Improving the management of sepsis patients through rapid metagenomic pathogen and antibiotic resistance detection in blood

Lluís Moragues Solanas

Submitted for the degree of Doctor of Philosophy

University of East Anglia

Quadram Institute

October 2023

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Acknowledgments.

I will always be grateful for Prof. Justin O'Grady to give me the opportunity to undertake this project in 2019, his mentorship was key in helping me understand all the necessary aspects of the clinical metagenomics world. Thanks to all Justin's group for welcoming me. Specially to Gemma, for showing me all the necessary lab techniques during the first year and having so much patience with my laboratory problems, and also to Alp and Alex, for sharing tips and endless scientific brainstorming sessions.

I am also grateful for Dr. Matthew Gilmour for accepting me in his new group and making the transition smooth. Everyone at Gilmour's group has been extremely helpful and great laboratory companions. Special thanks to the group post-docs, Ana Victoria, Maria and Nasmi, they always offered great mentorship and helped when I needed. This thesis would not be possible without their generosity. Special thanks to Thanh for his patience in helping me with my informatics problems, to Carl for being a great colleague to share scientific problems and go to conferences with and to Prof. Mark Webber, for his mentorship at the latest stages of the project.

Thanks to Maria and Ari for always being up when I needed a break and specially to Oleksii for being the best lunch companion, I'm sure you will do a great thesis. I am also thankful for all my friends back home and in the UK for their constant support, for giving me an opportunity to go away from my project problems and keeping me sane. Special thanks to my parents and my sister. They are an inspiration to me and I would not achieve any of this without their constant encouragement.

Most importantly, to Marta. It is impossible to describe how much you helped me to get here. Thanks for your constant support. I am lucky to have you by my side.

Abstract.

Metagenomics is a promising approach to improve the diagnosis of bloodstream infections (BSI) by offering precise and sensitive genomic information on the microbial presence in blood. Long-read sequencing technologies have successfully supported the use of clinical metagenomics for the diagnosis of conditions such as respiratory illness. However, their use in BSI diagnosis has proven more challenging because of the high levels of host DNA and the low number of BSI-causing microorganisms in blood samples. In this study, several methodological steps were developed and optimised to achieve the generation of a metagenomics pipeline with optimal sensitivity. These included the optimisation of a saponin-based host DNA depletion step and a whole genome amplification method, and the creation of a bioinformatics pipeline to analyse the nanopore sequencing data. The implementation of these steps resulted in a sensitive metagenomics pipeline that could detect, in proof-of-concept experiments, *E. coli* and *S. aureus* spiked at clinically relevant concentrations in under 9 hours. However, when the same pipeline was tested on clinical blood samples, its sensitivity was lower compared to results obtained with standard culture-based results. Consequently, an improved metagenomics pipeline was developed through the implementation of a rapid bacterial enrichment step, resulting in a 12-hour protocol with increased sensitivity. In proof-of-concept experiments, the improved metagenomics pipeline allowed the identification of the four most prevalent BSI causative species (and their antimicrobial resistance determinants) when these were spiked at concentrations as low as 1 CFU/mL in blood. The testing of this pipeline on clinical blood samples from BSI patients will confirm its suitability for diagnosis. Overall, this study has generated a clinical metagenomics pipeline that provides high levels of information in reduced times compared to standard blood cultures, having the potential to be used in the clinics and bring important advances in the BSI diagnosis field.

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.

Contribution Statement.

Some of the work described on chapter 4 features collaborative research involving Dr Daniel S. Lockhart and Miss Elinor McSorley from the company Momentum Bioscience Ltd. They helped on the experimental design as well as collaborating with the wet-lab experiments. The bioinformatics work described on chapter 4 was performed in close collaboration with Dr. Thanh Le-Viet at the Quadram Institute. He designed the bioinformatics pipeline, produced the sequencing results data and helped with their analysis.

Some of the Figures that appear in chapters 1 have been reproduced from articles and appear with permission from their respective journals.

Contents.

Acknowledgments....2
Abstract....3
Contribution Statement....4
List of Figures....11
List of Tables....14
List of abbreviations....15
CHAPTER ONE....16
1.1 Definitions: Bloodstream infection and sepsis....17
1.2 Pathology....18
1.3 Aetiology....21
1.4 Epidemiology, incidence and mortality....24
1.5 Clinical management of sepsis....26
1.5.1 Symptoms and early detection....27
1.5.2 Initial treatment....28
1.6 Blood culture diagnosis of BSI....29
1.6.1 Blood culture....29
1.6.2 Pathogen identification and antimicrobial susceptibility testing....30
1.6.3 The limitations of blood culture diagnosis....31
1.7 Current alternatives to blood culture methodologies....33
1.7.1 Fluorescent *in situ* hybridization-based method....33
1.7.2 Nucleic acid amplification-based methods....34
1.7.3 Metagenomics sequencing approach – Clinical metagenomics....35
1.7.3.1 Essential steps of a clinical metagenomics pipeline....37
1.7.3.1.1 Host depletion methods....37
1.7.3.1.2 Whole genome amplification methods....38
1.7.3.1.3 Sequencing technologies....40
1.7.3.1.4 Bioinformatic analysis for pathogen identification....41
1.7.3.*1*.5 Antimicrobial resistance analysis....42
1.7.3.1.6 Current clinical metagenomics challenges....44
1.7.3.2 Latest advances in shotgun clinical metagenomics pipelines for BSI detection....44
1.8 Summary....46
1.9 Hypothesis and study objectives....47
CHAPTER TWO....48
2.1 – Introduction....49

2.1.1 Blood sample collection tube. 49
2.1.2 Saponin-based host DNA depletion. 50
2.1.3 Microorganisms DNA extraction. 51
2.1.4 Whole genome amplification - multiple displacement amplification. 53
2.1.5 Nanopore library preparation, sequencing and bioinformatics analysis. 53
2.1.6 Aims and hypothesis. 55
2.2 – Methods. 58
2.2.1 Blood sample collection. 58
2.2.2 Blood samples spiking. 58
2.2.3 Blood culture and Sodium Polyanethole Sulfonate (SPS) experiments. 58
2.2.4 Bead beating and DNA extraction protocol. 59
2.2.5 qPCR assay. 60
2.2.6 Statistical analysis 61
2.2.7 Standard saponin-based depletion protocol. 61
2.2.8 Quality Control - DNA quantification and fragment seize measurement. 61
2.2.9 Whole Genome amplification. 62
2.2.10 Debranching. 64
2.2.11 Bead cleaning. 64
2.2.12 Nanopore library preparation and sequencing. 65
2.2.13 Host depletion methods and optimisation. 65
2.2.14 Limit of detection experiments. 67
2.2.15 Bioinformatics analysis 68
2.3 – Results. 69
2.3.1 Testing the host depletion method on different blood culture tubes. 69
2.3.1.1 Host depletion on EDTA-coated collection tubes. 69
2.3.1.2 Host depletion on commercial blood culture medium. 70
2.3.1.3 The effect of sodium polyanethole sulfonate on saponin host depletion reaction. 71
2.3.2 Optimisation of the DNA extraction protocol. 73
2.3.2.1 DNA extraction buffer. 73
2.3.2.2 Testing the efficiency of PBS buffer in the extraction protocol. 74
2.3.3 Whole genome amplification. 74
2.3.3.1 Benefits of an extra amplification step. 75
2.3.3.2 Finding the most efficient whole genome amplification kit. 76
2.3.3.3 Optimisation of Single Cell REPLI-g whole genome amplification protocol. 77
2.3.4 Amplified DNA debranching optimization. 80
2.3.4.1 Testing different debranching strategies. 81

2.3.5 Limit of detection experiments on optimized extraction and whole genome amplification methods. 82
2.3.6 Host depletion method optimisation. 84
2.3.6.1 Standard saponin-based host depletion and the mitochondria problem. 85
2.3.6.2 Testing the viability of different strategies to increase the human mitochondria depletion. 86
2.3.6.2.1 Negative selection of leukocytes using Dynabeads® CD45. 86
2.3.6.2.2 Generation of pores on mitochondria membranes using calcium. 87
2.3.6.2.3 Pneumolysin enzyme can generate pores on mitochondrial membranes. 88
2.3.6.2.4 Mitochondrial depletion using magnetic Miltenyi anti-TOM22 beads. 89
2.3.6.2.5 Phospholipase D (PLD) can disrupt cell membranes. 90
2.3.6.2.6 Differential centrifugation can separate bacteria and host cells in different fractions. 91
2.3.6.2.7 Modification of saponin concentration. 92
2.3.6.2.8 Final decision on how to improve the host DNA depletion method. 93
2.3.7 Limit of detection experiments on optimized saponin-based host depletion protocol. 94
2.3.8 Further optimisation of the depletion protocol 96
2.3.8.1 Scaling-up the saponin-based depletion volume protocol. 96
2.3.8.2 Limit of detection experiments third version (scaled-up volume) 97
2.4 – Discussion. 100
2.4.1 Saponin-based host depletion cannot be successfully applied to commercial blood cultures. 101
2.4.2 Optimised DNA extraction protocol caused no loss of bacterial DNA. 102
2.4.3 Optimised whole genome amplification resulted in a sensitive method that can amplify large quantities of DNA in a short time. 103
2.4.4 Whole genome amplification products decrease nanopore flow-cells pore activity over time. 104
2.4.5 Mitochondrial DNA remains in blood samples after saponin-based depletion. 105
2.4.6 Improvements on individual clinical metagenomics pipeline steps resulted in a final sensitivity of 1-5 CFU/mL for *E. coli* and 10-50 CFU/mL for *S. aureus.* 108
2.4.7 Conclusion. 110
CHAPTER THREE 111
3.1 – Introduction. 112
3.1.1 – Proof-of-concept implementation of a clinical metagenomics pipeline into a clinical setting. 112
3.1.2 – Clinical metagneomics pipeline, alternative approaches and experimental design. 116
3.1.3 Aims and hypothesis. 117
3.2 Methods. 119

3.2.1 Criteria for the selection and collection of clinical blood samples. 119
3.3.2 – Metagenomics pipeline. 121
3.3.2.1 Saponin-based host depletion and DNA extraction. 121
3.3.2.2 qPCR assay and statistical analysis. 121
3.3.2.3 Whole genome amplification and debranching. 122
3.3.2.4 Nanopore library preparation and sequencing. 122
3.3.2.5 Bioinformatics analysis. 122
3.3.3 – 16S amplification pipeline. 123
3.3.4 – Ethical approval. 123
3.3 – Results. 124
3.3.1 Host DNA depletion efficiency in clinical blood samples. 124
3.3.2 Sequencing results from clinical blood samples. 125
3.3.2.1 Identification of causative agents based on relative abundance. 125
3.3.2.2 Identification of causative agents based on total number of reads. 129
3.3.2.3 16S sequencing results. 132
3.3.2.4 Microbiology *vs* Metagenomics *vs* 16S. 134
3.3.3 qPCR detection compared to microbiology results. 136
3.4 – Discussion. 137
3.4.1 Host depletion efficiency decreased when tested on clinical blood samples. 137
3.4.2 Metagenomic abundance and 16S pipelines had poor positivity rate and sensitivity. 138
3.4.3 qPCR validation could only confirm microbiology results on one patient blood sample. 142
3.4.4 Conclusion. 142
CHAPTER FOUR 144
4.1 – Introduction. 145
4.1.1 Transitioning to a new blood anticoagulant collection tube. 145
4.1.2 Sepsi*PURE* capture and enrichment method to increase the CMg pipeline sensitivity. 146
4.1.3 Filtering of reads and taxonomic classification. 147
4.1.4 K-mer-based antimicrobial resistance (AMR) genes detection. 148
4.1.5 Aims and hypothesis. 148
4.2 – Methods. 150
4.2.1 Blood sample obtention. 150
4.2.2 Generation of mock specimens by spiking of pathogens into blood samples and colony count. 150
4.2.3 Saponin-based host depletion and anticoagulants tubes comparison. 150
4.2.4 Sepsi*PURE* enrichment method. 151
4.2.5 Statistical analysis 151

4.2.6 Whole genome amplification and debranching. 152
4.2.7 Nanopore library preparation and sequencing. 152
4.2.8 Bioinformatics pipeline. 152
4.3 – Results. 154
4.3.1 – Saponin-based depletion using EDTA or CPD. 154
4.3.2 – Sepsi*PURE* enrichment method optimisation. 155
4.3.2.1. First combination of Sepsi*PURE* with host DNA depletion method. 155
4.3.2.2 Effects of changing Sepsi*PURE* incubation medium. 156
4.3.2.3 Comparing Sepsi*PURE* medium and beads fractions. 158
4.3.2.4 Testing the optimised Sepsi*PURE* method. 159
4.3.3 –PCR-free library preparation approach – testing the pore activity. 161
4.3.4 – Limit of detection. 162
4.3.4.1 Limit of detection experiments with *E. coli* and *S. aureus* - analysis based on relative abundance. 162
4.3.4.2 *S. aureus* limit of detection experiments, checking for total number of reads and bioinformatics thresholds optimisation. 164
4.3.4.3 Limit of detection on '5mL enrichment protocol' and '1mL standard protocol' CMg pipeline. 169
4.3.4.3.1 *Escherichia coli* Limit of detection experiments. 169
4.3.4.3.2 *Staphylococcus aureus* limit of detection experiments. 171
4.3.4.3.3 *Klebsiella pneumoniae* limit of detection experiments. 173
4.3.4.3.4 *Enterococcus faecalis* limit of detection experiments. 175
4.3.4.4 Antimicrobial determinants detection: 1mL vs 5mL pipelines. 177
4.4 – Discussion. 181
4.4.1 Bacterial enrichment significantly improved host depletion efficiency. 182
4.4.2 Using a PCR-free library preparation method helped maintaining flow cell pore activity constant. 183
4.4.3 Changing to a total number of reads-over-abundance based criteria increased the sensitivity and specificity of the pipeline. 185
4.4.4 All spiked species were detected when present at clinically relevant concentrations. 187
4.4.5 Antimicrobial determinants were identified at all spiked concentrations. 190
4.4.6 Conclusion. 191
CHAPTER FIVE 193
5.1 Assets of the study and implementation of new technologies in the medical field. 195
5.2 Limitations of the study and future directions 197
5.3 Conclusion 201
Publications, conferences and collaborations. 203

Bibliography 204
Supplementary material. 228

List of Figures.

Figure 1. Brief representation of pathophysiological events that occur during sepsis. 21

Figure 2. Global incidence of sepsis cases standardised per 100,000 habitants. 24

Figure 3. Schematic diagram of the workflow for pathogen identification. 35

Figure 4. Schematic representation of a CMg pipeline. 36

Figure 5. Schematic diagram of the clinical metagenomics pipeline. 57

Figure 6. Standard saponin-based method on blood samples collected on EDTA-coated tubes. 70

Figure 7. Saponin-based host DNA depletion efficiency on commercial blood culture medium 71

Figure 8. SPS presence effect on saponin-based depletion method 72

Figure 9. Buffers effect on bead beating and DNA extraction efficiency 73

Figure 10. Assessment of bacterial DNA extraction efficiency on blood samples. 74

Figure 11. Effect on reads classified as *E. coli* when subjecting the sample to a WGA step. 75

Figure 12. Commercial WGA MDA methods comparison. 76

Figure 13. Comparison Standard vs Scaled-up WGA REPLI-g Single Cell protocols. 78

Figure 14. WGA *'optimised* protocol' amplification efficiency when spiked with *E.coli* gDNA. 79

Figure 15. WGA *'optimised protocol'* amplification efficiency when spiked with *S. aureus* gDNA. 80

Figure 16. Debranching strategies comparison effect on pore activity 82

Figure 17. Optimised CMg pipeline LoD experiment using samples spiked with *E. coli* 84

Figure 18. Host mitochondria DNA composition before and after saponin-based host depletion. 85

Figure 19. Dynabeads® effect on host DNA and spiked bacteria 87

Figure 20. Calcium ($CaCl_2$) effect on host DNA and spiked bacteria 88

Figure 21. Pneumolysin effect on host DNA and spiked bacteria 89

Figure 22. Miltenyi effect on host DNA and spiked bacteria 90

Figure 23. Phospholipase D effect on host DNA and spiked bacteria 91

Figure 24. A differential centrifugation effect on host DNA and spiked bacteria 92

Figure 25. Increasing saponin concentration effect on host DNA and spiked bacteria 93

Figure 26. Optimised CMg pipeline (including the new optimised saponin-based depletion method) LoD experiment suing samples spiked with *E. coli* 95
Figure 27. Coverage was evenly distributed across the reference genome 96
Figure 28. Scaled-up and standard saponin-based depletion methods effect on host DNA and spiked bacteria.....97
Figure 29. Optimised CMg pipeline (including scaled-up depletion protocol) LoD experiment using samples spiked with *E. coli* 98
Figure 30. Schematic diagram of clinical metagenomics pipeline. 108
Figure 31. Schematic representation of the QIB-NNUH collaboration for metagenomics detection of BSI..117
Figure 32. Process map of QIB-NNUH collaboration for the obtention of clinical blood samples. 120
Figure 33. Host depletion efficiency on clinical blood samples compared to non-infected and non-treated controls. 125
Figure 34. Human and microorganism reads abundance on clinical blood samples. 126
Figure 35. Taxonomic classification of samples and the respective negative controls. 127
Figure 36. Samples 15, 16 and 19 and their respective NTC taxonomic classification. 130
Figure 37. qPCR causative species identification on clinical blood samples. 136
Figure 38. Host DNA depletion efficiency on blood samples collected on EDTA-coated tubes and CPD-coated tubes. 154
Figure 39. Host depletion efficiency using Sepsi*PURE*.....156
Figure 40. Host depletion efficiency using Sepsi*PURE* with LB medium. 157
Figure 41. Beads and media fractions comparison after applying Sepsi*PURE* method. 159
Figure 42. Assessment of host DNA depletion and *E. coli* DNA enrichment when combining Sepsi*PURE* and saponin-based depletion methods. 160
Figure 43. Comparison of pore activity on samples subjected to PCR-based or PCR-free library preparation methods.....162
Figure 44. LoD abundance results on *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines spiked with *E. coli*. 163
Figure 45. LoD abundance results on *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *S. aureus*. 164

Figure 46. Total number of taxonomically classified *S. aureus* reads using the *'5mL enrichment protocol'*. 165

Figure 47. Total number of taxonomically classified reads as *S. aureus* when different *taxa score* were applied. 167

Figure 48. Taxonomic classification of reads obtained using *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines spiked with *E. coli*. 170

Figure 49. Taxonomic classification of reads obtained using *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *S. aureus*. 172

Figure 50. Taxonomic classification of reads obtained using *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *K. pneumoniae*..... 174

Figure 51. Taxonomic classification of reads obtained using *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *E. faecalis*..... 176

Figure 52. Schematic representation of steps involved in 1mL and 5mL CMg pipelines. 188

Figure 53. Schematic representation of a medical technology innovation cycle. 195

List of Tables.

Table 1. Frequency and rank of most common microorganisms causing BSI by region and time perio*d.* 15

Table 2. Summary of the current main CMg pipelines for the detection of BSIs. 45

Table 3. Sequencing metrics results obtained on optimised CMg pipeline using samples spiked with *E. coli* or *S. aureus* 99

Table 4. Clinical blood samples sequencing results based on abundance. 128

Table 5. Metagenomics sequencing results on clinical blood samples considering total number of classified reads. 131

Table 6. Metagenomics and 16S sequencing results on clinical blood samples. 133

Table 7. Microbiology blood culture compared to metagenomics and 16S sequencing results on clinical blood samples. 135

Table 8. Sequencing metrics on *S. aureus* reads using the *'5mL enrichment protocol'* 166

Table 9. Sequencing metrics on *S. aureus* reads using the *'5mL enrichment protocol'* and processed with different *taxa score* filtering. 168

Table 10. Sequencing metrics on samples spiked with E. coli and subjected to the '5mL enrichment protocol' and '1mL standard protocol' CMg pipelines. 171

Table 11. Sequencing metrics on samples spiked with S. aureus and subjected to the '5mL enrichment protocol' and '1mL standard protocol' CMg pipelines. 173

Table 12. Sequencing metrics on samples spiked with *K. pneumoniae* and subjected to the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. 175

Table 13. Sequencing metrics on samples spiked with *E. faecalis* and subjected to the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. 177

Table 14. AMR gene detection metrics after 3 hours of sequencing using the '5mL enrichment protocol' and '1mL standard protocol' CMg pipelines. 179

Table 15. AMR gene detection metrics after 24 hours of sequencing using the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. 180

List of abbreviations.

AMR – Antimicrobial Resistance

AST – Antimicrobial Susceptibility Testing

BSI – Bloodstream Infection

CFU – Colony Forming Unit

CPD – Citrate Phosphate Dextrose

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid

ESBLs – Extended-Spectrum Beta-Lactamases

ICU – Intensive Care Unit

KMA – KmerResistance

MALDI-TOF MS – Matrix Assisted Laser Desorption Ionization-Time Of flight Mass Spectrometry

MDA – Multiple Displacement Amplification

MRSA - Methicillin-Resistant *Staphylococcus aureus*

NHS – National Health Service

NETs – Neutrophil Extracellular Traps

NICE – The National Institute for Health and Care Excellence

NTC – Non-Template Control

NNUH – Norfolk and Norwich University Hospitals

LB – Luria-Bertani

LBA – Lysis Buffer A

ORSA – Oxacillin-Resistant *S. aureus*

PBS – Phosphate-Buffered Saline

PLD – Phospholipase D

QIB – Quadram Institute

RPB – Rapid PCR Barcoding Kit

RBK – Rapid Sequencing Kit

SPS – Sodium Polyanethole Sulfonate

WBC – White Blood Count

WGA – Whole Genome Amplification

CHAPTER ONE

INTRODUCTION

1.1 Definitions: Bloodstream infection and sepsis.

Bloodstream infection (BSI), characterised by the presence of live microorganisms in the bloodstream, is a major health concern worldwide as it affects roughly 50 million of people every year and approximately 20% of the cases result in death [1–4]. To diagnose a BSI, blood samples are collected from patients presenting clinical signs of infection and are subjected to a blood culture test. In this test, the blood culture must exhibit the growth of microorganisms [5] for the condition to be categorised as BSI.

The definitions of sepsis and BSI are two sides of the same coin, as sepsis primarily arises as a response to the presence of microorganisms in the bloodstream [6]. Sepsis is currently defined as a life-threatening reaction caused by the dysregulation of the host response to an infection that, if it is not detected at an early stage and managed quickly (within hours), can lead to end-stage organ malfunction and death [7]. The sequence of physiological changes that occur in response to infection, culminating in organ dysfunction, is commonly referred to as Systemic Inflammatory Response Syndrome (SIRS) [8,9].

For decades, the common symptoms observed during SIRS, such as changes in temperature, white blood cell count, or heart rate, have been crucial in suspecting or confirming sepsis. However, the newer definition of sepsis introduced in the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) [7] has moved away from the notion that SIRS is a prerequisite or outcome of sepsis, as previously defined in Sepsis-2 [10]. Instead, its new definition recognizes that sepsis can entail a more complex response, which ultimately leads to organ dysfunction.

Sepsis can be categorised according to numerous parameters. One of the most used classifications is based on whether the BSI was acquired at or outside the hospital. Hospital-onset sepsis is defined when a positive blood culture is obtained or organ dysfunction begins on or after

the third day of hospitalisation and it is also often referred to as nosocomial BSI. On the contrary, any patient meeting sepsis criteria before day 3 since hospitalisation is categorized as having community-onset sepsis [11]. Although hospital-onset sepsis accounts for only approximately 10-15% of cases, it is generally associated with higher morbidity compared to community-onset cases, due to a combination of factors such as the health state of the patients or the prevalence of antimicrobial resistant strains [12]. There are also important distinctions within sepsis cases in newborns, which can be classified as early-onset (when the infection is acquired within the first 72 hours of life, often related to intrapartum transition from mothers) and late-onset sepsis (when the infection is acquired after 3 days of life). Additionally, there is also the distinction between early-onset and neonatal sepsis, with the latter occurring within the first month of life [13,14].

1.2 Pathology.

Blood is the liquid tissue that carries and sustains several fundamental functions, which are carried out by specialised cells. In its homeostatic form, human blood is comprised of 54% plasma, 45% red blood cells (also called erythrocytes) and roughly 1% white blood cells (or lymphocytes). Depending on the health status of the person, there will be a variable amount of platelets (thrombocytes) present [15]. Erythrocytes are responsible for transporting oxygen to the different tissues, lymphocytes serve as an efficient monitoring system against invasive microorganisms, and thrombocytes play a crucial role in wound healing and blood coagulation [16]. Traditionally, blood has been viewed as a completely sterile environment, with the detection of microorganisms considered indicative of infection. However, in recent years, some authors argue the existence of a blood microbiome, although the information available is very limited and this concept is still not accepted among all experts in the field [16–18].

In a systematic review by Castillo *et al.* (2019) the authors suggested that as the technology for the detection of microorganisms continues to advance, there is an increase in studies that report the presence of microbial species (e.g. *E. coli* or *E. faecalis*) in blood from patients with no

sepsis symptoms [17]. The phenomenon of microorganisms being present in the blood without causing clinical symptoms is known as transient bacteraemia (in cases involving bacteria) [18]. The concept of transient bacteraemia was first described in 1923 by Prof. Lewis T. in patients with endocarditis. These patients had defective heart valves which became infected, and some microorganisms leaked into their bloodstream without triggering a host response [19]. Transient bacteraemia has also been extensively documented in dentistry, particularly in procedures involving tooth extraction, which can also results in infective endocarditis and transient bacterial presence in the bloodstream [20]. While the presence of transient bacteria or a blood microbiome remains uncertain, the authors from these studies concur that the presence of these microorganisms, even if transient, has the potential to lead to sepsis.

