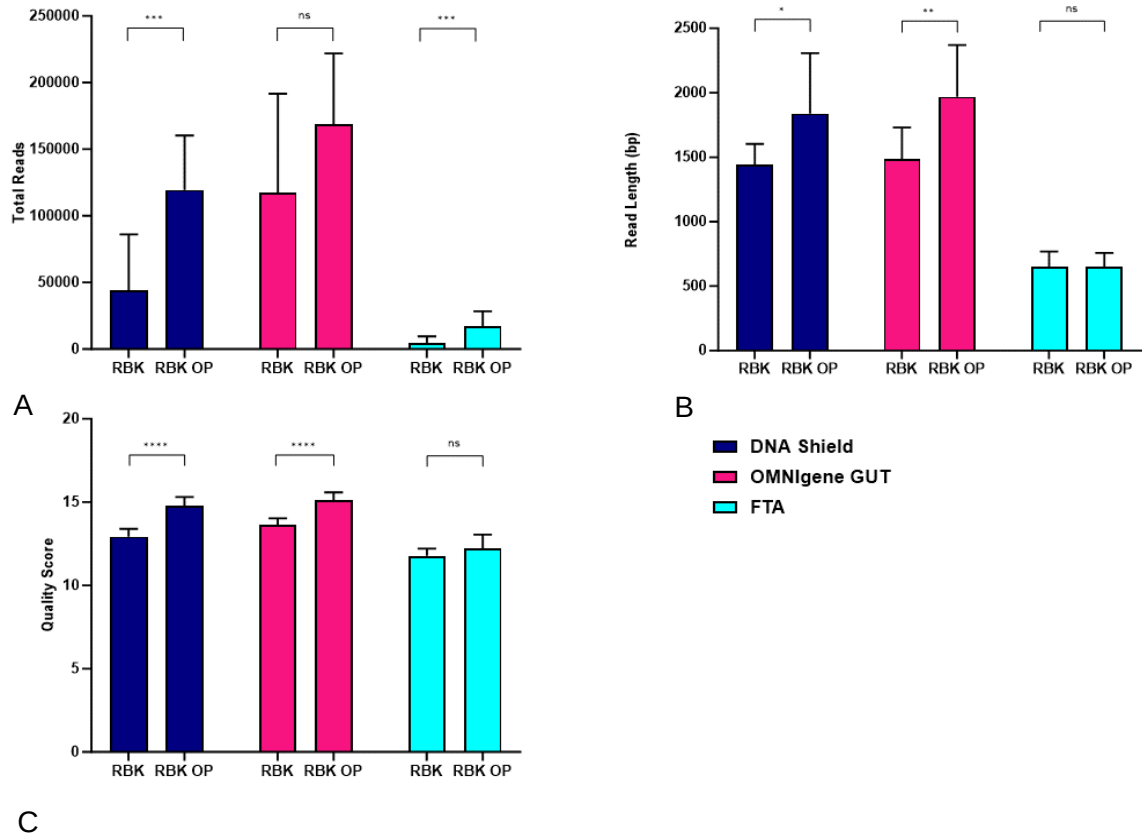


Figure 27. Comparison of the sequencing metrics from the RBK and RBK OP prepared libraries.

The mean total sequence reads (A), read lengths (B), and sequencing quality scores obtained from RBK and RBK OP prepared libraries from samples stored in each storage method. A cohort of stool samples were prepared for sequencing using the manufacturer's recommended RBK library preparation method and the optimised RBK OP method and the resulting sequencing data metrics were compared. RBK OP generated significantly higher total reads, read length and quality scores for DNA Shield samples, higher read length and quality score for OMNIgene GUT samples and higher total reads for FTA samples compared to RBK.

+SD. DNA Shield and OMNIgene GUT $n=10$. FTA cards $n=9$. ns $p>0.05$. * $p<0.05$. ** $p<0.01$. *** $p<0.001$. **** $p<0.0001$

Table 13. Comparison of the detection of *Cryptosporidium* by CMg using the RBK and RBK OP library preparation methods.

		<i>Cryptosporidium</i> reads (% of classified microbial reads)					
Sample	BioFire FilmArray GI panel result	DNA Shield		OMNIgene GUT		FTA	
		RBK	RBK OP	RBK	RBK OP	RBK	RBK OP
C1	<i>C. difficile</i> Norovirus	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
B1	<i>Cryptosporidium</i>	1 (0.012)	2 (0.031)	18 (0.023)	16 (0.035)	-	-
B2	<i>Cryptosporidium</i>	2 (0.0028)	1 (0.0020)	2 (0.0020)	4 (0.011)	0 (0)	0 (0)
B3	<i>Cryptosporidium</i> <i>C. difficile</i> Enteropathogenic <i>E. coli</i>	0 (0)	9 (0.050)	17 (0.023)	26 (0.043)	0 (0)	12 (0.33)
B10	Negative	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
B40	<i>Campylobacter</i>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
B44	<i>Campylobacter</i>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
B56	Enteropathogenic <i>E. coli</i>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
B80	Enteropathogenic <i>E. coli</i>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
B94	<i>G. lamblia</i>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

A cohort of stool samples stored using 3 storage methods were processed for CMg using the RBK and RBK OP methods and the number and proportion of classified microbial reads assigned to the *Cryptosporidium* genus were compared. % classified microbial reads = % of *Cryptosporidium* reads over the total number of taxonomically classified microbial reads obtained. RBK OP identified higher *Cryptosporidium* read numbers compared to RBK. There were no *Cryptosporidium* reads identified in BioFire FilmArray GI panel negative samples.

4.4.3 Sample processing

The remaining samples were processed for CMg using the RBK OP library preparation method. Due to the low total read yields and number of reads assigned to *Cryptosporidium*, samples were also processed using the RPCR library preparation method and the results were compared. Samples stored in the OMNIgene GUT tubes yielded a mean DNA concentration of 117.3 ng/μL per sample, significantly higher than the DNA Shield samples at 45.06 ng/μL and the FTA samples at 17.97 ng/μL. The FTA samples had significantly lower mean DNA fragment sizes at 9362 bp compared to 19001

bp for the DNA Shield and 19086 bp for the OMNIgene GUT samples. The DNA Shield and OMNIgene GUT samples generated higher total reads, read length, and quality scores compared to FTA samples from both RBK OP and RPCR prepared samples (Figure 28).

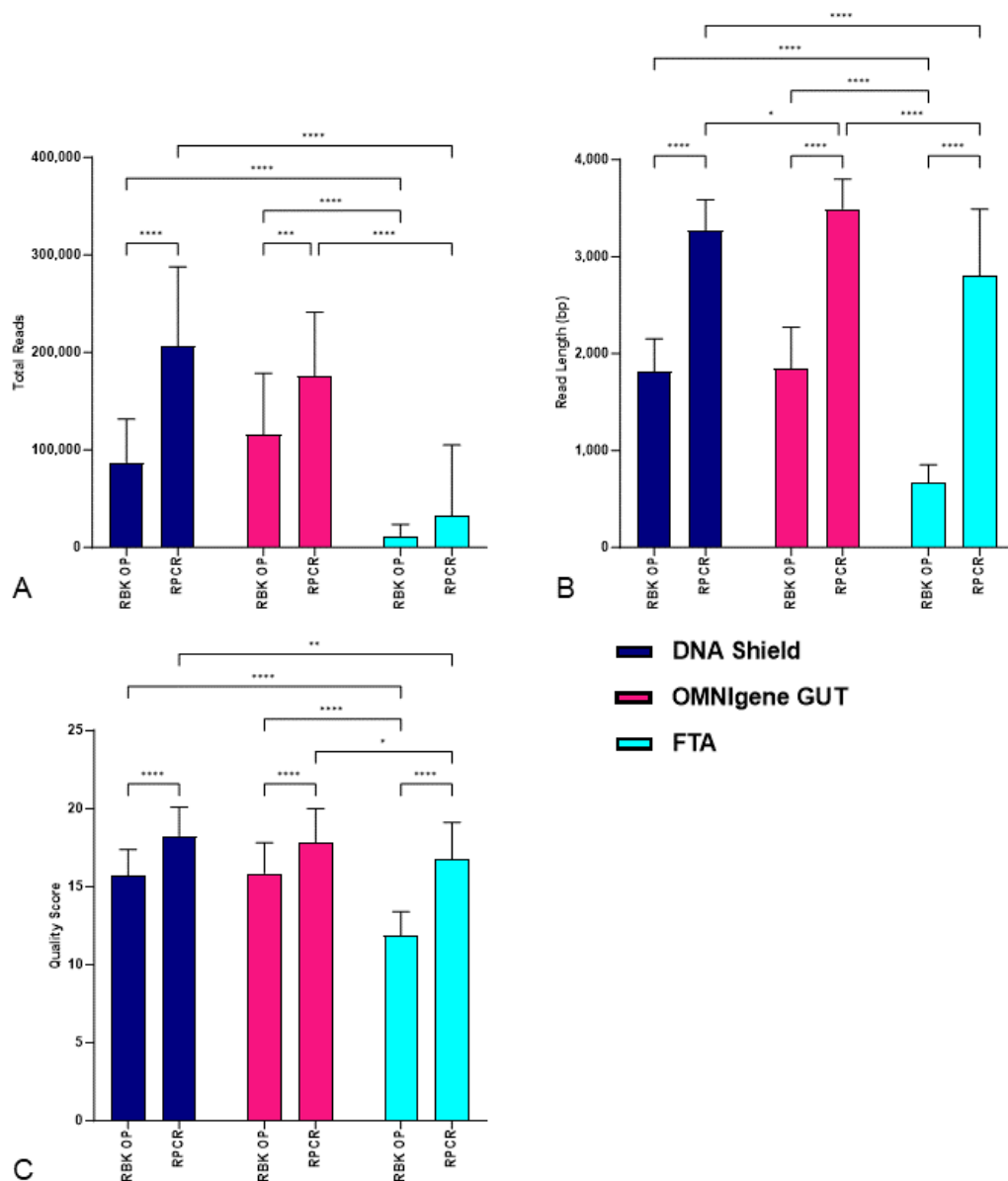


Figure 28. Comparison of the sequencing metrics from RBK OP and RPCR prepared libraries.

Comparison of the total number of reads (A), sequence read length (B) and sequencing quality score (C) obtained from each storage method using the RBK OP and RPCR library preparation methods. Stool samples stored using each storage method were prepared for nanopore-based DNA sequencing and the total reads, read length and sequencing quality scores compared. The RPCR method generated the higher total reads, read length and quality scores across all storage methods. Samples stored on FTA cards generated the lowest total reads, read lengths and quality scores.

+SD. ns $p > 0.05$ (not shown). * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. **** $p < 0.0001$

4.4.4 Detection of *Cryptosporidium*

The number of reads assigned to *Cryptosporidium* from positive samples were compared between the storage kits and the library preparation methods (Figure 29A). The DNA Shield samples prepared with the RPCR method generated significantly higher numbers of *Cryptosporidium* reads compared to the RPCR prepared FTA samples and the RBK prepared DNA Shield samples. There was no significant difference in the total *Cryptosporidium* reads between the RBK OP prepared samples for any storage method. *Cryptosporidium* reads were not identified in any of the BioFire FilmArray GI panel negative samples. Whilst the FTA samples generated significantly lower *Cryptosporidium* reads, there was no significant difference in the proportion of classified microbial reads assigned to *Cryptosporidium* between any storage method or library preparation method. Comparison of the average read length assigned to *Cryptosporidium* and the sequencing depth (Figures 29B and 29C) showed that there was no significant difference in read length between any storage method. However, samples prepared using RPCR generated significantly longer *Cryptosporidium* reads compared to RBK OP for all storage methods. The RPCR method also generated significantly higher *Cryptosporidium* sequencing depth compared to RBK OP. There was no significant difference in sequencing depth between any of the sample storage methods or preparation method for the OMNIgene GUT and FTA samples (Figure 29C).

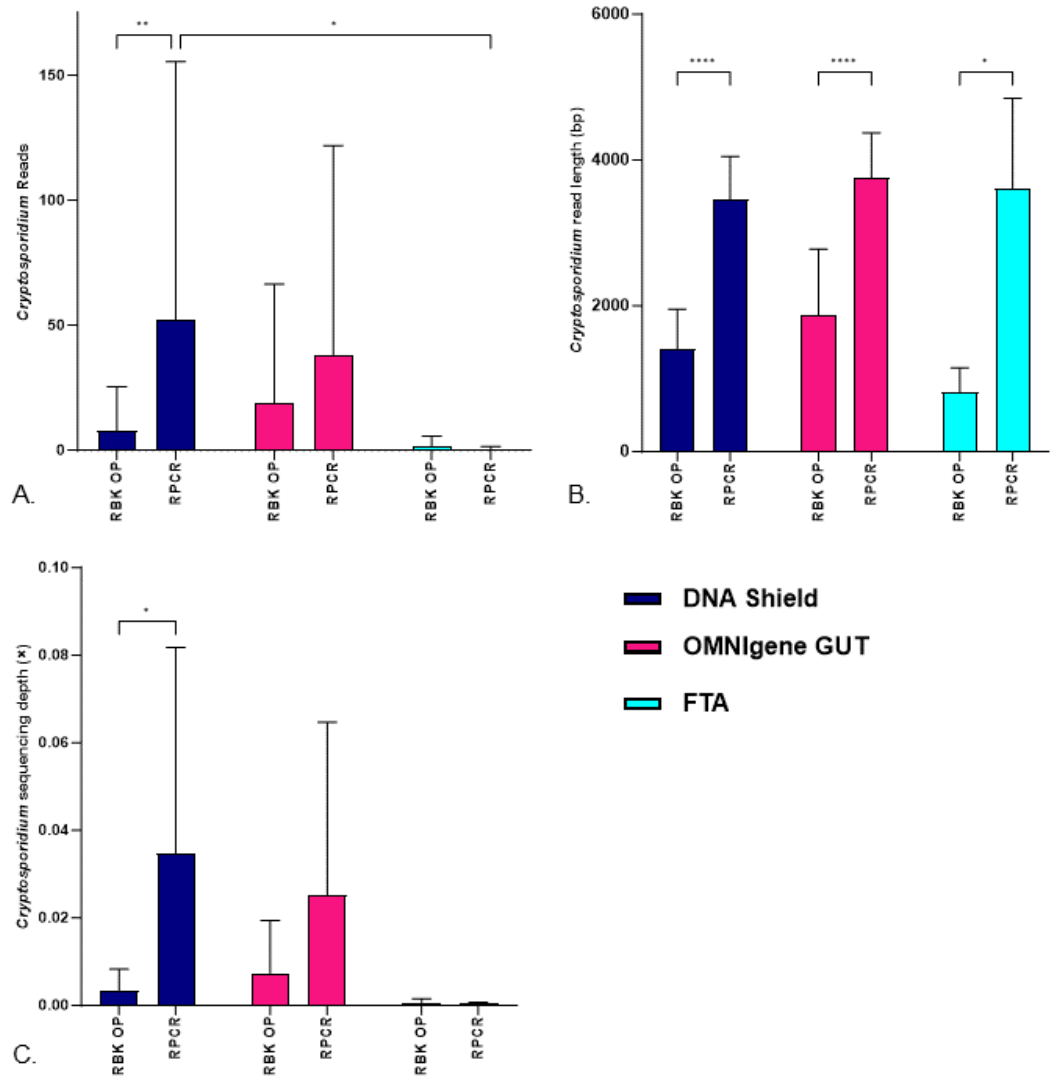


Figure 29. *Cryptosporidium* sequence reads from the BioFire FilmArray GI panel positive samples.

Stool samples stored in each storage method were prepared using the RBK OP or RPCR library preparation methods for nanopore-based sequencing. The total number of reads assigned to the *Cryptosporidium* genus were compared. The RPCR method from samples stored in DNA Shield samples generated higher *Cryptosporidium* reads compared to the RBK method and samples stored by FTA card.

DNA Shield and OMNIgene GUT $n=24$. FTA $n=20$. +SD. ns $p>0.05$ (not shown). * $p<0.05$. ** $p<0.01$ **** $p<0.0001$

To provide a quantitative measure of *C. hominis* DNA and the positivity rate in each sample following long-term storage, the samples were analysed by a species-specific *C. hominis* qPCR assay. There was a reduction in the Cq values obtained from the OMNIgene GUT samples compared to the DNA Shield samples. There was no significant difference between the DNA Shield and the FTA or the OMNIgene GUT and the FTA samples (Figure 30). Compared to the BioFire FilmArray GI panel results immediately following sample collection, there was a reduction in *C. hominis* qPCR positive samples (Cq<45) for all storage methods. Twenty-one DNA Shield samples tested positive by qPCR, compared to 24 by BioFire FilmArray GI Panel. The number of positive OMNIgene

GUT samples reduced to 23 from 24 positive samples and the number of positive FTA samples reduced to 17 from 20. The full list of the *Cryptosporidium* reads, and the proportion of classified microbial reads identified from each sample with the corresponding BioFire FilmArray GI panel result are presented in Table 14.

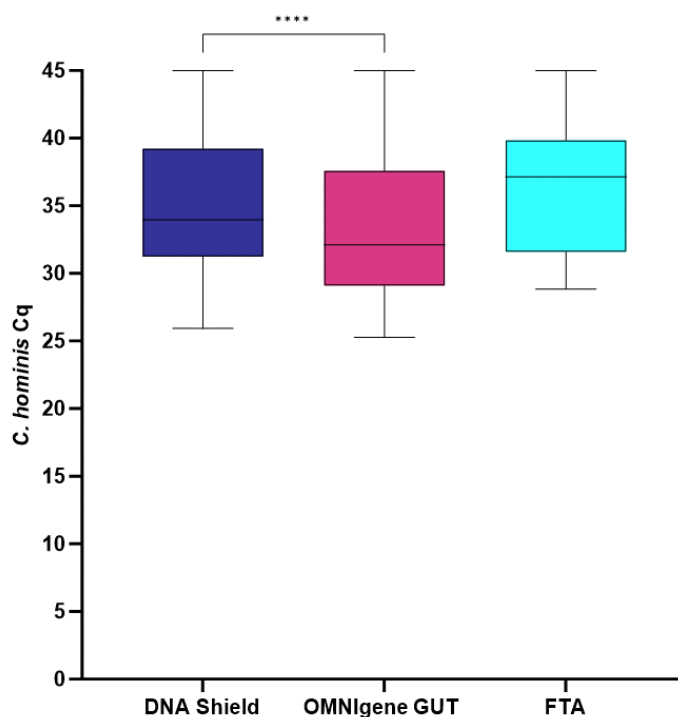


Figure 30. *C. hominis* Cq values from BioFire FilmArray GI panel positive samples stored using three methods.

DNA extracts were analysed by qPCR for the detection of *C. hominis* species-specific DNA. Cq values from samples stored in each storage method were compared. Cq values from *C. hominis* negative samples were not included in the analysis. There was a significant reduction in Cq values obtained from Omnigene GUT samples compared to the DNA Shield samples.

Error bars represent the minimum and maximum values. DNA Shield $n=24$. OMNigene GUT $n=24$. FTA $n=20$. ns $p>0.05$ (not shown). **** $p<0.0001$

Table 14. Comparison of CMg and qPCR for the detection of *Cryptosporidium*.

Sample Number	BioFire FilmArray GI panel result	<i>C. hominis</i> qPCR assay Cq			Total reads assigned to <i>Cryptosporidium</i> genus by CMg (proportion of classified microbial reads)					
					RBK			RPCR		
		DS	OG	FTA	DS	OG	FTA	DS	OG	FTA
C1	<i>C. difficile</i> toxin A/B Norovirus GI/GII	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B1	<i>Cryptosporidium</i>	31.95	30.93	-	2 (0.09)	16 (0.10)	-	324 (0.25)	41 (0.62)	-
B2	<i>Cryptosporidium</i>	33.08	30.70	31.21	1 (0.00)	4 (0.03)	0 (0.00)	16 (0.01)	6 (0.00)	0 (0.00)
B3	<i>Cryptosporidium</i> <i>C. difficile</i> toxin A/B	29.28	27.49	28.87	9 (0.09)	26 (0.07)	12 (1.29)	312 (0.21)	339 (0.21)	1 (1.27)
B10	Negative	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B40	<i>Campylobacter</i> (<i>C. jejuni</i> / <i>C. coli</i> / <i>C. upsaliensis</i>)	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B44	<i>Campylobacter</i> (<i>C. jejuni</i> / <i>C. coli</i> / <i>C. upsaliensis</i>)	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B56	Enteropathogenic <i>E. coli</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B80	Enteropathogenic <i>E. coli</i>	No detection	No detection	40.33	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B94	<i>G. lamblia</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B4	<i>Cryptosporidium</i>	32.14	29.61	30.99	0 (0.00)	4 (0.00)	1 (0.03)	7 (0.01)	11 (0.01)	1 (0.11)
B5	<i>Cryptosporidium</i>	33.85	31.41	32.76	1 (0.0)	1 (0.00)	0 (0.00)	3 (0.00)	10 (0.01)	1 (0.15)
B6	<i>Cryptosporidium</i>	34.99	32.21	34.46	1 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	5 (0.00)	1 (0.19)
B7	<i>Cryptosporidium</i>	33.96	32.02	33.52	1 (0.00)	1 (0.00)	0 (0.00)	3 (0.00)	5 (0.00)	2 (0.03)
B8	<i>Cryptosporidium</i>	33.97	37.49	39.84	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.00)	0 (0.00)	0 (0.00)
B9	<i>Cryptosporidium</i>	28.99	28.55	28.84	26 (0.07)	31 (0.06)	0 (0.00)	131 (0.19)	242 (0.24)	0 (0.00)
B11	<i>Cryptosporidium</i>	39.54	36.32	39.22	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B12	<i>Cryptosporidium</i>	32.92	30.77	34.08	0 (0.00)	3 (0.00)	0 (0.00)	4 (0.00)	7 (0.00)	0 (0.00)
B13	<i>Cryptosporidium</i>	31.01	28.92	-	0 (0.00)	4 (0.00)	-	10 (0.01)	18 (0.02)	-
B14	<i>Cryptosporidium</i>	28.39	27.51	-	72 (0.06)	23 (0.04)	-	149 (0.19)	50 (0.05)	-
B015	Negative	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B016	<i>Cryptosporidium</i>	25.94	25.28	28.85	42 (0.05)	200 (0.24)	16 (0.12)	260 (0.16)	127 (0.02)	4 (0.55)
B017	<i>Cryptosporidium</i>	27.73	26.22	-	30 (0.06)	135 (0.16)	-	29 (0.04)	47 (0.01)	-
B018	<i>Cryptosporidium</i>	38.16	37.92	37.92	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B020	<i>Cryptosporidium</i> <i>Salmonella</i>	43.12	38.93	39.26	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B021	Enteraggregative <i>E. coli</i> Enteropathogenic <i>E. coli</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B022	<i>Cryptosporidium</i>	36.18	32.69	36.39	1 (0.00)	0 (0.00)	1 (0.01)	2 (0.00)	1 (0.00)	0 (0.00)
B026	<i>Cryptosporidium</i> Enteropathogenic <i>E. coli</i> Enterotoxigenic <i>E. coli</i>	43.07	37.62	39.87	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

B032	Enteropathogenic <i>E. coli</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B033	<i>Cryptosporidium</i> Astrovirus	38.35	36.14	42.99	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B045	<i>Campylobacter</i> (<i>C. jejuni</i> / <i>C. coli</i> / <i>C. upsaliensis</i>) Enteropathogenic <i>E. coli</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B046	Negative	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B049	<i>Campylobacter</i> (<i>C. jejuni</i> / <i>C. coli</i> / <i>C. upsaliensis</i>)	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B057	<i>Cryptosporidium</i> <i>Campylobacter</i> (<i>C. jejuni</i> / <i>C. coli</i> / <i>C. upsaliensis</i>)	No detection	42.60	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B085	<i>Entamoeba Histolytica</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B081	<i>Cryptosporidium</i> Enteropathogenic <i>E. coli</i>	37.09	36.57	38.52	0 (0.00)	2 (0.01)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
C07b	<i>Cryptosporidium</i> Enteropathogenic <i>E. coli</i>	No detection	No detection	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
C012	<i>Cryptosporidium</i> Enteropathogenic <i>E. coli</i>	No detection	42.77	No detection	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
C013	Negative	No detection	No detection	No detection	No detection	No detection	No detection	No detection	No detection	No detection

DNA extracts were analysed by qPCR for the detection of *C. hominis* species-specific DNA. *C. hominis* Cq values obtained from all samples and the corresponding number of reads assigned to the *Cryptosporidium* genus obtained from CMg analysis and the BioFire FilmArray GI panel results performed on the primary samples are reported. There was no detection of *Cryptosporidium* by CMg in any BioFire FilmArray and qPCR negative samples.

The sensitivity, specificity, PPV and NPV for the detection of *Cryptosporidium* were calculated (Table 15 A-C). Compared to the BioFire FilmArray GI panel, the OMNIgene GUT samples achieved the highest sensitivity and NPV values by qPCR of 95.83% and 93.75 respectively (Table 15A). For CMg, the OMNIgene GUT samples processed using the RPCR method achieved the highest sensitivity and NPV of 54.17% and 57.69 respectively compared to the BioFire FilmArray GI panel testing carried out prior to long-term storage, whilst the FTA RPCR samples had the lowest, with no samples identifying *Cryptosporidium* above the threshold of ≥ 5 reads using the RPCR method (Table 15B). All the DNA Shield and OMNIgene GUT samples had a specificity of 100% and PPV of 100, with no false positive detection of *Cryptosporidium* in any sample.

Comparing CMg to the qPCR assay also carried out following long-term storage, the OMNIgene GUT RPCR samples achieved the highest sensitivity of 56.52% and a NPV of 61.54 (Table 15C). Reducing the threshold for a positive identification to ≥ 1 read assigned to *Cryptosporidium*, the highest sensitivity was achieved by the OMNIgene GUT and DNA Shield RPCR samples (66.67%) and a NPV of 68.18 compared to qPCR. The specificity

and PPV remained at 100% and 100.00 respectively when applying this threshold, with no false positive identifications.

Table 15 A-D. Sensitivity, NPV, specificity and PPV of CMg for the detection of *Cryptosporidium*.

Storage Method	Sensitivity (%)	NPV	Specificity (%)	PPV
DNA Shield	87.50	83.33	100.00	100.00
OMNIgene GUT	95.83	93.75	100.00	100.00
FTA	85.00	82.35	93.33	94.44

A

Method	Sensitivity (%)	NPV	Specificity (%)	PPV
DNA Shield RBK	20.83	44.12	100.00	100.00
DNA Shield RPCR	37.50	50.00	100.00	100.00
OMNIgene GUT RBK	25.00	45.45	100.00	100.00
OMNIgene GUT RPCR	54.17	57.69	100.00	100.00
FTA RBK	5.00	44.12	100.00	100.00
FTA RPCR	0.00	42.86	100.00	0.00

B

Method	Sensitivity (%)	NPV	Specificity (%)	PPV
DNA Shield RBK	23.81	52.94	100.00	100.00
DNA Shield RPCR	42.86	60.00	100.00	100.00
OMNIgene GUT RBK	26.09	48.49	100.00	100.00
OMNIgene GUT RPCR	56.52	61.54	100.00	100.00
FTA RBK	11.76	51.52	100.00	100.00
FTA RPCR	0.00	48.57	100.00	0.00

C

The sensitivity, NPV, specificity and PPV for the detection of *Cryptosporidium* following long-term storage by (A) qPCR compared to the BioFire FilmArray GI panel (B) CMg compared to the BioFire FilmArray GI panel and (C) CMg compared to qPCR.

To understand the concordance of CMg and the *C. hominis* species-specific qPCR assay, the correlation between the number of *C. hominis* reads identified by CMg per 100,000 classified microbial reads and the concentration of *C. hominis* genome copies per μL in the DNA extract was calculated. For the DNA Shield samples, both RBK OP and RPCR were not positively correlated (RBK: $R^2=0.1043$; $p=0.1532$. RPCR: $R^2=0.007858$; $p=0.7024$). The OMNIgene GUT RBK OP samples were positively correlated ($R^2=0.8652$; $p<0.0001$) but OMNIgene GUT RPCR samples were not ($R^2=0.0004506$; $p=0.9234$). Both the FTA RBK OP and RPCR samples were positively correlated (RBK: $R^2=0.2389$; $p=0.0465$. RPCR: $R^2=0.3686$; $p=0.0097$) (Figure 31).

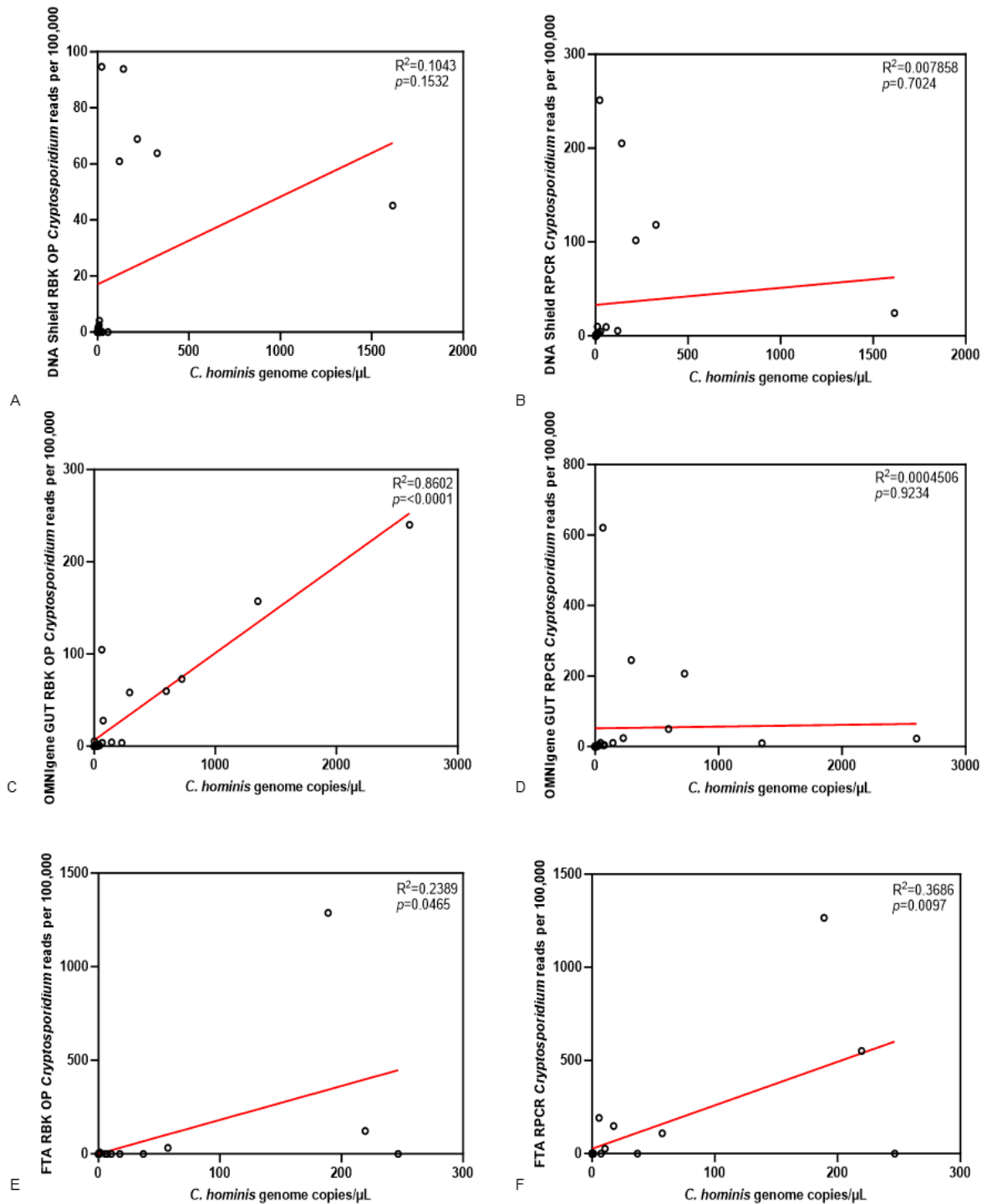


Figure 31. Correlation of qPCR and CMg for the detection of *Cryptosporidium*.

DNA extracts obtained from stool samples stored in each storage method were analysed by *C. hominis* species-specific qPCR with a standard curve and prepared using the RBK OP and RPCR methods for CMg. The concentration of *C. hominis* DNA was then plotted against the number of *Cryptosporidium* reads per 100,000 classified microbial reads identified by CMg to determine the R^2 value. The samples that were positive for *C. hominis* by qPCR (Cq value <45) are included. The DNA Shield RBK OP and RPCR *Cryptosporidium* reads per 100,000 reads were not positively correlated with the *C. hominis* genome copies/μL. The OMNIgene GUT RBK OP samples were positively correlated but OMNIgene GUT RPCR samples were not. Both the FTA RBK and RPCR samples were positively correlated.