Infectious agents can enter the bloodstream when the skin and mucous barriers are compromised (e.g. during surgery or open wounds) or from various organs within the host [21]. Infections originating from other organs normally happen as a result of an infection occurring in another part of the body (such as the gut, urinary tract or the lungs) trespassing into the bloodstream [22,23]. For instance, increased gut permeability (characterized by the abnormal transfer of molecules and fluids from the gut to other tissues, commonly known as leaky gut syndrome [24]), can lead to the translocation of viable bacteria from the gut microbiome into the bloodstream, resulting in a condition known as gut-derived sepsis [25]. The other main entry point would be during surgery or open wounds. Whittley *et al.* (2018) [26] conducted a metagenomic analysis of microbial data in blood from healthy individuals and discovered a strong presence of genetic information from bacterial species typically associated with skin and buccal flora. This supports the theory held by several authors, including Cogen *et al.* (2008) [27] and Castillo *et al.* (2019) [16], that the majority of transient microorganisms in the bloodstream (which can potentially lead to sepsis) originate from the mouth or skin and enter the bloodstream when the barriers in these tissues are compromised. Other rare ways in which pathogens could reach the bloodstream would be through vascular devices (e.g. needles) or via lymphatic vessels draining a

focus of infection (e.g. abscess) [28]. Finally, in the neonatal population, cases classified as early-onset sepsis are primarily attributed to bacteria transmitted from the mother's genital tract during birth [29].

For most individuals, the entry of microorganisms in the bloodstream triggers a rapid physiological and immunological response, leading to the elimination of the microorganisms and the restoration of homeostasis by the innate immune system [30]. However, in a small percentage of cases, this response is excessive, leading to sepsis or septic shock. To date, the factors that trigger this disproportionate immune response to BSI are still not well understood [31].

Upon the detection of microorganisms in blood, the innate immune system mobilizes innate immune cells that produce pro-inflammatory cytokines. Clinical deterioration is often associated with an overactivation of the immune system and the consequent release of high numbers of cytokines. This excessive cytokine signalling causes a systemic inflammatory response characterized by a significant presence of neutrophils and their release of neutrophil extracellular traps (NETs) [32]. The disproportionate presence of NETs can impact blood coagulation, endothelial permeability, and lead to the apoptosis of numerous host cells [33]. These abnormal inflammatory processes and progressive tissue damage culminate in multi-organ dysfunction [34], as illustrated in Figure 1. If not managed promptly (within the first day), nearly 40% of sepsis cases result in death within 28 days [30].

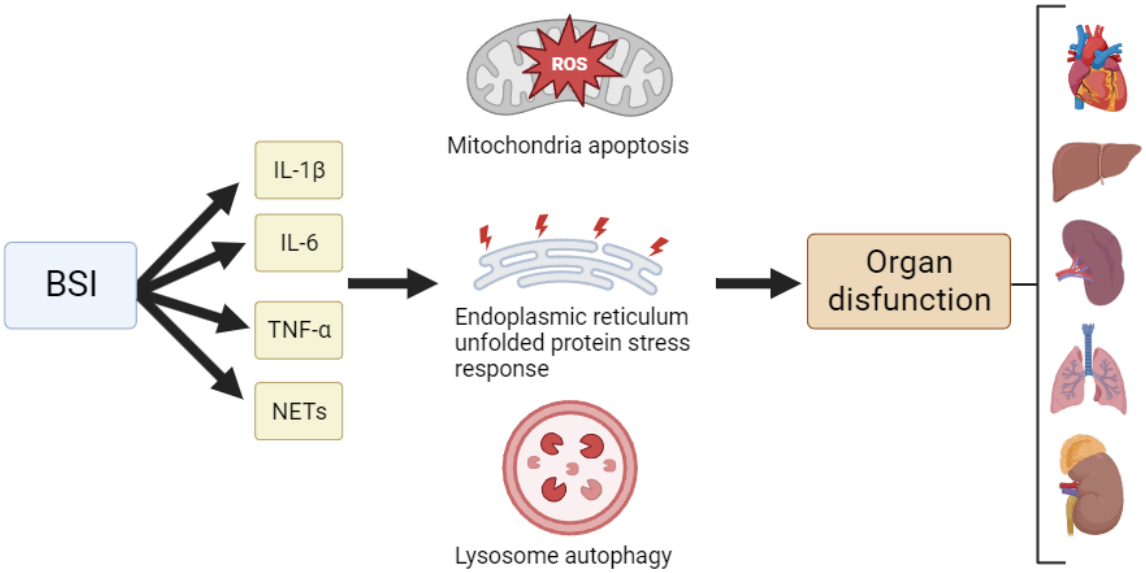


Figure 1. Brief representation of pathophysiological events that occur during sepsis. The detection of microorganisms in the bloodstream by the immune system can produce an abnormal innate immune cell reaction and a cascade of cytokines (IL-1β, IL-6 and TNF-α) and NETs that will cause the disfunction of multiple intracellular organs and ultimately organ disfunction (heart, liver, spleen, lung and kidney). Figure adapted from Huang *et al.* (2019).

1.3 Aetiology.

Any microorganism present in the bloodstream, including bacteria, fungi, parasites, or viruses, has the potential to trigger sepsis. Nonetheless, bacteria are the most commonly identified microorganism in BSI [35]. A comprehensive worldwide systematic review conducted by Diekema *et al.* in 2019 [36], analysing over 250,000 BSI cases spanning the last two decades, unveiled that the most frequently identified pathogens responsible for BSI were *Staphylococcus aureus* (20.7%) followed closely by *Escherichia coli* (20.5%), *Klebsiella pneumoniae* (7.7%), *Pseudomonas aeruginosa* (5.3%), and *Enterococcus faecalis* (5.2%) [35]. Notably, the study observed a consistent increase in the prevalence of Gram-negative bacteria over the last decade, and the frequency of pathogens varied across geographic regions (as outlined in Table 1).

Although the study by Diekema *et al.* did not include data from the United Kingdom, a report by Public Health England, covering the NHS (excluding Scotland) in 2017 [37] found very similar results to those obtained from other worldwide health systems: 95.8% of BSI were

attributed to bacterial agents, while fungal species accounted for 4.2% of the cases. The most frequently isolated species from BSI samples were, in descending order, *E. coli, S. aureus, K. pneumoniae, S. pneumoniae,* and *E. faecalis*.

North America	Latin America	Europe	Asia-Pacific
S. aureus	*E. coli*	*E. coli*	*E. coli*
E. coli	*S. aureus*	*S. aureus*	*S. aureus*
K. pneumoniae	*K. pneumoniae*	*K. pneumoniae*	*K. pneumoniae*
E. faecalis	*P. aeruginosa*	*P. aeruginosa*	*P. aeruginosa*
P. aeruginosa	*E. cloacae*	*E. faecalis*	*E. cloacae*
S. epidermidis	*A. baumannii*	*S. epidermidis*	*E. faecalis*
E. faecium	*S. epidermidis*	*E. faecium*	*A. baumannii*
E. cloacae	*E. faecalis*	*E. cloacae*	*E. faecium*
S. pneumoniae	*S. marcescens*	*A. baumannii*	*S. epidermidis*
S. agalactiae	*E. faecium*	*P. mirabilis*	*S. agalactiae*

Table 1. Rank of most common microorganisms causing BSI by region between 2013-2016. Table adapted from Diekema *at al.* (2019).

Interestingly, different bacterial species tend to be associated with certain underlying conditions. As outlined in a review by Viscoli in 2016 [1], in individuals with a robust immune system, a significant number of cases are attributed to various *Streptococcus* and other Gram-positive species. Conversely, among hosts with age-related compromised defences (e.g., newborns or the elderly), which represent the largest population affected by bloodstream infections, the majority of cases are caused by Gram-negative species such as *E. coli* and *K. pneumoniae*. Lastly, for septic patients affected by pathological or pharmacological conditions that render them susceptible to bloodstream infections, the causative agents can encompass both Gram-positive and Gram-negative species, with variations in outcomes observed across different studies.

In recent years, a notable trend has emerged with strains becoming increasingly resistant to antibiotics, particularly among Gram-negative rods [1]. The occurrence of multi-drug-resistant *Enterobacteriaceae* and Extended-spectrum beta-lactamases (ESBL-phenotype) *Klebsiella spp*. and *E. coli* causing BSI has doubled over the past two decades [36,38]. Conversely, the presence of

resistant strains among Gram-positive bacteria has been more consistent across the last decades. Key resistant phenotypes such as Methicillin-resistant *Staphylococcus aureus* (MRSA), Oxacillin-resistant *S. aureus* (ORSA), or vancomycin-resistant *Enterococcus spp*. have exhibited a decline in prevalence among BSI cases [36,39]. The rise of antimicrobial resistance (AMR) phenotypes poses a significant challenge to clinical healthcare systems, given that antibiotics are crucial for reducing the bacterial presence circulating in the bloodstream and are essential for the treatment of BSI [40].

Although less common than bacteraemia, fungal-induced BSI (fungaemia) is becoming an increasingly prevalent cause of sepsis, comprising approximately 5% of all sepsis cases [41]. From all the fungal-induced BSI cases, 80-90% are candidaemias or BSI caused by *Candida* and five specific *Candida* species (*Candida albicans, C. glabrata, C. tropicalis, C. parapsilosis,* and C*. krusei*) account for over 90% of candidemia cases. Other fungal BSI cases are primarily caused by *Aspergillus* species [41,42]. Despite its lower incidence compared to bacteraemia, candidemia is a frequent cause of hospital-onset sepsis, resulting in substantial morbidity and mortality. Delays in initiating antifungal treatments (often assuming a bacterial infection) and a consistent rise in resistance to triazoles and amphotericin B contribute to these negative outcomes [43].

Finally, while relatively uncommon, BSI can also occasionally be instigated by parasites (which is known as parasitaemia) or viruses (termed viremia). Parasitaemia cases have been primarily associated with infections caused by *Strongyloides spp.* (which causes strongyloidiasis) [44], *Toxoplasma gondii* (referred to as toxoplasmosis) [45], or indirectly induced by *Plasmodium falciparum* (sepsis developed as a secondary cause of malaria) [46–48]. Viruses can mianly lead to sepsis by initially replicating in specific organs and subsequently entering the bloodstream, resulting in what is referred to as secondary viremia [49]. A notable example of secondary viremia is the early-onset sepsis induced by the herpes simplex virus [50].

1.4 Epidemiology, incidence and mortality.

Assessing the global epidemiology of BSI is a complex undertaking. Challenges arise due to variations in research methodologies, differences in patient demographics, diverse healthcare systems, and the limitations of low-sensitivity diagnostic tests, all of which increase the difficulty to accurately determine the actual incidence and mortality rates of sepsis [51].

In a comprehensive metareview of sepsis and incidence data spanning from 1990 to 2017 conducted by Rudd *et al.* in 2020 [52], it was estimated that in 2017 alone, approximately 48.9 million individuals worldwide experienced sepsis, of which 11 million resulted in death (19.7%). The authors noted that the majority (>60%) of those sepsis cases occurred in low- and middle-income countries, and a significant proportion (67%) of the patients had other underlying health conditions. Encouragingly, the study revealed positive trends: there has been an 18.8% decrease in sepsis incidence and a remarkable 52.8% reduction in mortality since 1997 (Figure 2).

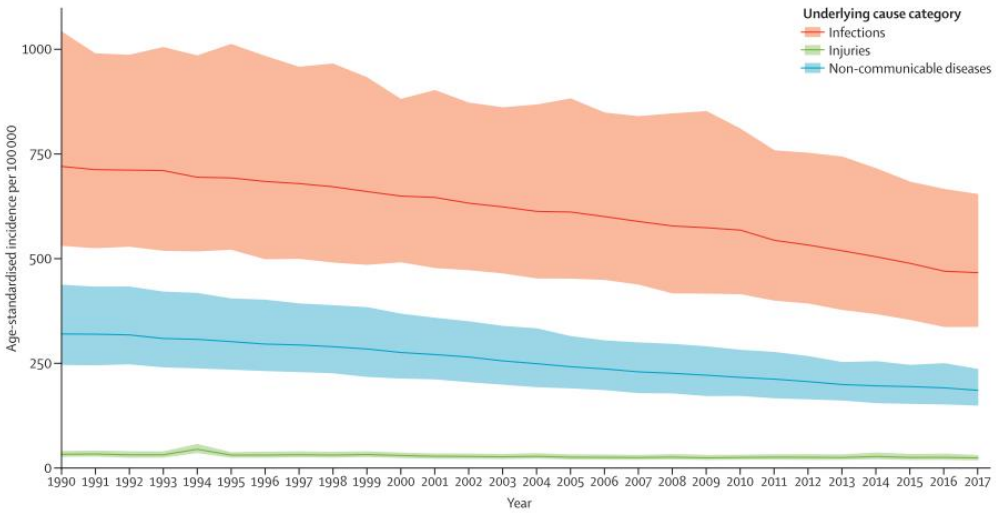


Figure 2. Global incidence of sepsis cases standardised per 100,000 habitants. Cases were grouped according to whether sepsis was caused as a result of underlying infections which caused a health loss (in red), underlying injuries (in green) or non-communicable diseases (unknown origin; in blue). Figure reproduced from Rudd *et al.* (2020) with permission.

Despite the overall reduction in sepsis cases, there is still a notable prevalence of early-onset sepsis and neonatal sepsis globally, affecting 1-100 per 1,000 live births [14]. Particularly in

low- and middle-income countries, neonatal sepsis stands as one of the primary causes of neonatal morbidity and mortality [53]. A systematic review conducted by Fleischmann *et al.* (2021) [54], with a focus on middle-income countries, estimated 2,800 cases of neonatal sepsis per 100,000 live births and a mortality rate of 17.6%. Notably, the study reported a 1.4-fold increase in the incidence of the cases over the past decade.

Furthermore, a significant number of BSI and sepsis cases worldwide are acquired within healthcare settings. Nosocomial BSI stands as one of the foremost hospital-acquired infections across all age groups [55] and has been associated with high mortality rates, prolonged hospitalization stays and long-term neurodevelopmental effects in neonates [56–59]. The prevalence of nosocomial BSI poses a significant challenge to healthcare systems worldwide, with nearly half of severe sepsis cases in intensive care units (ICUs) being acquired within hospital settings, and approximately 25% occurring during ICU stays. Notably, the mortality rate for these patients exceeds 40% (usually due to the deteriorated health of patients and AMR resistant strains presence) which is significantly higher that of other ICU patients [60].

An epidemiological study conducted by Shankar-Hari *et al.* in 2017 [10] covering almost 200 ICUs across England over a 5-year period found that among the 650,000 patients that were admitted, approximately 30% were identified as sepsis cases. However, this represents only a small fraction of the total number of individuals who are estimated to experience BSI annually in England, which is between 1.5 and 2 million people or cases per year, as reported by Singer *et al.* in 2019 [61]. According to data from NHS Digital Hospital Episodes Statistics [62], out of the estimated 1.5-2 million BSI cases, between 2017 and 2018, only 350,344 developed into sepsis. For the majority of individuals with BSI in the UK, the condition is promptly managed through either the immune system's natural response or short courses of antibiotic treatment. In England, sepsis mortality rates are relatively low, with the Office for National Statistics reporting 2,234 sepsis-related deaths per year [63]. However, this condition imposes a substantial economic burden to society. A study

conducted in 2017 estimated that the annual cost of sepsis in the UK alone amounted to £7.76 billion. Out of this total cost, £830 million was directly related to hospitalization and treatment expenses, with the remainder being attributed to loss of earnings [64]. Reducing the incidence of sepsis by enhancing early BSI diagnosis would result in significant cost savings for the NHS.

Overall, because of the incidence and mortality of sepsis and the economic burden it poses for the healthcare systems, both better diagnostics and treatments are needed to improve the management of BSI and sepsis cases.

1.5 Clinical management of sepsis.

Over the past two decades, there have been several revisions of clinical guidelines aimed at determining the most effective approach to managing sepsis cases. With the Surviving Sepsis Campaign guidelines from 2004 [65] and more recent guidelines published in 2021 [66] along with contributions like The UK Sepsis Trust's 2017 publication [67] significant progress has been made (such in the volume of blood drawn from the patient or the rapidness in offering broad-spectrum antibiotic treatment), leading to a notable reduction in mortality rates [68].

The early detection of sepsis is key for patient survival. A study performed in a murine model of sepsis by Kumar *et al.* in 2006 [69] demonstrated that the mortality rate of the condition reached 100% if appropriate antibiotic treatment was not administered after 18 hours from infection. Even when antibiotic treatment was administered at 15 hours from infection, only 13.3% of mice survived. In another study, the same group revealed that each hour of delay in commencing appropriate antimicrobial treatment resulted in an approximately 8% decrease in the survival rate [70]. Overall, these findings underscored the critical importance of receiving suitable antimicrobial therapy at least within the initial 10-15 hours from infection. Multiple studies on human patients data, including those by authors including Gaieski *et al.* in 2010 [71] and Weiss *et al.* in 2014 [72],

have also demonstrated that delays in the timely administration of antibiotics are associated with increased mortality rates.

1.5.1 Symptoms and early detection.

Currently, NHS healthcare professionals follow the National Institute for Health and Care Excellence (NICE) guidelines, which were initially published in 2016 and last updated in 2017, for the early recognition and management of sepsis [73]. Given the critical importance of rapid treatment, it is essential to identify patients with suspected sepsis promptly, with the goal of administering broad-spectrum antibiotics within the first hour [74]. A more targeted treatment is only possible using blood culture-based methods, which are required for the identification of the low number of microorganisms in blood that lead to infection (1-10 CFU/mL) [75,76].

Because of the low turnaround times of blood culture results and the need of rapid treatment, the NICE guidelines focus on the identification of risk factors that elevate the likelihood of sepsis before blood culture results become available. Examples of such risk factors include recent invasive procedures (e.g., surgery) and compromised immune function. Furthermore, in alignment with the report by the UK Sepsis Trust [67], the NICE guidelines establish specific criteria that hospital staff should consider when evaluating a patient for sepsis. These criteria mainly include:

- Altered behaviour and unresponsiveness.
- Respiratory rate of 25 breaths per minute or difficulty to maintain saturation at ≥92%.
- Heart rate at more than 130 beats per minute.
- Systolic blood pressure of ≤90 mmHg or >40 mmHg below normal.
- Cyanosis or non-blanching rash of skin.
- Not having passed urine in previous 18 hours.

Patients exhibiting any of these symptoms are classified as high-risk for sepsis [67], warranting immediate treatment.

1.5.2 Initial treatment.

In the UK, Sepsis UK Trust and NICE established guidelines on the steps to treat sepsis patients, which are summarised in the UK Sepsis Trust manual [67]:

- Administer oxygen to keep saturations above 94%.
- Take blood cultures – consider other cultures.
- Give intravenous (IV) antibiotics.
- Give IV fluids – 500mL crystalloid stat.
- Measure serial lactates – if lactate >4mmol/L recheck after every 10mL/kg IV fluid infusion contraindicated
- Measure urine output – may require urinary catheter.

Once a patient meets the criteria for suspected sepsis diagnostic, it is crucial to promptly collect a blood sample to subject to blood culture [77]. Ideally, these samples should be obtained before administering antibiotics to avoid potential negative impacts on blood culture sensitivity. However, guidelines emphasize the importance of initiating treatment rapidly; if it is not feasible to collect blood samples for culture within the first few hours, antibiotic therapy should not be delayed [74,78].

Multiple studies [69,71,72] have demonstrated that the delay in antibiotic administration increases the mortality rate of BSI. However, the prompt administration of the correct treatment can be a challenging task due to the time required to identify the causative agents [79]. Consequently, broad-spectrum antibiotics are often administered during the initial days of infection, with treatment de-escalation only becoming possible once definitive conclusions can be drawn from microbiology test results. Nevertheless, overusing broad spectrum antibiotics can lead to several problems for both healthcare systems and the patients, such as an escalation in antimicrobial resistance [80], the incorrect treatment of patients infected with species that are non-

targeted or that are resistant to the treatment [81]. To mitigate the adverse effects of the current system for BSI treatment, it is imperative to develop sensitive methods that enable rapid detection of the causative agents and antimicrobial susceptibility (AST). This, in turn, can contribute to the important de-escalation of broad-spectrum antibiotic therapy [82].

1.6 Blood culture diagnosis of BSI.

Despite the recommendations emphasizing the importance of collecting blood samples promptly for cultures [83], the relative influence of blood culture during the early stages of infection is often limited, particularly in more acute cases. Clinicians are compelled to initiate empiric treatment rapidly, being aware that it will take a minimum of 1-3 days for the results of blood culture tests to become available. However, the utilization of blood culture becomes pivotal for refining treatment and achieving optimal management of BSI, particularly in situations where patients continue to exhibit symptoms several days after the initial diagnosis and treatment [84].

1.6.1 Blood culture.

Blood culture serve as the current standard diagnostic tool for the identification of BSI, relying on the detection of viable microorganisms in blood samples. Following the recommended protocols outlined by Public Health England [78] and The UK Sepsis Trust [67], healthcare professionals typically draw 20-30 mL of blood from two different venipuncture sites (usually one per arm to minimise contamination issues) which is collected in one aerobic and one anaerobic bottle for incubation. In most clinical settings, these bottles are incubated in automated continuous-monitoring incubation systems that promptly alert personnel upon detection of microbial growth are used. The detection of growth in blood cultures is achieved through the monitoring of the presence of various metabolites, such as carbon dioxide [76].

Blood cultures typically include a combination of rich media, anticoagulants (usually Sodium polyanetholesulfonate or SPS), and a resin mixture that captures antibiotics to mitigate the

adverse effects on viable microorganisms [78,85]. In approximately 90% of cases, positive cultures indicating growth are identified within the initial 24 hours of incubation. Detection rate increases to 95% after 48 hours and to 99% after 72 hours [86]. Despite the low turnaround times, blood cultures have been the primary standard method for BSI detection for many decades. Blood culture offer the advantage of providing both pathogen identification and subsequent antimicrobial susceptibility testing (AST) in a single, cost-effective, and user-friendly method [87].

1.6.2 Pathogen identification and antimicrobial susceptibility testing.

Identifying microorganisms from unprocessed blood samples presents an immense challenge. This difficulty results from a combination of factors, including the presence of various biological components in the blood that can interfere with downstream processing [88] and the relatively low concentration at which microorganisms are found in bloodstream infections (1-10 CFU/mL) [75]. Consequently, current standard identification techniques heavily depend on cultures to allow the replication of pathogens to increase this concentration. Blood cultures effectively amplify the microbial count in the blood by a factor of 10^9, significantly facilitating identification [76].

If blood culture results are positive the samples are further cultured on solid media in order to identify the causative species [89,90]. This step involves additional incubation time, typically 18-24 hours. Microbial identification relies on assessing biochemical characteristics of cultured organisms and conducting antimicrobial susceptibility testing to guide appropriate antibiotic treatment [91]. Many of these identification tests are automated, with the latest versions offering not only pathogen identification but also AST information simultaneously. Commonly used systems in clinical laboratories include the VITEK®2 and BD Phoenix systems, which have significantly enhanced detection of antibiotic-resistant organisms [92,93]. Once the species have been identified, the medical staff then has a list of significant species that have to be immediately report even if they only grew in of the bottles collected from the patient (e.g. *S. aureus, P. aeruginosa* or

Brucella spp.) and a list of species that will only be considered positive if they were detected In the rest of the bottles from the same patient (e.g. *Staphylococcus* coagulase-negative) [94].

Another standard technique used in clinical microbiology for pathogen identification in positive blood cultures is MALDI-TOF-MS [76]. This method involves obtaining protein and other molecule profiles through mass spectrometry and their comparison against a database containing profiles of various species [95]. Notably, this method offers the advantage of being faster than other approaches, as it eliminates the need for an intermediate sub-culturing step. It also provides highly sensitive results for a wide range of species [96]. However, it remains reliant on positive blood cultures and does not offer information on antibiotic resistance or susceptibility [97]. For AST, other common standard procedures include methods such as the use of antibiotic disc diffusion or minimal inhibitory concentration tests [78]. Although effective, these methods are less favoured compared to the more standardized and automated options described earlier [98].

1.6.3 The limitations of blood culture diagnosis.

Despite the alarming global incidence and mortality rates of BSI, its diagnosis remains challenging. This is primarily due to the low sensitivity and relatively long turnaround times associated with current blood culture methods, which can be as long as 5 days [86,99]. Since there are no reliable diagnostic alternatives to compare the results, assessing the actual sensitivity of blood culture is challenging. Several authors have suggested that blood culture sensitivity falls short: in a review by Bochud *et al.* (2004), it was noted that blood culture only yielded positive results in 30-40% of patients with severe sepsis or septic shock [100]. More recently, Gupta *et. al.* (2016) presented a slightly more optimistic picture, indicating that only 28-49% of severe sepsis hospitalizations (with strong indications of infection) were culture-negative [101]. Since “gold standard” is defined as a reliable diagnosis method [102], blood culture will be described as the “service standard” method for the detection of BSI due its extended use in spite of its limitations.

The low sensitivity observed in blood culture can be attributed to a combination of factors:

1. BSI (and subsequent sepsis) typically arise from a low number of microorganisms circulating in the blood, usually between 1-10 CFU/mL [103]. This low concentration, together with the fact that their distribution in the tissue can be heterogeneous, means that there are instances where no bacteria are present during sample extraction from the patient [104,105].
2. Despite guidelines recommending antibiotic administration after blood samples are taken, a considerable number of patients are already on antibiotics at the time of blood collection. This further diminishes the already low quantities of causative agents present in blood samples [86], contributing to the previous factor.
3. The diagnostic capability is limited to a panel of pathogens that can grow in commercial blood culture media. This limitation means that infections caused by fastidious microorganisms or microorganisms that do not grow in these media, such as *Candida spp.,* often go undetected [42,86,106].

The combination of the extended turnaround times and the limited sensitivity of blood culture has profound implications for BSI patients. The prompt confirmation of sepsis and the subsequent administration of the appropriate antibiotic treatment are crucial for the good prognosis of the patient. However, the inability of the current service standard microbiology cultures and associated methods to consistently provide sensitive and timely results may be contributing to the high mortality rates and poor outcomes associated with BSI [107]. In many cases, the use of broad-spectrum antibiotics (before knowing the microbiology tests outcome) results in inappropriate treatment for a significant number of patients. A study conducted in 2009 by Kumar *et. al.* revealed that approximately 20% [108] of patients with severe sepsis received incorrect initial treatment, leading to a five-fold reduction in their survival rate. Overall, finding a

diagnostic method alternative to blood culture will contribute to more accurate diagnostics and more targeted treatments, leading to a decrease in patient mortality.

1.7 Current alternatives to blood culture methodologies.

Over the past ten years, numerous novel detection methods have emerged in response to the limitations associated with blood cultures. These techniques can be used on positive blood culture samples or directly on whole blood specimens [76]. The multiple molecular biology tests that have been developed for the detection of BSI can generally be categorized into three main groups: (1) methods based on *in situ* hybridization, (2) techniques utilizing nucleic acid amplification, and (3) approaches involving metagenomics sequencing.

1.7.1 Fluorescent *in situ* hybridization-based method.

The fluorescent *in situ* hybridization (FISH)-is a widely used technique for the location of specific DNA sequences. The method is based on the binding of a probe (a nucleic acid amplicon with a known identity that is labelled with a fluorescent dye) to its DNA target region. The binding of these two strands results in the emission of measurable fluorescence, which will indicate the presence of the target sequence [109]. To date, FISH-based methods have been exclusively applied to positive blood culture samples, as they require an initial enrichment step to increase the concentration of bacteria present in the blood sample [109,110]. While FISH-based methods are rapid and user-friendly, they require a high pathogen load, they have limitations concerning the range of detectable species, and they cannot provide information on antimicrobial susceptibility or resistance [111].

An excellent illustration of FISH-based methodology with notable sensitivity (>97%) and specificity (>90%) is provided by the QuickFISH® kits developed by OpGen®. These kits can detected species in 30 minutes but are limited to the detection of a panel of species (e.g. *Staphylococcus, Enterococcus or Candida*) [112,113].

1.7.2 Nucleic acid amplification-based methods.

Polymerase chain reaction (PCR) stands out as one of the most widely employed alternative techniques for pathogen identification in blood, and it can function both directly from whole blood and from positive blood culture samples [110]. The primary advantages of a PCR-based approach over conventional culture include its superior sensitivity, reduced sample volume requirements, and the fact that it is not affected by the presence of antibiotics as it does not require for the microorganism to be intact to detect their DNA [89]. Nonetheless, this approach is constrained by factors such as the presence of substances in the blood that can inhibit the reaction, a substantial risk of contamination, and the relatively high abundance of host DNA compared to bacterial DNA that is found in blood samples [88,114].

PCR-based methods can generally be categorized into two types: those that employ broad-range universal primers and multiplex PCR approaches. Broad-range methods typically target conserved regions of bacterial ribosomal DNA (16S regions) or fungal rDNA (18S regions). A notable example of this approach is the SepsiTest™ developed by Molzym. SepsiTest™ is a CE-marked assay that involves broad-range amplification applied directly to whole blood samples, followed by sequencing of the amplified product. This test offers quick turnaround times (8-12 hours) and has high sensitivity and specificity (87% and 85.8%, respectively) [115,116]. However, it is not suitable for the detection of antibiotic resistance genes [110]. Therefore, blood culture would still be required.

An example of a CE-marked PCR-based test utilizing a multiplex approach is the LightCycler® SeptiFast by Roche. This test can identify a panel of common BSI-causing species directly from whole blood in approximately 6 hours [117]. In a study conducted by Straub *et al.* using this method on neonatal sepsis patients, the authors reported specificity and sensitivity rates of 90.2% and 72.9%, respectively [118]. However, the method is limited due to the number of species that can detect (19 bacterial and 6 fungal) [117].

Overall, despite the potential limitations outlined, molecular-based methods can offer results significantly faster than service standard culture-based methods as illustrated in Figure 3.

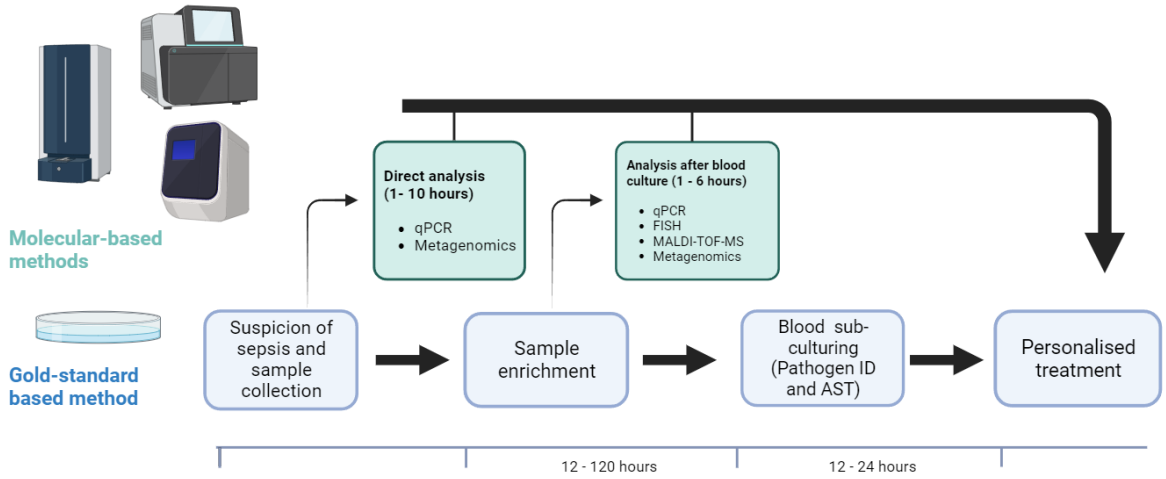


Figure 3. Schematic diagram of the workflow for pathogen identification. Figure adapted from Costa *et al.* (2022).

1.7.3 Metagenomics sequencing approach – Clinical metagenomics

Over the past decade, there have been remarkable advancements in next-generation sequencing technologies that have transformed the potential applications of metagenomic sequencing-based methods for diagnosing infectious diseases. These advancements are characterized by a significant reduction in the cost of reagents, faster processing times, and a substantial increase in data output. These improvements have empowered both research and clinical laboratories to embrace metagenomic-based approaches for diagnosing infections, conducting surveillance, and managing outbreaks [119,120]. Clinical Metagenomics (CMg) is the new field that has emerged in response to these rapid developments, and it is based on the generation of pipelines that use metagenomic sequencing techniques to genetically analyse all the genetic material present in a clinical sample, with the goal of diagnosing infectious diseases [110,120].

Several studies have demonstrated the potential of CMg in several tissues for the detection of various conditions: bone and join infections [121], ventilator-associated pneumonia [122],

meningitis and encephalitis [123] or BSI [124]. A typical CMg workflow is illustrated in Figure 4 [125].

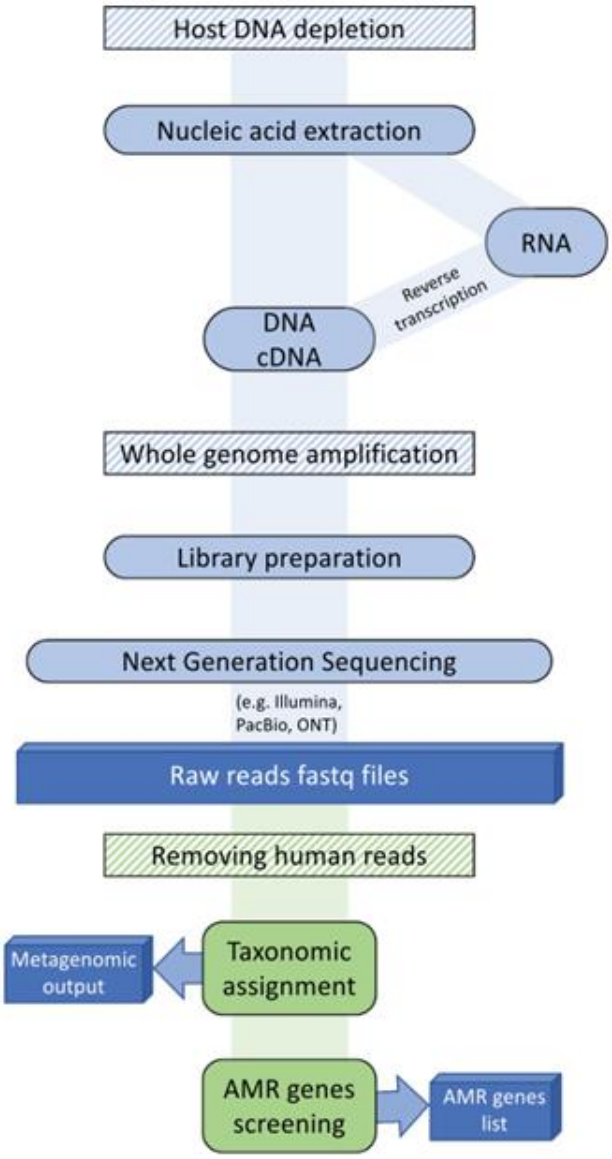


Figure 4. Schematic representation of a CMg pipeline. A generic CMg pipeline starts with an initial DNA extraction or RNA extraction followed by reverse transcription PCR. Optionally, some pipelines will introduce a host DNA depletion step before extraction and/or a whole genome amplification step after the extraction. Next, DNA will be prepared for sequencing and bioinformatic taxonomic assignment and antimicrobial resistance genes (AMR) screening will be performed. Figure reproduced from Moragues-Solanas *et al*. (2021) with permission.

A CMg approach can take one of two forms: it may involve amplifying specific regions, as in a 16S-23S ribosomal DNA amplification-based method, or it can entail the sequencing of the entire genome, known as shotgun genomics [110]. Amplicon-based methods have been effectively employed to detect BSI [126,127]. However, their diagnostics capabilities are limited to the detection of species containing the genes that are targeted [128]. Conversely, the shotgun metagenomics approach offers the advantage of not being restricted to particular pathogens. This,

together with the fact that it allows the identification of antimicrobial resistance and virulence genes [129], make it the most comprehensive approach.

1.7.3.1 Essential steps of a clinical metagenomics pipeline.

1.7.3.1.1 Host depletion methods.

Most biological samples used for the diagnosis of infection contain human cells, which can pose an issue for sequencing-based diagnostic tools. Upon DNA extraction, host cells yield a high amount of human DNA which masks that of the microbial infection-causing agent, which is present in the sample at very low concentrations [89,130]. It is estimated that there are approximately 10^9 white blood cells per mL of blood [131], while the presence of microorganisms circulating in blood causing BSI is minimal (1-10 CFU/mL [89]). Furthermore, the average genomic size of a human cell is about 100X times bigger than that of an average microorganism [132,133]. These factors create a massive imbalance on the amount of human genomic DNA that is sequenced when applying a CMg pipeline compared to microbial DNA. As such, removing white blood cells from blood samples before DNA extraction can significantly enhance the sensitivity of CMg tools. This depletion approach has been used by numerous researchers on various types of clinical samples (including blood [134], cerebrospinal fluid [135], and sputum [136]) with the objective of successfully alleviating the challenges associated with the high presence of host DNA.

Because the majority of BSI are caused by bacteria, host DNA depletion methods for BSI diagnostic applications have generally focused on the differential lysing of host cells, leaving the bacteria in the sample intact. This differential lysis is designed considering the distinct membrane properties of bacterial and host cells [137]. Thus, the method often involves the use of chemicals or enzymes that disrupt the phospholipid bilayers of eukaryotic cell membranes, exposing their DNA to degradation, without affecting the integrity of bacterial membranes. Some examples of chemicals capable of achieving this differential lysis effect include detergents like saponin, Triton-20, or Tween 20 [137,138]. Following membrane lysis, the most common strategy involves the

addition of a nuclease to digest the exposed host DNA. Frequently used nucleases for this purpose include DNAse I, Benzonase, or HL-SAN enzymes [137,139,140]. HL-SAN is particularly notable due to its high resistance to salt (high salt buffers can also play a key role at permeating cell membranes and aiding DNA digestion by breaking tertiary and quaternary DNA structures) and its exceptional efficiency in rapidly digesting nucleic acids in short incubation times [141,142].

While the inclusion of a host depletion step in a CMg pipelines optional, it is becoming an increasingly common step. Consequently, several commercial kits have been developed to serve this purpose. Important examples of such kits include NEB Next Microbiome DNA Enrichment kit, Molzym Ultra-Deep Microbiome Prep, QIAamp DNA Microbiome kit, and Zymo HostZERO microbial DNA kit. Many of these kits use the mentioned differential lysis approach. In a comparative experiment conducted by Heravi *et al.* (2020) using blood samples, all of these commercial kits demonstrated a significant reduction in host DNA presence, leading to improved identification of bacterial DNA in those samples [143] and proving their potential suitability for diagnostic methods.

Importantly, a significant drawback of using a differential lysis approach is the requirement for intact cell membranes in order to preserve the microbial genome from digestion. Thus, in cases where patients have received antibiotic treatment prior to sample collection or when the sample is not fresh, the membranes of the microorganism may not be intact, potentially resulting in the loss of microbial DNA yield.

1.7.3.1.2 Whole genome amplification methods.

The incorporation of a whole genome amplification (WGA) step in clinical CMg pipelines has gained popularity as it allows to exponentially amplify the DNA of a low number of microorganisms present in specific tissue samples [144,145]. WGA is a technology that non-selectively amplifies entire genome sequences to produce sufficient DNA templates for subsequent analyses and its application enables the generation of extensive genetic information from limited DNA quantities [146]. Generally, there are two distinct WGA approaches:

1. WGA PCR-based methods: These methods, including primer extension preamplification (PEP-PCR) or degenerate oligonucleotide methods (DOP-PCR), were among the first to be developed and typically involve the binding of random primers across the genome followed by exponential PCR amplification [147].
2. WGA isothermal methods: The most notable example of this approach is Multiple Displacement Amplification (MDA), a relatively recent technology that uses binding of random primers along the genome followed by amplification of the DNA in an isothermal reaction using the Phi29 polymerase [148]. This approach is highly advantageous for metagenomics applications due to its ability to generate extensive amounts of high-fidelity DNA from low biomass samples [144,149].

Despite the numerous advantages of using a WGA method in CMg pipelines, there are still a few limitations that deter some researchers from using it:

1. Multiple factors (such as the presence of GC-rich regions) can affect amplification and lead to uneven or non-specific amplification, resulting in amplification bias [150].
2. Some authors have reported that WGA methods have poor reproducibility between samples. Obtaining different results when using the same samples under the same conditions [146,151].
3. Certain amplification techniques, such as isothermal MDA, can generate a significant number of nucleic acid tertiary complex structures that may interfere with other downstream applications [152,153].

Despite the limitations of a WGA step, this approach can perform a crucial step in amplifying the very limited presence of bacterial in tissue samples. Without it, the microorganisms DNA yields could be too low to by detected by sequencing.

1.7.3.1.3 Sequencing technologies.

For many decades, Sanger sequencing held a dominant position in genetic analysis. However, with the new century, a wave of innovative sequencing technologies emerged, which brought significant technological advances and propelled the CMg field forward [120,154]. These new sequencing technologies are commonly referred to as Next Generation Sequencing (NGS) and prominent examples of NGS technologies used in CMg pipelines include Illumina, Ion Torrent, PacBio, and nanopore sequencing.

Illumina has stood as the market leader in sequencing technology for several years [155]. It offers cost-effective devices with highly accurate base calling and high-throughput capabilities. However, it does come with some limitations, such as extended running times, short-read sequencing (which complicates genome reconstruction), and the need to batch-process a large number of samples (as their cost per sequencing kit is priced for large sample batching) [156,157]. Several authors have reported the use of Illumina sequencing technology in CMg pipelines for the detection of BSI, as evidenced by works like those of Blauwkamp *et al*. (2019) [124] and Grumaz *et al.* (2016) [158], these approaches had a good sensitivity but slow turnaround times (>24 hours).

PacBio and Nanopore technologies are currently the only two producers of commercial long-read sequencing devices. PacBio uses a method known as "Single Molecule Real-Time" (SMRT) sequencing, in which a polymerase replicates circular DNA using fluorescently labeled nucleotides [156]. In contrast, Nanopore sequencing uses protein pores through which DNA or RNA passes, causing disruptions in the ion current and generating characteristic current signals based on the nucleotide sequence. Nanopore sequencing generally exhibits lower read accuracy compared to Illumina or PacBio [157]. However, authors such as Schmidt *et al.* (2017) [159] have described nanopore sequencing as the ideal technology for integration into CMg pipelines. This is because it is the only NGS technology that enables real-time data analysis, significantly reducing turnaround times. Moreover, nanopore sequencing has the advantage of not requiring sample batching,

particularly when using MinION devices, allowing for a paradigm shift where each patient sample can be sequenced as soon as it is drawn and processed. In 2015, Greninger *et al.* [160] were among the first to publish works on the application of nanopore sequencing to blood samples, reporting the detection of microorganisms in less than 6 hours which was a signinfcantly faster attempt compared to previous CMg pipelines (>24 hours).

1.7.3.1.4 Bioinformatic analysis for pathogen identification.

Using the optimal bioinformatics tool to correctly interpret CMg sequencing data is key for the accurate detection and identification of infectious causative agents. During the first International Conference on Clinical Metagenomics (ICCMg) on 2016 [161], the generation of more streamlined informatic pipelines was highlighted as significant step into the implementation of CMg-based tools in clinical set ups. The sheer volume of data generated by metagenomics sequencing requires highly sophisticated bioinformatic tools that are able to provide accessible information to non-expert users [162]. Currently, there is a considerable amount of bioinformatics tools that can be used to analyse metagenomics data to identify pathogens and detect AMR determinants [163]. Some examples of these informatic pipelines include SURPI+, EPI2ME, Bugseq, NanoOK and IDseq analytical tools [136,164–167].

The EPI2ME software is the cloud-based solution that Oxford Nanopore Technologies (ONT) provides for the quick analysis of nanopore sequencing data. This software uses the classification tool named Centrifuge to perform the taxonomic classification of the sequenced reads [168]. Centrifuge applies a *k*-mer-based analysis of individual reads to classify them into different species [169] [170] by analysing short fragments (*k*-mers) within each read and finding the matching lowest common ancestor[171]. Consequently, EPI2ME pipelines can process large amounts of data at a very high speed while maintaining good sensitivity [172].

Other alternative approaches to analyse the taxonomic composition of the sample would be the alignment- or assembly-based methods. These tools perform full alignment of the whole

read against an extended database or use all of the obtained reads to assemble a full genome [173]. These approaches tend to have better precision at identifying species but are slow and computationally demanding [174]. Minimap2 is an example of pairwise sequence aligner-based tool that can quickly map all metagenomics reads against large database references and that works very efficiently with long-reads such as those generated by nanopore sequencing [175], constituting an attractive alternative to EPI2ME.

1.7.3.*1*.5 Antimicrobial resistance analysis.

The constant increase in antibiotic resistance presented by infectious microorganisms is an evolving threat for healthcare systems. The rapid identification of antimicrobial resistance determinants and the profiling of the appropriate therapy in a timely manner can play a key role in the reduction of antibiotic overuse and in slowing the rise of these resistances. As AMR is determined by genes that confer resistance mechanisms (AMR determinants), an unbiased CMg approach has the potential to rapidly characterize these directly from clinical samples [120,176], providing valuable information on the antibiotics that will not be effective to treat the causative species.

Currently, for the most part, AMR surveillance relies solely on phonotypic experiments [177] and it has been suggested by many authors that the implementation of a CMg pipeline would still require phenotypic testing running in parallel to correlate the genomic presence of antibiotic resistance genes to the phenotypic resistance [178,179]. Establishing phenotypic correlations only based on sequencing results can be challenging, hence the application of one method can still benefit from the other (traditional culture-based AST methods) [180]. However, the unprecedented level of detail that can be achieved by only one experiment (sequencing) makes many believe that the emergence of sequencing pipelines can be the future of AMR surveillance [181]. Several authors, such as Zankari *et al.* (2013) [182] or Neuert *et al.* [183], successfully reported a concordance of almost 100% when testing the phenotypic and genomic prediction of antimicrobial

susceptibilities in parallel. Moreover, the constant generation of metagenomics data it is helping to predict AMR with a high degree of specificity (defined by the correct identification of the present AMR genes) by applying machine or deep learning techniques [184,185].

As per the taxonomic classification of aetiological agents, the computational analysis of the metagenomics results is crucial for the detection of correct AMR determinants. The substantial advances in wet lab sequencing pipelines have also been supported by the increase in different bioinformatics tools and resources for the detection of AMR determinants. Some example of these include ResFinder [186], AMRFinder [187] or MEGARes [188]. Usually, *in-silico* AMR detection is based on the alignment and presence of AMR determinants in the metagenomics sample which map against a contained AMR reference database. Besides the individual read alignment and identification of specific genes, an alignment- or assembly-based method can also be used for the AMR detection [189]. Some authors have described assembly-based approaches as optimal for the metagenomics surveillance, as it is vital for the generation of complete metagenome-assembled genomes (MAGs) for the more comprehensive and accurate antibiotic resistance predictions [190]. Some authors have described assembly-based approaches as optimal for the metagenomics surveillance, as it is vital for the generation of complete metagenome-assembled genomes (MAGs) for the more comprehensive and accurate antibiotic resistance predictions [190]. Alternately, there are also, *k*-mer based alignment technologies such as the ResFinder or the KmerResistance (KMA) tool [186,191].

Despite the potential of CMg pipelines to be used as a quick tool to identify AMR determinants and its phenotypic prediction capacities, the utilization of CMg for the detection and the surveillance of AMR genes is still in its infancy. To date, the successful application of this approach to detect AMR susceptibility in BSI blood samples has not been described.

1.7.3.1.6 Current clinical metagenomics challenges.

Despite the theoretical potential of the application of an unbiased CMg pipeline to detect BSI directly from blood samples, numerous challenges need to be addressed before its definitive implementation:

- BSI is typically caused by a very low number of microorganisms circulating in the blood [75], the DNA of which will be masked by the much higher presence of human DNA from the patient's white blood cells, making identification difficult [89,130].
- In particular, extracting DNA from the limited number of microorganisms within a sample with high quantities of host DNA [106] is especially problematic.
- CMg pipelines are susceptible to contamination from various sources (e.g. skin or reagent contamination), leading to the possibility of generating false positive results [192,193].
- Accurate interpretation of sequencing data requires precise identification of causative genes among the numerous identified taxa [194].

Despite these issues, new approaches and methods that resolve some of these challenges are constantly under research and improvement. As such, there are numerous successful examples in which a CMg pipeline for infection diagnostics has been successfully applied and showed the same or increased sensitivity when compared to service standard methodologies, such as those presented by Charalampous *et al.* in 2019 [141], Ruppe *et al.* [121] in 2017 or Leggett *et al.* (202) [195].

1.7.3.2 Latest advances in shotgun clinical metagenomics pipelines for BSI detection.

In the coming years, shotgun CMg pipelines have the potential to substitute traditional culture standard techniques based on blood culture. To date, several pipelines have been developed for the detection of causative agents in blood, but none of them have replaced blood culture. This is because they either struggle at achieving the same sensitivity and specificity levels

as blood culture or their turnaround times were too high to justify switching technologies [125,196]. A summary of the key characteristics of the main CMg tests for BSI that have been developed in academia or industry laboratories can be seen in Table 2.

	Sensitivity	Specificity	Turnaround time	Microorganisms detected	Sequencing technology	References
Fraunhoffer Institute pipeline (Illumina)	63%	28%	24–30 hours	Bacterial, viral and fungal species	Illumina HiSeq	[158,197]
Fraunhoffer Institute pipeline (Illumina)	Not reported	Not reported	<6 hours	Bacterial, viral and fungal species	ONT MinION References	[198]
Karius® Test	94%	63%	29 hours	Bacterial, viral and fungal species	Illumina NextSeq500	[124,199,200]
iDTECT™ Dx Blood test	82%	70%	2–3 days	Bacterial and viral	Life Technologies Ion Proton	[201]

Table 2. Summary of the current main CMg pipelines for the detection of BSIs.

Table reproduced from Moragues-Solanas et. al. (2021).