The linear regression lines are presented. DNA Shield $n=21$. OMNIgene GUT $n=23$. FTA $n=17$

4.4.5 Comparison of the detection of additional species and antimicrobial resistance genes

The detection of non-*Cryptosporidium* pathogens identified by the BioFire FilmArray GI panel at the species and genus level was evaluated. The detection of *E. coli* pathotypes was not evaluated, due to the challenges associated with differentiating between each pathotype and commensal *E. coli* species by CMg. The current CMg workflow was not developed to include the detection of RNA based viruses, therefore, detection of astrovirus and norovirus was also excluded. The comparison included the detection of *C. difficile*, *Campylobacter jejuni*, *Campylobacter coli*, *Campylobacter upsaliensis*, *Salmonella* and the enteric parasites, *G. lamblia* and *Entamoeba histolytica* from samples stored using the different storage methods and processed with the different library preparation methods. Of the ten samples that tested positive for these species, there was no difference in the sensitivity of detection between the DNA Shield and OMNIgene GUT samples, with six of the ten species and/or genus identified by the BioFire FilmArray GI panel comparable to the CMg result for both storage and preparation methods. Two of the ten identifications were identified in the FTA RBK samples, increasing to five of the ten from FTA RPCR samples. *Campylobacter*, which includes the species *C. upsaliensis*, *C. coli* and *C. jejuni* was the most common positive identification by the BioFire FilmArray GI panel amongst the cohort of samples, identified in five samples. The OMNIgene GUT samples resulted in the highest number of reads assigned to *C. jejuni* from the RBK OP and RPCR prepared samples respectively, although these increases were not significant. Similarly for *C. difficile*, the OMNIgene GUT samples identified the highest number of reads from both the RBK OP and RPCR prepared samples. There were no reads assigned to the enteric parasites *G. lamblia* and *E. histolytica* in any sample. The number of and proportion of classified microbial reads assigned to additional species are reported in Table 16.

Table 16. Detection of additional enteric pathogens by CMg.

Sample Number	BioFire FilmArray GI panel result	Total reads assigned to species/genus by CMg (proportion of classified microbial reads)					
		RBK OP			RPCR		
		DNA Shield	OMNIgene GUT	FTA	DNA Shield	OMNIgene GUT	FTA
C1	<i>C. difficile</i>	28 (0.10)	30 (0.14)	4 (0.52)	142 (0.11)	161 (0.14)	70 (0.26)
B3	<i>C. difficile</i>	15 (0.16)	71 (0.20)	3 (0.32)	269 (0.18)	342 (0.21)	0 (0.00)
B40	<i>C. upsaliensis</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	<i>C. coli</i>	4 (0.13)	0 (0.00)	0 (0.00)	39 (0.12)	25 (0.10)	86 (0.65)
	<i>C. jejuni</i>	60 (1.97)	40 (1.72)	5 (3.03)	1276 (3.83)	698 (2.76)	993 (7.47)
B44	<i>C. upsaliensis</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	<i>C. coli</i>	1 (0.01)	0 (0.00)	0 (0.00)	6 (0.01)	9 (0.01)	3 (0.02)
	<i>C. jejuni</i>	35 (0.19)	13 (0.07)	1 (0.31)	221 (0.21)	303 (0.37)	69 (0.37)
B94	<i>G. lamblia</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
B020	<i>Salmonella</i>	0 (0.00)	1 (0.01)	0 (0.00)	1 (0.00)	0 (0.00)	0 (0.00)
B045	<i>C. upsaliensis</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	<i>C. coli</i>	20 (0.38)	19 (0.30)	3 (0.10)	183 (0.40)	343 (0.69)	228 (1.16)
	<i>C. jejuni</i>	432 (8.25)	559 (8.76)	86 (2.73)	5992 (13.22)	10773 (21.70)	3932 (20.09)
B049	<i>C. upsaliensis</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	<i>C. coli</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	<i>C. jejuni</i>	1 (0.02)	0 (0.00)	0 (0.00)	2 (0.03)	1 (0.01)	2 (0.06)
B057	<i>C. upsaliensis</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	<i>C. coli</i>	1 (0.00)	5 (0.01)	0 (0.00)	11 (0.00)	9 (0.00)	40 (0.02)
	<i>C. jejuni</i>	67 (0.09)	98 (0.11)	0 (0.00)	432 (0.13)	245 (0.12)	552 (0.31)
B085	<i>E. histolytica</i>	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

The number and proportion of classified microbial reads assigned to the enteric pathogens identified by the BioFire FilmArray GI panel from the primary stool samples from each storage and library preparation method are reported. The OMNIgene GUT samples prepared using the RPCR method resulted in the highest mean number of reads assigned to the enteric pathogens.

The detection of AMR genes by CMg from samples stored in each storage method and prepared using the RBK OP library preparation method were compared. A higher mean number of AMR genes per sample were identified from samples stored in the DNA Shield and OMNIgene GUT samples, 8.6 and 10.2 respectively, compared to 0.3 identified from the FTA card samples (Figure 32). The majority of the samples stored on the FTA cards (25/35) did not identify the presence of an AMR gene.

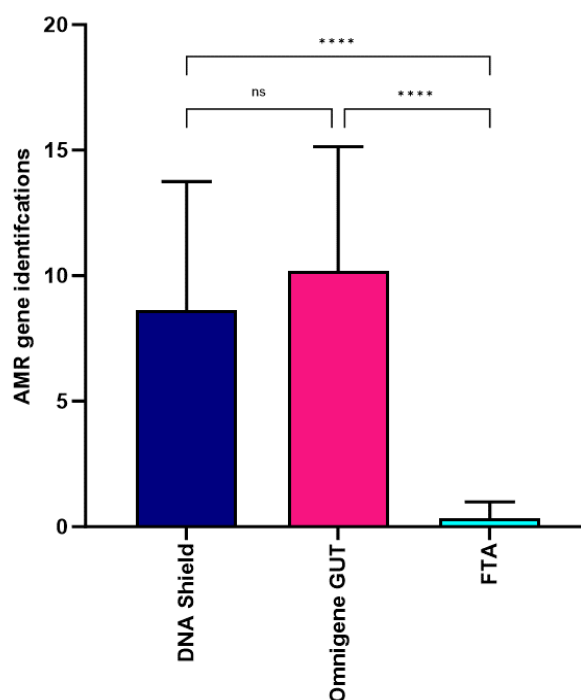


Figure 32. Frequency of AMR gene identifications.

AMR genes were identified using KMA (version 1.4.9) from the amrfinderplus database. AMR genes identified with a template coverage of $\geq 90\%$ and depth of coverage of $\geq 1\times$ were included. The mean number of AMR genes identified per sample were then compared. DNA Shield and OMNIgene GUT samples identified a higher number of AMR genes per sample compared to the FTA samples.

DNA Shield $n=39$. OMNIgene GUT $n=39$. FTA $n=35$. ns $p>0.05$. **** $p<0.0001$

AMR genes associated with resistance to tetracyclines were most commonly identified from all storage methods, followed by lincosamide and macrolides as the second and third most commonly identified drug class (Figure 33). The full list of AMR genes identified per sample are reported in Tables A2-A4 in the Annex.

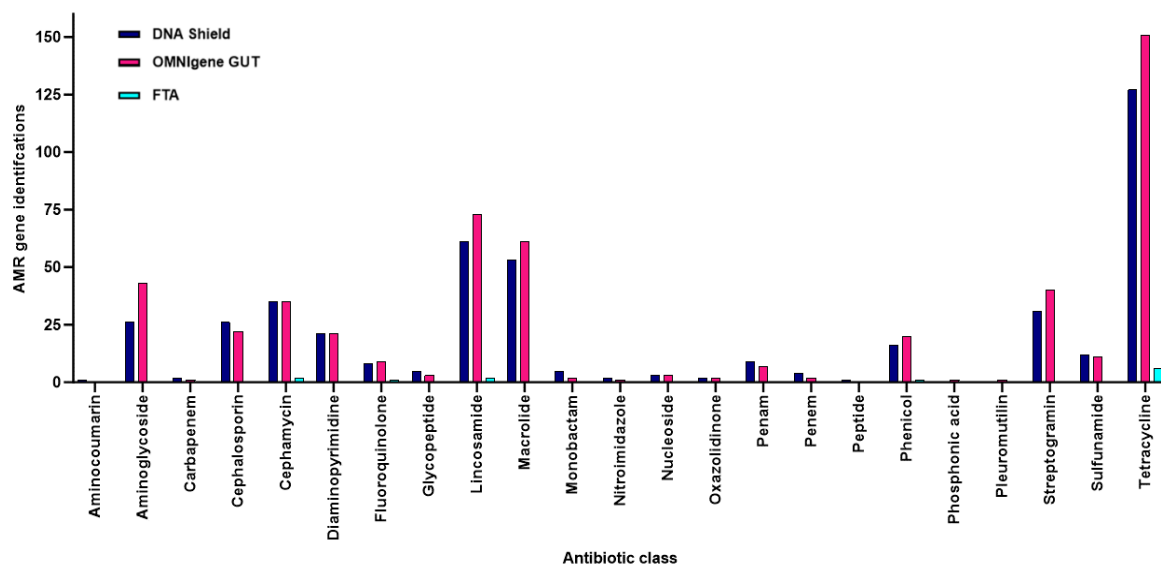


Figure 33. Identification of AMR genes associated with antibiotic drug classes.

The AMR genes were identified using KMA (version 1.4.9) from the amrfinderplus database. The AMR genes identified with a template coverage of $\geq 90\%$ and depth of coverage of $\geq 1\times$ were included. The antibiotic drug class for which each AMR gene was associated was determined according to the CARD. AMR genes associated with resistance to tetracyclines were most identified for all three storage methods.

DNA Shield $n=39$. OMNIgene GUT $n=39$. FTA $n=35$.

4.4.6 Operational considerations

The FTA cards required minimal space for storage at ambient temperatures. Both the DNA Shield and OMNIgene GUT samples were stored within re-sealable tubes. DNA was eluted directly from the FTA cards without the need for a DNA extraction step. DNA was extracted from the DNA Shield and OMNIgene GUT samples using a manual DNA extraction kit. However, the use of an automated platform would minimise the need for manual extraction steps. The RPCR method increased the method complexity and total time to result compared to the RBK OP method, requiring additional sample dilution prior to the PCR step and taking an additional 2.5 hours to reach the sequencing step. Overall, preparation of all sample types using the RBK OP method required approximately four hours to reach the sequencing step for the preparation of ten samples. This increased to 6.5 hours using the RPCR method.

4.4.7 Species diversity and microbiome composition

Further analysis of the sequencing data was undertaken to understand the impact of the storage and library preparation methods on the composition of the gut microbiome identified by CMg sequencing. The α -diversity was measured by calculating the Shannon diversity index at a species level. There was a significant reduction in species diversity obtained from the FTA samples compared to the DNA Shield and OMNIgene GUT

samples. There was no significant difference in species diversity between the DNA Shield and OMNIgene GUT samples (Figure 34). The α -diversity obtained from samples that tested positive for *C. hominis* by the BioFire FilmArray GI panel analysis was also compared. There was no change in α -diversity between *C. hominis* positive and negative samples in any storage media or library preparation method, except the FTA RPCR samples. In this case, there was a reduction in α -diversity in *C. hominis* positive samples compared to negative samples.

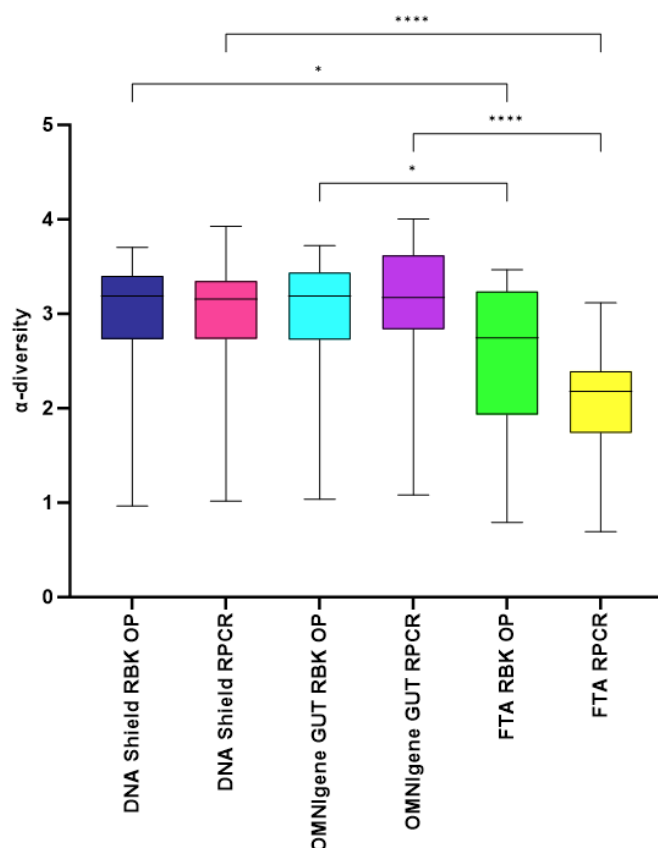


Figure 34. Microbial species α -diversity from stool samples.

The α -diversity identified by CMg was compared by calculating the Shannon diversity index at a species level, including all species identified at ≥ 5 reads. The diversity observed from samples stored in three storage types and prepared with the RBK OP and RPCR library preparation methods is reported. Four FTA samples that did not generate sufficient reads assigned to a microbial species were excluded. There was a significant reduction in species diversity from the FTA samples prepared with the RBK and RPCR methods.

DNA Shield and OMNIgene GUT $n= 39$. FTA $n=30$. Error bars represent the minimum and maximum values. ns $p>0.05$ (not shown) * $p<0.05$ **** $p<0.0001$

The proportion of reads assigned to the six most abundant phyla obtained from each storage and library preparation method were compared across all analysed samples (Figure 35). *Bacteroidota* and *Bacilotta* were the most abundant phyla across all storage methods. Use of the RPCR method resulted in an increased proportion of *Bacteroidota* in the DNA Shield and OMNIgene GUT samples and a reduction in *Bacilotta* and *Actinomycetota* compared to the RBK OP method. There was no significant alteration to the abundance of any phyla between any storage type prepared using the same method.

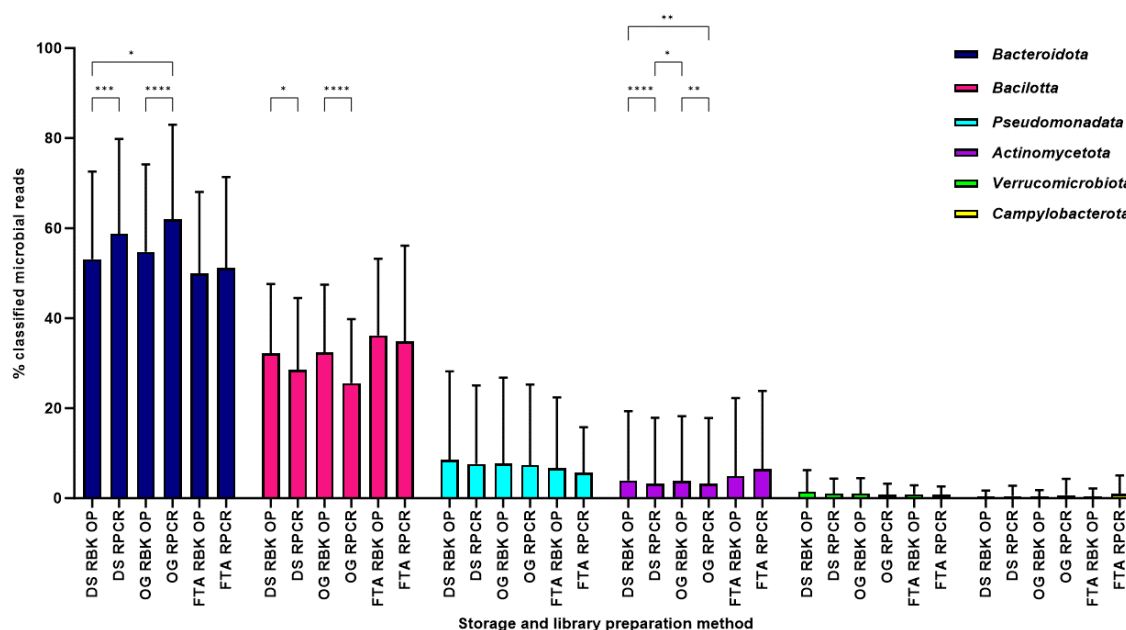


Figure 35. Proportion of classified microbial reads assigned to the most abundant phyla.

The effect of the storage and library preparation methods on the proportion of classified microbial reads assigned to the *Bacteroidota*, *Bacilotta*, *Pseudomonadota*, *Actinomycetota*, *Verrucomicrobiota* and *Campylobacterota* phyla in stool samples stored in the DNA Shield, OMNIgene GUT and FTA cards. The RPCR method resulted in a significant increase in the proportion of *Bacteroidota* and reduction of *Bacilotta* and *Actinomycetota* compared to the DNA Shield and OMNIgene GUT samples prepared using the RBK OP method.

+SD. DS = DNA Shield. OG = OMNIgene GUT. DS $n=39$. OG $n=39$. FTA $n=29$. ns $p>0.05$ (not shown)
* $p<0.05$. ** $p<0.01$. *** $p<0.001$. **** $p<0.0001$

There were differences in the composition of phyla observed between samples collected from different individuals. Samples B1, B49 and B57 contained a high proportion of *Pseudomonadota* related to the high abundance of species including *E. roggkampii*, *M. morganii* and *E. coli* in these samples. Sample B20 contained a high abundance of *Actinomycetota* due to the high abundance of *Bifidobacterium longum* in this sample (Figure 36).

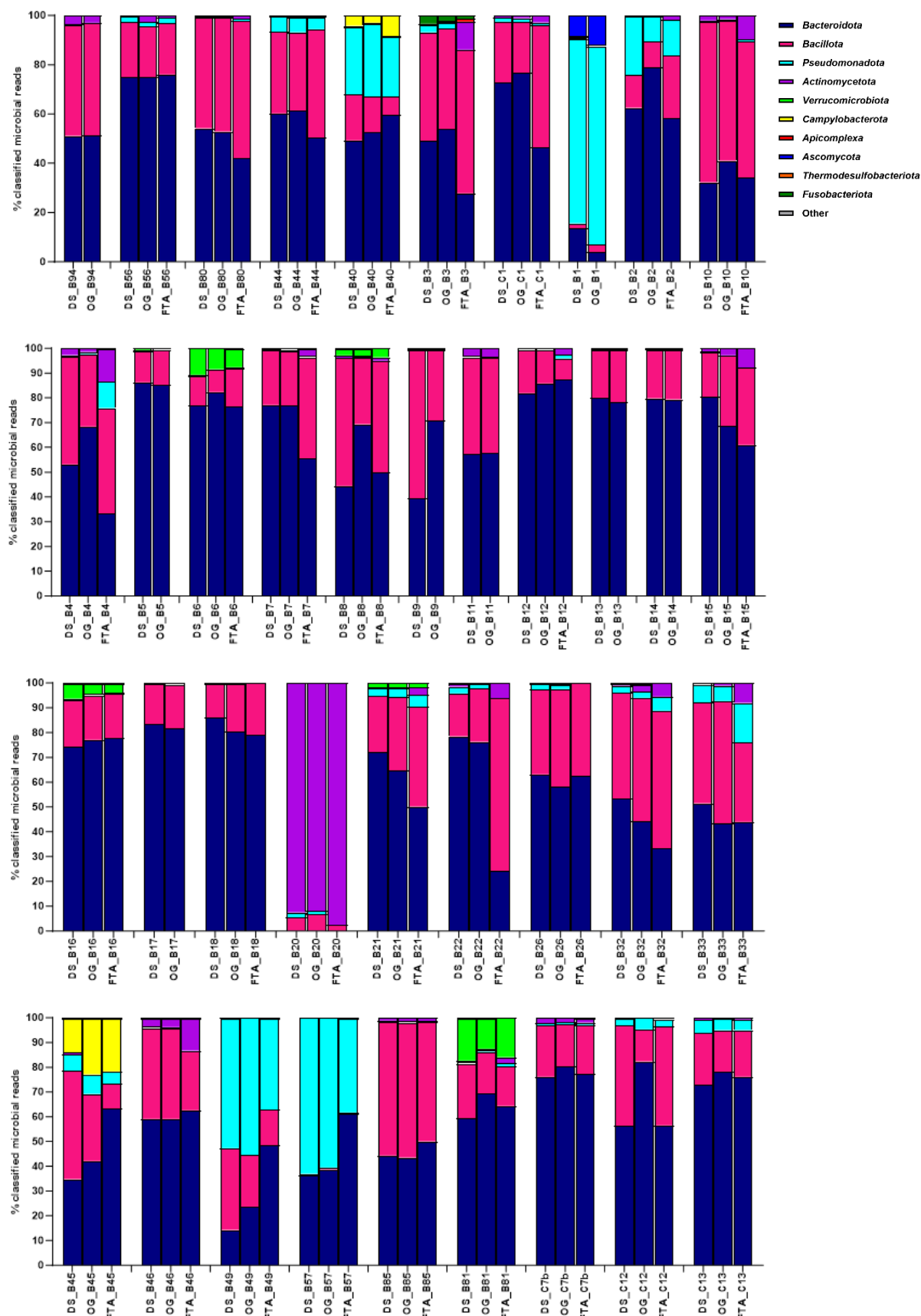


Figure 36. Proportion of classified microbial reads assigned to the most abundant phyla per individual from samples stored in three storage methods and prepared using the RPCR method.

Samples stored using each storage method were prepared for CMg. The proportion of classified microbial reads assigned to the most abundant phyla was then compared between individual samples.

DS = DNA Shield. OG = OMNIgene GUT. DS $n=39$. OG $n=39$. FTA $n=30$.

4.5 Discussion

CMg has the potential to be utilised as a future untargeted diagnostic test to complement the current methods available in deployed MTFs. However, CMg can only be applied effectively if the sample is collected and stored appropriately to preserve all genetic material, and the sample is prepared for sequencing using methods suitable for austere environments. This study aimed to develop stool sample CMg methods by comparing three sample storage and two nanopore-based DNA sequencing library preparation methods for the detection of a protozoan enteric pathogen, *C. hominis*. CMg was applied to stool samples from a retrospective cohort collected during an outbreak of TD amongst UK military personnel undergoing training exercises in Kenya. This work will help to inform the future deployment of CMg to improve the diagnosis of TD amongst deployed military personnel.

4.5.1 Stool sample storage method comparison

The addition of stool samples to preservatives that inactivate bacteria and prevent further growth are known to preserve the genetic material of microbial species, even at ambient temperatures, reducing the need for logistically challenging cold-chain storage. Previous studies have identified DNA Shield, OMNIgene GUT and FTA cards as effective methods for preserving microbial genetic material from stool samples (282, 283, 288) and these methods were compared as part of this study.

Samples stored in the OMNIgene GUT tubes had the highest sensitivity of *Cryptosporidium* detection by CMg analysis compared to the BioFire FilmArray GI panel analysis of the primary stool samples following an extended period of storage. FTA cards had the lowest sensitivity of *Cryptosporidium* detection, failing to identify *Cryptosporidium* sequence reads in the majority of known positive samples, irrespective of the library preparation method used. Despite the reduced sensitivity of *C. hominis* detection by sequencing from the FTA cards, qPCR analysis demonstrated that there was no significant increase in the *C. hominis* Cq values from the FTA card DNA. FTA cards are treated with chemicals to lyse and inactivate bacteria cells whilst capturing DNA in the card matrix, thereby preserving the DNA (289). It may be that a chemical component of the FTA card was co-eluted with the DNA and subsequently inhibited the preparation of the DNA for sequencing. In addition, DNA eluted from FTA cards was significantly more fragmented compared to the other methods, shown by the significantly smaller DNA fragment lengths, an indicator of DNA degradation when using this consumable. The increased fragmentation may also have impacted on the preparation of the DNA for long read nanopore sequencing, resulting in the reduced sequencing read yield. Increased fragmentation may not have had the same impact on the detection of *C. hominis* DNA by qPCR which amplifies a shorter DNA fragment (<200bp), leading to the differences in

sensitivity of detection between CMg and qPCR from the FTA card samples. Therefore, without further optimisation, FTA cards are not a suitable stool sample collection and long-term storage method for nanopore sequencing based CMg.

The sensitivity of *Cryptosporidium* detection by CMg was low across all storage methods compared to the BioFire FilmArray gastrointestinal panel testing of the primary stool samples. The stool samples were collected in February 2022 during the outbreak of TD for routine diagnostics. These samples were made available for CMg analysis following an extended period of storage of 14-16 months, beyond the manufacturer's recommended storage period for stool samples. The genetic material stored within the samples may have degraded during this time, reducing the sensitivity of detection by CMg. The results of the qPCR analysis corroborated the low proportion of *Cryptosporidium* reads identified by CMg, with only five samples returning Cq values of <30, indicating that the amount of *Cryptosporidium* genetic material in the DNA extracts was low. In the absence of the primary stool samples, which were not available for CMg analysis, it is not clear whether this loss of sensitivity is due to the length of storage time, or limitations associated with the CMg workflow. Future studies should aim to include the analysis of stool samples immediately following sample collection and include stool samples stored at -80°C, the current gold-standard for long-term stool sample storage. This would identify the cause of the loss of sensitivity and whether the low proportion of reads assigned to *Cryptosporidium* was because of poor pathogen genetic material load or loss of genetic material during long-term storage.

Whilst the sensitivity of *Cryptosporidium* detection was relatively poor for all methods, notably, there were no false positive identifications in any sample. This specificity was maintained even when the identification of one *Cryptosporidium* read was applied as the threshold for a positive identification. This finding has implications for the future application of CMg for the diagnosis of *Cryptosporidium*, suggesting that the identification of *Cryptosporidium* reads from stool samples, even at very low abundances, is indicative of an infection and requires intervention measures to monitor and prevent an outbreak.

A limitation associated with the application of CMg for the identification of *Cryptosporidium*, highlighted by this study, was that sufficient genome coverage was not obtained to identify the specific *C. hominis* genotype and subtype causing the outbreak. One of the key outcomes of the initial analyses performed on the primary stool samples following repatriation of the samples to the CRU was the identification of the unusual ImA13G1 subtype as the primary cause of the outbreak through amplification and sequencing of the *gp60* gene and the identification of microsatellite repeats downstream of this region (280). Using CMg, it was only possible to identify *Cryptosporidium* at the genus level, with reads incorrectly assigned to the *C. parvum* species by the bioinformatics analysis, likely due to the low number of reads identified by CMg and the

genetic similarities between *Cryptosporidium* species. Therefore, the methods evaluated in this study cannot, at this stage, provide the detailed identification of *Cryptosporidium* genotypes and subtypes that can be crucial for epidemiological investigations into outbreaks. Whilst CMg could provide the initial indication of the start of an outbreak, the repatriation of the samples to a fixed laboratory environment would still be required for subtyping. Enrichment for *Cryptosporidium*, either through the pre-extraction enrichment of the oocysts, similar to those previously developed for enteric pathogens (290), or post-extraction enrichment of *Cryptosporidium* reads through targeted sequence amplification could enable an increased number of reads assigned to *Cryptosporidium* to be sequenced, potentially allowing the detection of the *gp60* gene and microsatellite repeats with the sufficient coverage needed for subtyping. However, the addition of these steps would increase the complexity of the wet-lab methods and increase the time to a definitive result, in addition to potentially reducing the sensitivity of other untargeted pathogens. Moving forward, clinicians will need to inform the priorities for the deployed diagnosis of *Cryptosporidium* by CMg and determine whether the subtyping of *Cryptosporidium* would be advantageous.

4.5.2 Library preparation method comparison

The methods used to process samples for DNA sequencing are known to affect the sensitivity of species detection and the composition of microbial species identified, dependent upon factors including DNA input and sample type (291, 292). A key aim of this study was to develop the methods to enable the identification of enteric pathogens whilst being suitable for use in austere environments. Therefore, nanopore-based sequencing using the RBK OP and RPCR library preparation methods was evaluated, due to the reduced sample handling steps, increased equipment suitability for austere environments and lower total time to result compared to alternative technologies. The RPCR method achieved the highest sensitivity of *Cryptosporidium* detection, and the greatest number of reads assigned to additional enteric pathogens. However, this method added to the method complexity and an additional 2.5 hours to the total time to result. In addition, the required DNA input for the PCR amplification reaction was low enough to require dilution of the DNA samples, further adding to the method complexity. The addition of a PCR amplification step also adds potential for bias to metagenomics data sets that could interfere with the interpretation of the results (293). TD amongst military cohorts is unlikely to result in serious disease, being more likely to cause patient discomfort, lost working hours and increased pressure on healthcare personnel and facilities. Therefore, the clinical community may determine that the additional time to result associated with use of the RPCR method is worth the improved sensitivity of pathogen detection and the increased method complexity may be offset by increased workflow automation. The clinical community should therefore determine the priorities of method optimisation to

focus on their priorities for the diagnosis of TD and this should determine whether optimisation focuses on increasing sensitivity using the RBK OP method or reducing the time to result and complexity of the RPCR method.

4.5.3 Microbiome storage

When comparing the composition of the wider gut microbiome, the storage method did impact the α -diversity at a species level, with a lower species α -diversity observed in the FTA samples compared to the DNA Shield and OMNIgene GUT samples. However, the storage method did not impact the proportion of reads assigned to the most abundant phyla, perhaps surprisingly given the impact on the detection of *C. hominis* and species α -diversity. This suggests that the reduced total sequencing yield from the FTA card samples primarily impacted the detection of species present at low abundances, including *C. hominis*. In contrast, the use of different library preparation methods did impact the proportion of the most abundant phyla, with the RPCR method increasing the proportion of the *Bacteroidota* phylum and reducing the *Bacillota* phylum. However, the library preparation method had no significant impact on the observed α -diversity at a species level. This may be linked to the biases introduced by the PCR amplification step resulting in the over-representation of the more abundant phyla. However, it is not clear why the PCR step would result in increased abundance at the phylum level but not impact the observed species α -diversity. Careful consideration must be taken prior to selecting the library preparation method and further work is required to select the optimal method for the diagnosis of TD.

4.5.4 Detection of additional pathogens

A key advantage of CMg is its untargeted nature, potentially enabling the identification of any causative pathogen of TD. This included the additional pathogen species identified by the BioFire FilmArray gastrointestinal panel, including *C. difficile* and *Campylobacter* species. However, a limitation of the current workflow is that it does not enable the identification of RNA based viruses. Detection of RNA alongside DNA would require the addition of a reverse transcription (RT) step to convert the RNA to complementary DNA (cDNA) prior to sequencing. In addition, a number of studies have shown that the use of pre-extraction viral particle enrichment steps including all or a combination of centrifugation, filtration and nuclease treatments increase the sensitivity of viral pathogen detection by CMg (294, 295). However, this requires the separation of the sample into two workflows so that RNA and DNA-based pathogens can be identified, increasing the method complexity. The enrichment of viral sequences post-extraction has also been demonstrated as an effective method for enabling the detection of viral pathogens by CMg, including the use of large panels of hybridization probes to amplify known viral pathogen sequences, such as the Twist Target Enrichment method (296). This method

does not require the separation of workflows enabling the simultaneous detection of DNA and RNA based pathogens. Future work should focus on the incorporation of or development of a separate viral CMg workflow for stool samples to enable the full range of TD pathogens to be detected.

Whilst the untargeted nature of CMg is an advantage, it also complicates the interpretation of the results. In addition to the presence of pathogen species, the wide range of commensal species that make up most species in stool samples are also identified. In some cases, there may be overlap between these and it may not be clear whether a species is a pathogen or a commensal species from the sequencing data alone. Further work is required to understand the full range of pathogen species implicated in TD and the thresholds at which species should be considered as the causative pathogen of infection. Further understanding is also required of how the pathogen and commensal species interact within the gut microbiome, including the transfer of genetic material (297) and how this impacts on the detection of the aetiological agent of TD. The role of fungal and viral pathogens may also be of relevance to this and should be further investigated to understand the interpretation of metagenomics data for the diagnosis of TD (298). The list of TD pathogens should be compiled by and agreed by the wider research and clinical communities and continually updated as the pathogens involved in TD and wider gastrointestinal infections are more fully understood.

A further complication of the interpretation of CMg results is that some species, such as different pathotypes of pathogenic *E. coli*, require the identification of particular genes or virulence determinants to identify the specific pathotype (299). This includes the presence of specific enterotoxins such as the heat labile and heat stable toxins used to identify enterotoxigenic *E. coli* (300). Some of these pathotype determinants may not have been fully elucidated in the literature at this stage and may be revealed by further WGS and metagenomics studies of the gut microbiome. Therefore, the use of CMg to differentiate between different *E. coli* pathotypes is highly complex and beyond the scope of this study. Therefore, the ability of CMg to identify pathogenic *E. coli* pathotypes has not been evaluated and further work is required to understand the use of CMg to differentiate between pathogen strains and pathotypes in stool samples.