Recently, some CMg approaches using cell-free DNA (cfDNA) have been developed for the diagnosis of BSI [124,197]. These methods, compared to a metagenomics approach applied on whole blood, have an increased sensitivity and fast turn-around times as the amount of host cfDNA present in plasma is lower than in whole blood. However, this approach is unable to detect antimicrobial resistance (AMR) genes due to the lack of whole genome coverage and is prone to false positive results caused by the detection of cfDNA from species that are not the causative agent of infection [196,202–204]. Some notable examples would include the method developed by Grumaz *et al.* (2016) [158] or the Karius® Test [124]. Alternatively, CMg directly applied to whole blood, can overcome some of these issues. AMR genes can be detected as only intact bacterial cells from the infectious episode are sequenced, rather than cfDNA, which may not pertain to the

relevant aetiological agents [205]. Several methods that use direct sequencing from the whole blood approach are being developed in academic and commercial laboratories, but most of these are limited by the high levels of host genetic material (relative to pathogen) present in whole blood, which reduces their sensitivity [128,206,207]. The only commercially available or published method that uses this approach is the iDTECT™ Dx Blood test [201].

1.8 Summary

BSI and consequent sepsis rank amongst the top infectious diseases with the highest incidence and mortality. The appropriate treatment of BSI, which relies on the administration of antibiotics, its hampered by the standard blood culture-based diagnostic method that remains time-consuming and non-sensitive. There is an urgent need to find alternative diagnostics methodologies than can help patient management.

Despite the vast research and multiple alternatives presented in this chapter, new therapeutic approaches that can challenge blood culture are rare while the mortality rate in patients with septic shock continues to be unacceptably high. The application of a CMg pipeline holds the potential to solve some of the issues associated with blood culture. However, up to date, all the current attempts to apply a CMg pipeline for BSI detection present major drawbacks that are preventing their final implementation into a clinical set up.

For any alternative method to blood culture, including CMg pipelines, detecting as few as 1 colony forming unit per millilitre of blood is a tremendously challenging task. For the generation of a functional CMg pipeline, certain issues need to be addressed: the large amount of host DNA masking the bacterial DNA, the low amounts of microbial DNA which normally go undetected when sequencing, the high susceptibility of a sequencing approach to contamination or background reads presence. To achieve this, there are multiple opportunities and challenges to be tackled: (1) the overall sensitivity would benefit from increasing the volumes of processed blood (currently

approximately 1mL for most methods) similar to the amount of blood processed by blood culture (5-10mL), (2) host DNA methods could be applied to increase the probability of detecting the low number of microorganisms present, (3) WGA methods could be used to increase the low abundance of microorganisms DNA, (4) a rapid NGS sequencing approach could be used for the generation of a quick pipeline that can offer results faster than blood culture.

1.9 Hypothesis and study objectives

A paradox of sepsis and its diagnosis is the great imbalance between the number of human cells present in blood compared to the small number of microbial cells typically associated with infection. These host cells are predominately immunological and have roles in preventing and controlling infections. Yet, they present a significant diagnostic challenge by dominating the observable DNA during metagenomic sequencing and therefore masking the small proportion of microbial DNA present in the sample. In this study, to achieve a level of sensitivity and reliably to detect and classify microbial DNA in blood samples, it was proposed that structural aspects of host cells could be exploited to deplete host DNA, while also optimising biochemical facets of metagenomic methodologies to improve their molecular efficiency. It was hypothesized that this would be achieved by applying a differential lysis saponin-based host depletion method, combined with other methods that had the potential to help host DNA depletion. Importantly, the collaboration established with Momentum Bioscience Ltd, who partially sponsored the DTP grant that funded this project, allowed the exploration of the use of one of their methods (Sepsi*PURE*) for this purpose. Moreover, it was hypothesized that the detection of low number of microorganisms in the sample would benefit from a WGA which was tested by comparing different methods and optimizing a protocol for the exponential amplification of microorganisms DNA. The final product of combining all the mentioned steps would be sequenced using nanopore technology which would allow the generation of a fast CMg pipeline that could offer pathogen identification results in less than 12 hours, a significant advancement over current blood culture-based methods.

CHAPTER TWO

DEVELOPMENT OF INDIVIDUAL STEPS OF A CMG PIPELINE FOR BSI DIAGNOSTICS: HOST DEPLETION, WHOLE GENOME AMPLIFICATION AND NANOPORE SEQUENCING.

2.1 – Introduction

The main challenges faced by current service standard diagnostics for BSI could potentially be overcome by using a direct clinical metagenomics (CMg) approach on blood samples. CMg has the potential to offer a faster and more sensitive method for the identification of pathogenic species and antibiotic resistance gene determinants [208–210]. However, for CMg to achieve the necessary sensitivity in BSI diagnosis, the development of methodological solutions that address the high presence of human DNA and the low abundance of microorganisms in BSI clinical samples is required [103,130,211]. Moreover, other considerations such as the type of sample collection tube to be used or the most optimal DNA extraction method are also important for the overall performance of the pipeline.

2.1.1 Blood sample collection tube.

The selection of the vial that is used for the collection of blood samples can have large implications on the downstream application of molecular diagnostic based methods. Blood samples are usually collected in tubes containing anticoagulants, which will prevent clot formation and keep sample integrity [212]. Notable examples of anticoagulants used for this purpose in clinical settings include sodium citrate, heparin or ethylenediaminetetraacetic acid (EDTA) [213]. The choice of anticoagulant is key as these can have different chemical properties that can impact the quality of the samples and the molecular assays to which such samples will be subjected [214]. When performing downstream experiments that have the aim to analyse DNA from blood samples, EDTA-coated tubes are the preferred choice. This anticoagulant does not interact with DNA (unlike heparin, which can bind to it and consequently interfere with PCR and DNA extraction methods) and inhibits the activity of the DNAses present in the sample [215]. To date, saponin-based host DNA depletion on blood samples, which is a key method in this study, has only been successfully described in a patent by O'Grady *et al.* in 2021 [216] in which EDTA was chosen the anticoagulant of choice.

In clinical settings, samples are also often collected in blood culture media bottles. These culture bottles contain rich media and employ the anticoagulant Sodium Polyanethole Sulfonate (SPS) to preserve blood integrity and also to prevent the interaction of host cells with bacteria [78,85]. Given the bacterial enrichment process occurring in positive blood cultures, multiple authors have successfully described CMg approaches directly applied on these as a sensitive tool for quick pathogen identification and identification of AMR determinants [121,217–221]. However, the authors reported a large presence of host DNA masking the bacteria present [219,220], stressing the need of applying host DNA depletion methods, which had not been used in these studies. To date, no saponin-based host depletion method applied on blood cultures (containing media and blood) has been described.

2.1.2 Saponin-based host DNA depletion.

To overcome the challenge of host DNA outnumbering the limited microbial DNA in blood samples, the initial focus was put on using and optimising a saponin-based host depletion approach. In recent years, the saponin-based approach has been successfully used in CMg pipelines, including for respiratory samples (sputum) [222,223], urinary samples [224] and cerebrospinal fluid [225]. Host depletion involves the introduction of saponin to lyse or disrupt host cell membranes, followed by the addition of DNase to digest the host DNA [225]. The lytic effect of saponin on cell membranes is believed to occur through an interaction between the aglycon groups in saponin and the sterol molecules, mostly cholesterol, found in the membranes. This saponin interaction with the sterol groups on the membranes induces the formation of cholesterol-complex agents, ultimately leading to changes in the cell membrane structure and its permeability [226,227]. Saponin is especially valuable for CMg pipelines due to its ability to disrupt human cells, mainly due to the high cholesterol content in eukaryotic cells plasma membranes (approximately half of the lipids present) [228]. Consequently, saponin can generate holes and permeabilize eukaryotic membranes [229,230]. However, saponin has no effect on most infectious microbial

species due to the total absence of cholesterol in both Gram-negative and Gram-positive bacteria cell membranes [231], although it could potentially have a lytic effect on fungi due to their large ergosterol presence on their cell walls [232]. Overall, saponin stands as an excellent candidate for enriching clinical metagenomics samples due to the ability to disrupt human cells while leaving the majority of infectious pathogens intact.

Previous research group members had demonstrated that the saponin-based method was the most efficient at reducing host DNA compared to other host cell depletion methods when tested on sputum samples and in relation to lower respiratory infections [136]. Moreover, a streamlined version of the method suitable for blood samples has been described [216], wherein saponin is added at a final concentration of 1%, followed by the addition of an endonuclease (HL-SAN) to digest the host DNA exposed after saponin treatment. To our knowledge, the application of saponin-based host depletion in blood samples has only been described once by Israeli *et al*. in 2020 [142]. They used this method in a proof-of-concept CMg pipeline using spiked blood samples, however, detection of microbes was only possible in a 9-hour protocol when bacteria were present at concentrations of 10^3–10^5 CFU/mL. The saponin-based depletion method has primarily been optimised and described for sputum samples form respiratory tract infections, where causative agents can be found at 10^4-10^5 CFU/mL [233,234]. Thus, the need for further optimisation of the host depletion step when applied on blood samples was anticipated.

2.1.3 Microorganisms DNA extraction.

DNA extraction protocols are commonly used for molecular microbiology purposes, and notably, extraction protocols can have strong effects (e.g. enzymatic inhibition or poor DNA extracted yields) on downstream processes when applied to blood samples [235]. The sensitivity of the CMg pipeline depends greatly on having a DNA extraction protocol that yields high-quality and -quantity bacterial DNA from blood samples [236]. The extraction efficiency is especially relevant when extracting DNA from certain microorganisms, such as fungi or Gram-positive

bacteria, as their DNA extraction is normally hampered by thick, resistant cell walls [237,238]. To extract DNA from microorganisms, bacterial cells must be broken open to release their nucleic acids, while removing remaining proteins and other molecules [239]. A bead-beating method, employed as a preparative step prior to standard bacterial DNA extraction, has the potential to break or disrupt the microorganisms that are specially resistant, preparing them for subsequent DNA extraction methods [240]. Mechanical lysis techniques, like bead beating, are usually simpler and more efficient than enzyme-based approaches [241] which makes this method ideal for a universal DNA extraction method in a CMg pipeline. The bead beating method consists of mixing the sample with a lysing matrix (comprising beads of various sizes and materials) and rapidly agitating the samples to disrupt the cell membranes [242,243]. Normally, bacterial pellets are resuspended in a lysis buffer to assist in the physical lysing process and are then added to bead-beating tubes containing the lysing matrix. Using bead-beating tubes with a lysing matrix mixture of different sizes and materials, such as glass and ceramics, can provide flexibility for lysing different types of species and membranes [244]. After the bead beating step, samples are ready for DNA extraction using more conventional methods. These methods can be mainly divided as bead-based or silica membrane-based extraction methods, and recently, fully automated bead-based techniques have been considered the most suitable for routine clinical settings [245]. In this project, the Promega Maxwell® RSC robot was used for the automated extraction of DNA, as it was considered the most optimal DNA extraction robot available at the institute. The Maxwell utilizes bead-based cartridge extraction approach that includes a chemical lysis step followed by the addition of silica-coated paramagnetic beads that bind to the nucleic acids [246]. However, the research group had previously experienced poor extraction performance when using this robot on human tissue samples that were previously bead-beaten (personal communication). It was theorized that poor extraction outcomes could be due to an incompatibility between the lysis buffers used in the bead-beating tubes and the Promega Maxwell® RSC cartridges. As such, efforts

were directed towards improving the bead beating step by testing different lysis buffers to prevent the loss of bacterial DNA.

2.1.4 Whole genome amplification - multiple displacement amplification.

Whole Genome Amplification (WGA) methods are evolving rapidly as a powerful tool for exploring microbial communities in low-biomass samples, such as blood, with the goal of increasing the limited quantity of microbial DNA available [145]. Recently, the application of a non-PCR based isothermal WGA method – multiple displacement amplification (MDA) – has been successfully applied to low biomass DNA samples. When compared to other WGA techniques, MDA results in high amplification yields from low-input template while minimizing representation bias [144,247]. The MDA method is based on the annealing of random hexamers, which bind to and denature the present bacteria or host genomic DNA. Subsequently, phi29 polymerase attaches to the positions on the genome template where hexamers are bound, leading to strand-displacement DNA synthesis at a constant temperature [248]. As DNA is being amplified, strand displacement amplification gradually and exponentially increases the number of priming and amplification events occurring simultaneously, ultimately forming a complex network structure of hyperbranched DNA [153]. Generally, MDA protocols involve large incubation times (usually exceeding 10 hours [249,250]), rendering the introduction of a WGA MDA protocol into the CMg pipeline inefficient, as faster identification methods (within 12 hours when all the methods combined) are needed [69]. Nonetheless, other researchers have reported large increases in genome DNA amplified yields by employing shorter MDA incubation times [251,252]. Thus, efforts were focused on assessing the feasibility of introducing a WGA MDA method capable of rapidly amplifying DNA within hours.

2.1.5 Nanopore library preparation, sequencing and bioinformatics analysis.

Nanopore sequencing is one of the most suitable sequencing technologies to replace culture and other service standard diagnostics methods due to a high degree of portability and

speed of data acquisition [159,160]. Unlike other sequencing technologies, such as Torrent and Illumina, nanopore sequencing does not need to complete the run before analysis can start, which makes the sequencing results analysis significantly faster than the other competing technologies [253]. Several research groups have already reported successful implementation of shotgun metagenomics strategies using nanopore sequencing, leading to CMg pipelines that were able to report species identification in approximately 6 hours [254,255]. Moreover, manufacturers of this long-read technology offer different library preparation kit options, such as the Rapid PCR Barcoding kit (SQK-RPB004), that are especially suited for low inputs of DNA (to the level of nanograms) and allow for the multiplexed preparation of samples (up to 12). In this library preparation method, template DNA is enzymatically fragmented and tags the template DNA by binding adapters at the 3' and 5' ends of the fragmented amplicons. Then, primers which are complementary to the adapters amplify the DNA in a PCR reaction, which will further increase the DNA yield [256]. The SQK-RPB004 Barcoding kit has been successfully used on other CMg pipelines for the detection of microorganisms even at low concentrations [141,257,258]. Charalampous *et al*.(2019) demonstrated that this method could be modified to amplify larger quantities of DNA with no detrimental effect by increasing the reaction volume and the PCR cycle number [141]. Other non-PCR library approaches, such as the Rapid Barcoding kit (SQK-RBK004), offer shorter hands-on time protocols (approximately 15 minutes) compared to PCR-based approaches (approximately 2-3 hours). However, these non-PCR methods require larger DNA inputs (400-600ng) [259], which might not be applicable for blood samples due to the expected low bacterial yields.

The generated nanopore genomic libraries are normally sequenced on a MinION. The MinION, released by ONT in 2014, was the first ever sequencer of its kind and has since generated a large interest in the clinical metagenomics community. Its portability, speed and affordability make it an excellent technology for real-time applications [260]. While initially ONT flowcells on a MinION had the shortcoming of poor sequencing accuracy (approximately 60%) [261], recent

improvements in the flowcell pore technology and basecalling software have substantially improved raw read accuracy (95-100%) [260]. The generated fastq data can be analysed using the ONT desktop application EPI2ME, which includes software that can run a range of analysis pipelines for pathogen and AMR identification [262]. One of these pipelines, WIMP (What's In My Pot), is a read binning classification software tool based on Centrifuge, an engine that facilitates rapid quantification of reads down to the species level [263]. Several research groups have demonstrated that using a threshold based on the abundance of the total number of classified reads can positively discriminate between true and false positives [141,264]. Metagenomic results from infected tissue samples are known to contain a large number of background reads that originate from multiple sources [265,266]. Therefore, establishing a minimum number of microbial reads detected as a percentage of the total population of taxonomically classified reads is key for achieving the required specificity levels in the diagnostic pipeline.

Previous research by group members successfully demonstrated that an abundance threshold of ≥1% of the reads obtained for a microbial species was an effective strategy for identifying true infected samples as positive. This threshold was also successfully applied and described for the detection of respiratory tract infections by Charalampous *et al.* (2019) [141]. At the beginning of this study, this threshold was used to assess whether the species could be detected in the clinical metagenomics pipeline if their relative abundance over the total number of reads obtained were ≥1%.

2.1.6 Aims and hypothesis.

The first aim of the study was to establish which blood collection tube would be most compatible with the saponin-based depletion method and, consequently, with the overall CMg pipeline. This was tested using EDTA-coated tubes and commercial blood culture medium. Then, the aim of the study was to develop and optimise several of the laboratory steps that play a key

role in the clinical metagenomics pipeline. It was hypothesized that the introduction of a saponin-based depletion method followed by a WGA step, could play a key role in the bacteria identification present in blood samples.

To test this hypothesis, experiments on the best blood collection bottles to perform the saponin-based depletion method and DNA extraction protocols were conducted. Following that, three different commercial REPLI-g MDA kits (Single Cell, MIDI and UltraFast Mini; QIAGEN) were tested, which have been reported by other authors as very efficient MDA kits [144]. Upon the identification of the most efficient WGA MDA methods, the aim was to generate a protocol that could amplify bacterial DNA in the shortest possible time while requiring the lowest amount possible. Then, the application of the saponin-based host depletion step on blood samples was tested to assess its efficiency and potential for improvement.

As the methods were individually optimised, they were integrated into a CMg pipeline (Figure 5) to assess how each optimisation step impacted the overall pipeline sensitivity. When combining all the steps, the aim was to create a pipeline that would offer results (from sample collection) in under 12 hours. This timeframe was crucial, as it has been identified as a key survival point at which accurate detection of aetiological agents supports appropriate treatment to BSI patients [69].

To test the CMg pipeline performance, Limit of Detection (LoD) experiments were conducted by spiking different clinically relevant concentrations of *E. coli* or *S. aureus* into whole blood and subjecting them to the complete CMg pipeline. LoD experiments are critical tests that are routinely applied in clinical diagnostics for the assessment of the theoretical sensitivity and specificity of the CMg diagnostic pipeline [267]. After spiking the blood samples with *E. coli* or *S. aureus* samples were subjected to the CMg pipeline illustrated in Figure 5.

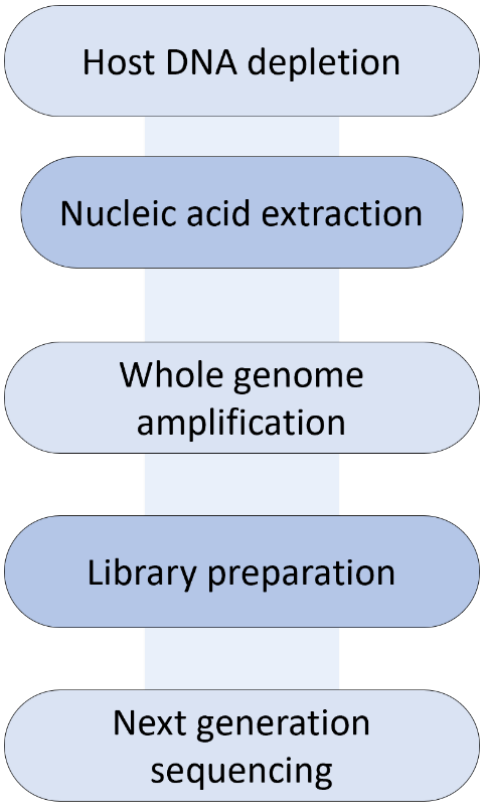


Figure 5. Schematic diagram of the clinical metagenomics pipeline. *E. coli* or *S. aureus* at different clinicaly relevant concentrations would be spiked into whole blood samples and subjected to all the optimised wet lab methods to finally be sequenced using nanopore sequencing. Samples would be considered positive for infection according to their abundance of taxnomically classified reads.

2.2 – Methods.

2.2.1 Blood sample collection.

Whole human blood samples were commercially acquired from Cambridge Bioscience. Blood samples were collected from healthy donors aged between 18-60 and were supplied in Ethylenediaminetetraacetic acid (EDTA) anticoagulants or with no anticoagulant and stored under chilled (4˚C) conditions.

2.2.2 Blood samples spiking.

Different volumes of blood corresponding to variants of the core protocol were spiked using serial dilutions (in PBS) of *Escherichia coli* (ATCC 10536 or NCTC 13441) or *Staphylococcus aureus* (ATCC 25923) in Luria-Bertani (LB) broth. For non-template control samples (NTC), blood samples were spiked with the same volume of PBS containing no bacterial species. Overnight spiked concentrations were calculated by plating the dilutions on LB agar plates and growing them overnight to determine the CFU/mL spiked into the samples.

2.2.3 Blood culture and Sodium Polyanethole Sulfonate (SPS) experiments.

The blood cultures BacT/ALERT® Standard Aerobic (Biomerieux), BacT/ALERT® PF Plus (Biomerieux) and BACTEC™ Plus Aerobic (BD) were tested by adding 8mL of no anticoagulant whole blood and spiked with 10 CFU of *E.coli*. After this, the cultures were put into a BACT/ALERT® 3D Microbial Detection System incubator (Biomerieux) at 37˚C. During the incubation 1mL aliquots were taken at 0, 2 and 16 hours and these were subjected to the standard saponin-based method described in section 2.2.7. The efficiency of the saponin-based host depletion method when applied to blood cultures was assessed by running these samples on a qPCR as described in section 2.2.5 for the detection of human chromosomal DNA. The obtained Ct values by qPCR were compared to ones on blood samples non subjected to saponin-based depletion and taken at the same time points.

SPS experiments were performed by first mixing SPS powder (Sigma-Aldrich) with water at 10% final concentration and then the mixture was used at 0.5%, 0.05%, 0.005% and 0.0005% final concentrations in 1mL whole blood spiked samples. After that, 1mL samples were subjected to the saponin-based depletion. Saponin-based depletion efficiency was assessed by running the samples containing different concentrations of SPS on a qPCR es described in section 2.2.5.

2.2.4 Bead beating and DNA extraction protocol.

DNA extraction protocol on whole blood samples for the efficient extraction of bacterial DNA was first tested by comparing different buffers to be used at the bead beating step. 50-100 CFU/mL of *E. coli* or *S. aureus* were spiked into 200µL of blood and centrifuged at 12,000xg for 3 minutes. Supernatants were discarded and pellets were resuspended in 1mL of PBS and transferred to 1.5mL Eppendorf tubes. Samples were pelleted again at 12,000xg for 3 minutes. Pellets were finally resuspended in 600µL PBS, Bacterial Lysis Buffer (BLB; Roche) or Lysis Buffer A (LBA; Promega). After this comparative experiment, all extractions were performed using PBS as resuspension buffer. The resuspended pellets were transferred into Lysing Matrix E 2.0mL tubes (MP Biomedicals) and bead-beaten at 6m/s for 60 seconds on a FastPrep-24 5G instrument (MP Biomedicals). After bead beating, samples were centrifuged at 20,000xg for 1 minute and the supernatants transferred into 1.5mL Eppendorfs containing 20µL of Proteinase K (Qiagen). Finally, the mixture was incubated at 65˚C for 5 minutes at 1000rpm. The DNA was extracted with the Maxwell RSC PureFood Pathogen kit (Promega) using the Maxwell RSC 48 (Promega) machine and finally eluted in 50µL of Elution Buffer, following the manufacturer's protocol with the following modification: samples were directly mixed with 300µL of Lysis buffer and then transferred into the cartridge. The DNA extraction efficiency was assessed by running the extracted DNA on a qPCR for the detection of *E. coli* and *S. aureus* DNA and comparing the fold change ratio values.

DNA extraction protocol efficiency was assessed by first running qPCR on pure gDNA yields from *E. coli* and *S. aureus* equivalent to 10^6 – 1 cells. The following formula was used to determine

the amount of DNA to add for the cell equivalences: 1 cell genome mass = dsDNA length x bp Dalton weight. Average bp was estimated to be 615.9 Daltons and a single Dalton $1.66x10^{-12}$pg [268]. To estimate the weight of single cells, *E. coli* and *S. aureus* were considered to have genome sizes of 5Mb and 2.8Mb respectively. Running these titrated equivalences on qPCR also resulted in the generation of a regression line. Then, blood samples were spiked with 10^1, 10^2 and 10^5 CFU of both *E. coli* and *S. aureus*, and saponin-based depletion and DNA extraction protocols were used. Extractions were run on qPCR and Cycle threshold (Ct) results were compared to the ones obtained from the gDNA regression line samples to check for possible loss of bacterial DNA during the extraction protocol.

2.2.5 qPCR assay.

SYBRGreen and probe-based qPCR assays were performed on samples which *E. coli* DNA (*cyaA* gene) and *S. aureus* (*eap* gene) (Supplementary Table 1). For all qPCR assays, each reaction contained a final concentration of 0.5µM of both reverse and forward primer, 0.2µM of probe primer (in probe-based assays for human chromosomal and *E. coli* DNA), 10µL of LightCycler 480 Probe or SYBR Green I Master Mix (2X, Roche), 5µL of DNA template and nuclease-free water was added to make a final reaction volume of 20µL. Where appropriate, qPCR Ct results were recalculated to obtain the approximate Ct values of the original samples. This was performed by assuming a 100% amplification efficiency in all assays and subtracting 3.3 Cts to the obtained results. qPCR assays were performed using the LightCycler 480 instrument (Roche). Program conditions were: pre-incubation at 95°C for 5min, amplification for 50 cycles at 95°C for 30s, 55°C for 30s and 72°C for 30s, with a final extension at 72°C for 5min. Ct values were determined using the Roche LightCycler 480 software. Host DNA depletion efficiency and bacterial presence were represented and measured as fold change ratio values. These were calculated from the ΔCt by normalizing all values to one of the untreated control biological replicate Ct. Normalized Ct values were converted to fold change from the equation $2^{-\Delta Ct}$.

2.2.6 Statistical analysis.

Data are expressed as means and ± SD and were analysed using GraphPad Prism 10 (version 10.0.2) software. When multiple comparison analysis were performed, p-values were obtained by performing an ordinary one-way ANOVA test with a post hoc Tukey's test. When only two samples were individually compared, results were obtained using a 2-tail unpaired t-test. Values were considered significant and represented in the graphics when *p ≤ 0.05, **p ≤ 0.01. ***p ≤ 0.001, ****p ≤ 0.0001. Importantly, non-significant differences (ns > 0.05) were not shown.

2.2.7 Standard saponin-based depletion protocol.

The standard host depletion protocol (saponin at 1% final concentration [269]) was initially performed to 200µL whole blood samples. First, 200µL of spiked blood (50-100 CFU/mL *E. coli*) were combined with saponin (in PBS) at a final concentration of 1%, 200µL of HLSAN Buffer (in water 5.5M NaCl and 100mM MgCl2) and 10µL of HLSAN DNase (ArticZymes) in a 1.5mL Eppendorf tube. Samples were incubated at 37˚C for 10 minutes at 1000rpm on a ThermoMixer (Eppendorf). Finally, DNA from the depleted samples was extracted by applying the DNA extraction protocol detailed on section 2.2.4.

2.2.8 Quality Control - DNA quantification and fragment seize measurement.

DNA was always quantified on the Qubit 4.0 (Thermo Fisher) using the high-sensitivity dsDNA kit (Thermo Fisher) and following the manufacturer's instructions. In brief, 1µL of each sample was mixed with 198µL of Qubit HS 1X buffer and 1µL of Qubit dsDNA HS Reagent (Thermo Fisher) in a Qubit Assay Tubes (Thermo Fisher). As per manufacturer's indications, every time DNA quantification on Quibit was performed, new calibration standards were measured first.

When necessary, quality and fragment seize measurement was assessed using the Genomic ScreenTape (Agilent Technologies) on the TapeStation 2200 (Agilent Technologies). Every time, 1µL of DNA was added to 10µL of sample buffer (Agilent Technologies). Samples were run

with Agilent genomic DNA and visualised using the TapeStation analysis software (Agilent Technologies).

2.2.9 Whole Genome amplification.

Initially, to assess the utility of a WGA step onto blood samples were spiked (50-100 CFU of *E. coli*) and subjected them to the standard saponin-based depletion protocol (section 2.2.7) and samples were DNA extracted (section 2.2.4). Purified samples were whole genome amplified using the Repli-g Single Cell kit (Qiagen) according to manufacturer's instructions and samples were compared to non-WGA controls (subjected to the rest of the steps). Samples were then subjected to Rapid PCR Barcoding (SQK-RPB004; ONT) and sequenced on a MinION (ONT) using nanopore sequencing (as described in section 2.2.12). The taxonomic classification of reads was obtained using EPI2ME software (What's In My Pot Pipeline; ONT) (as described in section 2.2.15). WGA sequencing results were further analysed to obtain the p-value in a 2-tail unpaired t test among the different replicate groups.

Whole genome amplification kits comparison experiments were performed using *E. coli* gDNA as template. *E. coli* overnights on LB were DNA extracted by pelleting the cells at 12,000g and the steps detailed in section 2.2.4 were followed. The extracted gDNA concentration was measured as detailed on section 2.2.8 and 500fg, 5pg and 50pg were added as template for the WGA REPLI-g Single Cell, REPLI-g MIDI and UltraFast Mini kits (QIAGEN) WGA kits. Then, all the amplifications were performed following manufacturer's instructions. After 2 hours and 30 minutes of amplification all the reactions were stopped and kits were compared by running them on a qPCR for the detection of *E. coli* DNA as detailed in section 2.2.5 and their *E. coli* amplification efficiencies were evaluated by comparing their fold change ratios.

After establishing the REPLI-g Single Cell (QIAGEN) as the best amplification kit for small concentrations of template DNA, it was first decided to adapt this kit to accommodate higher volumes of template from the recommended 2µL to 15µL. The manufacturer's WGA protocol was

compared to a new version of the protocol with the following adjustments to accommodate for the higher volume: 15µL of sample was mixed with 1µL DLB Buffer; polymerase master mix was prepared by mixing 29µL of Single Cell reaction buffer with 2µL of polymerase; amplification occurred at 37˚C for 2 hour and 30 minutes. This optimised version of the protocol was compared to the standard protocol containing less template volume by first taking 2µL of the WGA samples at 1, 1.5, 2 and 2.5 hours and the DNA concentrations were measured as described on 2.2.7. Also, the protocols were compared at the end of the 2.5 hours incubation by running them on a qPCR for the detection of *E. coli* DNA and comparing their fold change ratios as explained on section 2.2.4.

The WGA protocol was further optimised by introducing DTT to the reaction and shortening the incubation times: 15µL of sample was mixed with 0.16µL of 1M DTT and 1.84µL DLB Buffer at a final volume of 2µL per sample; polymerase master mix was prepared by mixing 29µL of Single Cell reaction buffer with 2µL of polymerase; amplification occurred at 37˚C for 2 hour and 30 minutes. To assess the WGA efficiency of the optimised protocol and compare it to the standard version, different concentrations of extracted gDNA of both *E. coli* and *S. aureus* were added as template for the WGA reaction. For this, *E. coli* or *S. aureus* overnights in LB were extracted by pelleting the cells at 12,000g and then proceeding with the DNA extraction as detailed in section 2.2.3. Then the concentrations on the extracted samples was measured and the gDNA equivalent to adding 10^3, 10^2, 10 and 1 cells were added. The amplification incubation time was 2 hours 30 minutes and samples were taken at 0, 1, 1.5, 2 and 2.5 hours and ran on qPCR for the detection of *E. coli* and *S. aureus* DNA as detailed on section 2.2.4. Ct results were converted to their cell equivalences of genomic material produced by using the following equation: $Cell\ quantity\ equivalence = 10^{\frac{Ct-b}{m}}$, *m* and *b* values were taken from the regression line equation obtained in the DNA extraction experiments detailed on section 2.2.3.

2.2.10 Debranching.

After WGA, artificial constructs that could interfere with nanopore sequencing were digested. The entire volume of the WGA reaction was treated with the enzymes T7 Endonuclease I (NEB) or the S1 Nuclease (Thermo Fisher) separately or combined. For treating the WGA reaction with T7 Endonuclease I, 20U of the enzyme and Buffer 2.0 (NEB) at 1X were added and then samples were incubating at 37˚C for 30 minutes. For treating the WGA reaction with S1 nuclease, 100U of the enzyme and buffer at 1X were added. Then, samples were incubating at room temperature for 30 minutes and the enzyme was heat inactivated at 65˚C during 10 minutes. For testing both enzymes combined, samples were first subjected to the S1 treatment as previously described and, after enzyme heat inactivation, 20U of the enzyme and Buffer 2.0 (NEB) at 1X were added and then samples were incubating at 37˚C for 30 minutes. After debranching, samples were cleaned with AMPure Beads and then samples underwent library preparation with Rapid PCR library kit (SQK-RPB004; ONT) and sequencing with nanopore sequencing. Debranching efficiency was evaluated by observing the pore activity decrease over sequencing time.

2.2.11 Bead cleaning.

Between some of the steps of the CMg pipeline it was necessary to clean the samples using the AMPure magnetic beads (Beckman Coulter). In a 1.5 ml Eppendorf DNA LoBind tube (Eppendorf) samples were mixed with the beads at different concentrations depending on the step and the desired resulting fragment length: (1) after DNA extracting saponin-based depleted blood samples and before proceeding with WGA, samples were mixed with beads at 1.8X ratio (sample : beads), (2) after the debranching step and before library preparation samples were mixed with beads at 1X ratio and (3) after library preparation samples were pooled together by cleaning them with a 0.6X ratio. After mixing the samples with the beads, tubes were on a HulaMixer (Life Technologies) for 5 minutes. Then beads pelleted using a DynaMag-2 magnetic rack (Thermo Fisher), supernatant was removed and beads underwent two 200µL 70% ethanol washes (in water).

Finally, DNA was eluted by resuspending the beads in nuclease-free water (Thermo Fisher), 15µL for the samples that had to be subjected to WGA and 10µL for the rest. Note that for pooling the samples after the library preparation protocol, the wash step was performed using 500µL (instead of 200µL) 70% ethanol washes (in water) and the DNA was eluted using 10µL of the MinION elution buffer (10mM Tris-HCl pH7.5-8.0 with 50mM NaCl).

2.2.12 Nanopore library preparation and sequencing.

Library preparation of WGA-debranched samples was completed using the Rapid PCR barcoding Kit SQK-RPB004 (Oxford Nanopore Technologies). Library preparations followed the manufacturer's instructions with the following modifications: up to 10ng DNA was used as template in a maximum total volume of 7.5µL with water, then 2.5µL FRM was added; reaction volumes were doubled, using 2µL RLB barcode, 50µL LongAmp Taq 2x Master mix (NEB) and 38µL nuclease-free water (100µL total reaction volume); thermocycling parameters were initial denaturation 3 min at 95°C (1 cycle), 25 cycles of denaturation 15 sec at 95°C , annealing 15 sec at 56°C, extension 4 min at 65°C, then a final extension 4 min at 65°C (1 cycle) held at 4°C. Once libraries were prepared, samples were quantified, pooled in equal concentrations, and cleaned with a 0.6X AMPure as detailed in 2.2.11. Approximately 100fmoles of pooled PCR library was loaded onto the MinION flow cell (R9.4.1, Oxoford Nanopore Technologies). For all samples, the sequencing run time was 3 hours.

2.2.13 Host depletion methods and optimisation.

Initially, the standard saponin-based host depletion protocol used was taken from Charalampous *et al.* in 2019 and 2021 [141,269] and described on section 2.2.7. Further improvement of the saponin-based method consisted of combining the saponin-based method with different strategies: CD45 Beads, Anti-TOM22 beads, Calcium (as $CalCl_2$), Pneumolysin, Phospholipase D, a differential centrifugation step or increased saponin concentration. Next is

detailed the different host DNA depletion protocols, note that for all samples, these methods were tested on 200µL of blood spiked with 50-100 CFU/mL of *E. coli* as described in section 2.2.2:

- Dynabeads™ CD45 (Invitrogen) were tested on spiked blood samples following the manufacturers protocol for the depletion of CD45+ cells. The blood supernatant resulting from the depletion (with the white cells removed) was subsequently processed using the saponin-based depletion protocol as described on section 2.2.7.
- Addition of calcium was evaluated by first generating different stock concentrations of $CaCl_2$ (Sigma-Aldrich) diluted in water (0mM, 5mM, 50mM and 500mM) and then adding 50µL of these dilutions onto 150µL of blood spike samples. Then samples were subjected to the standard saponin-based depletion protocol detailed on section 2.2.7.
- Pneumolysin effect on host cell membranes was evaluated by adding recombinant Ply (MyBioSource) at 0µg/mL, 1µg/mL and 2µg/ml at final concentrations into the spiked blood samples. Then, the samples were subjected to the saponin-based host depletion protocol detailed on section 2.2.7.
- Anti-TOM22 MicroBeads (Miltenyi Biotec) could potentially deplete mitochondria by adhering to their membranes and then magnetising and removing mitochondria from the sample, host cells by were lysed by applying the standard saponin-based host depletion method described on section 2.2.7. Then, depleted samples were pelleted at 12,000 g and the supernatants were discarded. Pellets were resuspended with 900µLof 1X separation buffer and 5µL of the beads (Miltenyi Biotec). The mixtures were transferred to a 2.0 mL Eppendorf tube and incubating on a Hula mixer (Life Technologies) during 1 hour at 4°C. Finally, samples went through the LS columns (Miltenyi Biotec) to retain the beads containing the mitochondria attached.
- Phospholipase D (PLD) in powder (Sigma-Aldrich) was first resuspended in water to a final concentration of 100 U/µL. Then, 0U, 200U or 400U of PLD were directly added onto the

spiked blood samples and were subjected to the saponin-based methods detailed on section 2.2.7.

- A differential centrifugation step separated plasma the rest of the blood components by first adding 200µL of PBS to the blood spiked samples and then centrifugating at 400g during 5 minutes. Approximately 190µL of plasma were recovered and it was subjected to the saponin-based depletion protocol detailed on section 2.2.7.
- The effects of increasing the saponin concentration were compared to the standard protocol described on section 2.2.7. Saponin was added (Sigma-Aldrich) (in PBS) at a final concentration of 1%, 3% and 6%, 200µL of HLSAN Buffer (in water 5.5M NaCl and 100mM MgCl2) and 10µL of HLSAN DNase (ArticZymes) in a 1.5mL Eppendorf tube. Samples were incubated at 37˚C for 10 minutes at 1000rpm on a ThermoMixer (Eppendorf).

After the application of all the detailed host depletion methods, all samples were DNA extracted as detailed in section 2.2.4 and then evaluated by qPCR to detect human chromosomal DNA (*RNA Pol. A* gene), human mitochondrial DNA (*MT-TL1* gene), *E. coli* DNA (*cyaA* gene) (Supplementary Table 1) as described in section 2.2.5.

Finally, a scaled-up saponin-based depletion protocol, which allowed an increase in sample volume processed, was optimized and compared to latest saponin-based protocol described with saponin at 3% final concentration. The 'scaled-up protocol' was tested by spiking 50-100 CFU/mL of *E. coli* into 1mL of blood and combined with saponin at a final concentration of 3%, 1mL of HLSAN Buffer and 10µL of HLSAN DNase on a 5.0mL Eppendorf tube. Samples were incubated at 37˚C for 30 minutes at 1000rpm on a ThermoMixer. Both samples were DNA extracted as described in section 2.2.3 and methods were compared by qPCR as described in section 2.2.5.

2.2.14 Limit of detection experiments.

Ultimately, saponin-based host depletion, DNA extraction, WGA, branching, library preparation and nanopore sequencing methods were combined on the same clinical metagenomics

CMg pipeline and the analytical sensitivity of the pipeline was tested on Limit of Detection (LoD) experiments. Blood samples were spiked with different concentrations of *E. coli* and *S. aureus* (as described in section 2.2.2) ranging from 1000 to 1 CFU. Then samples underwent saponin-based host depletion method (sections 2.2.7 and 2.2.13) and DNA extraction protocol (section 2.2.4) combined with a beads cleaning step to elute in a final volume of 15µL (section 2.2.11). Samples were next subjected to the optimised WGA protocol (section 2.2.8), debranching with T7 Endonuclease I (section 2.2.10) and cleaning with AMPure beads (section 2.2.11). Finally, DNA samples went through a PCR library preparation (section 2.2.12) and samples were pooled together using AMPure beads (section 2.2.11). The pooled library was ran on a TapeStation to check its quality (section 2.2.8) and it was nanopore sequenced with a MinION from Oxford Nanopore Technologies (ONT) (section 2.2.12).

2.2.15 Bioinformatics analysis.

Pathogens were identified in the sequence data using the EPI2ME software and the What's In My Pot pipeline (Oxford Nanopor Technologies). In order to determine the presence of bacteria within a sample, they were classified as positive for a mock bacterial infection if the total microorganisms reads were present at ≥1% over the total number of classified reads. EPI2ME results were used to calculate number of reads, abundance and average genome coverage (estimation calculated using *average read length*number of reads/genome seize*) of the spiked species. For sequencing reads QC and coverage analysis, fastq files were concatenated tail-to-head using cat (Galaxy Version 0.1.0) and then aligning them to the appropriate *E. coli* reference (ENA sample accession number: SAMEA2689781) using minimap2 (Galaxy version 2.17) with the default parameters for ONT data and bam files were obtained. Finally, QualiMap BamQC (Galaxy Version 2.2.2d) was used for bam file coverage analysis and visualisation.

2.3 – Results.

The CMg pipeline used in this study consisted into five main steps, as detailed in Figure 5: host depletion, DNA extraction, whole genome amplification, PCR library preparation and nanopore sequencing. Previous work performed by our research group focused on applying a saponin-based depletion method on blood samples. Based on preliminary data previously obtained by the group, the saponin method exhibited the capability to reduced large quantities of host DNA and thus, at the beginning the focus of the project was put in assessing in which blood collection tubes the saponin-based depletion was most efficient. Then, the DNA extraction and whole genome amplification steps were optimised.

2.3.1 Testing the host depletion method on different blood culture tubes.

2.3.1.1 Host depletion on EDTA-coated collection tubes.

First, the application of the standard saponin-based host depletion method was tested on spiked blood samples on EDTA-coated collection tubes.

The application of the saponin-based host depletion step caused a large reduction in host DNA material. After subjecting blood samples to the saponin-based method, DNA extraction and qPCR, the average fold ratio reduction on the target human chromosome was $5 \cdot 10^3$ fold, resulting in an approximate 99.9% reduction in host DNA presence. Additionally, at the same time (blood samples were spiked with 10^3-10^4 CFU/mL), did not cause any significant change in bacterial DNA presence (Figure 6).

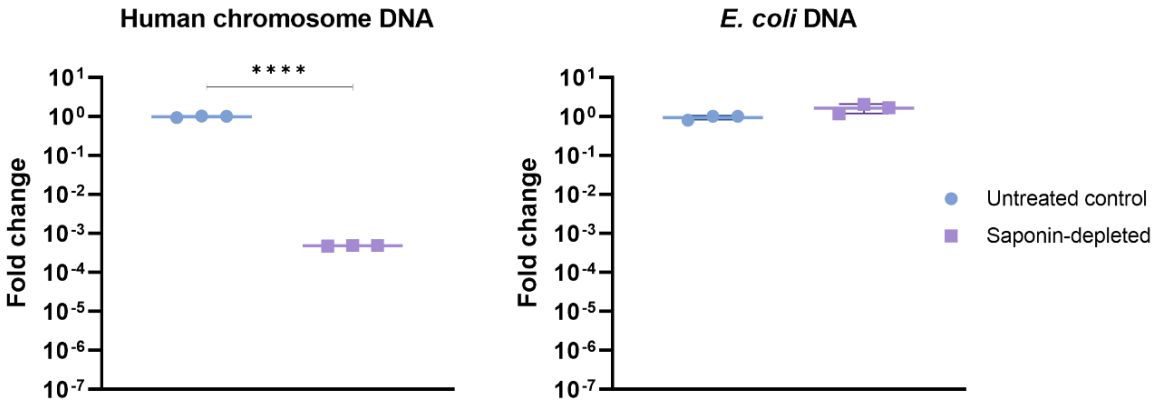


Figure 6. Standard saponin-based method on blood samples collected on EDTA-coated tubes. Fold change ratio results on human chromosome and *E. coli* DNA on samples subjected to standard saponin-based host depletion control vs non-depleted samples. The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), ***p ≤ 0.001

2.3.1.2 Host depletion on commercial blood culture medium.

To test the saponin-based depletion method on blood culture medium, blood samples were spiked *E. coli* and introduced into the following commercial blood culture medium tubes: BacT/ALERT® Standard aerobic and BacT/ALERT® PF Plus from Biomeriux and BD BACTEC™ Plus Aerobic. Then, cultures were incubated at 37°C for 16 hours and samples were taken at 0, 2 and 16 hour timepoints. Samples were then subjected to saponin-based depletion and DNA extraction protocols.

For the majority of blood culture samples tested, the application of the host depletion protocol did not significantly reduce the presence of host DNA (Figure 7). Although the difference was not always significant, the application of the host depletion protocol on depleted non-culture samples caused a higher reduction in chromosomal host DNA than in samples collected from blood cultures at all time points. Fold change values at 0 hours (Figure 7A), in which samples blood culture samples were directly subjected to host depletion before incubation) of the host depletion method on the two BacT/ALERT® cultures were not significantly different to the ones obtained on non-depleted samples. Only on blood samples coming from the BD BACTEC™ culture, the saponin-based method caused a significant reduction in host DNA. However, when the saponin method was

applied on non-culture samples, the fold change reduction was still 10^2-fold superior to BD BACTEC™ blood culture samples. For all the samples at both 2 and 16 hours (Figure 7B and C), the application of the saponin-based host depletion protocol on commercial cultures did not significantly reduce the host DNA presence when compared to no culture non-depleted controls. The amount of host DNA depleted was always higher when the protocol was applied on blood samples that were not mixed with culture.

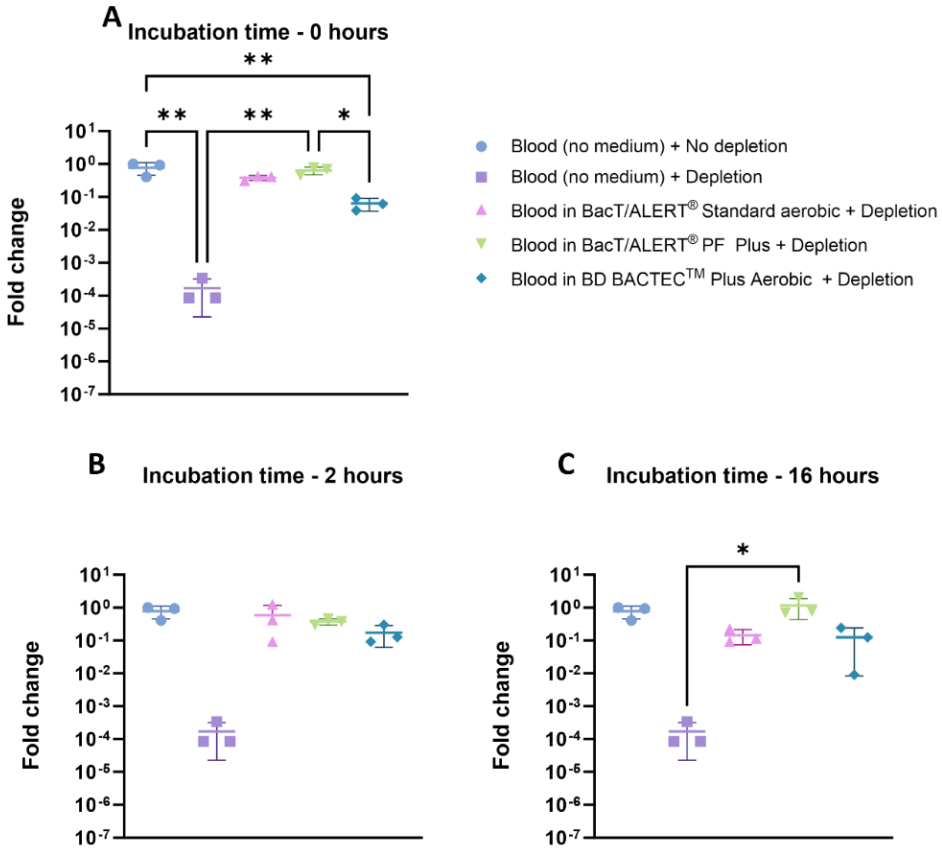


Figure 7. Saponin-based host DNA depletion efficiency on commercial blood culture medium. BacT/ALERT® Standard aerobic, BacT/ALERT® PF Plus from Biomeriux and BD BACTECTM Plus Aerobic blood culture media were mixed with blood and spiked with E. coli and were incubating for 0 **(A)**, 2 **(B)** and 16 **(C)** hours. Samples were then subjected to saponin-based host depletion protocol and ran on qPCR for the detection of host chromosomal DNA. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *p≤ 0.05, **p ≤ 0.01.

2.3.1.3 The effect of sodium polyanethole sulfonate on saponin host depletion reaction.

In order to find an explanation to the different results between the host depletion effectivity in commercial blood culture medium mixed with blood samples and blood samples

without the media, it was hypothesized the SPS, could be affecting the activity of the DNase. SPS, is the commonly used anticoagulant of commercial blood culture medium and this hypothesis was provided by other colleagues with extended experience using commercial blood culture medium and SPS. To assess if SPS could inhibit the saponin-based host depletion, different concentrations of SPS were added to blood samples which were then subjected to the host depletion protocol. The potential effect of the SPS on the host depletion efficiency was measured by calculating fold change values of the different depleted samples compared to non-depleted control samples. The addition of SPS, at all concentrations tested, prevented the saponin-based depletion method to eliminate host DNA (Figure 8). Despite the relatively low number of samples tested (n=2), the SPS effect was evident: all fold change values on samples containing SPS were close to the to the values obtained on the control non-depleted samples. As such, a decision was made that the metagenomics protocol was not suitable for its use on commercial blood cultures and from this point EDTA-coated tubes would be used as blood collection tubes.