The identification of AMR genes from CMg datasets must also be considered. Identifying the presence of pathogen species resistant to antibiotics is important for selecting the optimal treatment options. However, commensal species present in the gut microbiome are known to act as a reservoir for and horizontally transfer AMR genes to pathogenic species, which can include mobile genetic elements such as plasmids (301, 302). Therefore, it is not always clear whether the identification of an AMR gene is related to the resistance profile of the causative pathogen. In this study, AMR genes were identified in samples that tested negative by BioFire FilmArray gastrointestinal panel testing for

pathogens associated with TD. Assigning the AMR genes identified by CMg to a species through the analysis of the regions flanking the gene would help to better interpret the detection of each AMR species and whether it should be considered when selecting treatment options. This method of analysing the AMR gene flanking regions has been shown to provide useful information on the antibiotic-susceptibility phenotypes and there are analysis pipelines available for this (303). However, this method is bioinformatically complex, time consuming and is not always possible if the region flanking the gene is not identified, which can be the case if the species is present at a low abundance. In addition, if the AMR gene is present within a mobile genetic element, it may not be possible to assign this to a specific pathogen species based on the flanking region alone. Further understanding of the interpretation of the detection of AMR genes from stool samples is required before it can be used to inform the antibiotic treatment of TD.

4.6 Conclusions

The work carried out in this chapter compared three long-term stool sample storage methods and two nanopore-based DNA sequencing library preparation methods for the detection of *Cryptosporidium* by CMg. This work found that FTA cards, although convenient for sample collection, storage and transport, were not suitable for CMg due to the lower total sequencing reads and poor detection of low abundance species, including *Cryptosporidium*. Samples stored in OMNIgene GUT achieved the highest sensitivity of *Cryptosporidium* detection and contained higher concentrations of *C. hominis* DNA. Therefore, this method is recommended for the collection and storage of stool samples for CMg testing in deployed environments. The RBK OP library preparation method, although resulting in the lower total reads and sensitivity of *Cryptosporidium* detection compared to the RPCR library preparation method, is less complex and has a reduced time to result. The focus of future work should be determined by the priorities of the clinical community to address the limitations identified by this study, including the low sensitivity of *Cryptosporidium* detection. Ultimately, this work will contribute to informing the implementation of CMg for deployed TD diagnostics, including the optimal sample storage and preparation methods, improving the diagnosis of TD and reducing the impact of TD on deployed personnel.

4.6.1 Recommended stool sample clinical metagenomics workflow

The work carried out in this chapter resulted in the development of an optimised workflow for the collection, storage and processing of stool samples for the detection of enteric pathogens by CMg. This recommended workflow consists of the following steps:

- Sample collection and storage: samples should be collected and stored using OMNIgene GUT tubes. This method resulted in the highest concentration of DNA and the highest sensitivity of *Cryptosporidium* detection by CMg. In contrast, the

collection and storage of stool samples using FTA cards is not recommended, due to the poor sensitivity of *Cryptosporidium* detection and the low DNA extraction and sequencing yield.

- DNA extraction: in contrast to other clinical sample types including wounds, stool samples do not require a host DNA depletion step due to the relatively low proportion of host genetic material. Therefore, the sample is progressed directly to the DNA extraction step. In this study, the QIAamp PowerFecal Pro DNA extraction kit was used for the OMNIGene GUT and DNA Shield samples and a previously published bespoke method was applied to the FTA cards. However, these methods, although developed for stool samples, were carried out manually, increasing the number of hands-on steps. Future work should evaluate automated extraction platforms, developed for stool samples, to reduce the method complexity and increase the suitability for use in austere environments.
- Library preparation: due to the relatively high concentration of DNA extracted from stool samples, a WGA step was not required and the samples were progressed directly to the library preparation step. Despite the high DNA concentrations, the RPCR library preparation method resulted in the highest sensitivity of *Cryptosporidium* detection and sequencing yield and is recommended. However, future improvements to the sequencing yield of the RBK kit should be monitored, due to the reduced turnaround time and hands-on steps of this method.
- DNA sequencing: similarly to the wound swab workflow, nanopore-based DNA sequencing is recommended due to the potential for real-time data analysis and the reduced size and portability of the equipment. However, due to the high proportion of commensal bacteria in stool samples, and as demonstrated in this study, the causative pathogen of infection can be present at a low proportion of reads. Therefore, the number of reads generated per sample should be maximised to increase the sensitivity of detection of the pathogen species.

5 Evaluation of clinical metagenomics in a deployed medical treatment facility

5.1 Contribution statement

The work detailed in this chapter was carried out as a collaborative exercise with partners from the Defence Centre of Pathology. Major Colin Hudson (Royal Army Medical Corps) was responsible for the collation of respiratory samples collected as part of routine testing. Routine diagnostic testing was carried out by the microbiology laboratory at Frimley Park hospital. The processing of respiratory samples for CMg on board the Royal Fleet Auxiliary (RFA) Argus was carried out by Sergeant Jack Robson (Royal Air Force) and Surgeon Commander Matthew O'Shea (Royal Navy). Method development and testing, additional testing of the sputum sample DNA extracts by BioFire FilmArray multiplex PCR, analysis of the viral CMg sequencing data using the EPI2ME Antimicrobial Resistance Mapping Application (ARMA) pipeline and results interpretation was carried out by Carl Halford. The bioinformatics analysis of the bacterial CMg workflow sequencing data was carried out by Dr. Thanh Le-Viet (Quadram Institute Bioscience) in close collaboration with Carl Halford using a bioinformatics pipeline that Dr Thanh Le-Viet previously developed.

5.2 Introduction

5.2.1 Lower respiratory tract infections

Lower respiratory tract infections (LRTIs) are a significant cause of morbidity and mortality worldwide, responsible for 2.59 million deaths, 4.7% of the total all-cause deaths and the most deaths of any communicable disease, in 2019 (304). Pneumonia, an acute infection of the lung parenchyma by one or multiple co-infecting pathogens (305), and bronchiolitis, an acute viral infection of the lower respiratory tract (306) are considered to be the most important LRTIs. Pneumonia is caused by a wide range of aetiological agents, including bacterial, viral and fungal species. These vary according to geographic location (307), age and gender of the patient (308), existing comorbidities (309) and whether the patient is situated within a healthcare facility. For example, community acquired pneumonia in children under five years of age is commonly caused by viral pathogens, accounting for 30-67% of cases, whilst bacterial infections are more common in older children (310). Nosocomial pneumonia is commonly caused by bacterial species including MRSA, *P. aeruginosa*, extended-spectrum β -lactamase-producing *Enterobacteriaceae* spp., and multi drug-resistant *A. baumannii* (311). Bronchiolitis mainly affects young children, causing the hospitalisation of 2-3% of infants younger than one year. Respiratory syncytial virus is the main cause of Bronchiolitis, detected in 60-80% of cases (312).

LRTIs are prevalent amongst military personnel both during training exercises and when on deployment, with acute respiratory diseases being the most common cause of the

need for outpatient treatment and hospitalisation (313). Deployed military environments, associated with exposure to endemic respiratory pathogens, mixing of personnel in crowded living conditions and physical and psychological stress, create a high risk for the spread of respiratory diseases (314). The prevalence of respiratory disease is also high amongst new military recruits, with upwards of 80% becoming symptomatic and 20% requiring medical treatment within two months of training, placing a significant burden on healthcare facilities and resulting in the loss of working days and reduced operational readiness. Adenoviruses and influenza A and B viruses are amongst the most commonly identified pathogens associated with these infections (315).

More recently, the emergence and rapid development and spread of novel variants of COVID-19, caused by the SARS-CoV-2 virus (316-318), placed significant pressures on the ability of military populations across the globe. In April 2020, an outbreak of COVID-19 on board the French aircraft carrier Charles de Gaulle resulted in 1148 of the 1767 personnel on-board the ship testing positive for COVID-19 by RT-PCR. Whilst only 2.6% of positive cases resulted in severe symptoms and only three staff were hospitalised, this outbreak had a significant impact on the deployment. Personnel were required to undertake activities to reduce the spread of the infection including remote working, separation of personnel into cohorts with minimal mixing between cohorts, and lockdowns, all of which were difficult to achieve and manage within the confined space of a ship. These activities resulted in reduced staff availability and the interruption of military activities (319).

5.2.2 Deployed lower respiratory tract infections diagnostics

The impact of respiratory infections on military personnel and the effectiveness of the deployments means that the ability to diagnose LRTIs rapidly in deployed environments is highly important. Current methods available within role two MTFs rely upon the use of rapid molecular tests. This includes the BioFire FilmArray pneumonia plus panel, which enables the detection and quantification of 18 bacterial pathogens, nine viruses and seven AMR genes with good sensitivity and specificity (28). The BioFire FilmArray respiratory panel 2.1 plus panel is also available, testing for 18 viral and four bacterial targets, including five influenza viruses and respiratory syncytial virus with high sensitivity and specificity (26). Importantly, this panel has been updated to include the detection of SARS-CoV-2, providing a highly sensitive method of detection suitable for use in deployed environments (29). This automated platform tests one sample per run, extracting the DNA and carrying out the multiplex PCR reactions sequentially without the need for operator input once the sample has been loaded. Whilst the platform is simple to operate, the limitation of running one sample per run can limit the utility of the platform in outbreak situations in which hundreds of samples may require testing to limit the spread of an outbreak. Microbiology based tests are available at role two enhanced and role three

environments. However, as discussed previously, microbiological culture-based tests, whilst being the gold standard methods for diagnosis, require static incubator equipment and the supply (and periodic resupply) of reagents, proving to be logistically challenging. Further, the growth of microorganisms requires 18-36 hours to produce an actionable result, and the identification of fastidious organisms is limited. Even in well-resourced civilian diagnostics laboratories that have access to a range of methods including molecular and microbiological culture based tests, 25-47% of community acquired LRTIs go undiagnosed (320). The untargeted nature of CMg, with the potential to identify both the more common aetiological agents alongside the rarer agents that may not be expected or tested for using other methods, presents an opportunity to provide a diagnosis when other methods fail. Already, CMg has been applied to respiratory samples both in case reports (321) and more recently, as part of more routinized service evaluations within diagnostic laboratories (72, 322). The additional detection of genotypic AMR genes has the potential to further inform treatment options, with innovative methods of improving the sensitivity of AMR gene detection becoming available through enrichment of low abundance sequences (323), adding to and complementing the current diagnostic tests available in deployed environments.

5.2.3 Deploying clinical metagenomics workflows to the RFA Argus

The work carried out and described thus far in this project involved the development of CMg workflows for various sample types relevant to military patient cohorts with the aim of reducing the method complexity and operator hands on time whilst reliably identifying the aetiological agent. Ensuring that the methods were suitable for use in deployed military MTFs was crucial to demonstrating the potential for the future utility of CMg to the military clinical community, who would be the end-users of any diagnostic technology developed by the project. However, all this work was carried out within static laboratory environments and therefore, was not a true reflection of the difficulties associated with operating within austere environments. The only reliable method of better understanding these challenges was to process samples within a deployed environment and identify the technical and logistical challenges first-hand. Through links developed with the military clinical community, an opportunity to work directly with the military end-users to deploy and operate a CMg workflow within a role three military medical environment, the RFA Argus ship, became available.

The RFA Argus operates as a primary casualty receiving ship and is equipped with a role three MTF, including a 100-bed hospital with resuscitation and surgical facilities, a ten-bed critical care unit and a 20-bed high dependency unit. The ship is also equipped with fixed laboratory facilities for diagnostic activities. BMSs from all three services deploy with the RFA Argus to support the delivery of medical care on board within a clinical laboratory. This involves a range of activities, including performing diagnostic tests and the screening

of blood products. In-between deployments, BMSs embed into and support NHS diagnostic laboratories. Therefore, while BMSs have a range of related skills, the preparation of samples for DNA sequencing represents a divergence from their day-to-day activities and required additional training.

The work described in this chapter comprised of two proof of concept exercises, carried out in collaboration with the Defence Centre of Pathology, with the aims of identifying the technical and logistical requirements of deploying and operating a CMg workflow in a deployed MTF. The first exercise involved the processing of excess nasopharyngeal swabs collected during routine testing of military personnel for COVID-19. This exercise was organised with a rapid turnaround, with just six-weeks separating the inception of the project and delivery of the equipment to the RFA Argus. Due to the samples being collected as part of routine COVID-19 testing, the CMg workflow was adapted, incorporating methods previously developed for the detection of RNA viruses by CMg (294, 295, 324), to enable the detection of viral pathogens. Subsequently, the BMS was trained in the workflow procedures, including the DNA extraction, RT and library preparation and nanopore sequencing. The equipment and samples were then delivered onto the RFA Argus, and the samples were processed on board the RFA Argus by the BMS. Subsequently, a second opportunity to deploy a CMg workflow to the RFA Argus became available. Due to challenges associated with detection of viral pathogens, this exercise aimed to detect bacterial pathogens from excess sputum samples. A CMg workflow was adapted from a previously published method that successfully identified respiratory pathogens with good sensitivity and specificity using nanopore-based DNA sequencing within a fixed laboratory environment (62) and deployed to the RFA Argus to process excess sputum samples and the results compared to routine microbiology testing.

5.2.4 Aims

The aims of this chapter were to evaluate the operation of CMg workflows within a deployed MTF by a military BMS to identify the logistical and technical challenges and inform the future optimisation of CMg wet-lab methods. Whilst the development of a CMg workflow for respiratory samples was not intended to be a focus of this project, the processing of respiratory samples for CMg and compared to routine microbiology would provide a greater understanding of the interpretation of CMg outputs for the identification of the causative pathogen of a respiratory infection, building upon the work previously carried out by other research groups.

5.3 Methods

5.3.1 Exercise one: viral clinical metagenomics workflow

5.3.1.1 Sample collection

Ten nasopharyngeal swab samples were collected for routine COVID-19 testing from UK military personnel and tested using the BioFire FilmArray respiratory panel RP2.1.

Following the completion of routine testing, samples were anonymised and provided for CMg analysis as part of the proof of concept exercise. Permission for the use of the excess samples was provided by the Defence Centre of Pathology as part of a service evaluation exercise.

5.3.1.2 RNA extraction

Processing for CMg was carried out by a Royal Air Force BMS on board the RFA Argus within the role three MTF. The laboratory set-up, including the mobile workstation, thermal cyclers and MK1c sequencing device (Oxford Nanopore Technologies) is shown in Figure 37.



Figure 37. Image of the equipment and laboratory set-up used on board the RFA Argus to process samples for CMg.

Image credit: Sergeant Jack Robson.

Two hundred microliters of viral transport media was taken from each swab sample tube and added to a separate labelled 1.5 mL tube. One negative control consisting of DNase/RNase free water (Invitrogen) was included at this stage and processed simultaneously. Multiple viral particle enrichment steps were then applied, as recommended

by previous viral metagenomics studies for the removal of host genetic material and increasing the sensitivity of viral detection (294, 295). These viral enrichment steps were selected due to their compatibility with the equipment available for the exercise and due to constraints on the reagents available. For example, filtration of the sample through a 0.45 μM polyethersulfone filter was omitted due to delays to the delivery of the required reagents. Each tube was centrifuged at 13400 rpm for 12 min to pellet the human and bacteria cells and the supernatant was transferred to a new labelled 1.5 mL tube. RNA was then extracted using the Quick-RNA Viral Kit (Zymo Research) according to the manufacturer's instructions, selected due to the end-user's familiarity with the protocol. Four hundred microliters of Viral RNA Buffer were added to 200 μL of the supernatant and mixed by pipetting. The mixture was then transferred to a Zymo-Spin IC Column in a Collection Tube and centrifuged for 2 min at 13400 rpm. Four hundred microliters RNA Wash Buffer was added to the column and the column was centrifuged for 1 min and the flow-through discarded. The DNase I reaction mix was then prepared by mixing 35 μL DNA Digestion buffer with 5 μL reconstituted DNase I enzyme (1 U/ μL) and 40 μL of this mix was then added to the matrix of the Zymo-Spin IC Column and incubated at room temperature for 15 min. This DNase incubation step represented the second viral enrichment step. Five hundred microliters of RNA prep buffer were then added to the column and centrifuged at 13400 rpm for 1 min. The flow through was discarded and the treated sample proceeded to RNA purification. Five hundred microliters Viral Wash Buffer was added to the column, the column was centrifuged at 13400 rpm for 30 s and the flow through discarded and the step repeated. Five hundred microliters of 95% ethanol were added to the column and centrifuged at 13400 rpm for 1 min. The column was then transferred to a new 1.5 mL elution tube and 15 μL DNase/RNase free water was added to the column and centrifuged at 13400 rpm for 30 s to elute the RNA.

5.3.1.3 Whole Transcriptome Amplification (WTA)

The protocols for cDNA synthesis and WTA using the Qiagen WTA single cell kit (Qiagen) were adapted from methods described by Fomsgaard *et al.* 2022 (324) and Rosentierne *et al.* 2014 (325). Firstly, 11 μL of purified RNA was added to a 0.2 mL PCR tube. Two microliters of gDNA wipeout buffer were added to the RNA and incubated at 42°C for 10 min on a miniPCR® mini16 thermal cycler (miniPCR Bio). Samples were added to a 96-well cool block to cool. The RT mix was then prepared by mixing 4 μL of RT/Polymerase buffer, 1 μL Random Hexamers (25 ng/ μL), 1 μL Random Primer and 1 μL Quantiscript RT Enzyme Mix. Seven microliters of the RT mix were then added to the purified RNA and incubated at 42°C for 60 min followed by 95°C for 5 min on a thermal cycler. The tubes were then placed on a 96-well cool block to cool. The ligation mix was then prepared in a 0.2 mL PCR tube by mixing 8 μL Ligase Buffer and 2 μL Ligase Mix. Ten microliters of the ligation mix were added to the sample tube and incubated at 24°C for 30 min and 95°C for 5 min on a thermal

cycler and the tubes placed on a 96-well cool block to cool. The REPLI-g SensiPhi amplification mix was then prepared in a 1.5 mL tube by mixing 29 μ L REPLI-g SC Reaction Buffer and 1 μ L REPLI-g SensiPhi DNA Polymerase. Thirty microliters of the REPLI-g SensiPhi amplification mix was added to the ligation reaction from the previous step and incubated at 30°C for 2 h and 65°C for 5 min on a thermal cycler. The tubes were then placed on a 96-well cool block to cool.

5.3.1.4 DNA quantification

Total DNA concentration was quantified using the dsDNA HS assay kit on the Qubit 4 fluorometer according to the manufacturer's instructions, as previously described.

5.3.1.5 Library preparation and sequencing

Quantified samples were then prepared for sequencing using the Rapid Barcoding (SQK-RBK004) library preparation kit (Oxford Nanopore Technologies) according to the manufacturer's instructions. Ten samples plus the negative control were multiplexed on the sequencing run. Up to 400 ng DNA in a total volume of 7.5 μ L was added to a 0.2 mL PCR tube and 2.5 μ L Fragmentation Mix RB01-12 was added and the reaction incubated at 30°C for 1 min and then at 80°C for 1 min on the thermal cycler. The tubes were then placed on a 96-well cool block to cool. Five microliters from each library were pooled into a 1.5 mL LoBind tube and 10 μ L of the pooled libraries was transferred to a new 0.2 mL PCR tube. One microliter of Rapid Adapter was added to 10 μ L of the pooled libraries and the tube was incubated at room temperature for 5 min. A R9.4.1 Flow cell (Oxford Nanopore Technologies) was inserted into an MK1c sequencing device (Oxford Nanopore Technologies). The flow cell was then checked using the flow cell check programme. Once checked and if there were >800 pores available for sequencing, the flow cell was primed for loading. Thirty microliters of thawed and mixed Flush Tether were added directly to the tube of thawed and mixed Flush Buffer and mixed by vortexing at room temperature. Eight hundred microliters of this priming mix were then added to the priming port of the flow cell. The flow cell was then incubated on the MK1c device for 5 min at room temperature. Once incubated for 5 min, the pooled libraries were prepared for loading. Thirty-four microliters of sequencing buffer, 25.5 μ L of loading beads and 4.5 μ L of nuclease-free water were added directly to the tube containing the pooled libraries. The SpotOn sample port cover was lifted and 200 μ L of the priming mix was added to the priming port. Seventy-five microliters of the prepared library were loaded into the SpotOn sample port in a dropwise fashion. The SpotOn sample port and Priming port covers were then replaced. A new experiment was then set up using the MK1c MinKnow user interface. The library preparation kit was selected, and the sequencing run set for 12 h duration. The sequencing run was then started, and the data stored in Fast5 and Fastq format.

5.3.1.6 Bioinformatics analysis

Analysis of the sequencing data to determine the number of reads assigned to the viral pathogens was carried out using the EPI2ME ARMA pipeline (Oxford Nanopore Technologies) following transfer of the MK1c to Dstl Porton Down. The EPI2ME ARMA pipeline includes a species identification workflow, known as What's in my Pot (WIMP). WIMP uses Kraken to map kmers to the Refseq database, identifying at the strain or species level, or at a higher taxonomic level if the species is not present in the database (326). The additional ARMA pipeline then uses the CARD for AMR gene identification, applying the default threshold for identification of >75% alignment identity and >40% coverage.

5.3.2 Exercise two: bacterial clinical metagenomics workflow

5.3.2.1 Sample collection

Twenty sputum samples were collected from patients with a suspected LRTI as part of their routine care at Frimley Park Hospital, UK. Samples were stored at 4-8°C until the completion of routine microbiological culture testing, carried out by the Frimley Park Hospital diagnostic laboratory. Samples were anonymised and provided for CMg analysis as part of a service evaluation exercise.

5.3.2.2 Routine microbiology testing

Samples were processed for routine microbiology according to the procedures detailed by the UK SMI: Investigation of bronchoalveolar lavage, sputum and associated specimens (32). Plates were examined at 18-24 h and again at 48 h post-inoculation for growth with extended incubation if required. Potentially significant isolates were selected and identified according to colony morphology, growth characteristics, Gram stain and rapid biochemical tests. Confirmation of colony identity was carried out by MALDI-TOF-MS or by extended biochemical tests including API® test strips (Biomérieux). The level of identity reported was based on the clinical presentation, the isolate and following any specific clinical instructions. AST was carried out by disc diffusion assay. Pure isolates were streaked against antibiotic panels according to the isolate. Conduct of antibiotic susceptibility testing and interpretation of zones of inhibition was done according to EUCAST guidelines (36).

5.3.2.3 DNA extraction

Suitable containment level 3 laboratories were not available on board the RFA Argus at the time of the exercise. Therefore, samples underwent the host depletion and heat inactivation steps at Frimley Park hospital prior to delivery to the RFA Argus. An equal volume of sputum sample and Remel™ sputasol solution (ThermoFisher) were mixed and incubated at 37°C with periodic shaking until liquefied. Four hundred microliters of liquefied sample were

transferred into a 2 mL microcentrifuge tube and centrifuged at 11000 rpm for 5 min. The supernatant was removed, and the pellet resuspended in 200 μ L PBS. The sample then underwent the optimised one-pot saponin host depletion. Forty microliters of 30% saponin in PBS, 200 μ L HLSAN buffer (5.5 M NaCl and 100 mM $MgCl_2$ in molecular-grade water) and 10 μ L HLSAN DNase enzyme were added to the sample and mixed by pipetting. The tube was then incubated at 37°C for 10 min at 1000 rpm on a thermomixer. Nine hundred microliters PBS was added, and the tube centrifuged at 13400 rpm for 3 min. The supernatant was discarded, and the pellet re-suspended in 200 μ L PBS. The sample was then heat inactivated at 80°C for 30 min and then frozen at -20°C ready for transfer to the RFA Argus.

On board the RFA Argus, the DNA was extracted from the depleted samples using the ZYMO Quick-DNA Fungal/Bacterial Miniprep Kit (ZYMO Research) according to the manufacturer's instructions. Two hundred microliters of sample were added to a labelled ZR BashingBead™ Lysis Tube (0.1 mm & 0.5 mm). A negative control of nuclease-free water only was included at this stage and processed simultaneously. Seven hundred and fifty microliters of BashingBead Buffer was then added and the tubes were secured to a vortex with 2 mL adapter and vortexed at maximum speed for 20 min. The tubes were then centrifuged at 12200 rpm for 1 min and 400 μ L of supernatant was transferred to a Zymo-Spin™ III-F Filter unit placed in a collection tube and centrifuged at 11000 rpm for 1 min. The filter unit was removed and 1200 μ L of Genomic Lysis Buffer was added to the filtrate. Eight hundred microliters of the mixture were added to a new Zymo-Spin™ IICR Column placed in a new collection tube and centrifuged at 12200 rpm for 1 min. The flow through was discarded and the step repeated. The Zymo-Spin™ IICR Column filter unit was placed into a new filter unit and 200 μ L DNA Pre-Wash Buffer was added to the column and was centrifuged at 12200 rpm for 1 min. Five hundred microliters of g-DNA Wash Buffer were added to the column and centrifuged at 12200 rpm for 1 min. The column filter unit was added to a 1.5 mL microcentrifuge tube and 40 μ L of DNA Elution Buffer was added directly to the column matrix. The column was centrifuged at 12200 rpm for 30 s to elute the DNA. The 40 μ L DNA extract was then either stored at -20°C or progressed directly to library preparation.

5.3.2.4 DNA purification and quantification

The 40 μ L DNA extract was purified by AMPure XP bead purification according to the manufacturer's instructions. AMPure XP beads were added to the DNA extract at a ratio of 1.2 $\times$ beads to DNA extract in a 1.5 mL LoBind tube. Following the DNA purification, carried out as previously described, the DNA was eluted into 20 μ L nuclease-free water. The purified DNA was then quantified using the dsDNA HS assay on the Qubit 4 fluorimeter, as previously described.

5.3.2.5 DNA library preparation and sequencing

DNA was then prepared for sequencing using the Rapid PCR Barcoding kit. Firstly, 2.5 µL Fragmentation Mix was added to 10 ng DNA made up to 7.5 µL total volume with nuclease-free water, for a total volume of 10 µL and the reaction heated to 30°C for 1 min and 80°C for 1 min on the thermal cycler. The sample then underwent the PCR reaction with the following reagent volumes; 10 µL fragmented DNA sample, 2 µL Rapid Barcode, 50 µL LongAmp Taq 2 × Master mix (New England Biosciences) and 38 µL nuclease-free water for a total reaction volume of 100 µL. The PCR programme used was initial denaturation for 3 min at 95°C (1 cycle), denaturation for 15 s at 95°C, annealing for 15 s at 56°C, extension for 4 min at 65°C (25 cycles) followed by a final extension of 4 min at 65°C (1 cycle). Up to 11 libraries, including 1 × negative control, were then pooled and purified by 0.6 × AMPure XP bead purification as previously described. The purified DNA was then quantified by Qubit dsDNA HS assay as previously described and 370 ng of DNA was taken forward to the adapter ligation step in a total volume of 10 µL. The pooled libraries then underwent adapter ligation and library loading according to the manufacturer's instructions as previously described. Libraries were sequenced on an MK1c device on an R9.4.1 flow cell for 24 h and the data stored in Fast5 and Fastq formats.

5.3.2.6 Bioinformatics analysis

ONT Taxonomic Profiling, a custom bioinformatics pipeline for the taxonomic profiling of microbial species and for the identification of AMR genes, previously developed for the analysis of metagenomics datasets for infectious disease diagnostics was applied (63). Fast5 files were base-called and de-multiplexed using the guppy_basecaller Super Accuracy model and the guppy_barcode of the Guppy software v6.2.x, respectively and the minimum barcoding score was set to 65. Minimap2 v2.17 was used to remove human reads from the data by mapping against the *Homo sapiens* reference genome CHM13v2 and then to map to the remaining reads to the RefSeq database to identify the microbial species. The reads were then filtered based on the harmonic mean of identity and coverage of the mapped reads using the following equation:

$$(2 \times \frac{[identity + coverage]}{[identity \times coverage]})$$

Reads with a taxa score ≥ 85 were included for taxonomic profiling. The pipeline also included applying the KMA (version 1.4.9) tool to screen for AMR genes with the amrfinderplus (v3.10) database. Microbial species identification was presented as the total number of reads assigned to the species and the percentage of the total classified microbial reads aligned to the species. Species were concluded to be present if ≥ 5 reads, minus the number of reads of the same species identified in the negative control sample,

matched that species. AMR genes were concluded to be present if their template identity and coverage were $\geq 90\%$ and their depth of coverage reported by KMA were $\geq 1X$.

5.3.3 BioFire FilmArray pneumonia panel plus analysis

The DNA extracts were stored at -20°C on board the ship and were transferred to Dstl Porton Down for additional analysis using the BioFire FilmArray pneumonia panel plus. Firstly, the pouch was loaded onto the pouch loading station and hydrated using the hydration vial. The sample buffer was added to the sample injection vial and the DNA extract was added directly to the sample buffer and mixed by pipetting. The sample mix was then added to the pouch via the sample port. The prepared pouch was then loaded into the BioFire FilmArray 2.0 platform, and the run started using the BioFire FilmArray 2.0 software sputum protocol.

5.4 Results

5.4.1 Exercise one: Deployment of a viral pathogen clinical metagenomics workflow to the RFA Argus

Routine testing of nasopharyngeal samples collected as part of asymptomatic COVID-19 testing of military personnel was carried out and ten samples were provided for CMg analysis on board the RFA Argus. Eight samples reported the identification of ≥ 1 respiratory virus, including adenovirus, parainfluenza 2 and 3, influenza A and B virus and respiratory syncytial virus. The remaining two samples were reported as negative (Table 17). CMg sequencing produced a mean total reads of 13574 (range; 1019-25010) per sample. The mean number of human reads per sample was 10811 reads (range; 672-23259) compared to a mean number of viral reads of 110 (range; 4-247). Of the viruses reported by BioFire FilmArray respiratory panel RP2.1 testing, the same virus was not reported in any sample by CMg (Table 17).

Table 17. Results of the CMg analysis of nasopharyngeal swab samples processed on board the RFA Argus.

Sample	Total Reads	Bacteria reads	Viral reads	Human reads	BioFire FilmArray respiratory panel 2.1 report	Corresponding viral pathogen identified by CMg
1	1019	341	4	672	Adenovirus	None
2	21645	1346	247	19844	Adenovirus	None
3	25010	1350	240	23259	Parainfluenza 2	None
4	23935	1180	138	22540	Parainfluenza 3	None
5	9414	893	79	8401	Influenza A virus H1N1	None
6	14701	1201	113	13349	Influenza B virus	None
7	9696	610	105	8940	RSV and Parainfluenza 3	None
8	8272	711	107	7436	Influenza B virus and Parainfluenza 3	None
9	18468	14778	50	2950	Negative	None
10	3578	2790	20	715	Negative	None

Viral transport media from nasopharyngeal swab sample tubes collected as part of routine COVID-19 testing of military personnel was processed for CMg on board the RFA Argus as part of a proof of concept exercise. The sequencing data was analysed using the Epi2me ARMA pipeline. No viral pathogens identified by the BioFire FilmArray respiratory panel RP2.1 were identified by CMg.

5.4.2 Exercise two: deployment of a bacterial clinical metagenomics workflow to the RFA ARGUS

Twenty excess sputum samples were collected from patients with a suspected LRTI and 19 were processed for standard of care microbiological culture testing by an NHS laboratory. One sample was provided for CMg testing but had not undergone culture testing. Five out of nineteen (26%) samples reported the presence of upper respiratory tract commensal (URTC) species without a respiratory pathogen present. The remaining 14/19 (74%) samples reported the presence of ≥ 1 respiratory pathogens, identifying a total of 22 respiratory pathogens. Multiple pathogens were detected in 5/14 samples and pathogens were reported to be mixed with respiratory commensal species in six samples. *Haemophilus influenzae* and *P. aeruginosa* were the most identified respiratory pathogens ($n=4$) followed by *S. aureus* ($n=3$).

Working on board the RFA Argus within the role three MTF, the BMS successfully processed and generated sequencing reads from all samples. The mean total number of reads per sample was 346634. The mean number of human reads per sample was 61144 and the mean number of classified microbial reads was 285490. Of the 22 respiratory pathogens reported by routine microbiology, 14/22 were identified by CMg at $\geq 1\%$ of

classified microbial reads. An additional three reported pathogens were identified by CMg at $\geq 0.1\%$ classified microbial reads. From the culture positive samples, CMg identified additional respiratory pathogens in 9/14 samples. In the samples in which microbiology identified URTCs only, CMg agreed in 2/5 samples, with no respiratory pathogens identified. CMg identified additional pathogens not identified by microbiology in 3/5 of these samples. *S. pneumoniae* was the most identified respiratory pathogen by CMg ($n=7$), followed by *H. influenzae* ($n=3$).

To provide additional corroboration of the CMg results, the DNA extracts were repatriated from the ship and transferred to Dstl Porton Down for additional analysis using the BioFire FilmArray pneumonia panel plus v2.0. Fourteen of the 20 DNA extracts tested positive for the detection of ≥ 1 respiratory pathogen, reporting 20 respiratory pathogen species in total, including 19 bacteria and one virus. Of the 19 bacteria species reported, CMg identified 12/19 (63.2%) at $\geq 1\%$ classified microbial reads. This increased to 16/19 (84.2%) when applying the threshold of $\geq 0.1\%$ of classified microbial reads. Additional pathogens were identified by CMg in three samples that tested negative by BioFire FilmArray analysis, with all three being *S. pneumoniae*. The results of routine microbiology, CMg and BioFire FilmArray pneumonia panel plus analysis are presented in Table 18.