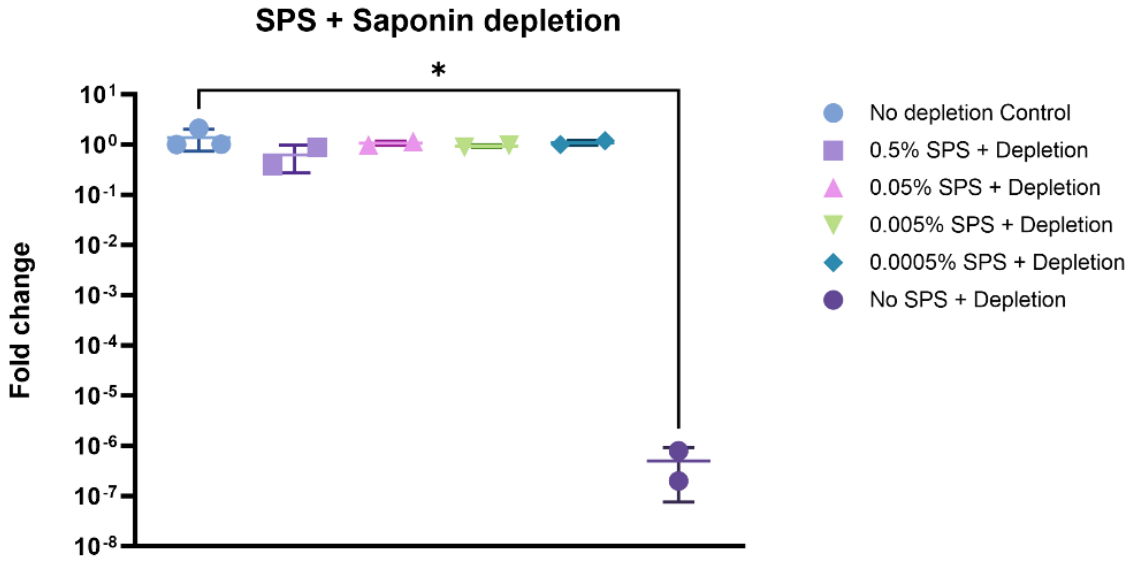


Figure 8. SPS presence effect on saponin-based depletion method. Blood was mixed with different concentrations of SPS and subjected to the saponin-based host depletion protocol. Samples were run on a qPCR for the detection of chromosomal DNA and compared to non-depleted and samples without SPS depleted. Data are means of biological replicates (n=2), ± SD. ns > 0.05 (not shown), *p≤ 0.05.

2.3.2 Optimisation of the DNA extraction protocol.

Next, the aim was to develop and optimise a DNA extraction protocol able to efficiently extract bacterial DNA even when microorganisms were present at low CFU numbers.

2.3.2.1 DNA extraction buffer.

The DNA extraction method commonly used by our research group involved pelleting the cells and then resuspending the pellet using a commercial bacterial lysis buffer (BLB) after the saponin-based depletion step. To optimise this step, different resuspension buffers were tested (Promega Lysis Buffer A or LBA, BLB and PBS) to assess which one would be most efficient for DNA extraction.

When using LBA and BLB buffers, the fold change results for both extracted *S. aureus* and *E. coli* DNA were significantly lower than their equivalent genomic DNA (gDNA), suggesting a loss of bacterial DNA. Conversely, PBS was the only extraction buffer which led to no statistically significant differences in the detection of bacterial DNA when compared to the controls (Figure 9).

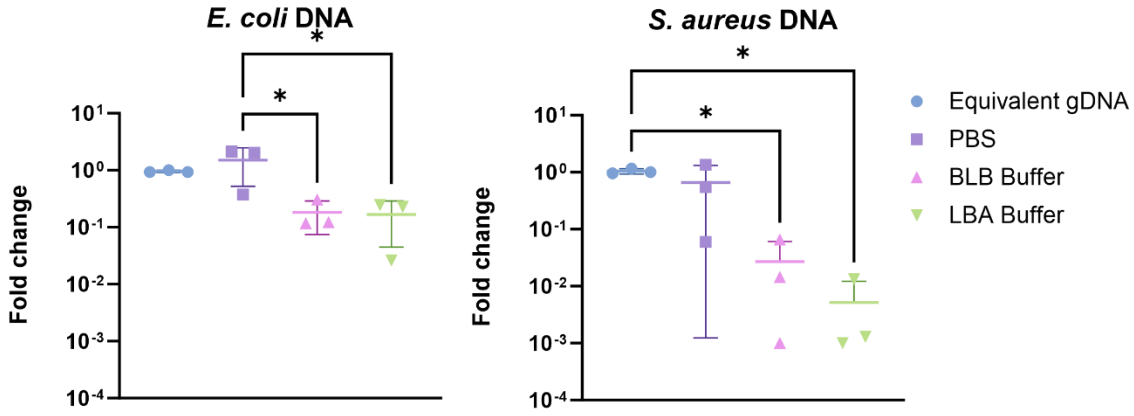


Figure 9. Buffers effect on bead beating and DNA extraction efficiency. Spiked blood samples with 10^2-10^3 CFU/mL of *E. coli* or *S. aureus* were subjected to saponin-based depletion and then the bacteria DNA was extracted by mixing the resulting cell pellets with LBA, BLB or PBS buffers and subjecting them to a bead beating step. Samples were then ran on qPCR for the detection of *E. coli* or *S. aureus* to check for possible bacterial DNA loss during the extraction process. The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *p≤ 0.05.

2.3.2.2 Testing the efficiency of PBS buffer in the extraction protocol.

In order to validate and assess the extraction protocol using PBS as resuspension buffer, blood samples were spiked with 10^1, 10^2 and 10^5 CFU of *E. coli* or *S. aureus* and then subjected to the extraction protocol involving PBS. Then the Ct values obtained by qPCR on extracted DNA were compared to the values on a control regression line. This control line was obtained by directly running on the same qPCR assay bacterial genome yields equivalent to $10^6 – 1$ cells CFU from *S. aureus* or *E. coli*. When using PBS in the protocol, no bacterial DNA was lost even when bacteria were present at concentrations as low as 1CFU/mL (Figure 10). Ct results were not significantly different compared to the values obtained from their gDNA equivalence regression line at any of the tested concentrations.

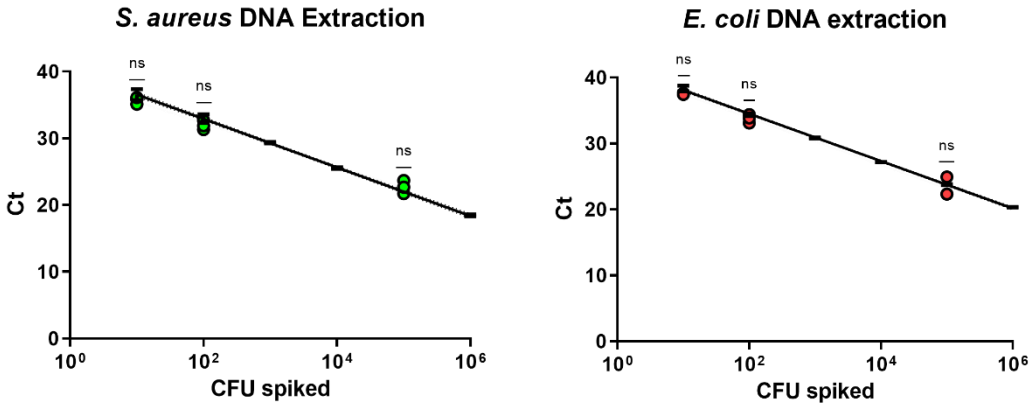


Figure 10. Assessment of bacterial DNA extraction efficiency on blood samples. Ct values obtained from blood samples spiked with 10^1, 10^2 and 10^5 of *E. coli* and *S. aureus* and subjected to DNA extraction. Individual biological replicates are represented by dots (n=3). Those were compared to the Ct results obtained on 10-fold titrated concentrations of gDNA equivalent to 10^0-10^6, which are represented as a regression line. ns > 0.05

2.3.3 Whole genome amplification.

So far, in the development of this CMg pipeline, blood samples were spiked with clinically relevant concentrations of DNA and subjected to the optimised saponin-based depletion and DNA extraction methods. However, upon measuring the concentration of the extracted DNA, it was found to be too low for use in nanopore sequencing library preparation (results not shown). Because of that, the introduction of an extra amplification step, such as a whole genome

amplification (WGA), was tested with the aim to indistinctly amplify and increase the DNA yields present in the samples after host depletion and DNA extraction.

2.3.3.1 Benefits of an extra amplification step.

The potential benefits of introducing a WGA step were tested in blood samples spiked with low concentrations of *E. coli* (50 – 100 CFU/mL). After applying the WGA step, samples were sequenced and *E. coli* reads were identified, and results were compared to those from samples without WGA. The introduction of the WGA protocol significantly increased the number of bacterial reads compared to samples without a WGA step (Figure 11). In the absence of WGA, an average of 4.8±3.8 reads were classified as *E. coli*. Subjecting the same samples to a WGA step resulted in an average of 7126.3±2460.0 reads taxonomically classified as *E. coli*, which constitutes a significant $<10^3$-fold increase in *E. coli* reads.

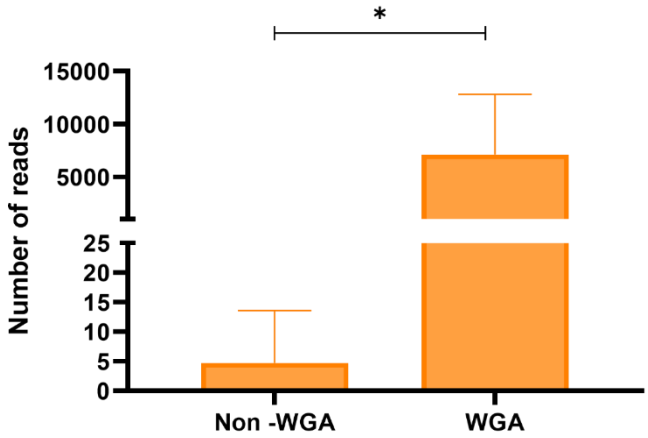


Figure 11. Effect on reads classified as *E. coli* when subjecting the sample to a WGA step. Sequencing results comparison of total number of reads taxonomically classified as *E. coli*. Blood samples spiked with 50-100 CFU of the bacteria, went through a WGA protocol and were compared to non-WGA samples. Data are means of biological replicates (n=4), ± SD. *p≤ 0.05.

Because of the low number of bacterial reads obtained from non-WGA samples, it was deduced that when bacteria are present at the lowest clinically relevant concentrations (1-5 CFU/mL), sequencing results would not able to detect any microorganism reads. This highlights the importance of incorporating an additional WGA step on the CMg pipeline.

2.3.3.2 Finding the most efficient whole genome amplification kit.

Once the need for the WGA step was established, an analysis and comparison of several commercial MDA whole genome amplification systems was conducted. This comparison aimed to determine which method could amplify the highest quantities of bacterial DNA in the shortest incubation time. For that, the REPLI-g Single Cell, REPLI-g MIDI and UltraFast Mini kits were tested by adding 500fg, 5pg and 50pg of *E. coli* gDNA (equivalent to approximately 10^2, 10^3 and 10^4 cells) as template. Overall, the REPLI-g Single Cell kit had the highest sensitivity and capability to amplify the highest quantities of bacterial DNA (Figure 12). At the highest concentration of template added (50pg), the presence of *E. coli* DNA increased approximately 10^4-fold when using either REPLI-g Single Cell or UltraFast Mini kits. At lower spiked concentrations of gDNA (500fg and 5pg), using the REPLI-g Single Cell kit always resulted in the highest amplification of *E. coli* DNA, with average fold change values of $5.2 \cdot 10^4$ and $2.3 \cdot 10^6$-fold for 500fg and 5pg samples, respectively. In contrast, the Ultrafast Mini kit, at these template concentrations, only increased the bacterial DNA reads between 10 and 100-fold. Using the REPLI-g MIDI kit, across all tested concentrations, resulted in no amplification of bacterial DNA (Figure 12).

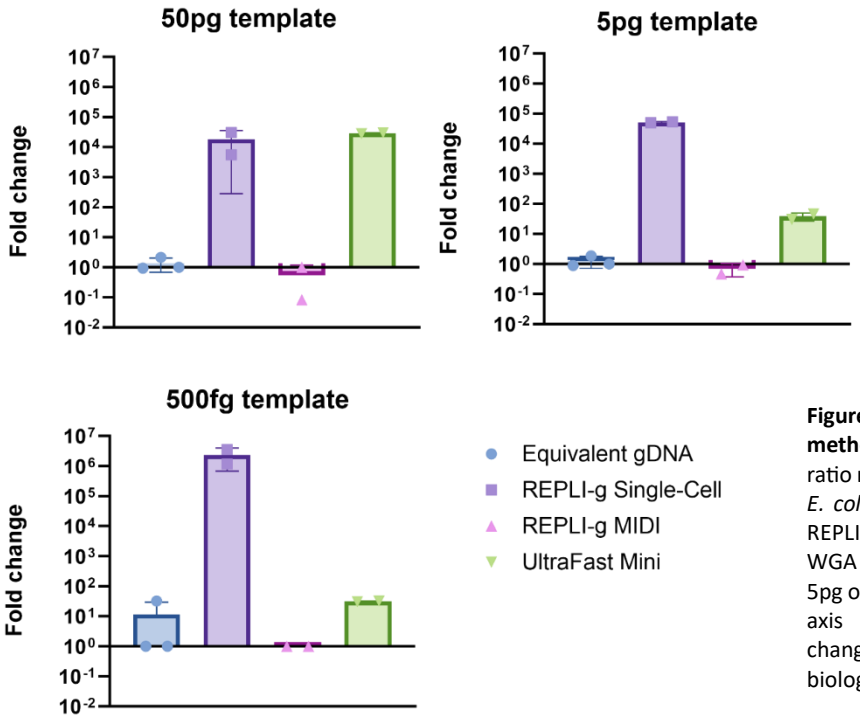


Figure 12. Commercial WGA MDA methods comparison. Fold change ratio results for the amplification of *E. coli* DNA on REPLI-g Single-cell, REPLI-g MIDI kit and UltraFast Mini WGA kits when spiked with 50pg, 5pg or 500fg of *E. coli* gDNA. The y-axis represents the log10 fold change. Data are means of biological replicates (n=2)

2.3.3.3 Optimisation of Single Cell REPLI-g whole genome amplification protocol.

Once the Single Cell REPLI-g kit was selected as the most efficient WGA protocol to amplify bacterial gDNA, especially when these are present at low concentrations, the protocol was optimised. First, template sample volume was increased from 2.5µl (as per the standard protocol described by the manufacturer) to 15µl. This modification allowed for more sample to be added to the reaction, consequently increasing the yields of bacterial DNA. To allow for such increase in sample volume, certain reagents had to be added at higher concentrations. By doing this, the aim was to generate a method that would allow to add more volume of sample and consequently increase the chances of microbial detection.

In order to test the efficacy of this change, samples with 50pg, 5pg and 500fg of extracted *E. coli* gDNA were subjected to the Single Cell REPLI-g kit using both the *standard protocol* and the new modified protocol with *increased volume protocol*. During the WGA amplification incubation, the two protocols were compared by taking samples at timepoints of 1, 1.5, 2 and 2.5 hours and measuring the gDNA concentration (only for 500fg and 5pg samples) (Figure 13A). Then, after 2.5 hours of amplification incubation time, the samples were also analysed via qPCR for the detection of *E. coli* DNA.

The concentrations obtained when using the new scaled-up protocol were similar to those obtained by the standard protocol at any of time points (Figure 13A). Moreover, at the end of the amplification incubation (2.5 hours), The average fold change ratio results were not different for the *standard protocol* and the *increased volume protocol* (Figure 13B). Thus, increasing the volume did not negatively affect the amplification efficiency of the protocol.

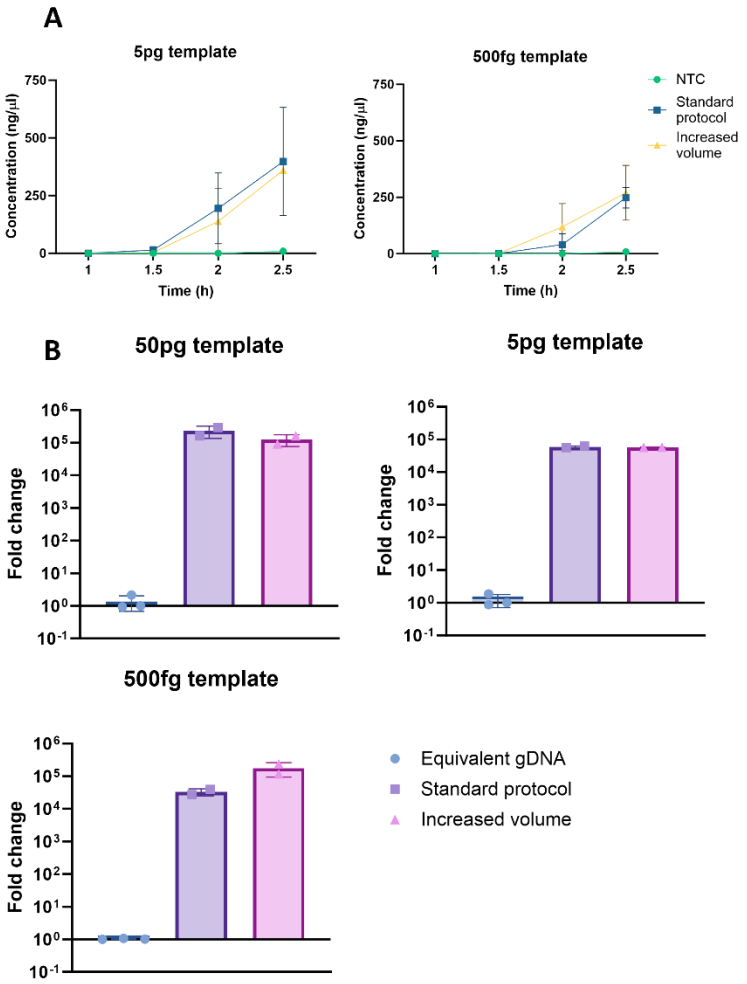


Figure 13. Comparison Standard vs Scaled-up WGA REPLI-g Single Cell protocols. Both protocols were spiked with 50pg-500fg of gDNA *E. coli* and amplification efficiency was assessed by measuring DNA concentration or fold change increased in DNA produced. (A) DNA concentrations at 1, 1.5, 2 and 2.5 hours when 5pg or 500fg of gDNA were added as template. (B) After incubating for 2.5 hours, samples from both protocols were ran on a qPCR for the detection of *E. coli* DNA. The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=2), ± SD

After successfully increasing the sample volume, further optimisation of the WGA method involved increasing the amount of denaturation buffer (and its DTT content) added to the WGA reaction. This adjustment aimed to increase the amount of DNA accessible to the polymerase (by denaturation of DNA tertiary structures), enabling the amplification of the lowest possible amount of bacterial DNA template into large yields in the shortest possible time. The WGA methodology containing the new denaturation buffer composition, which was combined with the increased volume of sample, was named *optimised protocol* and it was tested and compared to the *standard protocol*. For this experiment, different amounts of gDNA templates from *E. coli* and *S. aureus* equivalent to adding 10^3, 10^2, 10^1 and 1 CFU, were added to samples subjected to both protocols in parallel. Microorganism gDNA amplification was assessed at 0, 1, 1.5, 2 and 2.5 hours by qPCR. Finally, Ct results were converted to their CFU equivalences.

When the amount of *E. coli* gDNA template was equivalent to 10^3 *E. coli* cells, both protocols produced the same CFU equivalence values at all incubation time points. When the amount of gDNA template added was equivalent to 10^2 CFU and 10^1 CFU of *E. coli*, the optimised WGA protocol amplified DNA to approximately 10^4-fold increase compared to the initial input. Alternatively, results obtained with the standard protocol were significantly lower: for samples spiked with 10^2 CFU and 10^1 CFU the DNA were only amplified a further 10^2 and 10^1-fold, respectively. For both protocols the amount of amplified DNA reached its peak at the 1.5 hours timepoint; longer incubation times did not significantly increase the quantity of bacterial yield. Finally, amplification could not be detected with either protocol when samples were spiked with gDNA equivalent to 1 CFU (Figure 14).

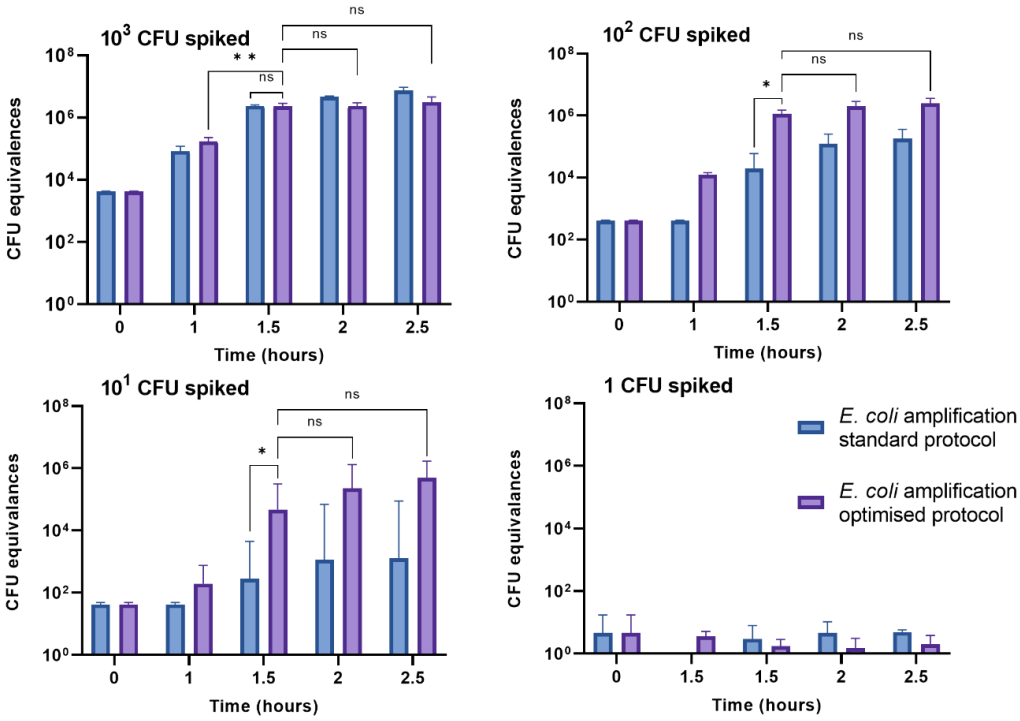


Figure 14. WGA *'optimised* protocol' amplification efficiency when spiked with *E.coli* gDNA. WGA protocols were spiked with *E. coli* gDNA equivalent to 10^3, 10^2, 10^1 and 1 CFU. At 0, 1, 1.5, 2 and 2.5 hours samples were ran on a qPCR to obtained the ampligfied CFU equivalences to compared the amplification efficiency of both protocols. Data are means of biological replicates (n=3), ± SD. ns > 0.05 *p≤ 0.05, **p ≤ 0.01.

In a similar experiment using gDNA equivalent to 10^3 cells of *S. aureus*, the *optimised protocol* produced significantly more (10-100X) gDNA material that the *standard protocol*, after 1.5

hours of incubation. After that, additional incubation time did not produce a significantly higher amount of amplified *S. aureus* gDNA. Initial template amounts equivalent to 10^2 CFU and 10^1 CFU were amplified approximately 10^3 and 10^2-fold, respectively, with the *optimised protocol* (Fig. 4B). Alternatively, amplification results obtained with the standard protocol were significantly lower with only the 10^2 CFU input samples DNA being amplified approximately 10X and no amplification at 10^1 and 1 CFU input added samples. Similar to *E. coli* - the amount of DNA amplified with the WGA optimised protocol reached its peak at the 1.5 hour for S. aureus. (Figure 15).

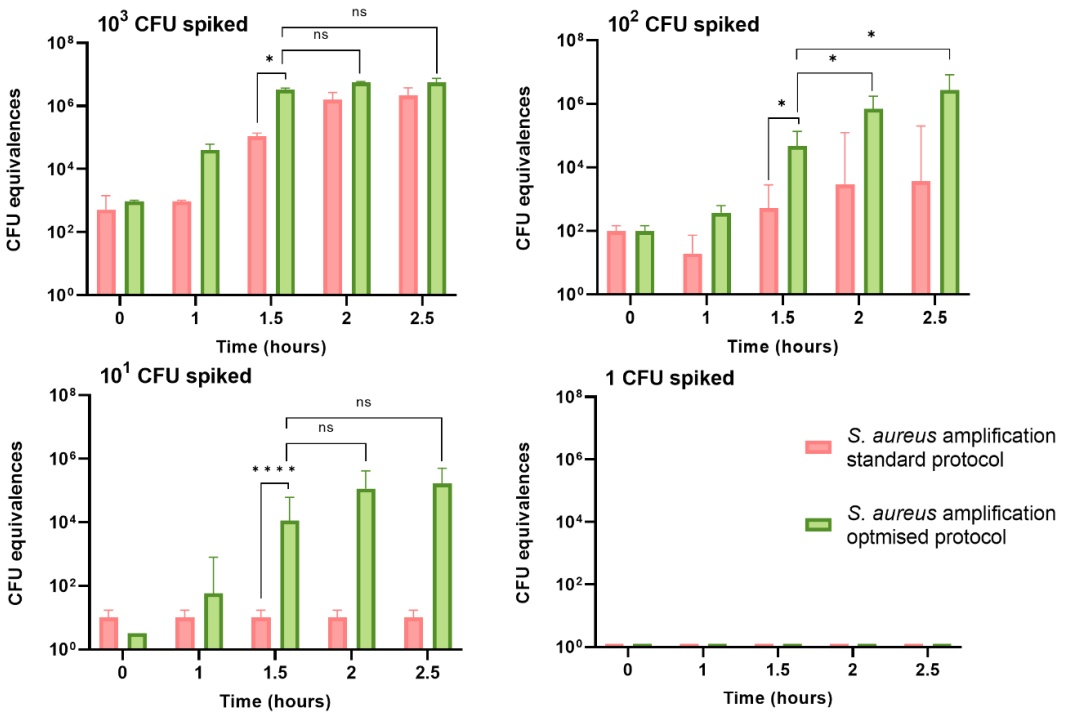


Figure 15. WGA *'optimised protocol'* amplification efficiency when spiked with *S. aureus* gDNA. WGA protocols were spiked with *S. aureus* gDNA equivalent to 10^3, 10^2, 10^1 and 1 CFU. At 0, 1, 1.5, 2 and 2.5 hours samples were ran on a qPCR to obtained the ampligfied CFU equivalences to compared the amplification efficiency of both protocols. Data are means of biological replicates (n=3), ± SD. ns > 0.05 *p≤ 0.05, **p ≤ 0.01. ***p ≤ 0.001, ****p ≤ 0.0001.