Table 18. Results of CMg from sputum samples processed on board the RFA Argus and additional routine microbiology and BioFire FilmArray pneumonia panel plus testing carried out in the UK.

Sample	Total Reads	Human Reads	Classified microbial reads	Organism identified by Culture	BioFire FilmArray Pneumonia panel plus report (copies/mL)	Respiratory pathogens identified by CMg	Respiratory pathogen reads	Proportion of classified microbial reads
1	339265	2307	233017	URTC	Negative	<i>S. pneumoniae</i>	29536	12.7
2	592078	590361	922	SG <i>S. aureus</i> . Mixed with respiratory flora.	<i>S. aureus</i> (10^4)	None	-	-
3	363967	33667	251502	URTC	Negative	None	-	-
4	494087	53065	342112	Light growth (LG) <i>H. influenzae</i>	<i>H. influenzae</i> (10^6)	<i>S. pneumoniae</i> <i>H. influenzae</i>	9154 2619	2.7 0.8
5	44565	25010	15595	Routine microbiology testing results not reported.	Negative	<i>S. pneumoniae</i>	288	1.8
6	511417	6495	355694	<i>S. oralis</i> mixed with URTCs	Negative	<i>S. pneumoniae</i>	7392	2.1
7	474205	8346	345198	Moderate growth (MG) <i>H. influenzae</i> . Mixed with respiratory flora.	<i>H. influenzae</i> (10^7)	<i>H. influenzae</i> <i>S. pneumoniae</i>	13893 5559	4.0 1.6
8	403305	79330	320988	MG <i>S. aureus</i>	<i>S. aureus</i> (10^6) <i>H. influenzae</i> (10^5)	<i>S. aureus</i> <i>H. influenzae</i>	314496 4128	98.0 1.3
9	686981	14564	521873	MG <i>H. influenzae</i> SG <i>K. pneumoniae</i> . Mixed with respiratory flora.	<i>Moraxella catarrhalis</i> (10^7)	<i>M. catarrhalis</i>	113544	21.8

10	583793	1693	441744	URTC	Parainfluenza virus	<i>Neisseria meningitidis</i> <i>Aggregatibacter</i> sp. oral taxon 513 <i>S. pneumoniae</i>	12804 8829 5938	2.9 2.2 1.5
11	60329	4772	46988	MG <i>S. pneumoniae</i>	<i>S. pneumoniae</i> (10 ⁶)	<i>S. pneumoniae</i>	172	0.4
12	69781	5757	55134	Heavy growth (HG) <i>Serratia</i> spp.	<i>Serratia marcescens</i> (10 ⁷)	<i>S. marcescens</i>	28547	51.6
13	61632	302	47578	HG <i>P. aeruginosa</i> MG yeast LG <i>S. marcescens</i>	<i>P. aeruginosa</i> (10 ⁷) <i>S. marcescens</i> (10 ⁶)	<i>P. aeruginosa</i> <i>Stenotrophomonas maltophilia</i> <i>Aggregatibacter aphrophilus</i> <i>S. marcescens</i>	10464 4974 1253 443	22.0 10.5 2.6 0.9
14	57857	81	53142	MG <i>M. catarrhalis</i> LG <i>P. aeruginosa</i> <i>Neisseria</i> spp. Mixed with respiratory flora.	<i>M. catarrhalis</i> (10 ⁷) <i>P. aeruginosa</i> (10 ⁷)	<i>M. catarrhalis</i> <i>Neisseria mucosa</i> <i>P. aeruginosa</i>	47419 3070 158	89.2 5.8 0.3
15	76487	4268	56412	URTC	<i>S. aureus</i> (10 ⁵)	None	-	-
16	274964	19220	177184	MG <i>M. catarrhalis</i> . Mixed with respiratory flora.	Negative	None	-	-
17	545312	1745	506232	MG <i>S. marcescens</i>	<i>S. marcescens</i> (10 ⁷) <i>H. influenzae</i> (10 ⁶)	<i>S. marcescens</i> <i>K. pneumoniae</i>	385530 13009	76.2 2.6
18	335856	8430	259281	LG <i>H. influenzae</i> Mixed with respiratory flora.	<i>H. influenzae</i> (10 ⁶)	<i>H. influenzae</i> <i>S. pneumoniae</i>	12030 5206	4.6 2.0
19	385404	18114	346898	MG <i>P. aeruginosa</i> LG <i>S. aureus</i>	<i>P. aeruginosa</i> (10 ⁷) <i>S. aureus</i> (10 ⁷)	<i>S. aureus</i> <i>P. aeruginosa</i>	242342 65793	69.9 19.0
20	571399	345354	192591	MG <i>P. aeruginosa</i> LG <i>Achromobacter</i> spp. SG yeast	<i>P. aeruginosa</i> (10 ⁷)	<i>P. aeruginosa</i> <i>S. maltophilia</i> <i>C. albicans</i> <i>Achromobacter xylosoxidans</i>	159910 8881 7804 7086	83.0 4.6 4.1 3.7

Excess sputum samples collected from NHS patients with a suspected LRTI were transferred to the RFA Argus for CMg testing, as part of a proof of concept exercise. DNA extracts were returned to Dstl Porton Down for BioFire FilmArray pneumonia panel plus analysis. Respiratory pathogen species identified by routine microbiology, BioFire FilmArray analysis and CMg are reported. % classified microbial reads = the number of reads assigned to a species over the total number of taxonomically classified microbial reads obtained.

The results of AST carried out as part of routine microbiology testing were provided for six samples. Of these six samples, AMR genes associated with resistance to the antibiotic class and species identified by routine microbiology AST was also identified by CMg in sample four *H. influenzae* was determined to be β -lactamase resistant by AST and CMg identified the *bla*_{TEM} AMR gene, associated with resistance to penam, penem, cephalosporin and monobactam antibiotics in *H. influenzae*, according to the CARD (216). However, AMR genes associated with resistance to tetracyclines and fluoroquinolones, amongst other antibiotic classes, were also identified by CMg which were not identified by AST. Notably, resistance to tetracyclines and macrolides were most identified by CMg but were not identified by AST. *MefA*, associated with resistance to macrolide antibiotics, was the most identified AMR gene, identified in 14 samples. Resistance genes associated with resistance to tetracycline were the most identified, identified 63 times in 14/19 samples, followed by resistance genes associated with resistance to macrolides, with 44 AMR gene identifications. BioFire FilmArray analysis did not report the presence of an AMR gene in any sample (Table 19).

Table 19. Results of AST and identification of AMR genes by BioFire FilmArray pneumonia plus panel and CMg.

Sample	Routine microbiology AST report	BioFire FilmArray Pneumonia panel plus AMR gene report	AMR gene identified by CMg	Associated Resistance to antibiotic drug class
1	AST result not reported	None	<i>tetA46</i> <i>pmrA</i> <i>mefA</i> <i>msrD</i> <i>tetM</i> <i>ermB</i>	Tetracycline Fluoroquinolone Macrolide Macrolide Tetracycline Streptogramin, macrolide, lincosamide
2	AST result not reported	None	None identified	-
3	AST result not reported	None	<i>tetA46</i> <i>mefA</i> <i>msrD</i> <i>tetW</i> <i>bla_{CtxA-3}</i> <i>tetB(46)</i>	Tetracycline Macrolide Macrolide Tetracycline Cephameycin Tetracycline
4	<i>H. influenzae</i> β-lactamase positive	None	<i>IsaC</i> <i>tet37</i> <i>tetQ</i> <i>tetA46</i> <i>pmrA</i> <i>bla_{TEM}</i> <i>mefA</i> <i>msrD</i> <i>tetM</i> <i>tetO</i> <i>ermB</i> <i>tetW</i> <i>bla_{CtxA-3}</i> <i>ermX</i> <i>tetB46</i>	Lincosamide, streptogramin, pleuromutilin Tetracycline Tetracycline Tetracycline Fluoroquinolone Penam, penem, cephalosporin, monobactam Macrolide Macrolide Tetracycline Tetracycline Streptogramin, macrolide, lincosamide Tetracycline Cephameycin Streptogramin, macrolide, lincosamide Tetracycline
5	Routine microbiology result not reported	None	<i>pmrA</i> <i>mefA</i> <i>msrD</i>	Fluoroquinolone Macrolide Macrolide
6	AST result not reported	None	<i>InuC</i> <i>IsaC</i> <i>tetQ</i> <i>tetA46</i> <i>pmrA</i> <i>mefA</i> <i>msrD</i> <i>tetM</i> <i>tetO</i> <i>bla_{CtxA}</i> <i>tetB46</i>	Lincosamide. Lincosamide, streptogramin, pleuromutilin Tetracycline Tetracycline Fluoroquinolone Macrolide Macrolide Tetracycline Tetracycline Oxacephem, cephamycin Tetracycline
7	<i>H. influenzae</i> β-lactamase negative	None	<i>IsaC</i> <i>mefA</i> <i>tetM</i> <i>tetQ</i> <i>tetA46</i> <i>tetA60</i> <i>pmrA</i> <i>hmrM</i> <i>msrD</i> <i>ermB</i> <i>aph(3')-IIIa</i> <i>ermF</i> <i>bla_{CtxA}</i> <i>tetO</i> <i>tetB46</i> <i>tetB60</i>	Lincosamide, streptogramin, pleuromutilin Macrolide Tetracycline Tetracycline Tetracycline Tetracycline Fluoroquinolone Fluoroquinolone Macrolide Streptogramin, macrolide, lincosamide Aminoglycoside Streptogramin, macrolide, lincosamide Cephameycin Tetracycline Tetracycline Tetracycline
8	AST result not reported	None	<i>bla_{R-1}</i> <i>tetM</i> <i>mepA</i> <i>hmrM</i> <i>blaz</i> <i>tet38</i> <i>lmrS</i>	Penam Tetracycline Tetracycline Fluoroquinolone Penam Tetracycline Oxazolidinones, trimethoprim, phenicol, macrolide, fusidane

9	AST result not reported	None	<i>IsaC</i> <i>tetA46</i> <i>tetA60</i> <i>pmrA</i> <i>mefA</i> <i>msrD</i> <i>aph(6)-Id</i> <i>aph(3'')-Ib</i> <i>tetB</i> <i>aph(3')-IIIa</i> <i>bla_{BRO}</i> <i>bla_{CfxA-3}</i> <i>ermX</i> <i>catA2</i> <i>bla_{TEM}</i> <i>tetB46</i> <i>tetM</i>	Lincosamide, streptogramin, pleuromutilin Tetracycline Tetracycline Fluoroquinolone Macrolide Macrolide Aminoglycoside Aminoglycoside Tetracycline Aminoglycoside Penam Cepharmycin. Streptogramin, macrolide, lincosamide Phenicol Penam, penem, cephalosporin, monobactam Tetracycline Tetracycline
10	AST result not reported	None	<i>mefA</i> <i>tetM</i> <i>tetA46</i> <i>tetA60</i> <i>pmrA</i> <i>hmrM</i> <i>msrD</i> <i>ermB</i> <i>tetW</i> <i>bla_{CfxA-3}</i> <i>bla_{OXA-85}</i> <i>tetB46</i> <i>tetB60</i>	Macrolide Tetracycline Tetracycline Tetracycline Fluoroquinolone Fluoroquinolone Macrolide Streptogramin, macrolide, lincosamide Tetracycline Cepharmycin Penam Tetracycline Tetracycline
11	AST result not reported	None	<i>tetM</i> <i>ermX</i>	Tetracycline Streptogramin, macrolide, lincosamide
12	<i>S. marcesans</i> AmpC positive	None	<i>sdeY</i>	Multidrug
13	<i>S. marcescens</i> β-lactamase positive	None	<i>IsaC</i> <i>mefA</i> <i>tetW</i> <i>catB7</i> <i>bla_{CfxA-3}</i> <i>ermX</i> <i>crpP</i>	Lincosamide, streptogramin, pleuromutilin Macrolide Tetracycline Phenicol. Cepharmycin Streptogramin, macrolide, lincosamide Fluoroquinolone
14	AST result not reported	None	<i>norM</i> <i>bla_{BRO}</i>	Multidrug Penam
15	AST result not reported	None	<i>IsaC</i> <i>tetM</i> <i>tetA46</i> <i>tetA60</i> <i>mefA</i> <i>msrD</i> <i>tetB46</i>	Lincosamide, streptogramin, pleuromutilin Tetracycline Tetracycline Tetracycline Macrolide macrolide Tetracycline
16	AST result not reported	None	<i>IsaC</i> <i>tetA46</i> <i>tetA60</i> <i>bla_{TEM-1}</i> <i>mefA</i> <i>msrD</i> <i>tetM</i> <i>ermB</i> <i>bla_{CfxA}</i> <i>tetB46</i> <i>tetB60</i>	Lincosamide, streptogramin, pleuromutilin Tetracycline Tetracycline Penam, monobactam, penem, cephalosporin Macrolide Macrolide Tetracycline Streptogramins, macrolide, lincosamide Oxacephem, cephamycin Tetracycline Tetracycline
17	<i>S. marcescens</i> AmpC positive	None	<i>pmrA</i> <i>sdeY</i> <i>tetM</i> <i>ermX</i>	Fluoroquinolone Multidrug Tetracycline Streptogramin, macrolide, lincosamide
18	<i>H. influenzae</i> β-lactamase negative	None	<i>IsaC</i> <i>tetM</i> <i>tetW</i> <i>tetA46</i> <i>tetA60</i> <i>pmrA</i> <i>hmrM</i> <i>mefA</i> <i>msrD</i> <i>tetO</i> <i>ermB</i> <i>tet(B)</i> <i>tetB(46)</i> <i>tetB(60)</i>	Lincosamide, streptogramin, pleuromutilin Tetracycline Tetracycline Tetracycline Tetracycline Fluoroquinolone Fluoroquinolone Macrolide Macrolide Tetracycline Streptogramins, macrolide, lincosamide Tetracycline Tetracycline Tetracycline

19	AST result not reported	None	<i>mepA</i> <i>mefA</i> <i>msrD</i> <i>tetM</i> <i>bla_z</i> <i>fosB-Saur</i> <i>ermB</i> <i>bla_{R-1}</i> <i>tet38</i> <i>lnuG</i> <i>catB7</i> <i>ermX</i> <i>lmrS</i> <i>crpP</i> <i>tetB46</i>	Tetracycline Macrolide Macrolide Tetracycline. Penam Phosphonic Streptogramins, macrolide, lincosamide Penam Tetracycline Lincosamide Phenicol Streptogramin, macrolide, lincosamide Oxazolidinone, trimethoprim, phenicol, macrolide, fusidane Fluoroquinolone Tetracycline
20	AST result not reported	None	<i>catA16</i> <i>mefA</i> <i>pmrA</i> <i>emrB-sm</i> <i>mexA</i> <i>catB7</i> <i>ermX</i> <i>crpP</i>	Phenicol Macrolide Fluoroquinolone Fluoroquinolone. Peptide, sulfonamide, diaminopyrimidine, carbapenem, tetracycline, macrolide, penam, aminocoumarin, fluoroquinolone, cephalosporin, phenicol, monobactam, cephamycin, penem Phenicol. Streptogramins, macrolide, lincosamide Fluoroquinolone

Results of routine microbiology AST from sputum samples collected from NHS patients with a suspected LRTI, the AMR genes identified by CMg and BioFire FilmArray pneumonia panel plus analysis are reported. The antibiotic class to which each AMR gene is associated was determined according to the CARD. AMR genes associated with resistance to the antibiotic drug classes identified by AST in the same sample are highlighted in bold. Six samples reported AST results, with CMg identifying a matching AMR gene in sample 4. BioFire FilmArray analysis did not report the identification of any AMR genes.

5.5 Discussion

This chapter describes two proof of concept exercises, carried out to better understand the technical and logistical challenges associated with operating a CMg workflow within a deployed MTF. Ultimately, this work will be used to inform the future development and optimisation of CMg workflows and the interpretation of the output of these workflows. The first exercise consisted of the processing of excess nasopharyngeal swabs, collected as part of routine COVID-19 testing, on board the RFA Argus ship by a military BMS. The BMS was trained in each step of the workflow prior to the start of the exercise, the equipment and reagents were delivered to the RFA Argus and the BMS set-up and operated the equipment within the laboratory. The 11 excess respiratory samples were processed on board the ship by the BMS, and the sequence data analysed for the detection of viral pathogens. The second exercise aimed to build upon the experience gained during the first, processing sputum samples collected from patients with a suspected LRTI and comparing the results to routine microbiology testing.

5.5.1 Exercise one: viral clinical metagenomics workflow

This exercise represented the first example of CMg being carried out on board a RFA ship for the diagnosis of infectious disease. However, there were several limitations and challenges associated with operating within the deployed environment, leading to the failure to identify the viral respiratory pathogens previously identified by routine testing.

Firstly, delays to the delivery of certain reagents prior to the start of the exercise meant that it was not possible to establish and optimise the workflow within a standard, fixed laboratory environment prior to the delivery of the equipment and reagents to the RFA Argus. Therefore, it was feasible that there were technical issues that occurred during the exercise that had not been previously encountered, leading to the failure to identify the viral pathogens. However, without the previous experience of carrying out the workflow it was not possible to conclusively identify the cause of the issues experienced during the exercise and these can only be speculated on.

The delay to the delivery of the reagents also meant that the filter centrifuge columns, intended for the removal of human and bacterial cellular material pre-RNA extraction, were not included in the protocol. This was likely to have increased the proportion of human genomic DNA in the sample, which represented the majority of reads relative to pathogen reads, preventing the detection of viral pathogens. A comparison of different viral particle enrichment strategies for metagenomics studies have found that the inclusion of a multi-step approach of centrifugation, syringe filtration and DNase treatment was the most effective method for reducing host genetic material whilst retaining viral genetic material (294). Future workflows, if targeting viral pathogens, must ensure that the pre-extraction viral enrichment steps are included to increase the sensitivity of virus detection.

The logistical requirements of delivering temperature sensitive reagents to the ship also presented a significant challenge. Due to last-minute changes to the departure location, the reagents and equipment were transported to several locations *en route* to the ship. It is likely that this led to the freeze-thaw of the frozen sequencing reagents and the incorrect storage of the nanopore flow cells, causing degradation to the reagents and damaging the protein nanopores of the flow cells. As a result, a flow cell that had been previously used for training purposes with fewer than the recommended 800 pores available for sequencing was used for the sequencing run. In a deployed environment, it may be necessary to re-locate the MTF for operational reasons or to carry out diagnostics within a mobile facility without fixed infrastructure. The use of reagents that are more robust to fluctuations in temperature would be advantageous in these situations. Oxford Nanopore Technologies have previously released a field sequencing library preparation kit with lyophilised reagents suitable for room temperature storage, although this kit is no longer available to purchase (327). Use of these reagents may have prevented the reagent degradation observed in this exercise. The use of temperature monitoring equipment, stored with the reagents to record and alarm during anomalous temperature events would also have prevented the incorrect storage of reagents or at least highlighted that the reagents had been stored incorrectly and allowed corrective action to be taken and should be included in future.

Once the reagents and equipment had been delivered, the BMS set up and carried out the workflow, providing feedback on any difficulties experienced during sample processing. It was reported that the methods were too complex for use on board a ship and that further optimisation was required to reduce the method complexity and the number of hands-on steps. It was therefore recommended that the number of manual pipetting steps were reduced, due to the potential for human error when operating on a ship environment, particularly during inclement weather in which the laboratory environment may be unstable. The DNA extraction methods used were manual, as a suitable automated platform was not available within the timeframe of the exercise. In addition, the library preparation methods, including quality control steps, were carried out manually. Automated DNA extraction and liquid handling solutions are commercially available and bespoke programmes for each step could be designed to reduce the manual handling steps. Ultimately, the aim should be to deploy a fully automated sample to result CMg platform, operated in a similar manner to platforms such as the BioFire FilmArray 2.0, that require no manual input once the sample has been loaded and produces a simple report as the output. Direct involvement of industry partners is likely to be required for this to develop a bespoke, ruggedized CMg platform suitable for use in austere deployed MTFs.

Despite these challenges, the first deployment of a CMg workflow to a role three MTF environment provided useful experience and greater understanding of the requirements of the military end user, space and equipment availability, user time availability and the opportunity for the BMS to provide feedback on the operation of the methods. This information was then incorporated into the method development for the second exercise.

5.5.2 Exercise two: bacterial clinical metagenomics workflow

The second exercise aimed to evaluate the processing of sputum samples for CMg in a deployed environment by a military BMS, applying the lessons learned from the first exercise. The resulting CMg data was evaluated for the identification of respiratory pathogens compared to routine microbiology testing. Following the logistical issues experienced when delivering the reagents during the first exercise, additional measures were put in place to ensure the correct storage of the reagents. These included temperature monitoring of the temperature sensitive reagents to ensure that these were maintained at the correct temperature range during transfer. An additional amendment in response to the results of the first exercise was that, instead of aiming to detect viral pathogens from nasopharyngeal swab samples, the workflow was developed to identify bacterial and fungal pathogens from sputum samples. This enabled the use of a saponin based host depletion method, previously demonstrated for respiratory sample CMg (62), to be incorporated into the workflow to reduce the proportion of human DNA in the samples. As a result, sequencing data was successfully generated for all samples, the

human proportion of the sequencing data was reduced, and respiratory pathogens were identified in good concordance with routine microbiology.

Similarly to the first exercise, it was not possible to evaluate the full CMg workflow in patient samples within a fixed laboratory environment, including the heat inactivation step, prior to the start of the exercise. The heat inactivation step of 80°C for 30 minutes, carried out following the host-depletion, was required due to the lack of a suitable containment level three laboratory on-board the RFA Argus. Heat inactivation steps alter the composition of the microbiomes identified by metagenomics sequencing, depending on the sample type, potentially preventing the detection of certain pathogen species (328). However, the impact of the heat inactivation step on the detection of bacterial pathogens from sputum samples was not fully understood prior to the exercise. The inclusion of a positive control mixture of a range of bacterial pathogens at known concentrations, spiked into known negative sputum samples and processed alongside the patient samples, could have enabled a better understanding of the impact of the heat inactivation step on the detection of certain species and is recommended for future exercises.

Despite representing a more successful attempt at carrying out CMg methods within a deployed environment compared to the first exercise, there were still several logistical and technical challenges. Firstly, the sensitivity of pathogen detection compared to routine microbiology, at 64%, was lower compared to previous examples of respiratory CMg workflows carried out in fixed hospital environments (230). For several samples, it was reported by the BMS that the viscosity of the sputum sample, despite the addition of sputasol solution, that the sample could not be adequately mixed during the host depletion step. As a result, it is likely that the efficiency of the host depletion was reduced, resulting in a high proportion of human reads. This was observed in sample two, in which scanty growth of *S. aureus* was reported by microbiology testing but was not identified by CMg. This is likely due to the low number of classified microbial reads at 922 reads, compared to 590361 human reads, indicating that the human depletion had not worked as expected. Additional methods for the liquefaction of the sputum samples should be investigated to ensure that sufficient human DNA depletion is achieved for all samples.

An additional challenge for the detection of lower respiratory pathogens was the significant proportion of classified microbial reads identified to be URTCs. This was also reflected by the reporting of the presence of URTCs by routine microbiology in 6/14 positive samples. The method for the collection of sputum, involving the patient coughing the sputum sample into a sterile container, carries considerable risk of upper respiratory tract contamination (329). The presence of URTCs in the sequencing data is likely to inhibit the detection of respiratory pathogens by reducing the proportion of reads assigned to the lower respiratory pathogen species. Attention should be paid during the sample collection process to ensure the adequate quality of sputum sample based on cytologic

criteria. Samples can be screened via the detection of squamous epithelial cells, indicating upper respiratory tract contamination, and inflammatory cells, such as polymorphonuclear leukocytes, that indicate the sample was collected from a site of active infection (330). Samples shown to be significantly contaminated should be rejected for CMg analysis and, if possible, an additional sample collected to ensure adequate sensitivity of pathogen detection.

The specificity of CMg compared to routine microbiology testing was reduced by the erroneous identification of *S. pneumoniae* in several microbiology and BioFire FilmArray negative samples. As discussed, the sequencing data identified significant proportions of URTCs, including *Streptococcal* species such as *Streptococcus mitis*. *S. mitis* was identified in nine samples by CMg, including every sample in which *S. pneumoniae* was incorrectly identified by CMg. *S. mitis* shares many genetic features with *S. pneumoniae* and is a reservoir for the exchange of genetic information with *S. pneumoniae* within the respiratory tract environment (331). Therefore, the genetic similarities between the two species may have led to the misidentification of *S. pneumoniae* by the bioinformatics analysis pipeline through the incorrect assigning of *S. mitis* reads. Future deployed analysis pipelines should aim to differentiate between *S. pneumoniae*, a recognised respiratory pathogen, and commensal *Streptococcal* species to improve the specificity of the workflow. Similar relationships involving the transfer of genetic material between known pathogens and commensal species should be investigated, including those in which species are mis-labelled within the database itself, and accounted for in the analysis process to prevent false positive identifications.

An additional challenge related to the identification of AMR genes. A significant number of AMR genes were identified per sample by CMg, most related to resistance to tetracycline. Although AST results from only six samples were reported, most AMR genes identified by CMg were not related to the resistance profile reported by AST. Whilst the untargeted nature of CMg is one of the key advantages over targeted tests, the drawback of this feature is that information that may not be relevant to the treatment of the infection is also reported. As with the reporting of respiratory pathogens, the analysis pipelines for deployed CMg workflows must produce an easily interpretable report that enables clinicians with limited experience of CMg to inform the treatment of the patient. The identification of AMR genes unrelated to the aetiological pathogen should not be reported. Previous respiratory sample CMg workflows reported the phenotypes from acquired-resistance elements, ensuring that only the relevant information was provided to the clinician (72). Applying this method, the relevant AMR genes reported by a deployed CMg workflow may be pre-determined prior to the deployment and be subject to updates based on the location of the deployment and intelligence gained from previous deployments to

better target the reporting of AMR genes and ensure that relevant information is not omitted.

Another issue associated with the data analysis was that the analysis was carried out off-board once the ship had returned to port in the UK. Access to the internet on board the ship was limited for operational reasons and there was no availability of a standalone device with the required bioinformatics pipelines and databases installed. To enable the full processing of samples for CMg whilst on deployment, a standalone data analysis platform is required.

Despite these challenges, this exercise also demonstrated a key advantage of CMg, through the identification of respiratory pathogens not identified by routine microbiology. This included the identification of *M. catarrhalis* in sample nine, which was subsequently confirmed by BioFire FilmArray testing but had not been identified by microbiology. The ability to identify pathogens that are not usually tested for or may be missed by standard diagnostic tests is a key advantage of CMg and has been demonstrated throughout this project. However, the ability to understand the context and significance of the identification of additional pathogens and whether their identification requires treatment must be improved. When operating in deployed environments, the BMSs and clinical staff are unlikely to have access to a specialist molecular biologist with an in-depth understanding of the relationship between the proportion of sequencing reads assigned to a pathogen species and the cause of the infection. For example, some additional respiratory pathogens identified in this exercise, including *S. maltophilia* identified in sample 20, constituted a low proportion of microbial reads. When identified at low proportions, it may not be clear to a clinician whether this represents an aetiological pathogen requiring treatment or is present as a commensal species. As CMg becomes a more commonly used approach, the understanding of the interpretation and reporting of CMg results amongst the clinical community will improve and this should be incorporated into the output of deployed CMg workflows.

5.5.3 The future of deployed diagnostics

These exercises have demonstrated that it is feasible for CMg methods to be performed in deployed military MTFs by BMSs as an additional diagnostic test. This work should now be applied to inform the future development of the deployed diagnostic pathway to ensure that military personnel receive the optimal treatments for the range of infectious disease that they are exposed to, reducing morbidity and mortality. In the short-term, it is expected that CMg could provide an additional diagnostic test complementing the current suite of diagnostic tests available to deployed clinicians, applied in instances where standard testing is not sufficient. Continued development of CMg methods, including improved automation and better interpretation of the results would then enable CMg to be utilised

more routinely, even replacing existing diagnostic tests in some cases. However, while this project has laid the groundwork for this, the limitations and challenges associated with deploying CMg demonstrate that there is much to do before CMg could be used routinely.

During the completion of this project, external research groups have published CMg workflows for the diagnosis of infectious disease, further demonstrating the potential of CMg. In a key development, this includes workflows that have combined the detection of viral and bacterial pathogens (230), representing another step towards CMg becoming a truly untargeted method enabling the detection of all potential pathogens from one test. Collaborations with these external research groups to evaluate RNA and DNA based CMg workflows in a range of different deployed environments, including land-based MTFs would allow CMg methods to continue to be developed for deployed diagnostics. Leading on from this project, the Royal Navy have now initiated a project for the procurement of the equipment required for CMg workflows so that additional evaluation of CMg in deployed environments can take place.

Further to the ongoing research into the detection of the aetiological agent of an infection, there is also considerable research focused on applying measurements of the host response to infection. Whilst most studies focus on the measurement of protein host biomarkers (332), host metatranscriptomics methods are also in development, as reviewed by Haslam, 2021 (333). A potential advantage of introducing CMg methods to deployed environments is that related tests measuring additional relevant data points could then be applied using the same equipment, or even simultaneously within the same workflow, reducing the logistical challenge and providing another method for clinicians to evaluate their patients. Rapid host biomarker systems are already in development to determine whether an infection is viral or bacterial, to reduce the use of antibiotics in cases of viral infection, showing good potential (334) and CMg could expand upon this progress.

Further additions are required for the CMg workflows described in this study to enable their routine use for the diagnosis of infectious disease where erroneous results can have disastrous consequences for the patient, with method standardisation being crucial to this. This includes the use of spike-in controls to identify method failure events and quantify exogenous contamination (151, 335), workflow automation to reduce the potential for human error and the production of a simple user interface and report format. The generation of the reports should include the automated interpretation of thresholds for the positive identification of pathogens according to the pathogen species and the sample type, as informed by the clinical and research community as the understanding of the detection of specific pathogens increases. The final step towards deploying CMg as a routine diagnostic test will be the completion of clinical validation tests consisting of larger scale studies to demonstrate the benefits of using CMg in terms of reduced healthcare

costs, reduced logistical requirements, and improved patient outcomes compared to current diagnostic methods. This is likely to pose a considerable challenge due to the untargeted nature of CMg and the ability to, in theory, identify any pathogen present in a sample. However, validated commercial CMg services have been established and are in operation (336) showing that it is possible, and the potential benefits to patients from the availability of rapid, untargeted diagnostics in a deployed environment mean that we should continue to work towards this goal.

5.6 Conclusions

The work described in this chapter consists of two proof of concept exercises in which CMg workflows were deployed to the RFA Argus MTF and samples were processed by a military BMS. These represent the first examples of CMg methods carried out on board an RFA ship for the diagnosis of infectious disease. Several key challenges were highlighted that should be addressed by future method development, including better monitoring of temperature sensitive reagents, reducing method complexity through increased automation, enabling on-board data analysis and producing simple result outputs. As a result of this work, the Royal Navy will expand the evaluation of CMg into new deployed environments, and if successful, CMg will become a key part of the deployed diagnostic pathway.

6 Concluding remarks

The overall aim of this project was to develop, optimise and evaluate CMg methods for a range of sample types relevant to deployed military personnel, building towards the ultimate goal of CMg being routinely used for the diagnosis of infectious disease in deployed MTFs. The work carried out in this project included the development of a novel CMg workflow for the identification of pathogens from wound swab samples, initially using contrived swab samples before applying the methods to wound swab samples collected from patients with a suspected wound infection. CMg was also applied to stool samples, stored using three storage methods, collected from UK military personnel during an outbreak of TD. This included the development of the CMg workflow and comparison of the three storage methods to help inform the future deployment of CMg to austere environments. Finally, CMg workflows were deployed to a role three MTF on board the RFA Argus and carried out by a military BMS, in the first example of CMg methods being used for the diagnosis of infectious disease in a UK deployed MTF.

One of the key outcomes of this work was the demonstration of the potential for CMg as a deployed diagnostic test to the wider military clinical community. Prior to this project, the knowledge of CMg and its applications for the diagnosis of infectious disease amongst the military clinical community was poor, understandably so given that CMg was a novel method that had not been used routinely within fixed healthcare facilities, let alone within deployed MTFs. Through collaborations with military clinicians during this project, the potential advantages of CMg over the current deployed diagnostic tests, including the identification of pathogens that were missed by conventional methods and the identification of additional clinically relevant information such as the presence of AMR genes without the need for additional testing, have been demonstrated.

The untargeted nature of CMg, a key advantage for infectious disease diagnostics, has also been demonstrated. The emergence of novel pathogens, through natural genetic mutation or intentional genetic engineering, presents a risk to deployed military personnel. If these changes to the pathogen genetic code affect the regions of the genome used for targeted detection methods, this can result in false-negative results. Untargeted CMg represents a solution to this limitation, enabling the detection of the wider genome, both for the identification of the pathogen and for the detection of clinically relevant genomic regions, including AMR genes. This advantage has been demonstrated to the military clinical community during this project through the identification of additional pathogen species and AMR genes not identified by routine testing.

This is crucial, as without the investment of the senior decision makers within the military clinical community, development of CMg towards a deployed diagnostic test would likely stall. As a direct result of the work detailed in Chapter Five, the Royal Navy have initiated

a research project to continue the evaluation of CMg in deployed environments, involving a wider cohort of BMSs and external academic collaborators. Therefore, whilst there are several technical hurdles that must be overcome before CMg could be used routinely, this project represents a major step towards the incorporation of CMg into the deployed diagnostic pathway.

Equally as important as demonstrating the potential of CMg for deployed diagnostics, this project has also highlighted specific situations in which CMg may not be the correct solution and areas that require improvement. In Chapter Four, the processing of stool samples stored using FTA cards by CMg resulted in poor sensitivity of pathogen detection, including *C. hominis*. As a result, the military clinical community have been advised against the use of FTA cards for the storage of stool samples. Similarly, the processing of nasopharyngeal swabs for the detection of viral pathogens highlighted several technical challenges, including poor sensitivity of pathogen detection and the complexity of the CMg workflow. As a result, it is likely that the development of CMg workflows, in the short-term, will focus on the detection of bacterial and fungal pathogens. These outcomes demonstrate the impact of this project on shaping the direction of the future development of CMg as a deployed diagnostic test.

Another key limitation identified at several points during this project was the bioinformatics analysis and interpretation of the output of CMg workflows. The identification of reportable pathogens using statistically robust reporting thresholds and determining the resistance of the pathogens to antimicrobial treatments from the plethora of information provided by CMg outputs represents the next major hurdle for the continued development of CMg and future work should aim to address these issues. Achieving this is likely to require further collaboration with external parties, including academic, healthcare and military partners, to generate larger CMg data sets alongside detailed sample metadata that can be used to inform the interpretation of CMg outputs.

This project represents significant progress towards the use of CMg in deployed MTFs for the diagnosis of infectious disease and included the first example of the operation of a CMg workflow within a UK deployed MTF. Whilst further method development and optimisation is required, CMg represents an opportunity to provide rapid and untargeted diagnostics, complementing and adding to the current suite of diagnostic tests and improving the treatment of infectious disease amongst deployed military personnel.

7 Annex

Table A 1. Detection of high priority wound pathogen species by microbiology testing, qPCR and CMg.

Sample Number	<i>E. coli</i>				<i>P. aeruginosa</i>				<i>S. aureus</i>				<i>S. pyogenes</i>			
	Species reported by microbiology	Cq	CMg Reads (*)	Average read length (bp)	Species reported by microbiology	Cq	CMg Reads (*)	Average read length (bp)	Species reported by microbiology	Cq	CMg Reads (*)	Average read length (bp)	Species reported by microbiology	Cq	CMg Reads (*)	Average read length (bp)
01	No	ND	935 (0.59)	900.04	No	ND	6 (0.01)	1823.17	ND				ND			
02		ND				ND			No	ND	26 (0.32)	1189.15		ND		
03		ND				ND			Yes	34.29	1555 (64.85)	793.02		ND		
04		ND				ND			No	ND	32 (0.37)	636.36		ND		
05	No	25.13	251 (0.10)	1536.50	No	34.45	0 (0.00)	-	No	34.15	854 (0.34)	1072.55		ND		
06		ND				ND				ND				ND		
07		ND				ND			No	ND	6 (0.04)	793.66	No	ND	39 (0.25)	1342.90
08	Yes	19.29	20441 (44.77)	1391.72		ND				ND			No	ND	8 (0.02)	1163.14
09		ND				ND				ND				ND		
10		ND				ND			Yes	24.91	1126 (41.17)	998.23	No	26.47	1542 (56.31)	1074.60
11		ND				ND			Yes	32.47	0 (0.00)	-		ND		
12	No	ND	120 (0.38)	594.44		ND				ND				ND		
13		ND				ND			No	36.18	0 (0.00)	-	No	ND	138 (0.60)	682.46
14		ND				ND			Yes	18.01	23082 (99.84)	878.67		ND		
15		ND				ND			Yes	19.13	15365 (99.81)	911.72		ND		
16		ND				ND			Yes	25.78	1933 (33.40)	745.26	No	ND	25 (0.43)	800.71
17	No	ND	20 (0.57)	1873.63		ND				ND				ND		
18	Yes	21.02	35004 (66.03)	1366.20		ND			No	ND	29 (0.06)	1336.12		ND		
19		ND				ND				ND				ND		

20	No	ND	23 (0.08)	976.61	No	ND	27 (0.09)	1342.40	Yes	27.51	186 (0.62)	1074.10	ND			
21	No	ND	36 (0.05)	735.15	ND				Yes	35.02	94 (0.14)	950.76	No	ND	9 (0.01)	1533.50
22	No	ND	72 (0.12)	1515.62	No	ND	26 (0.04)	1408.39	ND			No	ND	29 (0.05)	1294.08	
23	ND		ND				Yes	28.85	212 (0.56)	907.75	No	ND	10 (0.03)	831.67		
24	ND		ND				No	ND	42 (0.50)	1034.74	No	ND	6 (0.07)	775.50		
25	ND		ND				No	ND	32 (0.17)	1019.96	No	24.14	125 (0.66)	1052.53		
26	ND		ND				Yes	31.53	544 (6.45)	964.70	Yes	29.13	6994 (82.97)	1019.32		
27	No	36.06	0 (0.00)	-	ND				Yes	27.67	868 (7.93)	921.96	No	ND	28 (0.26)	966.19
28	ND		Yes		26.93	1294 (3.06)	1508.12	No	ND	10 (0.02)	996.80	No	ND	56 (0.13)	1151.87	
29	ND		ND				Yes	28.18	62 (65.26)	1014.03	ND					
30	ND		No		17.70	101154 (82.80)	1742.53	ND			ND					
31	ND		ND				ND			ND						
32	ND		ND				No	ND	5 (0.03)	1092.00	No	ND	54 (0.32)	785.56		
33	No	26.93	39 (0.04)	1544.59	No	ND	536 (0.56)	2493.68	No	34.76	11 (0.01)	1223.44	No	ND	54 (0.06)	1227.55
34	ND		Yes		26.48	1224 (9.37)	1061.42	ND			ND					
35	No	37.81	0 (0.00)	-	No	30.25	20 (0.11)	1234.06	Yes	34.82	252 (1.39)	575.16	No	ND	48 (0.27)	549.17
36	No	26.48	43 (0.18)	1030.63	No	22.44	314 (1.34)	1245.78	Yes	22.46	1377 (5.87)	728.68	No	ND	631 (2.69)	442.17
37	ND		No		34.57	0 (0.00)	-	Yes	29.39	23 (0.08)	691.43	No	ND	722 (2.63)	454.00	
38	No	ND	576 (1.78)	1372.05	No	22.35	463 (1.43)	1199.77	No	ND	9 (0.03)	491.22	No	ND	14 (0.04)	536.91
39	ND		ND				No	37.53	75 (0.95)	598.06	No	ND	103 (1.30)	674.48		
40	No	33.22	0 (0.00)	-	No	23.84	518 (2.27)	1150.95	No	34.67	85 (0.37)	1013.27	No	ND	145 (0.64)	652.39

The DNA extract from each additional wound swab sample was tested for four high priority pathogen species by qPCR. The routine microbiology testing report, qPCR Cq value and the number of reads assigned to the target species by CMg are reported. % classified microbial reads = % of species reads over the total number of taxonomically classified microbial reads obtained.

* % classified microbial reads. ND= No detection

Table A 2. Detection of AMR genes from DNA Shield RBK samples.

Sample Number	AMR genes identified by CMg sequencing	Template Coverage	Depth (×)	Associated antibiotic drug class
C1	<i>dfrF</i>	100	2	Diaminopyrimidine
	<i>mef(En2)</i>	101.16	3.42	Macrolide, lincosamide
	<i>ctr(C)</i>	99.04	3.96	Oxazolidinone, lincosamide, phenicol, streptogramin
	<i>aph(3')-IIIa</i>	100.25	4.72	Aminoglycoside
	<i>catA13</i>	100	5.72	Phenicol
	<i>tet(Q)</i>	100	13.7	Tetracycline
B1	<i>tet(A)</i>	100.75	1.19	Tetracycline
B2	<i>erm(F)</i>	100	1	Streptogramin, macrolide, lincosamide
	<i>lnu(AN2)</i>	100	1	Lincosamide
	<i>mexA</i>	100.35	1	Peptide, sulfonamide, diaminopyrimidine, carbapenem, tetracycline, macrolide, penam, aminocoumarin, fluoroquinolone, cephalosporin, phenicol, monobactam, cephamycin, penem
	<i>tet(O)</i>	100.05	1.07	Tetracycline
	<i>bla_{PDC}</i>	100.08	1.95	Cephalosporin, monobactam, carbapenem
	<i>crpP</i>	98.99	1.97	Fluoroquinolone
	<i>mdtM</i>	99.68	1.99	Nucleoside, fluoroquinolone, phenicol, lincosamide
	<i>aph(6)-Id</i>	100	3	Aminoglycoside
	<i>bla_{OXA}</i>	100	3	Penam
	<i>tet(Q)</i>	100	3	Tetracycline
	<i>tet(W)</i>	100	3.76	Tetracycline
	<i>dfrA14</i>	100	4	Diaminopyrimidine
	<i>sul2</i>	100	5.24	Sulfonamide
B3	<i>vanR-D</i>	100	1	Glycopeptide
	<i>erm(F)</i>	100.12	1	Streptogramin, macrolide, lincosamide
	<i>vanS-D</i>	100.26	1	Glycopeptide
	<i>tet(O)</i>	93.07	1.29	Tetracycline
	<i>catA13</i>	100.16	2	Phenicol
	<i>tet(Q)</i>	100.1	3.55	Tetracycline
B10	<i>tet(Q)</i>	100	5.41	Tetracycline
	<i>tet(W)</i>	100	6.69	Tetracycline
B40	<i>acrF</i>	100.03	1	Fluoroquinolone, cephalosporin, cephamycin, penam
	<i>tet(Q)</i>	100.1	2	Tetracycline
B44	<i>mef(En2)</i>	100.66	1.01	Macrolide, lincosamide
	<i>tet(O/32/O)</i>	100.05	1.85	Tetracycline
	<i>tet(32)</i>	100.47	1.85	Tetracycline
	<i>catP</i>	100	3	Phenicol
	<i>tet(Q)</i>	100	3.57	Tetracycline
B56	<i>tet(Q)</i>	89.2	1.7	Tetracycline
	<i>bla_{CtxA-6}</i>	100	12.29	Cephamycin
	<i>aph(6)-Id</i>	100	4.88	Aminoglycoside
	<i>sul2</i>	100.25	2	Sulfonamide
	<i>bla_{CtxA}</i>	101.14	1.81	Cephamycin
B80	<i>tet(Q)</i>	96.88	1.18	Tetracycline
	<i>tet(Q)</i>	100.16	5.15	Tetracycline
	<i>bla_{CtxA-6}</i>	100	10.52	Cephamycin
	<i>tet(W)</i>	100	4.31	Tetracycline
	<i>tet(32)</i>	100.1	1	Tetracycline
	<i>erm(F)</i>	100	9.61	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	101.5	1.9	Aminoglycoside
	<i>nimE</i>	100	3.78	Nitroimidazole
	<i>bla_{CtxA-5}</i>	101.04	1.33	Cephamycin
B94	<i>tet(Q)</i>	100.67	3.11	Tetracycline
	<i>dfrF</i>	100.2	1	Diaminopyrimidine
	<i>erm(F)</i>	100.5	1	Streptogramin, macrolide, lincosamide
B4	<i>mef(En2)</i>	100	4.48	Macrolide, lincosamide
	<i>tet(Q)</i>	100	6.26	Tetracycline
	<i>mdtM</i>	100.41	2.65	Nucleoside, fluoroquinolone, phenicol, lincosamide
	<i>erm(B)</i>	100	2	Streptogramin, macrolide, lincosamide
	<i>catP</i>	100.32	1	Phenicol
	<i>lnu(AN2)</i>	100.78	2.01	Lincosamide
	<i>bla_{CtxA}</i>	100	2.9	Cephamycin
	<i>mef(A)</i>	100	8.08	Macrolide
	<i>tet(40)</i>	81.65	2	Tetracycline
	<i>tet(O)</i>	91.3	1.94	Tetracycline
B5	<i>bla_{CtxA-2}</i>	100	12.44	Cephamycin
	<i>tet(Q)</i>	100	21.1	Tetracycline
	<i>bla_{CtxA-6}</i>	100.31	2	Cephamycin
	<i>sul2</i>	101.84	1.46	Sulfonamide
	<i>bla_{CblA}</i>	100	3.64	Cephalosporin
B6	<i>mef(En2)</i>	100.08	1	Macrolide, lincosamide
	<i>bla_{CepA}</i>	100.11	2	Cephalosporin
	<i>dfrF</i>	100.61	1.01	Diaminopyrimidine
	<i>tet(W)</i>	100	10.02	Tetracycline
	<i>tet(32)</i>	100.1	1	Tetracycline

	<i>Inu(AN2)</i>	100	3	Lincosamide
	<i>blaCblA</i>	102.58	1.34	Cephalosporin
	<i>tet(O)</i>	101.35	1.48	Tetracycline
B7	<i>tet(Q)</i>	100.26	4.78	Tetracycline
	<i>tet(W)</i>	100	6.8	Tetracycline
	<i>blaCblA-6</i>	100	9.08	Cepharmycin
	<i>tet(40)</i>	100.82	1.5	Tetracycline
	<i>tet(O)</i>	96.46	1.91	Tetracycline
	<i>blaCblA-5</i>	100.52	1.9	Cepharmycin
B8	<i>mef(En2)</i>	100	4.26	Macrolide, lincosamide
	<i>tet(Q)</i>	100	15.63	Tetracycline
	<i>tet(W)</i>	100	15.69	Tetracycline
	<i>tet(X2)</i>	100	3.57	Tetracycline
	<i>blaCblA-6</i>	100	7.7	Cepharmycin
	<i>blaTEM</i>	100.12	1	Penam, penem, cephalosporin, monobactam
	<i>sul2</i>	100.98	1.38	Sulfunamide
	<i>aph(3')-IIIa</i>	99.87	2.73	Aminoglycoside
	<i>aadE</i>	100	3	Aminoglycoside
	<i>Inu(B)</i>	100	3	Lincosamide
	<i>erm(B)</i>	100	3.58	Streptogramin, macrolide, lincosamide
	<i>dfrF</i>	100	2	Diaminopyrimidine
	<i>tet(32)</i>	101.3	1.76	Tetracycline
	<i>erm(F)</i>	100	8.88	Streptogramin, macrolide, lincosamide
	<i>blaOXA</i>	100.36	1	Penam
	<i>Inu(AN2)</i>	100	7.83	Lincosamide
	<i>blaCblA</i>	100	8.83	Cepharmycin
	<i>blaCblA</i>	100.34	2.3	Cephalosporin
	<i>tet(O)</i>	100	13.28	Tetracycline
	<i>tet(40)</i>	100	3.63	Tetracycline
B9	<i>tet(Q)</i>	100	4.95	Tetracycline
	<i>sul2</i>	100	8.68	Sulfunamide
	<i>dfrF</i>	100	1	Diaminopyrimidine
	<i>erm(F)</i>	100	3.97	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	6.22	Aminoglycoside
	<i>blaOXA</i>	100	3.23	Penam
	<i>blaCblA</i>	100	10.93	Cephalosporin
	<i>tet(X2)</i>	100.09	3.18	Tetracycline
	<i>tet(O)</i>	100.1	5.57	Tetracycline
	<i>erm(G)</i>	101.36	2.21	Streptogramin, macrolide, lincosamide
	<i>ere(D)</i>	100.08	7.16	Macrolide
	<i>tet(W)</i>	100.05	4.25	Tetracycline
B11	<i>Inu(C)</i>	100.61	2.01	Lincosamide
	<i>tet(O)</i>	101.09	1.42	Tetracycline
	<i>tet(Q)</i>	100	10.94	Tetracycline
	<i>tet(W)</i>	99.9	5.87	Tetracycline
	<i>blaCblA-6</i>	100	7.66	Cepharmycin
	<i>tet(M)</i>	101.82	1.75	Tetracycline
	<i>tet(L)</i>	99.64	1.81	Tetracycline
	<i>tet(32)</i>	100.1	1	Tetracycline
	<i>blaCblA</i>	100	3	Cephalosporin
	<i>tet(W)</i>	99.48	1.16	Tetracycline
	<i>mef(A)</i>	100	6.11	Macrolide
	<i>blaCblA-5</i>	100	8.57	Cepharmycin
B12	<i>tet(M)</i>	102.66	1.87	Tetracycline
	<i>tet(Q)</i>	100	12.58	Tetracycline
	<i>cfr(C)</i>	101.32	2.01	Oxazolidinone, lincosamide, phenicol, streptogramin
	<i>erm(B)</i>	100	1	Streptogramin, macrolide, lincosamide, streptogramin
	<i>tet(W)</i>	100	18.03	Tetracycline
	<i>erm(F)</i>	100.12	3	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	5.16	Aminoglycoside
	<i>dfrA14</i>	100.21	2	Diaminopyrimidine
	<i>blaCblA</i>	100.93	2.89	Cepharmycin
	<i>tet(X2)</i>	100.26	3.66	Tetracycline
	<i>tet(O)</i>	100	13.04	Tetracycline
B13	<i>erm(G)</i>	100.82	1.99	Streptogramin, macrolide, lincosamide
	<i>mef(En2)</i>	103.23	2.3	Macrolide, lincosamide
	<i>tet(Q)</i>	100	12.78	Tetracycline
	<i>tet(W)</i>	100.42	3.04	Tetracycline
	<i>erm(F)</i>	100	14.33	Streptogramin, macrolide, lincosamide
	<i>tet(Q)</i>	100.05	1	Tetracycline
	<i>blaCblA</i>	100.67	1.45	Cephalosporin
	<i>tet(X2)</i>	100	9.36	Tetracycline
	<i>tet(O)</i>	101.41	2.69	Tetracycline
	<i>tet(40)</i>	100.74	1.42	Tetracycline
B14	<i>Inu(C)</i>	100.81	2.01	Lincosamide
	<i>mef(En2)</i>	100	5.32	Macrolide, lincosamide
	<i>tet(Q)</i>	100	10.37	Tetracycline
	<i>tet(Q)</i>	100	15.76	Tetracycline
	<i>aph(6)-Id</i>	101.19	2.01	Aminoglycoside
	<i>sul2</i>	100	2.78	Sulfunamide
	<i>aadA1</i>	100	3	Aminoglycoside
	<i>tet(W)</i>	100	27.88	Tetracycline
	<i>Inu(AN2)</i>	100.19	2.72	Lincosamide
	<i>blaCblA</i>	100	3	Cepharmycin
	<i>blaCblA</i>	100	5	Cephalosporin
	<i>tet(40)</i>	100	3.55	Tetracycline
	<i>mef(A)</i>	100	6.34	Macrolide

	<i>dfrA14</i> <i>tet(O)</i>	100 100.1	3 17.65	Diaminopyrimidine Tetracycline
B015	<i>mef(En2)</i> <i>tet(Q)</i> <i>Inu(AN2)</i> <i>bla_{CtxA-3}</i> <i>tet(O)</i>	102.16 99.95 100.19 100 100	1.53 4.24 3 8.34 5	Macrolide, lincosamide Tetracycline Lincosamide Cepharmycin Tetracycline
B016	<i>mef(En2)</i> <i>tet(Q)</i> <i>tet(X2)</i> <i>aadE</i> <i>tet(Q)</i> <i>catP</i> <i>tet(W)</i> <i>erm(F)</i> <i>aadS</i> <i>Inu(AN2)</i> <i>bla_{CtxA}</i> <i>bla_{CblA}</i> <i>tet(O)</i> <i>tet(40)</i>	100 100.16 100.17 100 100.1 100 100 100 100 100 100 100 100 100 100	2 41.7 1 2.81 2.59 2 6.43 12.4 6.94 3 11.89 5 9.8 10.52	Macrolide, lincosamide Tetracycline Tetracycline Aminoglycoside Tetracycline Phenicol Tetracycline Streptogramin, macrolide, lincosamide Aminoglycoside Lincosamide Cepharmycin Cephalosporin Tetracycline Tetracycline
B017	<i>erm(G)</i> <i>mef(En2)</i> <i>tet(Q)</i> <i>tet(Q)</i> <i>erm(B)</i> <i>dfrF</i> <i>tet(M)</i> <i>tet(W)</i> <i>bla_{CtxA}</i> <i>bla_{CepA}</i> <i>bla_{CblA}</i> <i>tet(40)</i> <i>mef(A)</i> <i>tet(O)</i>	100.27 100.33 100 100 100.27 104.85 100.47 100 100.41 100.11 100 100.33 99.92 100	1.96 2.55 8.35 2.54 1 1.94 1 15.55 1 1 4.85 2.33 3.78 14.87	Streptogramin, macrolide, lincosamide Macrolide, lincosamide Tetracycline Tetracycline Streptogramin, macrolide, lincosamide Diaminopyrimidine Tetracycline Tetracycline Cepharmycin Cephalosporin Cephalosporin Tetracycline Macrolide Tetracycline
B018	<i>erm(G)</i> <i>mef(En2)</i> <i>tet(Q)</i> <i>sul2</i> <i>erm(T)</i> <i>tet(Q)</i> <i>erm(F)</i> <i>Inu(AN2)</i> <i>bla_{CblA}</i> <i>tet(X2)</i> <i>tet(O)</i> <i>tet(40)</i> <i>tet(W)</i> <i>bla_{CtxA-4}</i>	100 100 100 100 100.14 100 100 100 103.03 100.09 100 100 100 101.41 100	1 7.26 6.5 9.49 2 16.22 21.74 5.52 1.95 2.35 19.82 11.88 1.97 3.3	Streptogramin, macrolide, lincosamide Macrolide, lincosamide Tetracycline Sulfunamide Streptogramin, macrolide, lincosamide Tetracycline Streptogramin, macrolide, lincosamide Lincosamide Cephalosporin Tetracycline Tetracycline Tetracycline Tetracycline Cepharmycin
B020	-	-	-	-
B021	<i>mef(En2)</i> <i>tet(Q)</i> <i>bla_{CtxA-6}</i> <i>dfrA1</i> <i>sul2</i> <i>dfrF</i> <i>tet(W)</i> <i>Inu(AN2)</i> <i>bla_{CtxA}</i> <i>bla_{CepA}</i> <i>tet(O)</i>	98.76 100 100 100.21 100 101.21 100 100 100 100 100	1.48 8.01 14.4 1 2 1.89 5.53 2 2.73 1 4.64	Macrolide, lincosamide Tetracycline Cepharmycin Diaminopyrimidine Sulfunamide Diaminopyrimidine Tetracycline Lincosamide Cepharmycin Cephalosporin Tetracycline
B022	<i>bla_{EC}</i> <i>dfrA1</i> <i>tet(Q)</i> <i>tet(W)</i> <i>tet(X2)</i> <i>vanY-D</i> <i>aph(6)-Id</i> <i>sul2</i> <i>dfrF</i> <i>aad9</i> <i>erm(F)</i> <i>tet(Q)</i> <i>Inu(AN2)</i> <i>bla_{CblA}</i> <i>tet(40)</i> <i>mef(A)</i> <i>tet(O)</i> <i>tet(O/32/O)</i> <i>emrD</i>	100.79 100 100 96.41 100 101.44 100 100 101.21 100 100 100.42 100 100.11 100 99.92 97.76 100.1 100	2.01 1 16.73 3.38 2.96 1.01 2.96 2.39 2.01 1 2 2 1 4.27 4.43 2.62 2.27 2.8 3	Cephalosporin Diaminopyrimidine Tetracycline Tetracycline Tetracycline Glycopeptide Aminoglycoside Sulfunamide Diaminopyrimidine Aminoglycoside Streptogramin, macrolide, lincosamide Tetracycline Lincosamide Cephalosporin Tetracycline Macrolide Tetracycline Tetracycline Phenicol
B026	<i>tet(Q)</i> <i>bla_{CtxA-6}</i> <i>dfrF</i> <i>erm(F)</i>	100 100 100.4 100.12	6.99 14.28 2 3	Tetracycline Cepharmycin Diaminopyrimidine Streptogramin, macrolide, lincosamide

	<i>aadS</i>	100	4.19	Aminoglycoside
	<i>tet(32)</i>	98.33	1.86	Tetracycline
	<i>bla_{CtxA-3}</i>	100	5.67	Cephameycin
B032	<i>tet(Q)</i>	100	11.98	Tetracycline
	<i>tet(X1)</i>	100.44	1.36	Tetracycline
	<i>msr(D)</i>	100	6.55	Streptogramin, macrolide
	<i>tet(A)</i>	100.25	1	Tetracycline
	<i>aadE</i>	100	1.97	Aminoglycoside
	<i>erm(B)</i>	100	5.87	Streptogramin, macrolide, lincosamide, streptogramin A, streptogramin B
	<i>dfrF</i>	100.2	2	Diaminopyrimidine
	<i>aad9</i>	100	3	Aminoglycoside
	<i>tet(W)</i>	100	6.94	Tetracycline
	<i>erm(F)</i>	100	4.98	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	2	Aminoglycoside
	<i>tet(Q)</i>	100.16	4.04	Tetracycline
	<i>bla_{CblA}</i>	101.91	1.92	Cephalosporin
	<i>tet(W)</i>	100.21	1	Tetracycline
	<i>mef(A)</i>	100	7.08	Macrolide
	<i>tet(O)</i>	100	12.15	Tetracycline
	<i>erm(G)</i>	100	7.25	Streptogramin, macrolide, lincosamide
	<i>bla_{CtxA-4}</i>	100	8.29	Cephameycin
B033	<i>mef(En2)</i>	100.17	1.36	Macrolide, lincosamide
	<i>tet(Q)</i>	100.05	2.4	Tetracycline
	<i>tet(Q)</i>	100	7.86	Tetracycline
	<i>bla_{CtxA-6}</i>	100.83	1.85	Cephameycin
	<i>catP</i>	100	2	Phenicol
	<i>tet(W)</i>	100.1	5.58	Tetracycline
	<i>tet(O)</i>	101.61	2.85	Tetracycline
	<i>tet(40)</i>	100.66	2.21	Tetracycline
B045	-	-	-	-
B046	<i>tet(Q)</i>	90.91	1.45	Tetracycline
	<i>bla_{CtxA-6}</i>	100.31	2.1	Cephameycin
B049	<i>sul1</i>	100.36	2	Sulfonamide
	<i>bla_{TEM}</i>	100	1	Penam, penem, cephalosporin, monobactam
	<i>mph(A)</i>	100	3	Macrolide
	<i>aac(3)-IId</i>	100.23	1	Aminoglycoside
	<i>dfrA12</i>	100.4	2	Diaminopyrimidine
	<i>aadA2</i>	100.38	1.87	Aminoglycoside
B057	<i>tet(Q)</i>	100.26	3.4	Tetracycline
	<i>acrF</i>	100	13.28	Fluoroquinolone, cephalosporin, cephameycin, penam
	<i>emrD</i>	100	27.3	Phenicol
	<i>bla_{TEM}</i>	100	25.2	Penam, penem, cephalosporin, monobactam
	<i>aph(6)-Id</i>	100	19.75	Aminoglycoside
	<i>sul2</i>	100	24.92	Sulfonamide
	<i>aph(3'')-Ib</i>	100	15.43	Aminoglycoside
	<i>tet(B)</i>	101.58	2.02	Tetracycline
	<i>dfrA14</i>	100	10.89	Diaminopyrimidine
	<i>bla_{EC}</i>	100	23.81	Cephalosporin
B085	<i>tet(Q)</i>	100.36	2.53	Tetracycline
	<i>vanG2</i>	99.14	1.35	Glycopeptide
	<i>catP</i>	101.28	1.01	Phenicol
	<i>tet(W)</i>	100.1	3.02	Tetracycline
	<i>bla_{CtxA}</i>	100.21	2	Cephameycin
B081	<i>tet(Q)</i>	100.05	2	Tetracycline
	<i>tet(W)</i>	99.95	1.17	Tetracycline
	<i>aadE</i>	100	2	Aminoglycoside
	<i>erm(F)</i>	100.37	1.96	Streptogramin, macrolide, lincosamide
	<i>tet(Q)</i>	99.95	1.24	Tetracycline
	<i>bla_{CtxA-3}</i>	100	4.02	Cephameycin
	<i>nimE</i>	100	1	Nitroimidazole
	<i>vanR-G</i>	100.14	1	Glycopeptide
C07b	<i>tet(Q)</i>	100	7.74	Tetracycline
	<i>aadE</i>	100.12	1	Aminoglycoside
	<i>tet(32)</i>	100	3.6	Tetracycline
	<i>erm(F)</i>	100.37	2	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	3	Aminoglycoside
	<i>bla_{CtxA}</i>	100.1	2.89	Cephameycin
	<i>mef(A)</i>	100.16	2	Macrolide
	<i>tet(W)</i>	100	7.14	Tetracycline
	<i>tet(O)</i>	100	8.96	Tetracycline
C012	<i>dfrA17</i>	101.69	1.02	Diaminopyrimidine
	<i>mef(En2)</i>	100.25	1.63	Macrolide, lincosamide
	<i>tet(Q)</i>	100	15.53	Tetracycline
	<i>bla_{Ctx-M}</i>	100	5	Cephalosporin
	<i>aadA5</i>	100	1	Aminoglycoside
	<i>tet(W)</i>	100	3.87	Tetracycline
	<i>erm(B)</i>	100	1.96	Streptogramin, macrolide, lincosamide
	<i>qnrS</i>	100	3	Fluoroquinolone
	<i>dfrF</i>	100.2	3	Diaminopyrimidine
	<i>erm(F)</i>	100.12	4.75	Streptogramin, macrolide, lincosamide
	<i>lnu(AN2)</i>	100	1	Lincosamide
	<i>bla_{CtxA}</i>	100	1	Cephameycin
	<i>tet(O)</i>	100	8.79	Tetracycline

C013	<i>tet(Q)</i>	100	4.65	Tetracycline
	<i>tet(W)</i>	100.16	3.25	Tetracycline
	<i>tet(X2)</i>	100.34	1.74	Tetracycline
	<i>bla_{CtxA-6}</i>	102.08	2.02	Cepharmycin
	<i>bla_{EC}</i>	100.26	1	Cephalosporin
	<i>mdtM</i>	100	2	Nucleoside, fluoroquinolone, phenicol, lincosamide
	<i>dfrF</i>	100	1	Diaminopyrimidine
	<i>bla_{CtxA}</i>	100.31	1.41	Cepharmycin
	<i>tet(40)</i>	100	1.38	Tetracycline
	<i>emrD</i>	100.08	1.55	Phenicol

AMR genes identified from samples stored in DNA Shield™ and prepared using the RBK library preparation method are reported. AMR genes were identified using KMA (version 1.4.9) from the amrfinderplus database. AMR genes identified with a template coverage of ≥90% and depth of coverage of ≥1× were included.

Table A 3. Detection of AMR genes from OMNigene GUT RBK samples.