2.3.4 Amplified DNA debranching optimization.

While performing sequencing runs on samples that had been subjected to the optimised WGA protocol, an unusual decrease in the number of active pores was observed on the flow cell after a few minutes into the sequencing run (results not shown). It was hypothesised that this

decrease could be caused by obstruction of the pores resulting from the branching of DNA during WGA. Thus, a DNA debranching step was introduced into the pipeline.

2.3.4.1 Testing different debranching strategies.

To test the pore blockage hypothesis and solve this issue, experiments were conducted involving the incubation of amplified DNA with different digestion enzymes before library preparation. Samples containing *E. coli* gDNA were subjected to WGA protocol and then subjected to one of three different debranching protocols: incubation with T7 endonuclease I, incubation with nuclease S1, and incubation with a combination of T7 and S1 nucleases. The debranching efficiency was measured by annotating the percentage of active pores every 5 minutes during 50 minutes of sequencing time.

The lack of a debranching step in post-WGA samples caused a drastic decrease in active pores after 50 minutes of sequencing, dropping from an initial 20.6% to only 7.9%, representing a 12.7% reduction. This phenomenon was also observed in all the debranched samples that suffered from drastic reductions in pore activity in short time (Figure 16A). All debranching methods were compared by normalising the percentage of active pores of each sample against the percentage of active pores recorded at minute 5 (Figure 16B). After 50 minutes of sequencing, the samples treated with Nuclease S1, T7 Endonuclease I and both nucleases experienced reductions in active pores of 20.7%, 10.1% and 9.4%, respectively. These results were very similar to the ones obtained from non-debranched samples, in which the number of active pores experienced a decrease of 12.7% after 50 minutes. Note that all the debranched strategies were only tested once (n=1) and consequently results should be interpreted with caution. The application of a debranching step with the T7 Endonuclease I resulted in a more moderate reduction in pore blocking compared to the non-treated control. Because of that, the use of T7 Endonuclease I as a debranching method

was implemented for the rest of the study as a quick and easy method with potential benefits to mitigate the pore blocking issue caused by the WGA step.

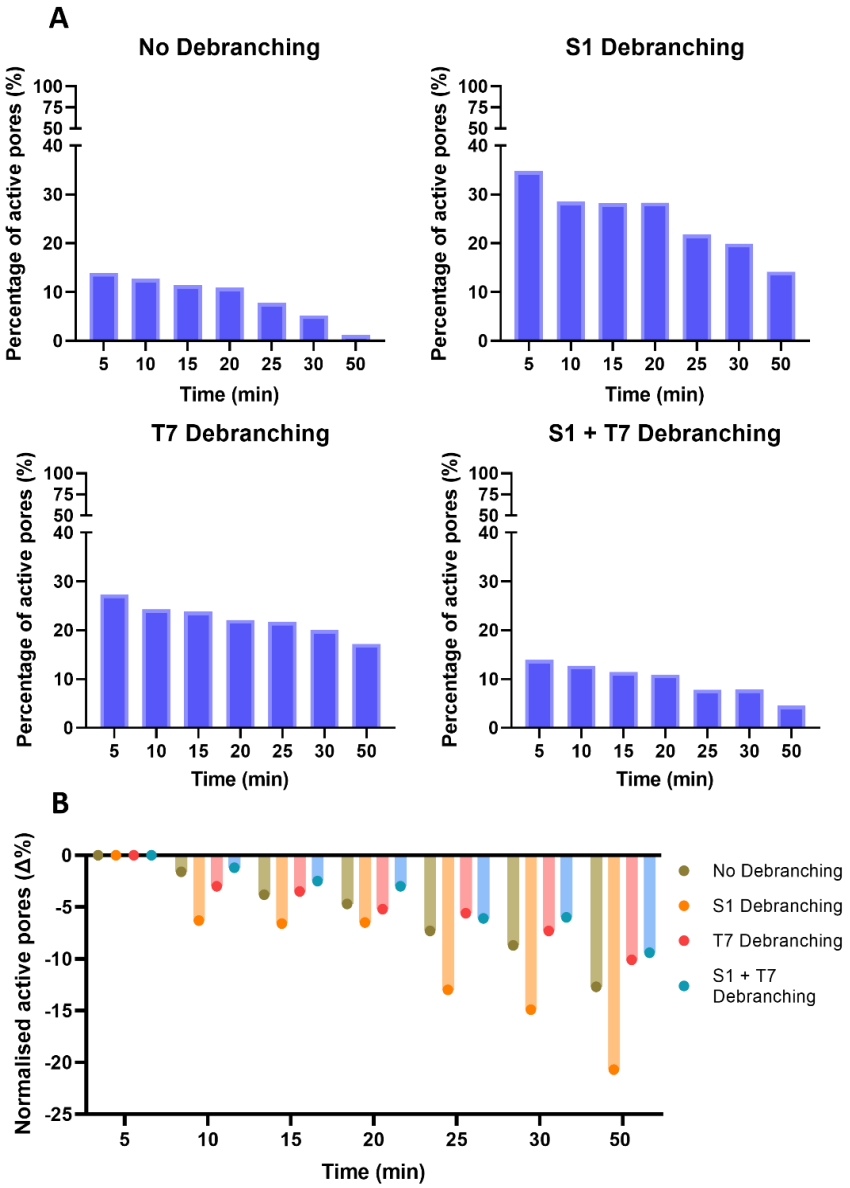


Figure 16. Debranching strategies comparison effect on pore activity. Samples that underwent the WGA protocol were subjected to three different debranching protocols (Nuclease S1, T7 Endonuclease I and a combination of both nucleases) or no protocol. **(A)** Samples were sequenced and the percentage of sequecning active pores was recorded every 5 minures during a 50 minutes sequencing run. **(B)** Δ% of pore activity obtained by comparing the different % obtained at different timepoints to % obtained at the first 5 minutes. n=1

2.3.5 Limit of detection experiments on optimized extraction and whole genome amplification methods.

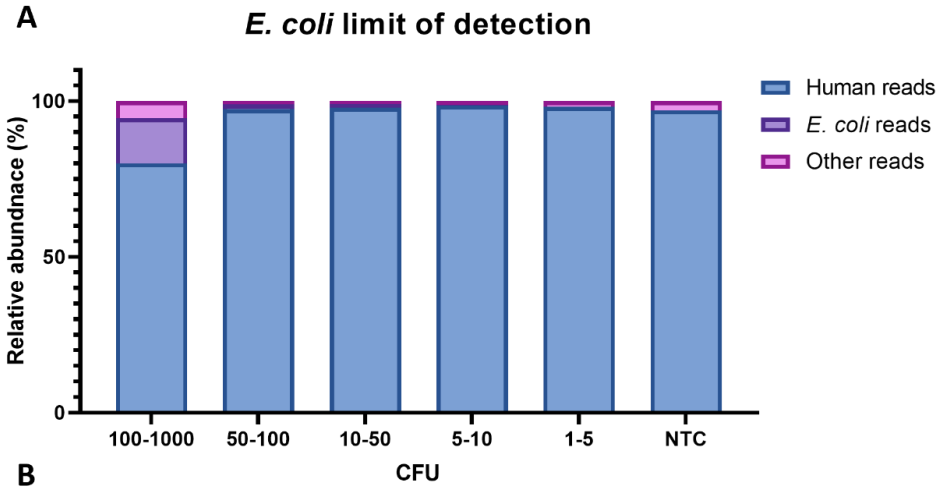
Limit of detection (LoD) experiments were performed to assess the overall sensitivity of the pipeline when bacterial samples were present at low clinically relevant concentrations in whole blood samples. This was a way to assess how the different described optimisation steps impacted the overall sensitivity of the CMg pipeline. The first iteration of the pipeline consisted in subjecting

the samples to the standard depletion method and the optimized DNA extraction and WGA previous protocols. Then, samples were subjected to T7 endonuclease I debranching protocol and library preparation. Since the DNA yields (from WGA and debranching) were relatively low (1-100ng/µL), the samples had to be subjected to a PCR-based library preparation (SQK-RPB004). Finally, the libraries were sequenced on a MinION for 3 hours and the results were collected and analysed. Samples were deemed positive for infection if the abundance *E. coli* reads was ≥1% over the total number of classified reads.

When blood samples were spiked with *E. coli*, this microbial species reads were detected on all spiked samples, but the abundance threshold ≥1% was only met by the samples spiked with 100-1000 (14.5%), 50-100 (1.5%) and 10-50 CFU (1.3%). This demonstrated a limit of detection of 50-250 CFU/mL (Figure 17A). Despite the fact that the research group had previously demonstrated that the standard saponin method was capable of depleting more than 99.9% of chromosomal host DNA [136], on this experiment samples there was still an overrepresentation of human reads compared to bacterial reads. Except for the sample with the highest spiked amount of bacteria (100-1000 CFU), human reads represented >90% of the total number of classified reads (Figure 17A).

Moreover, there was a total absence of reads classified as *E. coli* in the no-template control (NTC). Instead, the NTC samples had bacterial reads typically associated with environmental contamination, such as species from the *Acinetobacter*, *Shigella* and *Staphylococcus* genera (data not shown). While these non-*E. coli* species were also found in the rest of the samples, none of them were detected as infectious agent (none of them accounted for ≥1% of the total reads). In terms of *E. coli* abundance, number of reads and genomic coverage, the results correlated with the spiked inoculum , even for samples that did not meet the abundance threshold (Figure 17B).

A



B

Sample (CFU/mL)	Total number of classified reads	Microbe Reads	Total number *E. coli* reads	*E. coli* reads abundance (%)	Genome Coverage (X)
100-1000	3727	746	539	14.5	0.04
50-100	5272	86	81	1.5	0.006
10-50	1580	23	20	1.3	0.004
5-10	1613	7	6	0.4	0.0002
1-5	730	2	1	0.1	0.00008
NTC	146	40	0	0	-

Figure 17. Optimised CMg pipeline LoD experiment using samples spiked with *E. coli*. 200µl blood samples were spiked with 100-100, 50-100, 10-50, 5-10, 5-10 CFU of *E. coli* and non-spiked (NTC). Then samples were subjected to the standard saponin-based host depletion method, DNA extraction and WGA optimised protocols. Samples were then subjected to library preparation and sequencing. **(A)** Taxonomically classified reads were classified on 'human', '*E. coli*' and any bacterial read that was not *E. coli* ('other reads'). Relative abundance of each of the three categories were obtained and samples were classified as positive for bacterial infection if any of the species detected relative abundance reads was ≥1%. **(B)** Table of reads metric obtained at the different spiked samples: microorganism abundances results and average coverage against the *E. coli* reference. n=1.

2.3.6 Host depletion method optimisation.

In initial proof-of-concept LoD experiments, the bacterial abundance values were still too low to be considered positive for infection when bacteria were spiked at clinically relevant concentrations. The low sensitivity results highlighted the importance of improving the host depletion step protocol as the large abundance of human reads was preventing the obtention of more bacterial reads. This was addressed by first testing the saponin-based host depletion method previously used by members of the group on blood samples and detailed in section 2.3.1.1. Initial

efforts were put into detecting the limitations of this method and increasing the amount of host DNA being depleted to ultimately increase the overall sensitivity of the method.

2.3.6.1 Standard saponin-based host depletion and the mitochondria problem.

The same samples that were subjected to the standard saponin-based host depletion method in section 2.3.1.1 were sequenced in order to investigate further the composition of the remaining host DNA present in the sample. The results were evaluated by taxonomically classifying the obtained reads. Non-depleted blood samples contained 99.1% host DNA of the total population of reads. The saponin-based host depletion method reduced the host DNA to 31.1% of the total population of reads , while the proportion of bacteria reads increased from 0.86% to 68.8% (Figure 18A). Despite the large increase in the proportional representation of bacterial DNA in host-depleted samples, host DNA still constituted an extensive component of the processed samples. The composition of the remaining human DNA was investigated in long-read sequencing data and categorised as chromosomal or human mitochondrial DNA. Comparing samples before and after saponin-based depletion, 88.8% of the total population of reads matched the DNA of the human mitochondrion in depleted samples, whereas human mitochondrial DNA represented only 0.11% of the classified reads in non-depleted samples (Figure 18B).

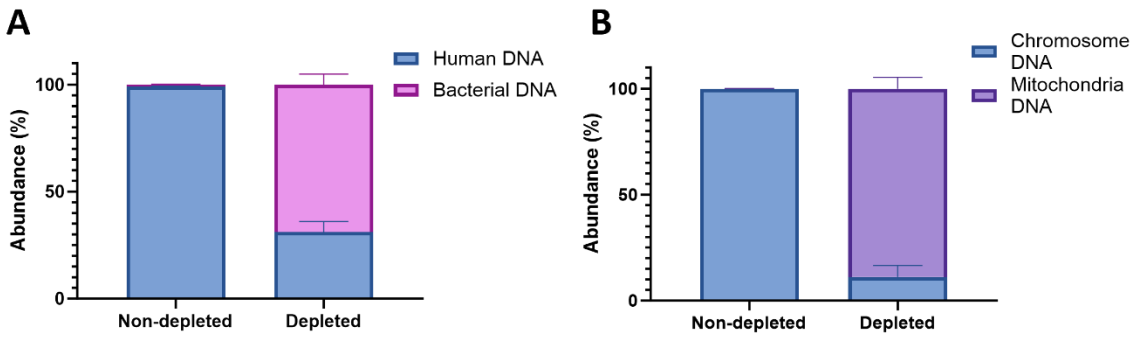


Figure 18. Host mitochondria DNA composition before and after saponin-based host depletion. Samples were spiked with 50-100 CFU/mL *E. coli* and subjected to the standard saponin-based depletion method, followed by DNA extraction and nanopore sequencing. All sequenced reads were taxonomically classified. (A) Relative abundance (%) of human chromosomal and bacterial (any species) DNA in samples after host depletion compared to control samples not subjected to host depletion (B) Relative abundance of human chromosomal and mitochondrial DNA after host depletion compared to control samples not subjected to host depletion. Data are means ± SD. n= 3.

2.3.6.2 Testing the viability of different strategies to increase the human mitochondria depletion.

In initial results using the saponin-based method on human blood tissue samples, there was still a large abundance of both chromosomal and mitochondrial DNA. In order to increase the method's overall sensitivity, four different strategies were tested in combination with the saponin-based host depletion method to reduce the presence of chromosomal and mitochondrial DNA: (1) beads that selectively bind to leukocytes (CD45 Beads) or mitochondria (Anti-TOM22 beads), (2) addition of calcium to open pores on the mitochondria membrane, (3) addition of enzymes (Pneumolysin or Phospholipase D) that are able to digest the mitochondria membrane, and (4) the separation of blood fractions to reduce the presence of leukocytes in the samples. All strategies were tested on blood spiked with ~10^3 -10^4 CFU of *E. coli* in combination with the saponin-based depletion method. DNA was extracted from samples both before and after the application of these approaches, and qPCR was run to obtain fold change ratio values for the presence of human chromosomal DNA, human mitochondrial DNA and *E. coli* DNA.

2.3.6.2.1 Negative selection of leukocytes using Dynabeads® CD45.

Dynabeads® (anti-CD45 magnetic beads) can selectively bind to leukocytes from human fluid tissue samples. The application of these beads onto blood could reduce the mitochondria-containing leukocyte population prior to saponin-based depletion. However, the application of the Dynabeads® had no added effect on the chromosomal DNA depletion, whether alone or with the saponin-based depletion. When looking at mitochondrial DNA, the application of the beads alone significantly reduced its presence when compared to the untreated blood control (approximately a 10-fold reduction; Figure 19). However, when combined with the saponin-based method, the beads did not have an increased effect in comparison to samples treated with saponin-only. Finally, adding

the beads alone or in combination with the saponin method had no detrimental effect on *E. coli* DNA (Figure 19).

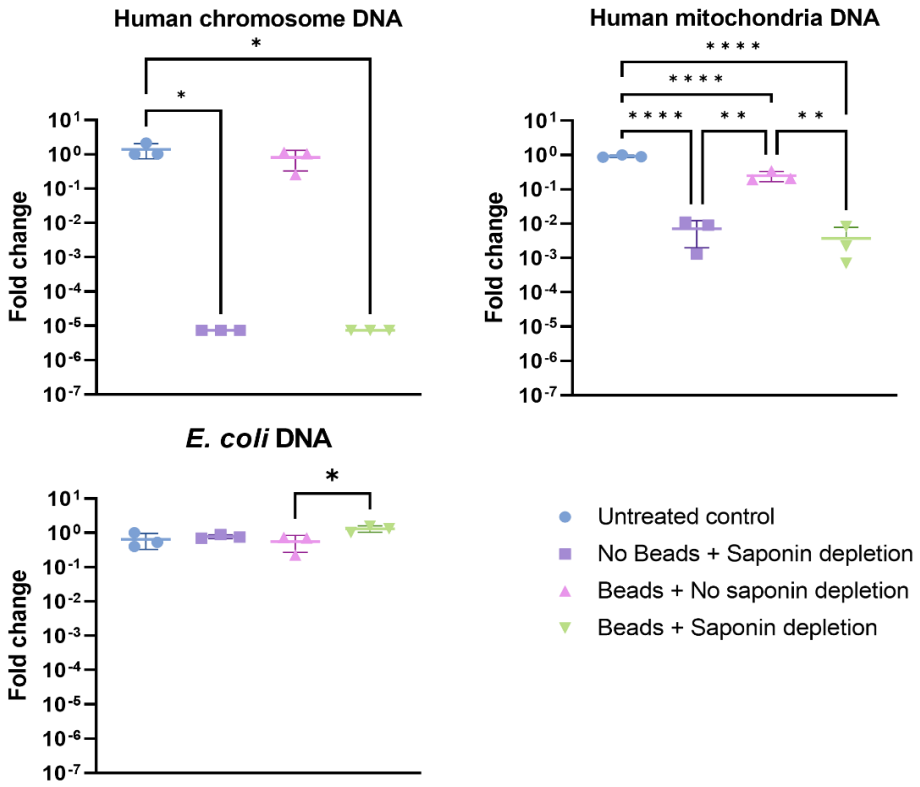


Figure 19. Dynabeads® effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected to the saponin-based host depletion or Dynabeads® protocols on its own or combined. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the log_{10} fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *p≤ 0.05, **p ≤ 0.01. ***p ≤ 0.001, ****p≤ 0.0001.

2.3.6.2.2 Generation of pores on mitochondria membranes using calcium.

Calcium has also been reported as a molecule than can disrupt the mitochondrial membrane via the generation of pores, through which mitochondrial DNA is released. The addition of calcium to generate the pores during the saponin-based method (which includes a DNase treatment), could potentially reduce the mitochondrial DNA presence. Because of that, calcium was added at different concentrations as $CaCl_2$ (5, 50 and 500 mM) into the saponin-based depletion method. As expected, the addition of calcium to the saponin-based method had no significant effect on the chromosol DNA reduction (Figure 20). However, the addition of 5mM and

50mM of $CaCl_2$ also had no effect on mitochondrial DNA when compared to depletion-only controls. Moreover, the addition of calcium at the highest concentration (500mM) caused a significant reduction in the efficacy of the host mitochondrial DNA depletion.

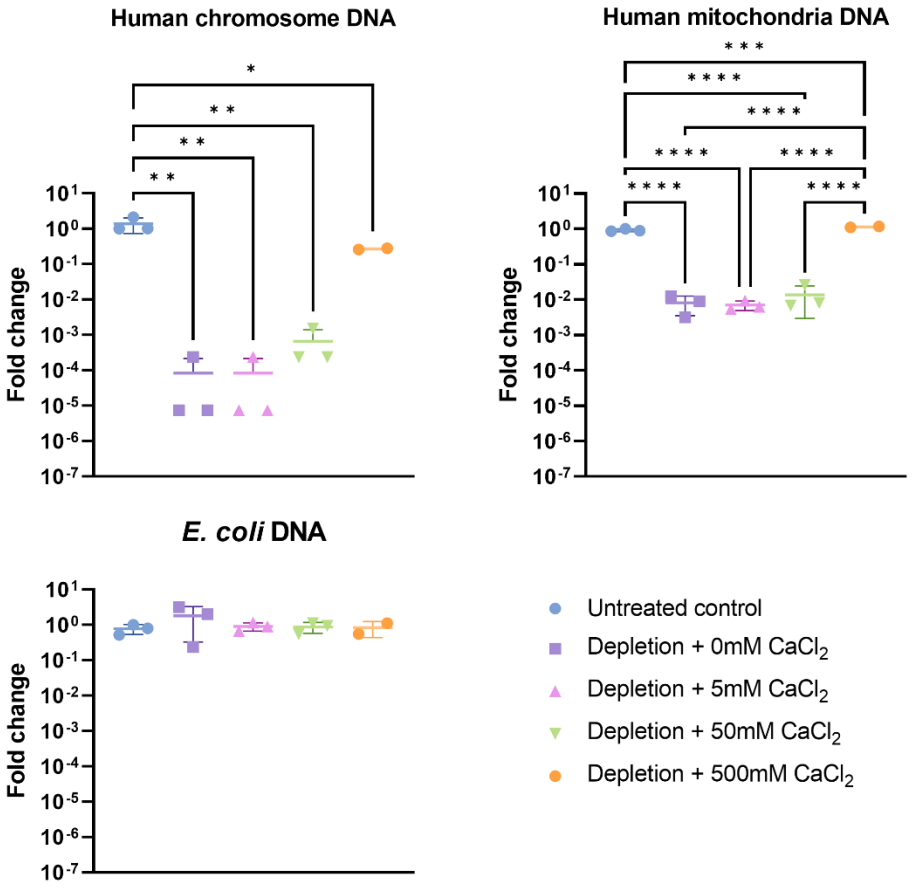


Figure 20. **Calcium ($CaCl_2$) effect on host DNA and spiked bacteria.** Spiked blood samples were untreated or subjected to the saponin-based host depletion mixed with 0, 5, 50 or 500mM of $CaCl_2$. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the log_{10} fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *p≤ 0.05, **p≤ 0.01. ***p ≤ 0.001, ****p ≤ 0.0001

2.3.6.2.3 Pneumolysin enzyme can generate pores on mitochondrial membranes.

Pneumolysin is a cholesterol-dependent cytolysin protein with the capacity to selectively form pores on the mitochondrial and cellular membranes of host cells. To take advantage of this property and improve human DNA depletion in blood samples, Pneumolysin was added at concentrations of 1µg/mL or 2µg/mL in combination with the saponin-based solution. None of the added concentrations lead to an improvement in host chromosomal or mitochondrial DNA

depletion compared to saponin-only treatment (Figure 21). No significant differences were observed in bacterial DNA detection under any of the conditions.

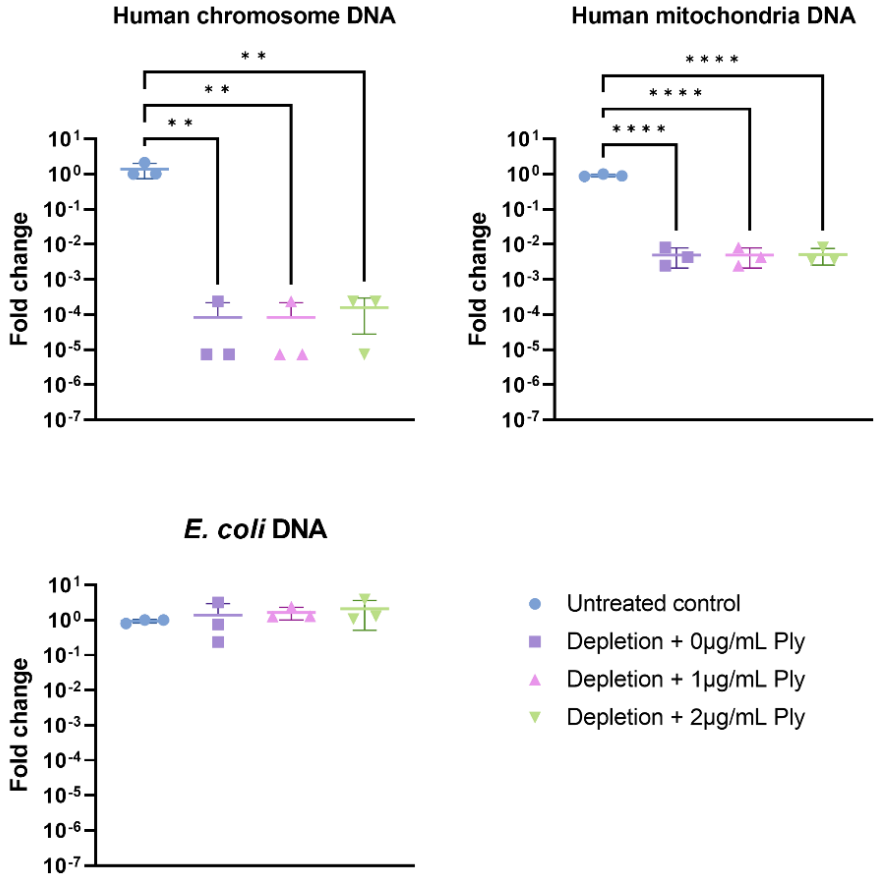


Figure 21. Pneumolysin effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected to the saponin-based host depletion alone or combined with 1µg/mL or 2µg/mL of Pneumolysin (Ply). Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01. ***p ≤ 0.001, ****p ≤ 0.0001.