Sample Number	AMR genes identified by CMg sequencing	Template Coverage	Depth (×)	Associated antibiotic drug class
C1	<i>mef(En2)</i>	100.17	2.46	Macrolide, lincosamide
	<i>tet(Q)</i>	100	12.2	Tetracycline
	<i>cfi(C)</i>	98.25	3.11	Oxazolidinone, lincosamide, phenicol, streptogramin
	<i>aph(3')-IIIa</i>	96.98	2.59	Aminoglycoside
	<i>aadA1</i>	100.25	1	Aminoglycoside
	<i>aadE</i>	100	1	Aminoglycoside
	<i>tet(W)</i>	101.15	1.77	Tetracycline
	<i>lnu(AN2)</i>	101.17	1.01	Lincosamide
B1	<i>catA13</i>	100	3	Phenicol
	<i>mef(En2)</i>	100.17	2.46	Macrolide, lincosamide
	<i>tet(Q)</i>	100	12.2	Tetracycline
	<i>cfi(C)</i>	98.25	3.11	Oxazolidinone, lincosamide, phenicol, streptogramin
	<i>aph(3')-IIIa</i>	96.98	2.59	Aminoglycoside
	<i>aadA1</i>	100.25	1	Aminoglycoside
	<i>aadE</i>	100	1	Aminoglycoside
	<i>tet(W)</i>	101.15	1.77	Tetracycline
B2	<i>lnu(AN2)</i>	101.17	1.01	Lincosamide
	<i>catA13</i>	100	3	Phenicol
	<i>mef(En2)</i>	103.4	1.83	Macrolide, lincosamide
	<i>tet(Q)</i>	100	3.58	Tetracycline
	<i>mdtM</i>	101.05	2.01	Nucleoside, fluoroquinolone, phenicol, lincosamide
	<i>aph(6)-Id</i>	100	2.8	Aminoglycoside
	<i>sul2</i>	100.25	2	Sulfonamide
	<i>tet(W)</i>	100	12.63	Tetracycline
B3	<i>erm(F)</i>	100	1	Streptogramin, macrolide, lincosamide
	<i>fosA_PA112</i>	100	1	Phosphonic acid
	<i>tet(Q)</i>	93.46	1.06	Tetracycline
	<i>lnu(AN2)</i>	100	2.9	Lincosamide
	<i>dfrA14</i>	100.21	2	Diaminopyrimidine
	<i>tet(Q)</i>	100.05	5.07	Tetracycline
	<i>erm(F)</i>	100	9.53	Streptogramin, macrolide, lincosamide
	<i>tet(O)</i>	100.21	2	Tetracycline
B10	<i>catA13</i>	100	7.96	Phenicol
	<i>tet(Q)</i>	100	6.78	Tetracycline
	<i>tet(W)</i>	100	5.94	Tetracycline
	<i>msr(D)</i>	101.09	1.51	Streptogramin, macrolide
	<i>tet(32)</i>	68.07	1.02	Tetracycline
	<i>mef(A)</i>	100.08	1	Macrolide
	<i>tet(O)</i>	77.14	1.75	Tetracycline
B40	<i>bla_{CtxA-6}</i>	100	1	Cepharmycin
	<i>dfrA8</i>	100.2	2	Diaminopyrimidine
B44	<i>tet(Q)</i>	100.05	2.94	Tetracycline
	<i>dfrF</i>	100.2	1	Diaminopyrimidine
	<i>catP</i>	100.32	2	Phenicol
B56	<i>bla_{CtxA-6}</i>	100	9.19	Cepharmycin
	<i>aph(6)-Id</i>	100.12	2	Aminoglycoside
	<i>sul2</i>	103.68	1.64	Sulfonamide
	<i>qnrS</i>	100	1	Fluoroquinolone
	<i>tet(Q)</i>	101.19	1.24	Tetracycline
	<i>tet(O)</i>	91.98	1.44	Tetracycline
B80	<i>dfrA14</i>	101.27	2.01	Diaminopyrimidine
	<i>tet(M)</i>	99.17	2.22	Tetracycline
	<i>tet(Q)</i>	100	22.45	Tetracycline
	<i>tet(Q)</i>	98.23	1.87	Tetracycline
	<i>tet(W)</i>	100	13.12	Tetracycline
	<i>bla_{CtxA-6}</i>	100	20.15	Cepharmycin
	<i>emrD</i>	100.08	1	Phenicol
	<i>tet(O)</i>	100	15.09	Tetracycline
	<i>tet(A)</i>	80.75	1.35	Tetracycline
	<i>sul2</i>	100	5.67	Sulfonamide

	<i>dfrF</i>	100	3	Diaminopyrimidine
	<i>erm(F)</i>	100	79.78	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	101.27	1.86	Aminoglycoside
	<i>nimE</i>	100	31.47	Nitroimidazole
	<i>bla_{CblA}</i>	100.11	1	Cephalosporin
	<i>tet(40)</i>	100	2.63	Tetracycline
	<i>bla_{CtxA-5}</i>	100	6.27	Cephameycin
B94	<i>tet(Q)</i>	100	8.04	Tetracycline
	<i>tet(W)</i>	100.26	2.3	Tetracycline
	<i>aadE</i>	98.62	1.86	Aminoglycoside
	<i>erm(B)</i>	100.27	1	Streptogramin, macrolide, lincosamide
	<i>dfrF</i>	100.61	1.01	Diaminopyrimidine
	<i>erm(F)</i>	107.62	2.08	Streptogramin, macrolide, lincosamide
	<i>erm(X)</i>	100.66	1.01	Streptogramin, macrolide, lincosamide
	<i>tet(O)</i>	100	7.8	Tetracycline
	<i>tet(40)</i>	100	4.22	Tetracycline
B4	<i>mef(En2)</i>	100	14.06	Macrolide, lincosamide
	<i>tet(Q)</i>	100	9.71	Tetracycline
	<i>tet(W)</i>	100	7.04	Tetracycline
	<i>tet(O)</i>	100.57	1.92	Tetracycline
	<i>erm(B)</i>	100	5.38	Streptogramin, macrolide, lincosamide, streptogramin
	<i>dfrF</i>	100.2	1	Diaminopyrimidine
	<i>tet(32)</i>	100.68	2.65	Tetracycline
	<i>erm(F)</i>	100	1	Streptogramin, macrolide, lincosamide
	<i>Inu(AN2)</i>	100	15.66	Lincosamide
	<i>bla_{CtxA}</i>	100	3	Cephameycin
	<i>bla_{CblA}</i>	102.02	1.98	Cephalosporin
	<i>tet(40)</i>	100.9	2.24	Tetracycline
	<i>emrD</i>	100.17	2	Phenicol
B5	<i>tet(Q)</i>	100	25.67	Tetracycline
	<i>tet(W)</i>	103.12	1.65	Tetracycline
	<i>bla_{CtxA-6}</i>	100	4.62	Cephameycin
	<i>cfiA</i>	100	2.08	Carbapenem
	<i>tet(O)</i>	99.43	1.93	Tetracycline
	<i>sul2</i>	100	1	Sulfunamide
	<i>aph(3')-IIIa</i>	101.01	2.01	Aminoglycoside
	<i>tet(Q)</i>	93.15	1.49	Tetracycline
	<i>erm(F)</i>	104.49	2.74	Streptogramin, macrolide, lincosamide
	<i>bla_{CtxA}</i>	100	12.06	Cephameycin
	<i>bla_{CblA}</i>	100.11	7.44	Cephalosporin
B6	<i>mef(En2)</i>	100	4.55	Macrolide, lincosamide
	<i>tet(Q)</i>	100	12.79	Tetracycline
	<i>tet(Q)</i>	99.95	6.11	Tetracycline
	<i>aadE</i>	101.73	1.27	Aminoglycoside
	<i>dfrF</i>	100.61	1.85	Diaminopyrimidine
	<i>tet(32)</i>	100	3.76	Tetracycline
	<i>erm(F)</i>	99.5	3.27	Streptogramin, macrolide, lincosamide
	<i>Inu(AN2)</i>	100	3.98	Lincosamide
	<i>bla_{CblA}</i>	103.37	2.47	Cephalosporin
	<i>tet(O)</i>	100	3.17	Tetracycline
	<i>tet(W)</i>	100	12.82	Tetracycline
B7	<i>Inu(C)</i>	100	2	Lincosamide
	<i>tet(O)</i>	100.1	4.03	Tetracycline
	<i>tet(Q)</i>	100	32.23	Tetracycline
	<i>tet(W)</i>	100	14.2	Tetracycline
	<i>tet(X2)</i>	98.46	2.38	Tetracycline
	<i>bla_{CtxA-6}</i>	100	10.87	Cephameycin
	<i>aadE</i>	100	2.96	Aminoglycoside
	<i>tet(40)</i>	100.25	2.53	Tetracycline
	<i>bla_{CtxA-5}</i>	99.9	1.77	Cephameycin
B8	<i>aph(3')-IIIa</i>	100.5	2.56	Aminoglycoside
	<i>Inu(C)</i>	100.2	1	Lincosamide
	<i>mef(En2)</i>	100.25	4.02	Macrolide, lincosamide
	<i>tet(Q)</i>	100	29.81	Tetracycline
	<i>tet(X2)</i>	100	11.02	Tetracycline
	<i>bla_{CtxA-6}</i>	100.31	5.21	Cephameycin
	<i>sul2</i>	102.94	1.39	Sulfunamide
	<i>aadE</i>	100	3.27	Aminoglycoside
	<i>erm(B)</i>	101.22	1.78	Streptogramin, macrolide, lincosamide
	<i>spw</i>	100	1	Aminoglycoside
	<i>dfrF</i>	100	2.79	Diaminopyrimidine
	<i>tet(Q)</i>	99.01	1.35	Tetracycline
	<i>tet(W)</i>	100	14.11	Tetracycline
	<i>erm(F)</i>	100	12.72	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	1	Aminoglycoside
	<i>Inu(AN2)</i>	100	4	Lincosamide
	<i>bla_{CtxA}</i>	100	10.79	Cephameycin
	<i>bla_{CblA}</i>	95.17	1.81	Cephalosporin
	<i>tet(W)</i>	100	2.61	Tetracycline
	<i>tet(O)</i>	100	16.43	Tetracycline
	<i>tet(40)</i>	100	6.36	Tetracycline
	<i>tet(O/32/O)</i>	99.95	1.51	Tetracycline
B9	<i>tet(Q)</i>	100	6.34	Tetracycline
	<i>tet(W)</i>	100	4.8	Tetracycline
	<i>sul2</i>	100	7.45	Sulfunamide
	<i>catP</i>	100.48	1	Phenicol
	<i>erm(F)</i>	100.12	2.71	Streptogramin, macrolide, lincosamide

	<i>aadS</i>	100	3.58	Aminoglycoside
	<i>bla_{OX}A</i>	100	4.47	Penam
	<i>bla_{CtxA}</i>	100.93	2.01	Cephameycin
	<i>bla_{CblA}</i>	100.45	2.71	Cephalosporin
	<i>tet(X2)</i>	100.26	2.33	Tetracycline
	<i>tet(O)</i>	100.1	4.44	Tetracycline
	<i>ere(D)</i>	101.63	2.52	Macrolide
B11	<i>lnu(C)</i>	100	3	Lincosamide
	<i>tet(Q)</i>	100	14.01	Tetracycline
	<i>tet(Q)</i>	100.16	6.32	Tetracycline
	<i>tet(W)</i>	100	5.57	Tetracycline
	<i>bla_{CtxA-6}</i>	100	4.82	Cephameycin
	<i>catP</i>	100.96	1.01	Phenicol
	<i>erm(F)</i>	100	5.63	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	101.74	4.74	Aminoglycoside
	<i>bla_{CtxA}</i>	100	11.78	Cephameycin
	<i>bla_{CblA}</i>	100	2.92	Cephalosporin
	<i>tet(40)</i>	100	4	Tetracycline
B12	<i>mef(En2)</i>	100	5.5	Macrolide, lincosamide
	<i>tet(M)</i>	100.1	2.51	Tetracycline
	<i>tet(Q)</i>	100	31.09	Tetracycline
	<i>dfrF</i>	100.81	1.87	Diaminopyrimidine
	<i>tet(W)</i>	100	19.61	Tetracycline
	<i>erm(F)</i>	100	2.73	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	99.54	2.19	Aminoglycoside
	<i>lnu(AN2)</i>	100	4	Lincosamide
	<i>bla_{CblA}</i>	102.24	1.31	Cephalosporin
	<i>tet(X2)</i>	100	6.38	Tetracycline
	<i>tet(40)</i>	102.21	1.64	Tetracycline
	<i>tet(O)</i>	100	14.54	Tetracycline
B13	<i>erm(G)</i>	100	7.74	Streptogramin, macrolide, lincosamide
	<i>mef(En2)</i>	100	8.23	Macrolide, lincosamide
	<i>tet(Q)</i>	100	33.43	Tetracycline
	<i>tet(W)</i>	100	6.16	Tetracycline
	<i>msr(D)</i>	102.32	2.35	Streptogramin, macrolide
	<i>aadE</i>	100.23	2.83	Aminoglycoside
	<i>dfrF</i>	100	3.59	Diaminopyrimidine
	<i>catP</i>	99.84	1.84	Phenicol
	<i>erm(F)</i>	100	25.35	Streptogramin, macrolide, lincosamide
	<i>tet(Q)</i>	100.83	1.14	Tetracycline
	<i>bla_{CblA}</i>	99.78	1.46	Cephalosporin
	<i>tet(X2)</i>	100	26.48	Tetracycline
	<i>tet(40)</i>	100	7.1	Tetracycline
	<i>mef(A)</i>	100	3.51	Macrolide
	<i>tet(O)</i>	100	8.02	Tetracycline
B14	<i>tet(Q)</i>	100.1	4.27	Tetracycline
	<i>tet(Q)</i>	100	14.6	Tetracycline
	<i>aph(6)-Id</i>	100	1	Aminoglycoside
	<i>aadA1</i>	100.13	1	Aminoglycoside
	<i>tet(W)</i>	100	11.72	Tetracycline
	<i>bla_{CblA}</i>	100	5.82	Cephalosporin
	<i>mef(A)</i>	99.92	2.55	Macrolide
	<i>dfrA14</i>	100	1	Diaminopyrimidine
	<i>tet(O)</i>	100.1	5.56	Tetracycline
B015	<i>tet(Q)</i>	100	7.15	Tetracycline
	<i>lsa(E)</i>	100.34	2.28	Lincosamide, pleuromutilin, streptogramin
	<i>lnu(B)</i>	100	1	Lincosamide
	<i>tet(W)</i>	100	1	Tetracycline
	<i>lnu(AN2)</i>	100.19	1	Lincosamide
	<i>bla_{CtxA-3}</i>	100	10.34	Cephameycin
	<i>bla_{CblA}</i>	101.35	1.8	Cephalosporin
	<i>tet(X2)</i>	100.43	1	Tetracycline
	<i>mef(A)</i>	100.08	1.36	Macrolide
	<i>tet(O)</i>	100	5.15	Tetracycline
B016	<i>lnu(C)</i>	100.2	1	Lincosamide
	<i>mef(En2)</i>	100.17	2	Macrolide, lincosamide
	<i>tet(M)</i>	102.45	1.52	Tetracycline
	<i>tet(Q)</i>	100.16	26.68	Tetracycline
	<i>aadE</i>	100.81	2.01	Aminoglycoside
	<i>erm(B)</i>	100.14	1	Streptogramin, macrolide, lincosamide, streptogramin A, streptogramin B
	<i>catP</i>	100	1	Phenicol
	<i>aad9</i>	100	2	Aminoglycoside
	<i>erm(F)</i>	100	6.38	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	4.44	Aminoglycoside
	<i>lnu(AN2)</i>	100	2.95	Lincosamide
	<i>bla_{CtxA}</i>	100	14.87	Cephameycin
	<i>tet(O)</i>	100	6.73	Tetracycline
	<i>tet(40)</i>	100	7.12	Tetracycline
	<i>tet(O/32/O)</i>	100.1	1.39	Tetracycline-
B017	<i>mef(En2)</i>	100	6.79	Macrolide, lincosamide
	<i>tet(Q)</i>	100	9.95	Tetracycline
	<i>tet(Q)</i>	100	17.31	Tetracycline
	<i>msr(D)</i>	100.07	1.34	Streptogramin, macrolide
	<i>aph(6)-Id</i>	100	1	Aminoglycoside
	<i>tet(M)</i>	100.94	1.2	Tetracycline
	<i>aadE</i>	100.12	2	Aminoglycoside
	<i>dfrF</i>	100	4	Diaminopyrimidine

	<i>tet(Q)</i>	100	3.2	Tetracycline
	<i>tet(W)</i>	100	12.23	Tetracycline
	<i>tet(32)</i>	100.05	4.45	Tetracycline
	<i>tet(Q)</i>	100.36	2	Tetracycline
	<i>lnu(AN2)</i>	100	5	Lincosamide
	<i>bla_{CfXA}</i>	100	3.97	Cephameycin
	<i>tet(O)</i>	100.1	22.97	Tetracycline
B018	<i>erm(G)</i>	100	1	Streptogramin, macrolide, lincosamide
	<i>mef(En2)</i>	100	12.97	Macrolide, lincosamide
	<i>tet(Q)</i>	100	48.88	Tetracycline
	<i>tet(Q)</i>	100.21	2.78	Tetracycline
	<i>msr(D)</i>	101.57	1.62	Streptogramin, macrolide
	<i>sul2</i>	100	20.87	Sulfonamide
	<i>erm(B)</i>	103.79	1.75	Streptogramin, macrolide, lincosamide, streptogramin
	<i>dfrF</i>	100	3	Diaminopyrimidine
	<i>tet(Q)</i>	100	15.16	Tetracycline
	<i>tet(W)</i>	100	8.05	Tetracycline
	<i>erm(F)</i>	100	52.65	Streptogramin, macrolide, lincosamide
	<i>lnu(AN2)</i>	100	13.51	Lincosamide
	<i>bla_{CfXA}</i>	100	4.11	Cephameycin
	<i>bla_{CbIA}</i>	101.01	2.01	Cephalosporin
	<i>tet(X2)</i>	100	13.7	Tetracycline
	<i>mef(A)</i>	100	1	Macrolide
	<i>tet(O)</i>	100	31.5	Tetracycline
	<i>tet(40)</i>	100	16.76	Tetracycline
B020	<i>lnu(AN2)</i>	100	1	Lincosamide
	<i>emrD</i>	100.25	1	Phenicol
B021	<i>mef(En2)</i>	100	3.95	Macrolide, lincosamide
	<i>tet(Q)</i>	100	10.97	Tetracycline
	<i>bla_{CfXA-6}</i>	100	12.37	Cephameycin
	<i>aadE</i>	100.23	2	Aminoglycoside
	<i>dfrF</i>	103.84	1.46	Diaminopyrimidine
	<i>tet(W)</i>	100	3.18	Tetracycline
	<i>erm(F)</i>	100.25	1	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100	1	Aminoglycoside
	<i>tet(Q)</i>	100.36	2	Tetracycline
	<i>lnu(AN2)</i>	100	3	Lincosamide
	<i>bla_{CfXA}</i>	100	2.97	Cephameycin
	<i>bla_{CbIA}</i>	100	3	Cephalosporin
	<i>tet(X2)</i>	100	6.56	Tetracycline
	<i>tet(O)</i>	100.45	2	Tetracycline
B022	<i>lnu(C)</i>	100.81	2.01	Lincosamide
	<i>tet(Q)</i>	100	23.05	Tetracycline
	<i>acrF</i>	100.35	2.08	Fluoroquinolone, cephalosporin, cephamycin, penam
	<i>aph(6)-Id</i>	100	4.69	Aminoglycoside
	<i>sat4</i>	100.37	1	Nucleoside
	<i>tet(A)</i>	100.17	2	Tetracycline
	<i>sul2</i>	100	3	Sulfonamide
	<i>aadE</i>	100.46	1.86	Aminoglycoside
	<i>dfrF</i>	100	1	Diaminopyrimidine
	<i>aad9</i>	100.13	1	Aminoglycoside
	<i>aadS</i>	100	3	Aminoglycoside
	<i>tet(32)</i>	101.04	1.24	Tetracycline
	<i>bla_{OXA}</i>	103.88	1.47	Penam
	<i>bla_{CbIA}</i>	100	3	Cephalosporin
	<i>tet(O)</i>	100.08	5.75	Tetracycline
	<i>tet(40)</i>	100.42	2.71	Tetracycline
	<i>dfrA14</i>	100.32	2	Diaminopyrimidine
B026	<i>tet(Q)</i>	100.16	4.46	Tetracycline
	<i>bla_{CfXA-6}</i>	100	11.47	Cephameycin
	<i>dfrF</i>	100	4.88	Diaminopyrimidine
	<i>tet(W)</i>	100.21	1	Tetracycline
	<i>erm(F)</i>	100	3.19	Streptogramin, macrolide, lincosamide
	<i>bla_{CfXA}</i>	100	6.27	Cephameycin
	<i>tet(O)</i>	100.16	6.78	Tetracycline
B032	<i>tet(Q)</i>	100	4.56	Tetracycline
	<i>vanR-Cd</i>	100	1	Glycopeptide
	<i>bla_{CfXA-6}</i>	100.83	2.01	Cephameycin
	<i>msr(D)</i>	100	14.85	Streptogramin, macrolide
	<i>erm(B)</i>	100.27	2	Streptogramin, macrolide, lincosamide
	<i>aadE</i>	101.15	1.82	Aminoglycoside
	<i>tet(W)</i>	100	7.23	Tetracycline
	<i>aadS</i>	100	3.29	Aminoglycoside
	<i>tet(44)</i>	100.47	1.13	Tetracycline
	<i>tet(Q)</i>	100	6.59	Tetracycline
	<i>tet(X2)</i>	100.77	1.31	Tetracycline
	<i>mef(A)</i>	100	7.14	Macrolide
	<i>erm(G)</i>	100	14.63	Streptogramin, macrolide, lincosamide
	<i>tet(O)</i>	100	8.07	Tetracycline
	<i>bla_{CfXA-4}</i>	100	19.14	Cephameycin
B033	<i>mef(En2)</i>	100.91	1.28	Macrolide, lincosamide
	<i>tet(Q)</i>	100	12.73	Tetracycline
	<i>bla_{CfXA-6}</i>	100	8.16	Cephameycin
	<i>aacA-ENT1</i>	100	1	Aminoglycoside
	<i>tet(W)</i>	100	3.43	Tetracycline
	<i>tet(32)</i>	101.46	2	Tetracycline

	<i>lnu(AN2)</i>	100.39	1.34	Lincosamide
B045	<i>dfrF</i>	100	1	Diaminopyrimidine
B046	<i>tet(Q)</i>	100.05	1.73	Tetracycline
	<i>bla_{CfIA-6}</i>	99.69	1.23	Cepharmycin
	<i>erm(B)</i>	100	2	Streptogramin, macrolide, lincosamide
	<i>aadE</i>	100.12	2	Aminoglycoside
	<i>bla_{CfIA}</i>	100	2.26	Cepharmycin
B049	<i>tet(Q)</i>	100.1	1	Tetracycline
	<i>acrF</i>	100.13	2.41	Fluoroquinolone, cephalosporin, cephamycin, penam
	<i>emrD</i>	100.08	1	Phenicol
	<i>bla_{TEM}</i>	100	1	Penam, penem, cephalosporin, monobactam
	<i>dfrA12</i>	100	2.73	Diaminopyrimidine
B057	<i>mef(A)</i>	100.33	1	Macrolide
	<i>tet(Q)</i>	100.16	3.5	Tetracycline
	<i>aph(3'')-Ib</i>	100	28.82	Aminoglycoside
	<i>acrF</i>	100	16.94	Fluoroquinolone, cephalosporin, cephamycin, penam
	<i>emrD</i>	100.13	17.99	Phenicol
	<i>sslE</i>	100	40.7	Phenicol
	<i>bla_{TEM}</i>	100	31.81	Penam, penem, cephalosporin, monobactam
	<i>aph(6)-Id</i>	100	2.86	Aminoglycoside
	<i>catA2</i>	100	40.75	Phenicol
	<i>sul2</i>	100	4.87	Sulfonamide
	<i>tet(B)</i>	100	31.06	Tetracycline
	<i>qnrS</i>	100	36.41	Fluoroquinolone
	<i>dfrA14</i>	100	32.29	Diaminopyrimidine
	<i>bla_{EC}</i>	100	21.75	Cephalosporin
B085	<i>tet(Q)</i>	100	14.39	Tetracycline
	<i>vanG2</i>	100	4.95	Glycopeptide
	<i>vanXY-G2</i>	100	5.87	Glycopeptide
	<i>tet(W)</i>	100	3.49	Tetracycline
	<i>erm(B)</i>	100	7.15	Streptogramin, macrolide, lincosamide
	<i>tet(Q)</i>	100.05	2.73	Tetracycline
	<i>tet(32)</i>	100.05	1.21	Tetracycline
	<i>erm(F)</i>	100	3	Streptogramin, macrolide, lincosamide
	<i>bla_{CfIA}</i>	100	5.83	Cepharmycin
	<i>tet(40)</i>	100.16	1	Tetracycline
	<i>mef(A)</i>	100.08	1	Macrolide
	<i>tet(O)</i>	100	3.79	Tetracycline
B081	<i>tet(Q)</i>	92.26	2.54	Tetracycline
	<i>tet(W)</i>	100.1	2.05	Tetracycline
	<i>tet(Q)</i>	99.95	1.22	Tetracycline
	<i>bla_{CfIA-3}</i>	100	3.02	Cepharmycin
	<i>bla_{CblA}</i>	100	1	Cephalosporin
C07b	<i>tet(Q)</i>	100	16.68	Tetracycline
	<i>tet(Q)</i>	100.21	2	Tetracycline
	<i>aph(6)-Id</i>	100.36	1	Aminoglycoside
	<i>sul2</i>	100.25	1	Sulfonamide
	<i>aph(3')-IIIa</i>	100.13	1	Aminoglycoside
	<i>tet(Q)</i>	100	1.6	Tetracycline
	<i>tet(W)</i>	100	8.67	Tetracycline
	<i>tet(32)</i>	100	5.77	Tetracycline
	<i>erm(F)</i>	100.75	2.01	Streptogramin, macrolide, lincosamide
	<i>aadS</i>	100.12	1	Aminoglycoside
	<i>lnu(AN2)</i>	100.19	2	Lincosamide
	<i>bla_{CfIA-3}</i>	100	10.14	Cepharmycin
	<i>bla_{CblA}</i>	100.11	2	Cephalosporin
	<i>tet(O)</i>	100	20.63	Tetracycline
C012	<i>mef(En2)</i>	100	2.69	Macrolide, lincosamide
	<i>sul1</i>	100.12	2	Sulfonamide
	<i>tet(Q)</i>	100	5.93	Tetracycline
	<i>erm(F)</i>	100.37	1.77	Streptogramin, macrolide, lincosamide
	<i>lnu(AN2)</i>	100	1	Lincosamide
C013	<i>mef(En2)</i>	100	6.98	Macrolide, lincosamide
	<i>tet(Q)</i>	100	6.64	Tetracycline
	<i>tet(Q)</i>	100	4.53	Tetracycline
	<i>tet(X2)</i>	100	4.94	Tetracycline
	<i>bla_{CfIA-6}</i>	100.62	52.96	Cepharmycin
	<i>mdtM</i>	100	2.24	Nucleoside, fluoroquinolone, phenicol, lincosamide
	<i>tet(W)</i>	100	4.42	Tetracycline
	<i>erm(F)</i>	100	2.9	Streptogramin, macrolide, lincosamide
	<i>tet(Q)</i>	100	3.13	Tetracycline
	<i>lnu(AN2)</i>	100	5.72	Lincosamide
	<i>bla_{CfIA}</i>	100	7.74	Cepharmycin
	<i>tet(40)</i>	100	3.41	Tetracycline
	<i>tet(O)</i>	100.21	1.1	Tetracycline
	<i>emrD</i>	100	1	Phenicol

AMR genes identified from samples stored in OMNIgene GUT and prepared using the RBK library preparation method are reported. AMR genes were identified using KMA (version 1.4.9) from the amrfinderplus database. AMR genes identified with a template coverage of $\geq 90\%$ and depth of coverage of $\geq 1\times$ were included.

Table A 4. Detection of AMR genes from FTA RBK samples.

Sample Number	AMR genes identified by CMg sequencing	Template Coverage	Depth	Associated antibiotic resistance category
C1	-	-	-	-
B2	-	-	-	-
B3	-	-	-	-
B10	-	-	-	-
B40	-	-	-	-
B44	-	-	-	-
B56	-	-	-	-
B80	<i>tet(40)</i>	100.57	1.01	Tetracycline
B94	-	-	-	-
B4	<i>bla_{CtxA}</i>	100.1	1	Cephameycin
B5	-	-	-	-
B6	-	-	-	-
B7	<i>tet(Q)</i>	102.65	1.63	Tetracycline
B8	-	-	-	-
B9	-	-	-	-
B11	-	-	-	-
B12	-	-	-	-
B13	-	-	-	-
B14	-	-	-	-
B015	<i>tet(40)</i>	93.78	2.18	Tetracycline
B016	<i>tet(Q)</i>	95.28	3.08	Tetracycline
B018	<i>tet(O)</i>	103.28	1.41	Tetracycline
B020	-	-	-	-
B021	<i>bla_{CtxA-6}</i>	100.1	2	Cephameycin
B022	<i>lnu(C)</i>	100	1	Lincosamide
	<i>tet(40)</i>	101.15	1.6	Tetracycline
	<i>emrD</i>	100	1	Phenicol
B026	-	-	-	-
B032	-	-	-	-
B033	-	-	-	-
B045	-	-	-	-
B046	-	-	-	-
B049	<i>lin</i>	100.06	1	Lincosamide
B057	<i>qnrS</i>	100.3	1.69	Fluoroquinolone
B085	-	-	-	-
B081	-	-	-	-
C07b	-	-	-	-
C012	-	-	-	-
C013	-	-	-	-

AMR genes identified from samples stored on FTA cards and prepared using the RBK library preparation method are reported. AMR genes were identified using KMA (version 1.4.9) from the amrfinderplus database. AMR genes identified with a template coverage of $\geq 90\%$ and depth of coverage of $\geq 1\times$ were included.

8 References

1. Van Seventer JM, Hochberg NS. Principles of infectious diseases: transmission, diagnosis, prevention, and control. *International encyclopedia of public health*. 2017:22.
2. Bartlett A, Padfield D, Lear L, Bendall R, Vos M. A comprehensive list of bacterial pathogens infecting humans. *Microbiology*. 2022;168(12).
3. Cleaveland S, Haydon DT, Taylor L. Overviews of Pathogen Emergence: Which Pathogens Emerge, When and Why? In: Childs JE, Mackenzie JS, Richt JA, editors. *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 85-111.
4. Fonkwo PN. Pricing infectious disease. The economic and health implications of infectious diseases. *EMBO Rep*. 2008;9 Suppl 1(Suppl 1):S13-S7.
5. Murray CK, Horvath LL, Ericsson CD, Hatz C. An Approach to Prevention of Infectious Diseases during Military Deployments. *Clinical Infectious Diseases*. 2007;44(3):424-30.
6. White DW, Feigley CE, McKeown RE, Hout JJ, Hebert JR. Association Between Barracks Type and Acute Respiratory Infection in a Gender Integrated Army Basic Combat Training Population. *Military Medicine*. 2011;176(8):909-14.
7. de Santi VP, Nicand E, Lagathu G, Michel R, Rosenmayr G, Raingeval JB, et al. Incidence, Etiology, and Determinants Associated With Acute Diarrhea Among French Forces Deployed to Chad. *Journal of Travel Medicine*. 2011;18(2):115-20.
8. CIMINERA P, BRUNDAGE J. MALARIA IN U.S. MILITARY FORCES: A DESCRIPTION OF DEPLOYMENT EXPOSURES FROM 2003 THROUGH 2005. *The American Journal of Tropical Medicine and Hygiene*. 2007;76(2):275-9.
9. Mossong J, Bonert C, Weicherding P, Opp M, Reichert P, Even J, et al. Mumps outbreak among the military in Luxembourg in 2008: epidemiology and evaluation of control measures. *Eurosurveillance*. 2009;14(7):19121.
10. Tribble DR, Murray CK, Lloyd BA, Ganesan A, Mende K, Blyth DM, et al. After the Battlefield: Infectious Complications among Wounded Warriors in the Trauma Infectious Disease Outcomes Study. *Military Medicine*. 2019;184(Supplement_2):18-25.
11. Turner MD, Sapp J. Failure to Plan: The Disease That Cost an American Empire. *Military Medicine*. 2023;188(7-8):171-3.
12. Wax RG. Manipulation of human history by microbes. *Clinical Microbiology Newsletter*. 2007;29(2):9-16.
13. Bailey MS. A brief history of British military experiences with infectious and tropical diseases. *Journal of the Royal Army Medical Corps*. 2013;159(3):150-7.
14. Cox AL ZM, Tadi P. Weil Felix Test. <https://www.ncbi.nlm.nih.gov/books/NBK559225/>: StatPearls Publishing; 2023.
15. Biselli R, Nisini R, Lista F, Autore A, Lastilla M, De Lorenzo G, et al. A Historical Review of Military Medical Strategies for Fighting Infectious Diseases: From Battlefields to Global Health. *Biomedicines*. 2022;10(8).
16. Riley P, Ben-Nun M, Turtle J, Bacon D, Owens AN, Riley S. COVID-19: On the Disparity in Outcomes Between Military and Civilian Populations. *Military Medicine*. 2021;188(1-2):311-5.
17. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*. 2020;25(3).

18. Weller SA, Armstrong SR, Bailey S, Burnell HT, Burt EL, Cant NE, et al. Development and operation of the defence COVID-19 lab as a SARS-CoV-2 diagnostic screening capability for UK military personnel. *BMJ Military Health*. 2022:e002134.
19. Routledge M, Lyon J, Vincent C, Clarke AG, Shawcross K, Turpin C, et al. Management of a large outbreak of COVID-19 at a British Army training centre: lessons for the future. *BMJ Military Health*. 2021:e001976.
20. Services DM. Defence Medical Services
<https://www.gov.uk/government/groups/defence-medical-services#overview>: UK Government; [
21. (NSO) NSO. ALLIED JOINT DOCTRINE FOR MEDICAL SUPPORT. C Version 1 + UK national Elements ed2019.
22. Perez KK, Olsen RJ, Musick WL, Cernoch PL, Davis JR, Peterson LE, et al. Integrating rapid diagnostics and antimicrobial stewardship improves outcomes in patients with antibiotic-resistant Gram-negative bacteremia. *Journal of Infection*. 2014;69(3):216-25.
23. Dimaio MA, Pereira IT, George TI, Banaei N. Performance of BinaxNOW for diagnosis of malaria in a U.S. hospital. *J Clin Microbiol*. 2012;50(9):2877-80.
24. Mistry DA, Wang JY, Moeser M-E, Starkey T, Lee LYW. A systematic review of the sensitivity and specificity of lateral flow devices in the detection of SARS-CoV-2. *BMC Infectious Diseases*. 2021;21(1):828.
25. Buss SN, Leber A, Chapin K, Fey PD, Bankowski MJ, Jones MK, et al. Multicenter Evaluation of the BioFire FilmArray Gastrointestinal Panel for Etiologic Diagnosis of Infectious Gastroenteritis. *Journal of Clinical Microbiology*. 2015;53(3):915-25.
26. Leber AL, Everhart K, Daly JA, Hopper A, Harrington A, Schreckenberger P, et al. Multicenter Evaluation of BioFire FilmArray Respiratory Panel 2 for Detection of Viruses and Bacteria in Nasopharyngeal Swab Samples. *Journal of Clinical Microbiology*. 2018;56(6):10.1128/jcm.01945-17.
27. Tansarli GS, Chapin KC. Diagnostic test accuracy of the BioFire® FilmArray® meningitis/encephalitis panel: a systematic review and meta-analysis. *Clinical Microbiology and Infection*. 2020;26(3):281-90.
28. Edin A, Eilers H, Allard A. Evaluation of the Biofire Filmarray Pneumonia panel plus for lower respiratory tract infections. *Infectious Diseases*. 2020;52(7):479-88.
29. Berry GJ, Zhen W, Smith E, Manji R, Silbert S, Lima A, et al. Multicenter Evaluation of the BioFire Respiratory Panel 2.1 (RP2.1) for Detection of SARS-CoV-2 in Nasopharyngeal Swab Samples. *Journal of Clinical Microbiology*. 2022;60(5):e00066-22.
30. Creager HM, Cabrera B, Schnaubelt A, Cox JL, Cushman-Vokoun AM, Shakir SM, et al. Clinical evaluation of the BioFire® Respiratory Panel 2.1 and detection of SARS-CoV-2. *Journal of Clinical Virology*. 2020;129:104538.
31. England NHS. Improving the blood culture pathway. A national review of blood culture pathway processes to support better antimicrobial stewardship and improved patient safety.; 2023 8 March 2023. Contract No.: PAR686.
32. Standards Unit MS, PHE. UK Standards for Microbiology Investigations. Investigation of bronchoalveolar lavage, sputum and associated specimens. In: PHE, editor. 2019. p. 1-38.
33. Costa SP, Carvalho CM. Burden of bacterial bloodstream infections and recent advances for diagnosis. *Pathogens and Disease*. 2022;80(1).

34. R. John Looney RTS. Role of the Vi antigen of *Salmonella typhi* in resistance to host defense in vitro. *Translational Research*. 1986;108(5):498-505.
35. O'Toole GA. Classic Spotlight: How the Gram Stain Works. *Journal of Bacteriology*. 2016;198(23):3128-.
36. Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. *Clinical Microbiology and Infection*. 2014;20(4):O255-O66.
37. Seok H, Jeon JH, Park DW. Antimicrobial Therapy and Antimicrobial Stewardship in Sepsis. *Infect Chemother*. 2020;52(1):19-30.
38. Ahmed N, Chowdhary A. Comparison of different methods of detection of enteric pathogenic protozoa. *Indian Journal of Medical Microbiology*. 2013;31(2):154-60.
39. Saez AC, Manser MM, Andrews N, Chiodini PL. Comparison between the midi parasep and midi parasep solvent free (SF) faecal parasite concentrators. *Journal of clinical pathology*. 2011;64(10):901-4.
40. ABDEL AZIZ MO, ABD EL-RAMAN EM, EL-BAHAIE ES. Evaluation of Midi Parasep® faecal parasite concentrator for the detection of intestinal parasitic infections: a comparative study. *Journal of the Egyptian Society of Parasitology*. 2020;50(2):423-30.
41. Giemsa G. Eine Vereinfachung und Vervollkommnung meiner Methylenblau-Eosin-Färbemethode zur Erzielung der Romanowsky-Nocht'schen Chromatinfärbung. *Centralblatt für Bakteriologie*. 1904;32:307-13.
42. Opoku Afriyie S, Addison TK, Gebre Y, Mutala A-H, Antwi KB, Abbas DA, et al. Accuracy of diagnosis among clinical malaria patients: comparing microscopy, RDT and a highly sensitive quantitative PCR looking at the implications for submicroscopic infections. *Malaria Journal*. 2023;22(1):76.
43. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977;74(12):5463-7.
44. Chen F, Dong M, Ge M, Zhu L, Ren L, Liu G, et al. The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology. *Genomics, Proteomics & Bioinformatics*. 2013;11(1):34-40.
45. Pfeiffer F, Gröber C, Blank M, Händler K, Beyer M, Schultze JL, et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports*. 2018;8(1):10950.
46. He B, Zhu R, Yang H, Lu Q, Wang W, Song L, et al. Assessing the impact of data preprocessing on analyzing next generation sequencing data. *Frontiers in Bioengineering and Biotechnology*. 2020;8:817.
47. Sohn J-i, Nam J-W. The present and future of de novo whole-genome assembly. *Briefings in bioinformatics*. 2018;19(1):23-40.
48. Surrey LF, MacFarland SP, Chang F, Cao K, Rath KS, Akgumus GT, et al. Clinical utility of custom-designed NGS panel testing in pediatric tumors. *Genome Med*. 2019;11(1):32.
49. Luthra R, Patel KP, Routbort MJ, Broaddus RR, Yau J, Simien C, et al. A Targeted High-Throughput Next-Generation Sequencing Panel for Clinical Screening of Mutations, Gene Amplifications, and Fusions in Solid Tumors. *J Mol Diagn*. 2017;19(2):255-64.
50. Zhong Y, Xu F, Wu J, Schubert J, Li MM. Application of Next Generation Sequencing in Laboratory Medicine. *alm*. 2021;41(1):25-43.
51. Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, Ogilvy-Stuart AL, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant

Staphylococcus aureus: a descriptive study. *The Lancet Infectious Diseases*. 2013;13(2):130-6.

52. Patrick Van Roey PD, William J. Wolfgang, Ph.D., Lisa Mingle, Ph.D., Matthew Shudt, Kara Mitchell, Samantha Wirth, Michelle Dickinson. Next Generation Sequencing Implementation Guide. In: Laboratories AoPH, editor. 2016.
53. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*. 2016;17(1):239.
54. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*. 2014;14(6):1097-102.
55. Munro R, Santos R, Payne A, Forey T, Osei S, Holmes N, et al. minoTour, real-time monitoring and analysis for nanopore sequencers. *Bioinformatics*. 2021;38(4):1133-5.
56. Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Scientific Reports*. 2017;7(1):18022.
57. Goordial J, Altshuler I, Hindson K, Chan-Yam K, Marcoléfas E, Whyte LG. In Situ Field Sequencing and Life Detection in Remote (79°26'N) Canadian High Arctic Permafrost Ice Wedge Microbial Communities. *Frontiers in Microbiology*. 2017;8.
58. Zheng P, Zhou C, Ding Y, Liu B, Lu L, Zhu F, et al. Nanopore sequencing technology and its applications. *MedComm*. 2023;4(4):e316.
59. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019;20(6):341-55.
60. Solé C, Guilly S, Da Silva K, Llopis M, Le-Chatelier E, Huelin P, et al. Alterations in Gut Microbiome in Cirrhosis as Assessed by Quantitative Metagenomics: Relationship With Acute-on-Chronic Liver Failure and Prognosis. *Gastroenterology*. 2021;160(1):206-18.e13.
61. Cameron SJS, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, et al. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. *PLOS ONE*. 2017;12(5):e0177062.
62. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature Biotechnology*. 2019;37(7):783-92.
63. Moragues-Solanas L, Le-Viet T, McSorley E, Halford C, Lockhart DS, Aydin A, et al. Development and proof-of-concept demonstration of a clinical metagenomics method for the rapid detection of bloodstream infection. *BMC Medical Genomics*. 2024;17(1):71.
64. Royer C, Patin NV, Jesser KJ, Peña-Gonzalez A, Hatt JK, Trueba G, et al. Comparison of metagenomic and traditional methods for diagnosis of *E. coli* enteric infections. *mBio*. 2024;15(4):e03422-23.
65. Olausson J, Brunet S, Vracar D, Tian Y, Abrahamsson S, Meghadri SH, et al. Optimization of cerebrospinal fluid microbial DNA metagenomic sequencing diagnostics. *Sci Rep*. 2022;12(1):3378.
66. Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, Spangler J, Sims EH, et al. Rapid 16S rRNA Next-Generation Sequencing of Polymicrobial Clinical Samples for Diagnosis of Complex Bacterial Infections. *PLOS ONE*. 2013;8(5):e65226.
67. Wagner K, Springer B, Pires VP, Keller PM. Molecular detection of fungal pathogens in clinical specimens by 18S rDNA high-throughput screening in comparison to ITS PCR and culture. *Scientific Reports*. 2018;8(1):6964.

68. Wellinghausen N, Kochem A-J, Disqué C, Mühl H, Gebert S, Winter J, et al. Diagnosis of Bacteremia in Whole-Blood Samples by Use of a Commercial Universal 16S rRNA Gene-Based PCR and Sequence Analysis. *Journal of Clinical Microbiology*. 2009;47(9):2759-65.
69. Sipos R, Székely AJ, Palatinszky M, Révész S, Márialigeti K, Nikolausz M. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol Ecol*. 2007;60(2):341-50.
70. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*. 2017;35(9):833-44.
71. Grumaz S, Grumaz C, Vainshtein Y, Stevens P, Glanz K, Decker SO, et al. Enhanced Performance of Next-Generation Sequencing Diagnostics Compared With Standard of Care Microbiological Diagnostics in Patients Suffering From Septic Shock. *Critical Care Medicine*. 2019;47(5):e394-e402.
72. Charalampous T, Alcolea-Medina A, Snell LB, Alder C, Tan M, Williams TGS, et al. Routine Metagenomics Service for ICU Patients with Respiratory Infection. *American Journal of Respiratory and Critical Care Medicine*. 2024;209(2):164-74.
73. Kujiraoka M, Kuroda M, Asai K, Sekizuka T, Kato K, Watanabe M, et al. Comprehensive Diagnosis of Bacterial Infection Associated with Acute Cholecystitis Using Metagenomic Approach. *Frontiers in Microbiology*. 2017;8(685).
74. Hernando-Amado S, Coque TM, Baquero F, Martínez JL. Defining and combating antibiotic resistance from One Health and Global Health perspectives. *Nature Microbiology*. 2019;4(9):1432-42.
75. Ahmad M, Khan AU. Global economic impact of antibiotic resistance: A review. *Journal of Global Antimicrobial Resistance*. 2019;19:313-6.
76. Antimicrobial Resistance Division (AMR) NAPaMaEN. Global action plan on antimicrobial resistance. 2016. Contract No.: ISBN: 9789241509763.
77. Organization WH. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug resistant bacterial infections, including tuberculosis. . World Health Organization; 2017. Contract No.: WHO/EMP/IAU/2017.12.
78. Street TL, Sanderson ND, Kolenda C, Kavanagh J, Pickford H, Hoosdally S, et al. Clinical Metagenomic Sequencing for Species Identification and Antimicrobial Resistance Prediction in Orthopedic Device Infection. *Journal of Clinical Microbiology*. 2022;60(4):e02156-21.
79. de Nies L, Lopes S, Busi SB, Galata V, Heintz-Buschart A, Laczny CC, et al. PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. *Microbiome*. 2021;9(1):49.
80. Schubert AM, Rogers MAM, Ring C, Mogle J, Petrosino JP, Young VB, et al. Microbiome Data Distinguish Patients with *Clostridium difficile* Infection and Non-*C. difficile*-Associated Diarrhea from Healthy Controls. *mBio*. 2014;5(3):10.1128/mbio.01021-14.
81. Khanna S, Pardi DS, Kelly CR, Kraft CS, Dhere T, Henn MR, et al. A Novel Microbiome Therapeutic Increases Gut Microbial Diversity and Prevents Recurrent *Clostridium difficile* Infection. *The Journal of Infectious Diseases*. 2016;214(2):173-81.
82. Grumaz S, Stevens P, Grumaz C, Decker SO, Weigand MA, Hofer S, et al. Next-generation sequencing diagnostics of bacteremia in septic patients. *Genome Medicine*. 2016;8(1):73.

83. Choutko V, Lazarevic V, Gaïa N, Girard M, Renzi G, Leo S, et al. Rare Case of Community-Acquired Endocarditis Caused by *Neisseria meningitidis* Assessed by Clinical Metagenomics. *Frontiers in Cardiovascular Medicine*. 2019;6(112).
84. Ji X-C, Zhou L-F, Li C-Y, Shi Y-J, Wu M-L, Zhang Y, et al. Reduction of Human DNA Contamination in Clinical Cerebrospinal Fluid Specimens Improves the Sensitivity of Metagenomic Next-Generation Sequencing. *Journal of Molecular Neuroscience*. 2020;70(5):659-66.
85. Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*. 2011;480(7376):241-4.
86. Lallement C, Pasternak C, Ploy M-C, Jové T. The role of IS CR1-borne POUT promoters in the expression of antibiotic resistance genes. *Frontiers in microbiology*. 2018;9:2579.
87. Kumar R, Yadav B, Singh R. Genetic determinants of antibiotic resistance in *Staphylococcus aureus* isolates from milk of mastitic crossbred cattle. *Current microbiology*. 2010;60:379-86.
88. Deekshit VK, Srikumar S. 'To be, or not to be'—The dilemma of 'silent' antimicrobial resistance genes in bacteria. *Journal of Applied Microbiology*. 2022;133(5):2902-14.
89. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Research*. 2019;29(5):831-42.
90. Murray CK, Yun HC, Markelz AE, Okulicz JF, Vento TJ, Burgess TH, et al. Operation United Assistance: Infectious Disease Threats to Deployed Military Personnel. *Military Medicine*. 2015;180(6):626-51.
91. David Keast DA, Dot Weir, Emily Haesler, Karen Ousey, Keryln Carville, Nicola Waters, Patricia Idensohn, Terry Swanson, Thomas Bjarnsholt. International Wound Infection Institute (IWII) Wound Infection in Clinical Practice. *Wounds International*. 2022.
92. Marchukov D, Li J, Juillerat P, Misselwitz B, Yilmaz B. Benchmarking microbial DNA enrichment protocols from human intestinal biopsies. *Frontiers in Genetics*. 2023;14.
93. Hasan MR, Rawat A, Tang P, Jithesh PV, Thomas E, Tan R, et al. Depletion of Human DNA in Spiked Clinical Specimens for Improvement of Sensitivity of Pathogen Detection by Next-Generation Sequencing. *Journal of Clinical Microbiology*. 2016;54(4):919-27.
94. Baumann E, Stoya G, Völkner A, Richter W, Lemke C, Linss W. Hemolysis of human erythrocytes with saponin affects the membrane structure. *Acta Histochemica*. 2000;102(1):21-35.
95. Bruggeling CE, Garza DR, Achouiti S, Mes W, Dutilh BE, Boleij A. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *MicrobiologyOpen*. 2021;10(3):e1191.
96. Charalampous T. Development and application of clinical metagenomics for the diagnosis and characterisation of lower respiratory infections. <https://ueaeprints.uea.ac.uk/id/eprint/80196/>: University of East Anglia; 2020.
97. Farraj SA, El-Kafrawy SA, Kumosani TA, Yousef JM, Azhar EI. Evaluation of Extraction Methods for Clinical Metagenomic Assay. *Microorganisms*. 2020;8(8):1128.
98. Dahm R. Friedrich Miescher and the discovery of DNA. *Developmental Biology*. 2005;278(2):274-88.

99. Miescher F. Letter i; to wilhelm his; tübingen, february 26th, 1869. Die histochemischen und physiologischen arbeiten von Friedrich Miescher-aus dem wissenschaftlichen Briefwechsel von F Miescher. 1869;1:33-8.
100. Franklin RE, Gosling RG. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*. 1953;172(4369):156-7.
101. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*. 1953;171(4356):737-8.
102. Auer GK, Weibel DB. Bacterial Cell Mechanics. *Biochemistry*. 2017;56(29):3710-24.
103. Gow NAR, Latge J-P, Munro CA. The Fungal Cell Wall: Structure, Biosynthesis, and Function. *Microbiology Spectrum*. 2017;5(3):10.1128/microbiolspec.funk-0035-2016.
104. Zhang B, Brock M, Arana C, Dende C, van Oers NS, Hooper LV, et al. Impact of Bead-Beating Intensity on the Genus- and Species-Level Characterization of the Gut Microbiome Using Amplicon and Complete 16S rRNA Gene Sequencing. *Frontiers in Cellular and Infection Microbiology*. 2021;11.
105. Lim MY, Song E-J, Kim SH, Lee J, Nam Y-D. Comparison of DNA extraction methods for human gut microbial community profiling. *Systematic and Applied Microbiology*. 2018;41(2):151-7.
106. Ciulla TA, Sklar RM, Hauser SL. A simple method for DNA purification from peripheral blood. *Analytical Biochemistry*. 1988;174(2):485-8.
107. Kravchenko AV, Chetverina EV, Chetverin AB. Preservation of nucleic acid integrity in guanidine thiocyanate lysates of whole blood. *Russian Journal of Bioorganic Chemistry*. 2006;32(6):547-51.
108. Ganz T. LYSOZYME. In: Laurent GJ, Shapiro SD, editors. *Encyclopedia of Respiratory Medicine*. Oxford: Academic Press; 2006. p. 649-53.
109. Li X, Bosch-Tijhof CJ, Wei X, de Soet JJ, Crielaard W, Loveren Cv, et al. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. *Journal of International Medical Research*. 2020;48(5):0300060520925594.
110. Tian H, Hühmer AF, Landers JP. Evaluation of silica resins for direct and efficient extraction of DNA from complex biological matrices in a miniaturized format. *Analytical biochemistry*. 2000;283(2):175-91.
111. J Shetty P. The Evolution of DNA Extraction Methods. *American Journal of Biomedical Science & Research*. 2020;8(1):39-45.
112. Wang X, Liu Y, Liu H, Pan W, Ren J, Zheng X, et al. Recent advances and application of whole genome amplification in molecular diagnosis and medicine. *MedComm*. 2022;3(1):e116.
113. Moghaddaszadeh-Ahrabi S, Farajnia S, Rahimi-Mianji G, Nejati-Javaremi A. A short and simple improved-primer extension preamplification (I-PEP) procedure for whole genome amplification (WGA) of bovine cells. *Anim Biotechnol*. 2012;23(1):24-42.
114. Lasken Roger S. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochemical Society Transactions*. 2009;37(2):450-3.
115. Sabina J, Leamon JH. Bias in whole genome amplification: causes and considerations. *Whole genome amplification: Methods and protocols*. 2015:15-41.
116. Lasken RS, Stockwell TB. Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC biotechnology*. 2007;7:1-11.

117. Dadzie FA, Beaudry MS, Deyanov A, Slanis H, Duong MQ, Turner R, et al. Evaluating the benefits and limits of multiple displacement amplification with whole-genome Oxford Nanopore Sequencing. *bioRxiv*. 2024.
118. Street TL, Barker L, Sanderson ND, Kavanagh J, Hoosdally S, Cole K, et al. Optimizing DNA Extraction Methods for Nanopore Sequencing of *Neisseria gonorrhoeae* Directly from Urine Samples. *Journal of Clinical Microbiology*. 2020;58(3):10.1128/jcm.01822-19.
119. Tyler AD, McAllister J, Stapleton H, Gauci P, Antonation K, Thirkettle-Watts D, et al. Field-based detection of bacteria using nanopore sequencing: Method evaluation for biothreat detection in complex samples. *PLOS ONE*. 2023;18(11):e0295028.
120. de Vries JJC, Brown JR, Fischer N, Sidorov IA, Morfopoulou S, Huang J, et al. Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples. *Journal of Clinical Virology*. 2021;141:104908.
121. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-100.
122. Kalikar S, Jain C, Vasimuddin M, Misra S. Accelerating minimap2 for long-read sequencing applications on modern CPUs. *Nature Computational Science*. 2022;2(2):78-83.
123. Rigden DJ, Fernández XM. The 2023 Nucleic Acids Research Database Issue and the online molecular biology database collection. *Nucleic Acids Research*. 2023;51(D1):D1-D8.
124. Du N, Chen J, Sun Y. Improving the sensitivity of long read overlap detection using grouped short k-mer matches. *BMC Genomics*. 2019;20(2):190.
125. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome research*. 2016;26(12):1721-9.
126. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder Tool and Resistance Gene Database by Using Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates. *Antimicrobial Agents and Chemotherapy*. 2019;63(11):10.1128/aac.00483-19.
127. Sanderson ND, Swann J, Barker L, Kavanagh J, Hoosdally S, Crook D, et al. High precision *Neisseria gonorrhoeae* variant and antimicrobial resistance calling from metagenomic Nanopore sequencing. *Genome research*. 2020;30(9):1354-63.
128. Bevan ER, Jones AM, Hawkey PM. Global epidemiology of CTX-M β -lactamases: temporal and geographical shifts in genotype. *Journal of Antimicrobial Chemotherapy*. 2017;72(8):2145-55.
129. Gordon N, Price J, Cole K, Everitt R, Morgan M, Finney J, et al. Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *Journal of clinical microbiology*. 2014;52(4):1182-91.
130. Ruppé E, d'Humières C, Armand-Lefèvre L. Inferring antibiotic susceptibility from metagenomic data: dream or reality? *Clinical Microbiology and Infection*. 2022;28(9):1225-9.
131. Van Camp P-J, Haslam DB, Porollo A. Prediction of antimicrobial resistance in Gram-negative bacteria from whole-genome sequencing data. *Frontiers in microbiology*. 2020;11:530987.
132. Gardner SE, Frantz RA, Saltzman CL, Hillis SL, Park H, Scherubel M. Diagnostic validity of three swab techniques for identifying chronic wound infection. *Wound Repair and Regeneration*. 2006;14(5):548-57.

133. Nurk S, Koren S, Rhie A, Rautiainen M, Bizikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science*. 2022;376(6588):44-53.
134. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics*. 2018;19(1):307.
135. Penders J, Vink C, Driessen C, London N, Thijs C, Stobberingh EE. Quantification of *Bifidobacterium* spp., *Escherichia coli* and *Clostridium difficile* in faecal samples of breast-fed and formula-fed infants by real-time PCR. *FEMS Microbiology Letters*. 2005;243(1):141-7.
136. Roschanski N, Fischer J, Guerra B, Roesler U. Development of a multiplex real-time PCR for the rapid detection of the predominant beta-lactamase genes CTX-M, SHV, TEM and CIT-type AmpCs in Enterobacteriaceae. *PLoS One*. 2014;9(7):e100956.
137. Wang HY, Kim S, Kim J, Park SD, Uh Y, Lee H. Multiplex real-time PCR assay for rapid detection of methicillin-resistant staphylococci directly from positive blood cultures. *J Clin Microbiol*. 2014;52(6):1911-20.
138. McConnell MJ, Perez-Ordóñez A, Perez-Romero P, Valencia R, Lepe JA, Vazquez-Barba I, et al. Quantitative real-time PCR for detection of *Acinetobacter baumannii* colonization in the hospital environment. *J Clin Microbiol*. 2012;50(4):1412-4.
139. He YH, Ruan GJ, Hao H, Xue F, Ma YK, Zhu SN, et al. Real-time PCR for the rapid detection of *vanA*, *vanB* and *vanM* genes. *J Microbiol Immunol Infect*. 2020;53(5):746-50.
140. Wood C, Sahl J, Maltinsky S, Coyne B, Russakoff B, Yagüe DP, et al. SaQuant: a real-time PCR assay for quantitative assessment of *Staphylococcus aureus*. *BMC Microbiology*. 2021;21(1):174.
141. Köppel R, Ruf J, Rentsch J. Multiplex real-time PCR for the detection and quantification of DNA from beef, pork, horse and sheep. *European Food Research and Technology*. 2011;232(1):151-5.
142. Herrera LJ, Raja S, Gooding WE, El-Hefnawy T, Kelly L, Luketich JD, et al. Quantitative Analysis of Circulating Plasma DNA as a Tumor Marker in Thoracic Malignancies. *Clinical Chemistry*. 2005;51(1):113-8.
143. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*. 2018;6(1):42.
144. Buckwalter SP, Sloan LM, Cunningham SA, Espy MJ, Uhl JR, Jones MF, et al. Inhibition controls for qualitative real-time PCR assays: are they necessary for all specimen matrices? *J Clin Microbiol*. 2014;52(6):2139-43.
145. de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P. Improved detection of microbial DNA after bead-beating before DNA isolation. *Journal of Microbiological Methods*. 2010;80(2):209-11.
146. Scharf S, Bartels A, Kondakci M, Pfeffer K, Henrich B, Haas R. Introduction of a bead beating step improves fungal DNA extraction from selected patient specimens. *International Journal of Medical Microbiology*. 2020;310(6):151443.
147. Cantera JL, White H, Diaz MH, Beall SG, Winchell JM, Lillis L, et al. Assessment of eight nucleic acid amplification technologies for potential use to detect infectious agents in low-resource settings. *PLOS ONE*. 2019;14(4):e0215756.
148. Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquarson FM, Louis P, et al. The Impact of Different DNA Extraction Kits and Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing. *PLOS ONE*. 2014;9(2):e88982.

149. AbuOun M, Navickaite I, Wilkes TI, Di Domenico M, Garofolo G, Bloemen B, et al. Report on initial findings and protocols for on-site diagnostic tests and IT solutions for human clinical, veterinary and environmental samples for early warning of emerging resistant pathogens: Deliverable D-JRP12-WP 3.2, 3.3, 3.4 & 3.5. One Health European Joint Programme (OHEJP); 2022.
150. Li R, Xie M, Dong N, Lin D, Yang X, Wong MHY, et al. Efficient generation of complete sequences of MDR-encoding plasmids by rapid assembly of MinION barcoding sequencing data. *Gigascience*. 2018;7(3):1-9.
151. Zinter MS, Mayday MY, Ryckman KK, Jelliffe-Pawlowski LL, DeRisi JL. Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome*. 2019;7(1):62.
152. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Yao J, Chia N, Hanssen AD, et al. Impact of Contaminating DNA in Whole-Genome Amplification Kits Used for Metagenomic Shotgun Sequencing for Infection Diagnosis. *J Clin Microbiol*. 2017;55(6):1789-801.
153. Direito SO, Zaura E, Little M, Ehrenfreund P, Roling WF. Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. *Environ Microbiol*. 2014;16(3):643-57.
154. Theunissen GMG, Rolf B, Gibb A, Jäger R. DNA profiling of sperm cells by using micromanipulation and whole genome amplification. *Forensic Science International: Genetics Supplement Series*. 2017;6:e497-e9.
155. Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Yao J, Chia N, Hanssen AD, et al. Impact of contaminating DNA in whole-genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis. *Journal of clinical microbiology*. 2017;55(6):1789-801.
156. Wurzer P, Guillory A, Parvizi D, Clayton RP, Branski LK, Kamolz L-P, et al. Human herpes viruses in burn patients: A systematic review. *Burns*. 2017;43(1):25-33.
157. Chrzastek K, Lee D-h, Smith D, Sharma P, Suarez DL, Pantin-Jackwood M, et al. Use of Sequence-Independent, Single-Primer-Amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses. *Virology*. 2017;509:159-66.
158. Kapel N, Kalimeris E, Lumley S, Decano A, Rodger G, Lopes Alves M, et al. Evaluation of sequence hybridization for respiratory viruses using the Twist Bioscience Respiratory Virus Research panel and the OneCodex Respiratory Virus sequence analysis workflow. *Microbial Genomics*. 2023;9(9).
159. Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, et al. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *Journal of Virological Methods*. 2015;213:139-46.
160. Brooks JP, Edwards DJ, Harwich MD, Rivera MC, Fettweis JM, Serrano MG, et al. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*. 2015;15(1):66.
161. Liu J, Chen D, Peters BM, Li L, Li B, Xu Z, et al. Staphylococcal chromosomal cassettes mec (SCCmec): A mobile genetic element in methicillin-resistant *Staphylococcus aureus*. *Microbial Pathogenesis*. 2016;101:56-67.
162. Woodford N, Ward ME, Kaufmann ME, Turton J, Fagan EJ, James D, et al. Community and hospital spread of *Escherichia coli* producing CTX-M extended-spectrum β -lactamases in the UK. *Journal of Antimicrobial Chemotherapy*. 2004;54(4):735-43.
163. Terry Swanson SWH, Warrnambool, Victoria (Australia), Donna Angel RPH, Perth (Australia), Geoff Sussman MU, Melbourne (Australia), Rose Cooper CMU, Cardiff (UK),

Emily Haesler CUaANU, Canberra (Australia), Karen Ousey IoSlaiP, Huddersfield (UK), et al. International Wound Infection Institute (IWII) Wound infection in clinical practice. Wounds International 2016. 2016.

164. Bessa LJ, Fazii P, Di Giulio M, Cellini L. Bacterial isolates from infected wounds and their antibiotic susceptibility pattern: some remarks about wound infection. *Int Wound J*. 2015;12(1):47-52.

165. Bowler PG, Duerden BI, Armstrong DG. Wound microbiology and associated approaches to wound management. *Clin Microbiol Rev*. 2001;14(2):244-69.

166. Raoofi S, Pashazadeh Kan F, Rafiei S, Hosseinipalangi Z, Noorani Mejareh Z, Khani S, et al. Global prevalence of nosocomial infection: A systematic review and meta-analysis. *PLOS ONE*. 2023;18(1):e0274248.

167. Esmail MAM, Abdulghany HM, Khairy RM. Prevalence of Multidrug-Resistant *Enterococcus faecalis* in Hospital-Acquired Surgical Wound Infections and Bacteremia: Concomitant Analysis of Antimicrobial Resistance Genes. *Infectious Diseases: Research and Treatment*. 2019;12:1178633719882929.

168. Xu L, McLennan SV, Lo L, Natfaji A, Bolton T, Liu Y, et al. Bacterial Load Predicts Healing Rate in Neuropathic Diabetic Foot Ulcers. *Diabetes Care*. 2007;30(2):378-80.

169. Sears ED, Wu L, Waljee JF, Momoh AO, Zhong L, Chung KC. The Impact of Deep Sternal Wound Infection on Mortality and Resource Utilization: A Population-based Study. *World Journal of Surgery*. 2016;40(11):2673-80.

170. Church D, Elsayed S, Reid O, Winston B, Lindsay R. Burn wound infections. *Clin Microbiol Rev*. 2006;19(2):403-34.

171. Penel N, Lefebvre JL, Cazin JL, Clisant S, Neu JC, Dervaux B, et al. Additional direct medical costs associated with nosocomial infections after head and neck cancer surgery: a hospital-perspective analysis. *International Journal of Oral and Maxillofacial Surgery*. 2008;37(2):135-9.

172. C. E, A. C, C. B, G. MC. Early infection after hip fracture surgery. *The Journal of Bone and Joint Surgery British volume*. 2008;90-B(6):770-7.

173. Goldberg MS. Death and Injury Rates of U.S. Military Personnel in Iraq *MILITARY MEDICINE*. 2010;175.

174. Belmont PJ, Owens BD, Schoenfeld AJ. Musculoskeletal Injuries in Iraq and Afghanistan: Epidemiology and Outcomes Following a Decade of War. *JAAOS - Journal of the American Academy of Orthopaedic Surgeons*. 2016;24(6):341-8.

175. Owens BD, Kragh JFJ, Wenke JC, Macaitis J, Wade CE, Holcomb JB. Combat Wounds in Operation Iraqi Freedom and Operation Enduring Freedom. *Journal of Trauma and Acute Care Surgery*. 2008;64(2):295-9.

176. Breeze J, Allanson-Bailey LS, Hepper AE, Midwinter MJ. Demonstrating the effectiveness of body armour: a pilot prospective computerised surface wound mapping trial performed at the Role 3 hospital in Afghanistan. *Journal of the Royal Army Medical Corps*. 2015;161(1):36-41.

177. Staruch RMT, Hettiaratchy S. Warzone trauma and surgical infections. *Surgery (Oxford)*. 2019;37(1):58-63.

178. Stewart L, Shaikh F, Bradley W, Lu D, Blyth DM, Petfield JL, et al. Combat-Related Extremity Wounds: Injury Factors Predicting Early Onset Infections. *Military Medicine*. 2019;184(Supplement_1):83-91.

179. Eardley WGP, Brown KV, Bonner TJ, Green AD, Clasper JC. Infection in conflict wounded. *Philos Trans R Soc Lond B Biol Sci*. 2011;366(1562):204-18.

180. Mende K, Stewart L, Shaikh F, Bradley W, Lu D, Krauss MR, et al. Microbiology of combat-related extremity wounds: Trauma Infectious Disease Outcomes Study. *Diagnostic Microbiology and Infectious Disease*. 2019;94(2):173-9.
181. Soderstrom MA, Blyth DM, Carson L, Campbell WR, Yabes J, Shaikh F, et al. 1388. Seasonality of Microbiology of Combat-related Wound Infections in Afghanistan. *Open Forum Infectious Diseases*. 2022;9(Supplement_2).
182. Rodriguez CJ, Ganesan A, Shaikh F, Carson ML, Bradley W, Warkentien TE, et al. Combat-Related Invasive Fungal Wound Infections. *Military Medicine*. 2022;187(Supplement_2):34-41.
183. Bielka K, Kuchyn I, Horoshko V. Intensive care units during the Ukraine war: challenges and opportunities. *Intensive Care Medicine*. 2023.
184. PHE. UK Standards for Microbiology Investigations: Investigation of swabs from skin and superficial soft tissue infections. In: Standards Unit MS, Public Health England, editor. 2018.
185. Ludwik K, Branski AA-M, Haidy Rivero, Marc G. Jeschke, Arthur P. Sanford, and David N. Herndon. Emerging Infections in Burns. *Surgical Infections*. 2009;10(5):389-97.
186. Revathi G, Puri J, Jain BK. Bacteriology of burns. *Burns*. 1998;24(4):347-9.
187. Stevens DL, Bisno AL, Chambers HF, Everett ED, Dellinger P, Goldstein EJC, et al. Practice Guidelines for the Diagnosis and Management of Skin and Soft-Tissue Infections. *Clinical Infectious Diseases*. 2005;41(10):1373-406.
188. Puca V, Marulli RZ, Grande R, Vitale I, Niro A, Molinaro G, et al. Microbial Species Isolated from Infected Wounds and Antimicrobial Resistance Analysis: Data Emerging from a Three-Years Retrospective Study. *Antibiotics*. 2021;10(10):1162.
189. Emerging Infections in Burns. *Surgical Infections*. 2009;10(5):389-97.
190. Ressler RA, Murray CK, Griffith ME, Rasnake MS, Hospenthal DR, Wolf SE. Outcomes of Bacteremia in Burn Patients Involved in Combat Operations Overseas. *Journal of the American College of Surgeons*. 2008;206(3):439-44.
191. Ramakrishnan M, Putli Bai S, Babu M. Study on biofilm formation in burn wound infection in a pediatric hospital in Chennai, India. *Ann Burns Fire Disasters*. 2016;29(4):276-80.
192. Defez C, Fabbro-Peray P, Cazaban M, Boudemaghe T, Sotto A, Daurès JP. Additional direct medical costs of nosocomial infections: an estimation from a cohort of patients in a French university hospital. *Journal of Hospital Infection*. 2008;68(2):130-6.
193. Maslova E, Eisaiankhong L, Sjöberg F, McCarthy RR. Burns and biofilms: priority pathogens and in vivo models. *npj Biofilms and Microbiomes*. 2021;7(1):73.
194. Mende K, Akers KS, Tyner SD, Bennett JW, Simons MP, Blyth DM, et al. Multidrug-Resistant and Virulent Organisms Trauma Infections: Trauma Infectious Disease Outcomes Study Initiative. *Mil Med*. 2022;187(Suppl 2):42-51.
195. Li S, Renick P, Senkowsky J, Nair A, Tang L. Diagnostics for Wound Infections. *Adv Wound Care (New Rochelle)*. 2021;10(6):317-27.
196. L.R. Copeland-HalperinMD GSR, A.J. KaminskyMD MPH, General Surgery Resident, N. BluefeldRN CWOCN, R. MiraliakbariMD BCPS, Associate Professor. Sample procurement for cultures of infected wounds: a systematic review. *Journal of Wound Care*. 2016;25(Sup4):S4-S10.
197. (PHE) PHE. UK Standards for Microbiology Investigations: Identification of *Staphylococcus* species, *Micrococcus* species and *Rothia* species. In: (PHE) PHE, editor. 2020.

198. Rajakaruna L, Hallas G, Molenaar L, Dare D, Sutton H, Encheva V, et al. High throughput identification of clinical isolates of *Staphylococcus aureus* using MALDI-TOF-MS of intact cells. *Infection, Genetics and Evolution*. 2009;9(4):507-13.
199. Parta M, Goebel M, Matloobi M, Stager C, Musher DM. Identification of Methicillin-Resistant or Methicillin-Susceptible *Staphylococcus aureus* in Blood Cultures and Wound Swabs by GeneXpert. *Journal of Clinical Microbiology*. 2009;47(5):1609-10.
200. Ellis S, Lin EJ, Tartar D. Immunology of Wound Healing. *Current Dermatology Reports*. 2018;7(4):350-8.
201. Sproston NR, Ashworth JJ. Role of C-Reactive Protein at Sites of Inflammation and Infection. *Frontiers in Immunology*. 2018;9(754).
202. Wu T-L, Tsao K-C, Chang CPY, Li C-N, Sun C-F, Wu JT. Development of ELISA on microplate for serum C-reactive protein and establishment of age-dependent normal reference range. *Clinica Chimica Acta*. 2002;322(1):163-8.
203. Andrew Kingsley VJ. Diagnosing wound infection: the use of C-reactive protein. *Wounds*. 2008;4(4):32-46.
204. Wang W, Mai Z, Chen Y, Wang J, Li L, Su Q, et al. A label-free fiber optic SPR biosensor for specific detection of C-reactive protein. *Scientific Reports*. 2017;7(1):16904.
205. Lam Le MB, Patrick Briggs, Neal Bullock, Windy Cole, Daniel DiMarco, Rachel Hamil, Khristina Harrell, Maria Kasper, Weili Li, Keyur Patel, Matthew Sabo, Kerry Thibodeaux, and Thomas E. Serena. Diagnostic Accuracy of Point-of-Care Fluorescence Imaging for the Detection of Bacterial Burden in Wounds: Results from the 350-Patient Fluorescence Imaging Assessment and Guidance Trial. *Advances in Wound Care*. 2020;0(0):null.
206. Daulton E, Wicaksono A, Bechar J, Covington JA, Hardwicke J. The Detection of Wound Infection by Ion Mobility Chemical Analysis. *Biosensors*. 2020;10(3):19.
207. Qiao B, Pang Q, Yuan P, Luo Y, Ma L. Smart wound dressing for infection monitoring and NIR-triggered antibacterial treatment. *Biomaterials Science*. 2020;8(6):1649-57.
208. Bradshaw CE. An in vitro comparison of the antimicrobial activity of honey, iodine and silver wound dressings. *Bioscience Horizons: The International Journal of Student Research*. 2011;4(1):61-70.
209. NHS. Wound Infection Guidance 2023 [
210. Gardner SE, Hillis SL, Frantz RA. Clinical Signs of Infection in Diabetic Foot Ulcers With High Microbial Load. *Biological Research For Nursing*. 2009;11(2):119-28.
211. Chakraborti C, Le C, Yanofsky A. Sensitivity of superficial cultures in lower extremity wounds. *J Hosp Med*. 2010;5(7):415-20.
212. Johnson AC, Buchanan EP, Khechoyan DY. Wound infection: A review of qualitative and quantitative assessment modalities. *Journal of Plastic, Reconstructive & Aesthetic Surgery*. 2022;75(4):1287-96.
213. Zhao-Fleming HH, Wilkinson JE, Larumbe E, Dissanaik S, Rumbaugh K. Obligate anaerobes are abundant in human necrotizing soft tissue infection samples – a metagenomics analysis. *APMIS*. 2019;127(8):577-87.
214. Farmer AR, Murray CK, Driscoll IR, Wickes BL, Wiederhold N, Sutton DA, et al. Combat-Related *Pythium aphanidermatum* Invasive Wound Infection: Case Report and Discussion of Utility of Molecular Diagnostics. *J Clin Microbiol*. 2015;53(6):1968-75.

215. Wang Q, Miao Q, Pan J, Jin W, Ma Y, Zhang Y, et al. The clinical value of metagenomic next-generation sequencing in the microbiological diagnosis of skin and soft tissue infections. *International Journal of Infectious Diseases*. 2020;100:414-20.
216. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya Amogelang R, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*. 2022;51(D1):D690-D9.
217. Reimer LC, Sardà Carbasse J, Koblitz J, Ebeling C, Podstawka A, Overmann J. BacDive in 2022: the knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Research*. 2021;50(D1):D741-D6.
218. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health*. 2017;5:307.
219. Feranchuk S, Belkova N, Potapova U, Kuzmin D, Belikov S. Evaluating the use of diversity indices to distinguish between microbial communities with different traits. *Research in Microbiology*. 2018;169(4):254-61.
220. Blanchard AC, Rooney AM, Yau Y, Zhang Y, Stapleton PJ, Horton E, et al. Early detection using qPCR of *Pseudomonas aeruginosa* infection in children with cystic fibrosis undergoing eradication treatment. *Journal of Cystic Fibrosis*. 2018;17(6):723-8.
221. Dunne EM, Marshall JL, Baker CA, Manning J, Gonis G, Danchin MH, et al. Detection of group a streptococcal pharyngitis by quantitative PCR. *BMC Infectious Diseases*. 2013;13(1):312.
222. Lim KT, Hanifah YA, Yusof MYM, Thong KL. *ermA*, *ermC*, *tetM* and *tetK* are essential for erythromycin and tetracycline resistance among methicillin-resistant *Staphylococcus aureus* strains isolated from a tertiary hospital in Malaysia. *Indian Journal of Medical Microbiology*. 2012;30(2):203-7.
223. Charalampous T, Alcolea-Medina A, Snell LB, Alder C, Tan M, Williams TGS, et al. Routine Metagenomics Service for ICU Patients with Respiratory Infection. *American Journal of Respiratory and Critical Care Medicine*. 2023;209(2):164-74.
224. Jing Q, Leung CHC, Wu AR. Cell-free DNA as biomarker for sepsis by integration of microbial and host information. *Clinical Chemistry*. 2022;68(9):1184-95.
225. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *New England Journal of Medicine*. 2019;380(24):2327-40.
226. Gardner SE, Hillis SL, Heilmann K, Segre JA, Grice EA. The Neuropathic Diabetic Foot Ulcer Microbiome Is Associated With Clinical Factors. *Diabetes*. 2013;62(3):923-30.
227. Price LB, Liu CM, Melendez JH, Frankel YM, Engelthaler D, Aziz M, et al. Community Analysis of Chronic Wound Bacteria Using 16S rRNA Gene-Based Pyrosequencing: Impact of Diabetes and Antibiotics on Chronic Wound Microbiota. *PLOS ONE*. 2009;4(7):e6462.
228. Misic AM, Gardner SE, Grice EA. The Wound Microbiome: Modern Approaches to Examining the Role of Microorganisms in Impaired Chronic Wound Healing. *Adv Wound Care (New Rochelle)*. 2014;3(7):502-10.
229. Schwartz DJ, Langdon AE, Dantas G. Understanding the impact of antibiotic perturbation on the human microbiome. *Genome medicine*. 2020;12:1-12.
230. Charalampous T, Alcolea-Medina A, Snell LB, Alder C, Tan M, Williams TGS, et al. Routine Metagenomics Service for ICU Patients with Respiratory Infection. *Am J Respir Crit Care Med*. 2024;209(2):164-74.

231. Pembaur A, Sallard E, Weil PP, Ortelt J, Ahmad-Nejad P, Postberg J. Simplified Point-of-Care Full SARS-CoV-2 Genome Sequencing Using Nanopore Technology. *Microorganisms*. 2021;9(12):2598.
232. Holzschuh A, Lerch A, Fakih BS, Aliy SM, Ali MH, Ali MA, et al. Using a mobile nanopore sequencing lab for end-to-end genomic surveillance of *Plasmodium falciparum*: A feasibility study. *PLOS Global Public Health*. 2024;4(2):e0002743.
233. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature*. 2016;530(7589):228-32.
234. DUPONT HL. Systematic review: the epidemiology and clinical features of travellers' diarrhoea. *Alimentary Pharmacology & Therapeutics*. 2009;30(3):187-96.
235. Glaesser D, Kester J, Paulose H, Alizadeh A, Valentin B. Global travel patterns: an overview. *Journal of Travel Medicine*. 2017;24(4).
236. Leung KCA, Leung AMA, Wong HCA, Hon LK. Travelers' Diarrhea: A Clinical Review. *Recent Patents on Inflammation & Allergy Drug Discovery*. 2019;13(1):38-48.
237. Behrens RH, Cramer JP, Jelinek T, Shaw H, von Sonnenburg F, Wilbraham D, et al. Efficacy and safety of a patch vaccine containing heat-labile toxin from *Escherichia coli* against travellers' diarrhoea: a phase 3, randomised, double-blind, placebo-controlled field trial in travellers from Europe to Mexico and Guatemala. *The Lancet infectious diseases*. 2014;14(3):197-204.
238. Cartwright R. Food and waterborne infections associated with package holidays. *Journal of Applied Microbiology*. 2003;94(s1):12-24.
239. Jagai JS, Smith GS, Schmid JE, Wade TJ. Trends in gastroenteritis-associated mortality in the United States, 1985–2005: variations by ICD-9 and ICD-10 codes. *BMC gastroenterology*. 2014;14:1-8.
240. Bayne-Jones S. The evolution of preventive medicine in the United States Army, 1607-1939: Office of the Surgeon General, Department of the Army; 1968.
241. Putnam SD, Sanders JW, French RW, Monteville M, Riddle MS, Rockabrand DM, et al. Self-Reported Description of Diarrhea Among Military Populations in Operations Iraqi Freedom and Enduring Freedom. *Journal of Travel Medicine*. 2006;13(2):92-9.
242. Deb B, Sircar B, Sengupta P, De S, Mondal S, Gupta D, et al. Studies on interventions to prevent eltor cholera transmission in urban slums. *Bulletin of the World Health Organization*. 1986;64(1):127.
243. Evans DP. Non-pharmacotherapeutic interventions in travellers diarrhoea (TD). *Journal of Travel Medicine*. 2018;25(suppl_1):S38-S45.
244. Bwanaisa LL, Heyderman RS, Molyneux EM. The challenges of managing severe dehydrating diarrhoea in a resource-limited setting. *International Health*. 2011;3(3):147-53.
245. Butler T. Loperamide for the Treatment of Traveler's Diarrhea: Broad or Narrow Usefulness? *Clinical Infectious Diseases*. 2008;47(8):1015-6.
246. Mullish BH, Williams HRT. *Clostridium difficile* infection and antibiotic-associated diarrhoea. *Clinical Medicine*. 2018;18(3):237-41.
247. Porter CK, Olson S, Hall A, Riddle MS. Travelers' Diarrhea: An Update on the Incidence, Etiology, and Risk in Military Deployments and Similar Travel Populations. *Military Medicine*. 2017;182(suppl_2):4-10.
248. Walters WA, Reyes F, Soto GM, Reynolds ND, Fraser JA, Aviles R, et al. Epidemiology and associated microbiota changes in deployed military personnel at high risk of traveler's diarrhea. *PLOS ONE*. 2020;15(8):e0236703.

249. López-Vélez R, Lebens M, Bundy L, Barriga J, Steffen R. Bacterial travellers' diarrhoea: A narrative review of literature published over the past 10 years. *Travel Medicine and Infectious Disease*. 2022;47:102293.
250. Burd EM, Hinrichs BH. Gastrointestinal Infections. *Molecular Pathology in Clinical Practice*. 2015:707-34.
251. Mark S. Riddle JWS, Shannon D. Putnam, and David R. Tribble. Incidence, etiology, and impact of diarrhea among long-term travellers (US military and similar populations): a systematic review. *The American Society of Tropical Medicine and Hygiene*. 2006;74(5):891–900.
252. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nature reviews microbiology*. 2010;8(3):207-17.
253. Gomes TA, Elias WP, Scaletsky IC, Guth BE, Rodrigues JF, Piazza RM, et al. Diarrheagenic *Escherichia coli*. *Brazilian journal of microbiology*. 2016;47:3-30.
254. Jesser KJ, Levy K. Updates on defining and detecting diarrheagenic *Escherichia coli* pathotypes. *Curr Opin Infect Dis*. 2020;33(5):372-80.
255. Premaratne JMKJK, Satharasinghe DA, Huat JTY, Basri DF, Rukayadi Y, Nakaguchi Y, et al. Impact of human *Campylobacter* infections in Southeast Asia: The contribution of the poultry sector. *Critical reviews in food science and nutrition*. 2017;57(18):3971-86.
256. Lääveri T, Antikainen J, Mero S, Pakkanen SH, Kirveskari J, Roivainen M, et al. Bacterial, viral and parasitic pathogens analysed by qPCR: Findings from a prospective study of travellers' diarrhoea. *Travel Medicine and Infectious Disease*. 2021;40:101957.
257. Patel MM, Hall AJ, Vinjé J, Parashar UD. Noroviruses: A comprehensive review. *Journal of Clinical Virology*. 2009;44(1):1-8.
258. Butler D, McLoughlin R, Flaherty GT. Travel-related gastrointestinal diseases: Assessment and management. *Public Health Challenges*. 2022;1(4):e30.
259. Schultsz C, Van den Ende J, Cobelens F, Vervoort T, Van Gompel A, Wetsteyn J, et al. Diarrheagenic *Escherichia coli* and acute and persistent diarrhea in returned travelers. *Journal of clinical microbiology*. 2000;38(10):3550-4.
260. Kim YS, Lee SH, Kim SH, Joo Y, Bahk GJ. Investigation of the experience of foodborne illness and estimation of the incidence of foodborne disease in South Korea. *Food Control*. 2015;47:226-30.
261. Standards Unit NIS, PHE. UK Standards for Microbiology Investigations: Gastroenteritis. 2020.
262. Alexander CL, Currie S, Pollock K, Smith-Palmer A, Jones BL. An audit of *Cryptosporidium* and *Giardia* detection in Scottish National Health Service Diagnostic Microbiology Laboratories. *Epidemiology and Infection*. 2017;145(8):1584-90.
263. Chalmers RM, Campbell BM, Crouch N, Charlett A, Davies AP. Comparison of diagnostic sensitivity and specificity of seven *Cryptosporidium* assays used in the UK. *Journal of Medical Microbiology*. 2011;60(11):1598-604.
264. Robinson G, Chalmers RM. *Cryptosporidium* diagnostic assays: Microscopy. *Cryptosporidium: Methods and Protocols*. 2020:1-10.
265. Swierczewski B, Simons M. Diagnostics in a Forward Deployed Setting. *Mil Med*. 2017;182(S2):11-6.
266. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science*. 2016;352(6285):565-9.

267. Boolchandani M, Blake KS, Tilley DH, Cabada MM, Schwartz DJ, Patel S, et al. Impact of international travel and diarrhea on gut microbiome and resistome dynamics. *Nature Communications*. 2022;13(1):7485.
268. Jiang P, Wu S, Luo Q, Zhao X-m, Chen W-H. Metagenomic Analysis of Common Intestinal Diseases Reveals Relationships among Microbial Signatures and Powers Multidisease Diagnostic Models. *mSystems*. 2021;6(3):10.1128/msystems.00112-21.
269. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine*. 2019;25(6):968-76.
270. Zhou Y, Wylie KM, Feghaly REE, Mihindukulasuriya KA, Elward A, Haslam DB, et al. Metagenomic Approach for Identification of the Pathogens Associated with Diarrhea in Stool Specimens. *Journal of Clinical Microbiology*. 2016;54(2):368-75.
271. Ahlinder J, Svedberg AL, Nystedt A, Dryselius R, Jacobsson K, Hägglund M, et al. Use of metagenomic microbial source tracking to investigate the source of a foodborne outbreak of cryptosporidiosis. *Food and Waterborne Parasitology*. 2022;26:e00142.
272. Mthethwa-Hlongwa NP, Amoah ID, Gomez A, Davison S, Reddy P, Bux F, et al. Profiling pathogenic protozoan and their functional pathways in wastewater using 18S rRNA and shotgun metagenomics. *Sci Total Environ*. 2024;912:169602.
273. Clinton White A, Jr. Cryptosporidiosis and the Ears of the Hippopotamus. *Clinical Infectious Diseases*. 2010;50(10):1373-4.
274. Checkley W, White AC, Jr., Jaganath D, Arrowood MJ, Chalmers RM, Chen X-M, et al. A review of the global burden, novel diagnostics, therapeutics, and vaccine targets for cryptosporidium. *The Lancet Infectious Diseases*. 2015;15(1):85-94.
275. Bouzid M, Hunter PR, Chalmers RM, Tyler KM. Cryptosporidium pathogenicity and virulence. *Clin Microbiol Rev*. 2013;26(1):115-34.
276. Ryan UM, Feng Y, Fayer R, Xiao L. Taxonomy and molecular epidemiology of Cryptosporidium and Giardia – a 50 year perspective (1971–2021). *International Journal for Parasitology*. 2021;51(13):1099-119.
277. Pinto DJ, Vinayak S. Cryptosporidium: Host-Parasite Interactions and Pathogenesis. *Current Clinical Microbiology Reports*. 2021;8(2):62-7.
278. Troeger C, Forouzanfar M, Rao PC, Khalil I, Brown A, Reiner RC, et al. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet infectious diseases*. 2017;17(9):909-48.
279. Dong S, Yang Y, Wang Y, Yang D, Yang Y, Shi Y, et al. Prevalence of Cryptosporidium Infection in the Global Population: A Systematic Review and Meta-analysis. *Acta Parasitologica*. 2020;65(4):882-9.
280. Toriro R, Pallett S, Woolley S, Bennett C, Hale I, Heylings J, et al. Outbreak of Diarrhea Caused by a Novel Cryptosporidium hominis Subtype During British Military Training in Kenya. *Open Forum Infectious Diseases*. 2024;11(1).
281. Khalil S, Mirdha BR, Panda A, Singh Y, Makharia G, Paul J. Cryptosporidium species subtypes and associated clinical manifestations in Indian patients. *Gastroenterol Hepatol Bed Bench*. 2017;10(4):311-8.
282. Kim JH, Jeon J-Y, Im Y-J, Ha N, Kim J-K, Moon SJ, et al. Long-term taxonomic and functional stability of the gut microbiome from human fecal samples. *Scientific Reports*. 2023;13(1):114.

283. Choo JM, Leong LEX, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports*. 2015;5(1):16350.
284. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. *mSystems*. 2016;1(3):10.1128/msystems.00021-16.
285. Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, et al. A framework for human microbiome research. *Nature*. 2012;486(7402):215-21.
286. Bolt Botnen A, Bjørnsen MB, Alberdi A, Gilbert MTP, Aizpurua O. A simplified protocol for DNA extraction from FTA cards for faecal microbiome studies. *Heliyon*. 2023;9(1):e12861.
287. Moniot M, Nourrisson C, Faure C, Delbac F, Favennec L, Dalle F, et al. Assessment of a Multiplex PCR for the Simultaneous Diagnosis of Intestinal Cryptosporidiosis and Microsporidiosis: Epidemiologic Report from a French Prospective Study. *The Journal of Molecular Diagnostics*. 2021;23(4):417-23.
288. Lalani T, Tisdale MD, Liu J, Mitra I, Philip C, Odundo E, et al. Comparison of stool collection and storage on Whatman FTA Elute cards versus frozen stool for enteropathogen detection using the TaqMan Array Card PCR assay. *PLOS ONE*. 2018;13(8):e0202178.
289. FTATMprovides W, editor FTA cards : high-quality media for storage and transport of DNA2015.
290. Moss JA, Gordy J, Snyder RA. Effective Concentration and Detection of *Cryptosporidium*, *Giardia*, and the Microsporidia from Environmental Matrices. *Journal of Pathogens*. 2014;2014(1):408204.
291. Gaulke CA, Schmeltzer ER, Dasenko M, Tyler BM, Thurber RV, Sharpton TJ. Evaluation of the Effects of Library Preparation Procedure and Sample Characteristics on the Accuracy of Metagenomic Profiles. *mSystems*. 2021;6(5):10.1128/msystems.00440-21.
292. Poulsen CS, Ekstrøm CT, Aarestrup FM, Pamp SJ. Library Preparation and Sequencing Platform Introduce Bias in Metagenomic-Based Characterizations of Microbiomes. *Microbiology Spectrum*. 2022;10(2):e00090-22.
293. Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, et al. Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Research*. 2019;26(5):391-8.
294. Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, Strydom H, et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods*. 2014;195:194-204.
295. Lewandowska DW, Zagordi O, Geissberger F-D, Kufner V, Schmutz S, Böni J, et al. Optimization and validation of sample preparation for metagenomic sequencing of viruses in clinical samples. *Microbiome*. 2017;5:1-13.
296. Castellot A, Camacho J, Fernández-García MD, Tarragó D. Shotgun metagenomics to investigate unknown viral etiologies of pediatric meningoencephalitis. *Plos one*. 2023;18(12):e0296036.
297. McInnes RS, McCallum GE, Lamberte LE, van Schaik W. Horizontal transfer of antibiotic resistance genes in the human gut microbiome. *Current Opinion in Microbiology*. 2020;53:35-43.
298. Pérez JC. Fungi of the human gut microbiota: Roles and significance. *International Journal of Medical Microbiology*. 2021;311(3):151490.

299. Riley LW. Distinguishing Pathovars from Nonpathovars: *Escherichia coli*. *Microbiology Spectrum*. 2020;8(4):10.1128/microbiolspec.ame-0014-2020.
300. Scotland S, Day N, Cravioto A, Thomas L, Rowe B. Production of heat-labile or heat-stable enterotoxins by strains of *Escherichia coli* belonging to serogroups O44, O114, and O128. *Infection and immunity*. 1981;31(1):500-3.
301. Lee K, Raguideau S, Sirén K, Asnicar F, Cumbo F, Hildebrand F, et al. Population-level impacts of antibiotic usage on the human gut microbiome. *Nature Communications*. 2023;14(1):1191.
302. Crits-Christoph A, Hallowell HA, Koutouvalis K, Suez J. Good microbes, bad genes? The dissemination of antimicrobial resistance in the human microbiome. *Gut Microbes*. 2022;14(1):2055944.
303. Matlock W, Lipworth S, Constantinides B, Peto TEA, Walker AS, Crook D, et al. Flanker: a tool for comparative genomics of gene flanking regions. *Microbial Genomics*. 2021;7(9).
304. Department of Data and Analytics DoD, Analytics and Delivery for Impact, World Health Organization, Geneva. World Health Organization methods and data sources for life tables 1990–2019. World Health Organization Geneva; 2020.
305. Mackenzie G. The definition and classification of pneumonia. *Pneumonia*. 2016;8(1):14.
306. Dalziel SR, Haskell L, O'Brien S, Borland ML, Plint AC, Babl FE, et al. Bronchiolitis. *The Lancet*. 2022;400(10349):392-406.
307. Sader HS, Castanheira M, Arends SJR, Goossens H, Flamm RK. Geographical and temporal variation in the frequency and antimicrobial susceptibility of bacteria isolated from patients hospitalized with bacterial pneumonia: results from 20 years of the SENTRY Antimicrobial Surveillance Program (1997–2016). *Journal of Antimicrobial Chemotherapy*. 2019;74(6):1595-606.
308. Gutiérrez F, Masiá M, Mirete C, Soldán B, Carlos Rodríguez J, Padilla S, et al. The influence of age and gender on the population-based incidence of community-acquired pneumonia caused by different microbial pathogens. *Journal of Infection*. 2006;53(3):166-74.
309. Cillóniz C, Polverino E, Ewig S, Aliberti S, Gabarrús A, Menéndez R, et al. Impact of Age and Comorbidity on Cause and Outcome in Community-Acquired Pneumonia. *Chest*. 2013;144(3):999-1007.
310. Manikam L, Lakhanpaul M. Epidemiology of community acquired pneumonia. *Paediatrics and Child Health*. 2012;22(7):299-306.
311. Jean S-S, Chang Y-C, Lin W-C, Lee W-S, Hsueh P-R, Hsu C-W. Epidemiology, Treatment, and Prevention of Nosocomial Bacterial Pneumonia. *Journal of Clinical Medicine*. 2020;9(1):275.
312. Pickles RJ, DeVincenzo JP. Respiratory syncytial virus (RSV) and its propensity for causing bronchiolitis. *The Journal of Pathology*. 2015;235(2):266-76.
313. Center AFHS. Surveillance snapshot: influenza reportable events, service members and other beneficiaries, 2009-2010. *MSMR*. 2010;17(1):17.
314. Korzeniewski K, Nitsch-Osuch A, Chciałowski A, Korsak J. Environmental factors, immune changes and respiratory diseases in troops during military activities. *Respir Physiol Neurobiol*. 2013;187(1):118-22.
315. Korzeniewski K, Nitsch-Osuch A, Konior M, Lass A. Respiratory tract infections in the military environment. *Respir Physiol Neurobiol*. 2015;209:76-80.

316. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*. 2020;382(13):1199-207.
317. Happi AN, Ugwu CA, Happi CT. Tracking the emergence of new SARS-CoV-2 variants in South Africa. *Nature Medicine*. 2021;27(3):372-3.
318. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-3.
319. de Laval F, Chaudet H, Gorgé O, Marchi J, Lacrosse C, Dia A, et al. Investigation of a COVID-19 outbreak on the Charles de Gaulle aircraft carrier, March to April 2020: a retrospective cohort study. *Eurosurveillance*. 2022;27(21):2100612.
320. Torres A, Lee N, Cilloniz C, Vila J, Van der Eerden M. Laboratory diagnosis of pneumonia in the molecular age. *European Respiratory Journal*. 2016;48(6):1764-78.
321. Chen X, Cao K, Wei Y, Qian Y, Liang J, Dong D, et al. Metagenomic next-generation sequencing in the diagnosis of severe pneumonias caused by *Chlamydia psittaci*. *Infection*. 2020;48(4):535-42.
322. Xie F, Duan Z, Zeng W, Xie S, Xie M, Fu H, et al. Clinical metagenomics assessments improve diagnosis and outcomes in community-acquired pneumonia. *BMC Infectious Diseases*. 2021;21:1-9.
323. Serpa PH, Deng X, Abdelghany M, Crawford E, Malcolm K, Caldera S, et al. Metagenomic prediction of antimicrobial resistance in critically ill patients with lower respiratory tract infections. *Genome Medicine*. 2022;14(1):74.
324. Fomsgaard A, Rasmussen M, Spiess K, Fomsgaard A, Belsham G, Fonager J. Improvement of field-deployable metagenomic virus detection by a simple pretreatment method 2022.
325. Rosenstierne MW, McLoughlin KS, Olesen ML, Papa A, Gardner SN, Engler O, et al. The Microbial Detection Array for Detection of Emerging Viruses in Clinical Samples - A Useful Panmicrobial Diagnostic Tool. *PLOS ONE*. 2014;9(6):e100813.
326. Juul S, Izquierdo F, Hurst A, Dai X, Wright A, Kulesha E, et al. What's in my pot? Real-time species identification on the MinION™. *bioRxiv*. 2015:030742.
327. Gargis AS, Cherney B, Conley AB, McLaughlin HP, Sue D. Rapid detection of genetic engineering, structural variation, and antimicrobial resistance markers in bacterial biothreat pathogens by nanopore sequencing. *Scientific Reports*. 2019;9(1):13501.
328. Boix-Amorós A, Piras E, Bu K, Wallach D, Stapylton M, Fernández-Sesma A, et al. Viral Inactivation Impacts Microbiome Estimates in a Tissue-Specific Manner. *mSystems*. 2021;6(5):10.1128/msystems.00674-21.
329. Rosón B, Carratala J, Verdaguer R, Dorca J, Manresa F, Gudiol F. Prospective study of the usefulness of sputum Gram stain in the initial approach to community-acquired pneumonia requiring hospitalization. *Clinical infectious diseases*. 2000;31(4):869-74.
330. Wong LK, Barry AL, Horgan SM. Comparison of six different criteria for judging the acceptability of sputum specimens. *Journal of clinical microbiology*. 1982;16(4):627-31.
331. Muzzi A, Donati C. Population genetics and evolution of the pan-genome of *Streptococcus pneumoniae*. *International Journal of Medical Microbiology*. 2011;301(8):619-22.

332. Fernandez-Carballo BL, Escadafal C, MacLean E, Kapasi AJ, Dittrich S. Distinguishing bacterial versus non-bacterial causes of febrile illness—a systematic review of host biomarkers. *Journal of Infection*. 2021;82(4):1-10.
333. Haslam DB. Future applications of metagenomic next-generation sequencing for infectious diseases diagnostics. *Journal of the Pediatric Infectious Diseases Society*. 2021;10(Supplement_4):S112-S7.
334. Tan CD, Van Den Broek B, Womersley RS, Kaforou M, Hagedoorn NN, Van Der Flier M, et al. A Novel Combination of Host Protein Biomarkers to Distinguish Bacterial From Viral Infections in Febrile Children in Emergency Care. *The Pediatric Infectious Disease Journal*. 2023;42(7):e235-e42.
335. Venkataraman A, Parlov M, Hu P, Schnell D, Wei X, Tiesman JP. Spike-in genomic DNA for validating performance of metagenomics workflows. *Biotechniques*. 2018;65(6):315-21.
336. Miller S, Naccache SN, Samayoa E, Messacar K, Arevalo S, Federman S, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome research*. 2019;29(5):831-42.