2.3.6.2.4 Mitochondrial depletion using magnetic Miltenyi anti-TOM22 beads.

Another approach consisted of the depletion of mitochondria from the sample after they had been released into the media due to leukocytes membrane disruption during the saponin-based treatment. Miltenyi beads, which can bind to the mitochondrial outer membrane thanks to their anti-TOM22 (mitochondrial membrane protein channel) antibodies, were used for this purpose. The use of the Miltenyi anti-TOM22 beads on saponin-depleted samples led to similar reduction in chromosomal DNA compared to use of saponin only. For mitochondrial DNA depletion,

addition of anti-TOM22 beads was approximately 5X more effective than the use of saponin-based depletion only (Figure 22). However, the application of the anti-TOM22 beads protocol also caused a significant reduction in *E. coli* DNA.

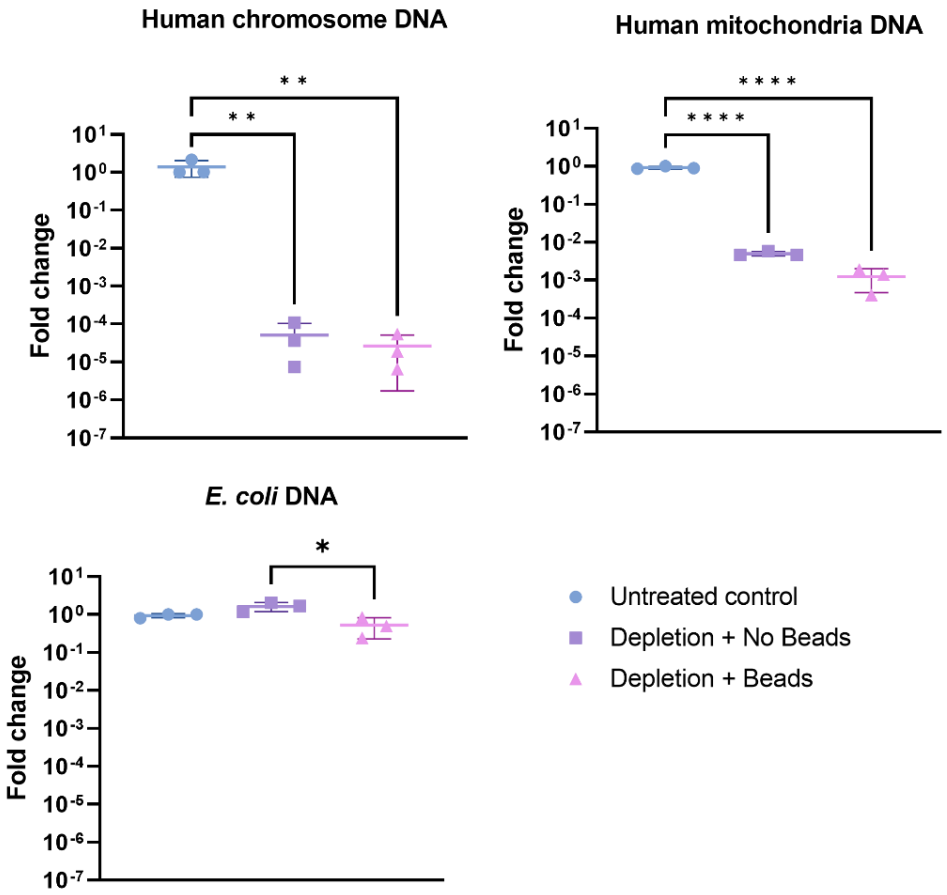


Figure 22. Miltenyi effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected to the saponin-based host depletion alone or combined with the anti-TOM22 beads. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *p≤ 0.05, **p≤ 0.01, ****p ≤ 0.0001.

2.3.6.2.5 Phospholipase D (PLD) can disrupt cell membranes.

Phospholipase D (PLD) enzyme can interact with the phospholipids present on eukaryotic membranes and disrupt them. Because of this, this enzyme was also tested as a potential addition to the saponin-based host depletion method to increase the depletion of host DNA. The addition of PLD to the saponin-based reaction did not cause a significant improvement in the depletion of

either chromosomal or mitochondrial DNA when compared to saponin-only treated samples. Adding PLD had no effect on *E. coli* DNA either (Figure 23).

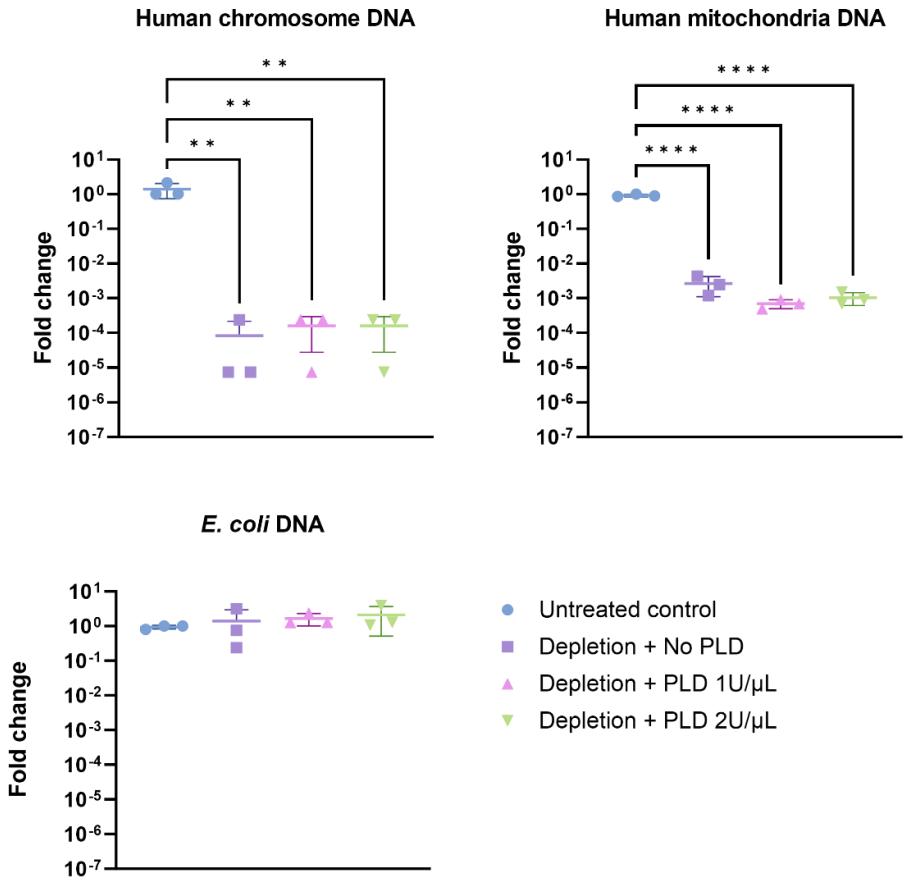


Figure 23. Phospholipase D effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected to the saponin-based host depletion alone or combined with 200U or 400U of Phospoholipade D (PLD). Samples were then DNA extracted and ran on qPCR for the detection of E. coli and host (mitochondrial and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01, ****p ≤ 0.0001.

2.3.6.2.6 Differential centrifugation can separate bacteria and host cells in different fractions.

A differential centrifugation procedure can separate whole blood into three fundamentally distinct layers: plasma fraction, buffy coat (including leukocytes and platelets), and red blood cell fraction (erythrocytes). To assess the effect of applying this method in our pipeline, whole blood samples spiked with bacteria were subjected to differential centrifugation. Only the plasma fraction was recovered, potentially eliminating a large proportion of leukocytes and platelets (and consequently host DNA). The retention of the plasma fraction caused an approximate 10X

reduction in chromosomal DNA and a 100X reduction in mitochondrial DNA when compared to untreated controls (Figure 24). However, differential centrifugation in combination with the saponin-based method did not significantly improve the reduction of host DNA achieved by the saponin-based method only. When comparing both methods applied individually, the saponin-based method was still more effective at depleting human chromosomal and mitochondrial DNA. The elimination of the buffy coat and red blood cell fractions alone did not reduce the presence of bacteria, as *E. coli* DNA fold change values were not significantly different from those obtained in untreated samples. Thus, after centrifugation, all the bacteria remained in the plasma fraction.

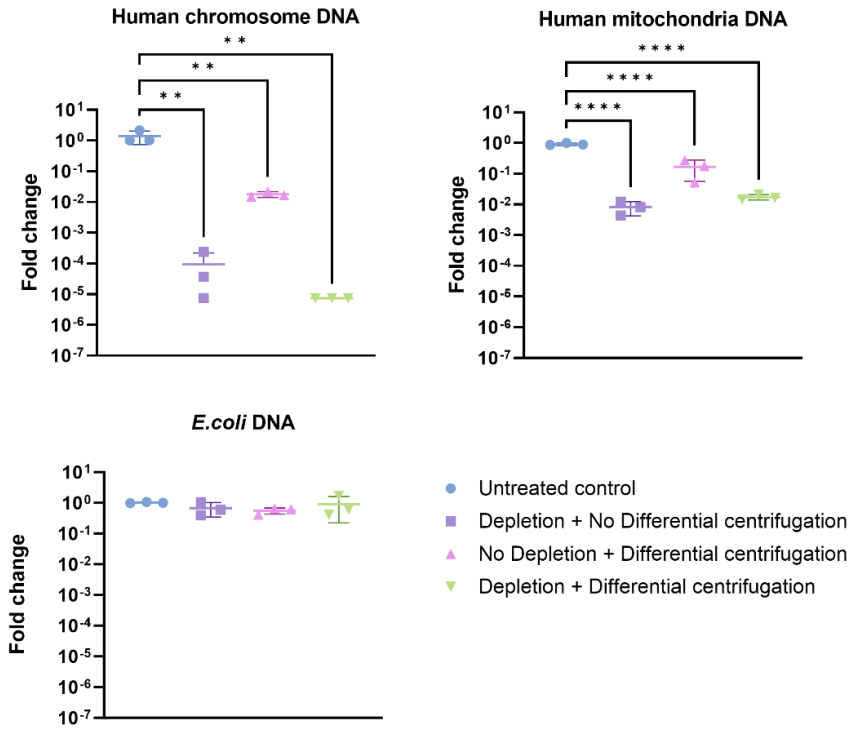


Figure 24. A differential centrifugation effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected to a differential centrifucation step. The plasma fraction was kept processed or subjected to the saponin-based depletion method. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01, ****p ≤ 0.0001.

2.3.6.2.7 Modification of saponin concentration.

Increasing the saponin concentration was also tested as a possible strategy to improve the depletion of host DNA. When compared to the standard 1% saponin treatment, 3% or 6% saponin

concentrations depleted the chromosomal DNA at a significantly similar efficiency. However, when increasing the concentration to 3% or 6% mitochondrial DNA presence was significantly decreased by approximately 10X when compared to the protocol with 1% saponin (Figure 25). When comparing the host DNA depletion of the 3% and 6% protocols, there were no significant differences for both chromosomal and mitochondrial DNA fold change values. However, increasing the saponin concentration to 6% caused a significant loss of bacterial DNA, *E. coli* DNA presence was reduced approximately 10-fold compared to the untreated control. At the same time, 3% saponin treatment had no negative effect on the detection of *E. coli* DNA.

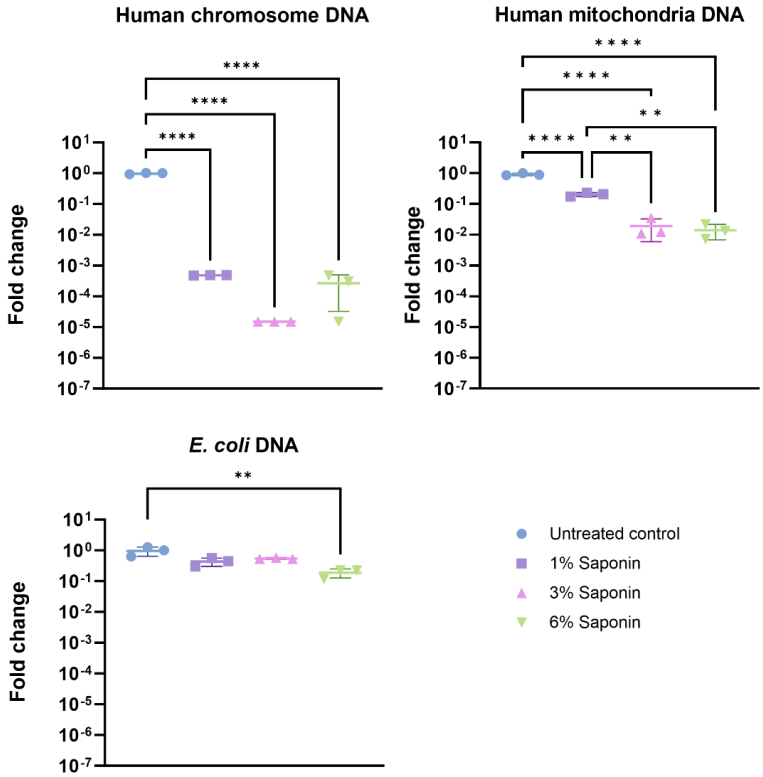


Figure 25. Increasing saponin concentration effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected saponin-based porotocl with saponin at three different concentrations: 1%, 3% and 6%. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01, ****p≤ 0.0001.

2.3.6.2.8 Final decision on how to improve the host DNA depletion method.

After assessing multiple different strategies to further reduce the presence of host DNA in combination with the saponin-based host DNA depletion protocol, the only successful strategy was

increasing the saponin concentration to 3%. This approach led to a decrease in host DNA concentration while not affecting bacterial DNA detection. Thus, treatment with saponin at a 3% concentration became the standard depletion method that was used for the rest of the study.

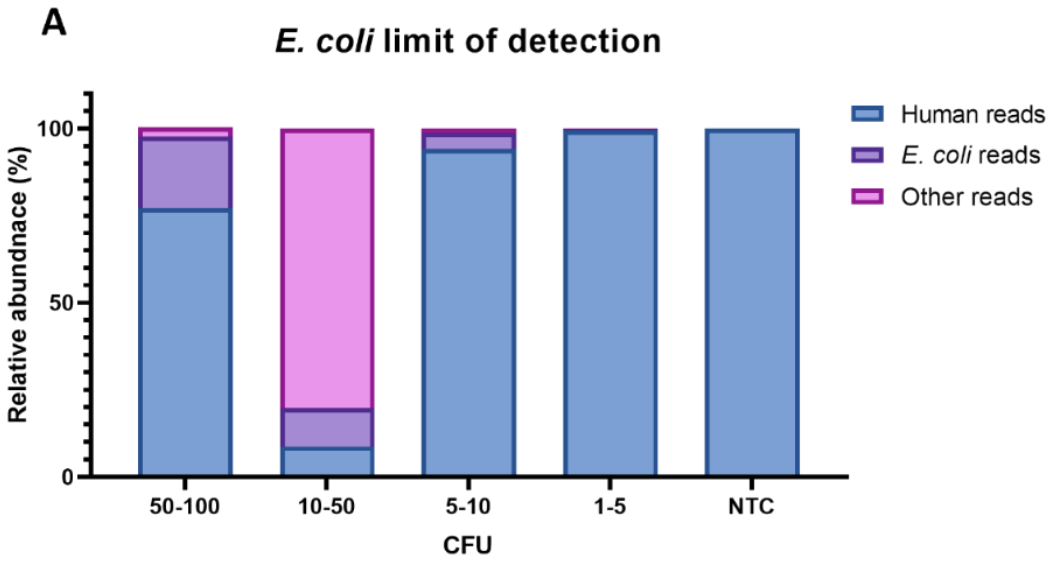
2.3.7 Limit of detection experiments on optimized saponin-based host depletion protocol.

With the initial proof of concept LoD experiments (section 2.3.5) in which all the protocols were combined (Figure 5), the CMg pipeline could detect taxonomically classified reads of spiked bacteria at all concentrations. Nonetheless, the abundance values were still too low to be considered positive for infection when bacteria were spiked at clinically relevant concentrations. To increase the sensitivity, the new optimised saponin-based depletion method was modified from the CMg pipeline, while the other methods remained unchanged. The effect of increasing the host depletion on the overall sensitivity of the pipeline was tested with an LoD experiment.

The improved host depletion protocol resulted in the detection of *E. coli* (abundance ≥1%) for all samples except the 1-5 CFU input samples, revealing a limit of detection of 25-50 CFU/mL (Figure 26A). The approximate 10X increase in mitochondrial depletion (observed in section 2.3.6.2.7) also resulted in an approximate 10X improvement in the CMg pipelines' s sensitivity when compared to previous iterations of the pipeline (limit of detection of 250-50 CFU/mL). Despite the notable improvement in the depletion of host DNA and the overall pipeline sensitivity, except for the 10-50 CFU spiked samples, there was still a large abundance of human DNA reads. In the 10-50 CFU spiked sample, approximately 78% of the reads were identified as *Serratia marcescens* (possible contamination). As per previous iteration, although the abundance results did not report all samples as positive*, E. coli* reads could be detected even at the lowest spiked concentration sample (1-5 CFU). At the same time, no *E. coli* reads were detected on the NTC sample (Figure 26B).

The decrease in the amount of host DNA also caused a noticeable increase in the number of reads classified as *E. coli* across all samples, ranging from approximately 11,000 reads at the highest concentration to 150 reads at the lowest (approximately 5-10X higher than previous LoD experiment in section 2.2.4). This increase in the number of obtained reads impacted the average genomic coverage, which was higher compared to the previous iteration. However, the coverage values were still noticeably low, as they were lower than 1X for all samples (Figure 26B).

A



B

Sample (CFU/mL)	Total number of classified reads	Microbe reads	Total number *E. coli* reads	*E. coli* reads abundance (%)	Genome coverage (X)
50-100	54727	12531	11119	20.3	0.88
10-50	27807	25367	3039	10.9	0.23
5-10	66674	3839	2835	4.5	0.19
1-5	57902	297	150	0.25	0.013
NTC	66045	183	0	0	-

Figure 26. Optimised CMg pipeline (including the new optimised saponin-based depletion method) LoD experiment suing samples spiked with *E. coli.* 200µl blood samples were spiked with 50-100, 10-50, 5-10, 5-10 CFU of *E. coli* and non-spiked (NTC). Samples underwent optimsed 3% saponin depletion method , DNA extraction and WGA optimised protocols. Samples were then subjected to library preparation and sequencing. **(A)** Taxonomically classified reads were classified on 'human', *'E. coli'* and any bacterial read that was not *E. coli* ('other read's). Relative abundance of each of the three categories were obtained and samples were classified as positive for bacterial infection if any of the species detected relative abundance reads was ≥1%. **(B)** Table of reads metric obtained at the different spiked samples: microorganism abundances results and average coverage against the *E. coli* reference. n=1**.**

Despite the low coverage results obtained in all samples, a closer inspection to the coverage map obtained at different spiked concentrations (Figure 27) indicated evenly distributed (non-bias) coverage across the whole *E. coli* genome and no bias towards a specific genomic region observed under any of the concentrations.

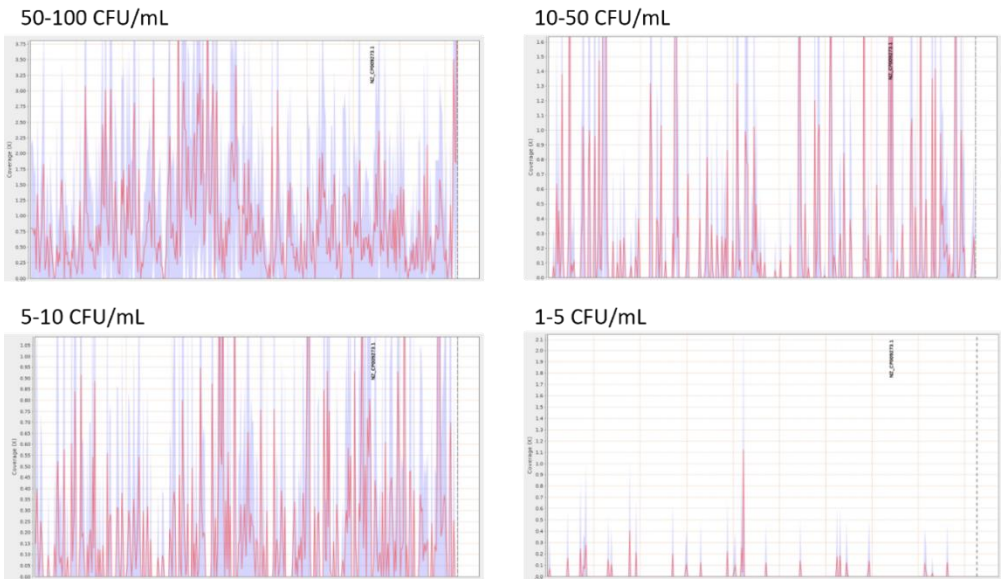


Figure 27. Coverage was evenly distributed across the reference genome. Blood samples that were spiked with 1-100 CFU and subjected to the iteration of the pipeline with the optimsed (3%) saponin-based depletion step. Sequecning results of the different concentrations tested were mapped against their *E. coli BW25113* reference using minimap2 to visualize the coverage obtained across the whole genome (in blue). n=1.

2.3.8 Further optimisation of the depletion protocol

During the LoD experiments using the new pipeline version, *E. coli* reads were detected in blood samples when it was present at clinically relevant levels. However, further enhancements were necessary to amplify the abundance of bacterial reads in spiked samples.

2.3.8.1 Scaling-up the saponin-based depletion volume protocol.

In order to enhance the chances of detecting bacteria in blood samples, which could improve the overall sensitivity of the pipeline, the standard depletion protocol (saponin at a 3% final concentration) was modified to increase the processed volume of blood from 200µl to 1mL. Untreated samples were used as controls. When compared to the standard protocol, this

modification did not cause any statistically significant differences in the detection of host DNA (chromosomal and mitochondrial) or *E. coli* DNA (Figure 28).

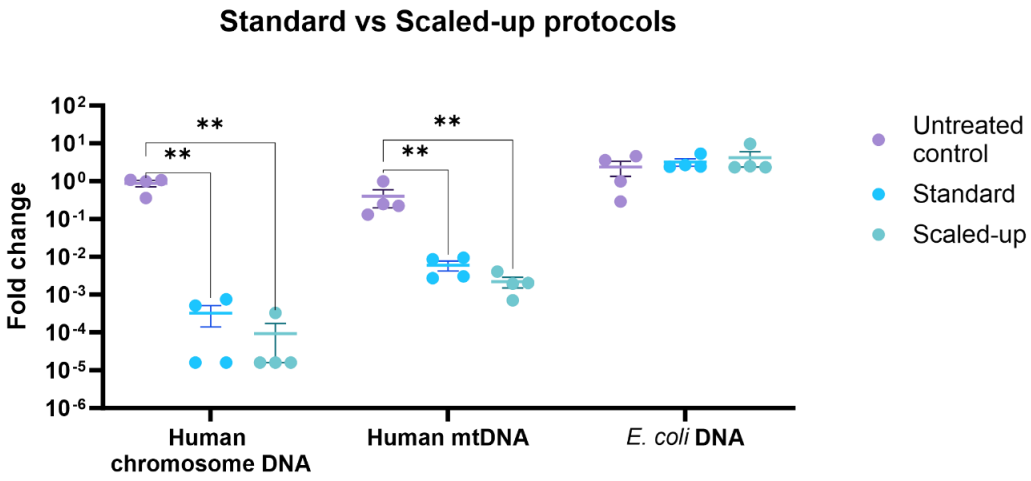


Figure 28. Scaled-up and standard saponin-based depletion methods effect on host DNA and spiked bacteria. Spiked blood samples were untreated or subjected to the standard or scaled-up saponin-based depletion methods. Samples were then DNA extracted and ran on qPCR for the detection of E. coli and host (mitochondrial (mt) and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=4), ± SD. ns > 0.05 (not shown), **p ≤ 0.01.

2.3.8.2 Limit of detection experiments third version (scaled-up volume)

The saponin-based host depletion was further improved by scaling-up the reaction to allow 1mL of blood, potentially enhancing the CFU/mL sensitivity levels of the CMg pipeline. This method modification was combined into the pipeline alongside the previously optimised protocols and tested on blood samples spiked with either *E. coli* or *S. aureus* in separate experiments.

Using the new version of the pipeline, spiked samples were positive for *E. coli* detection at all concentrations (reads ≥1%) and for *S. aureus* at all concentrations except for 1-5 CFU/mL. This translated into a proof-of-concept analytical sensitivity of 1-5 CFU/mL and 10-50 CFU/mL, respectively (Figure 29). For the samples spiked with *E. coli*, taxonomically classified reads had an average abundance of 63.67%, 17.32% and 5.01% for the 50-100, 5-10 and 1-5 CFU/mL samples, respectively, while NTC samples had an *E. coli* average abundance of 0.87%. For the samples spiked with *S. aureus*, taxonomically classified reads had an average abundance of 32.62%, 3.20% and

0.1% in the 50-100, 5-10 and 1-5 CFU/mL samples, respectively, while NTC samples had an *S. aureus* average abundance of 0.11%.

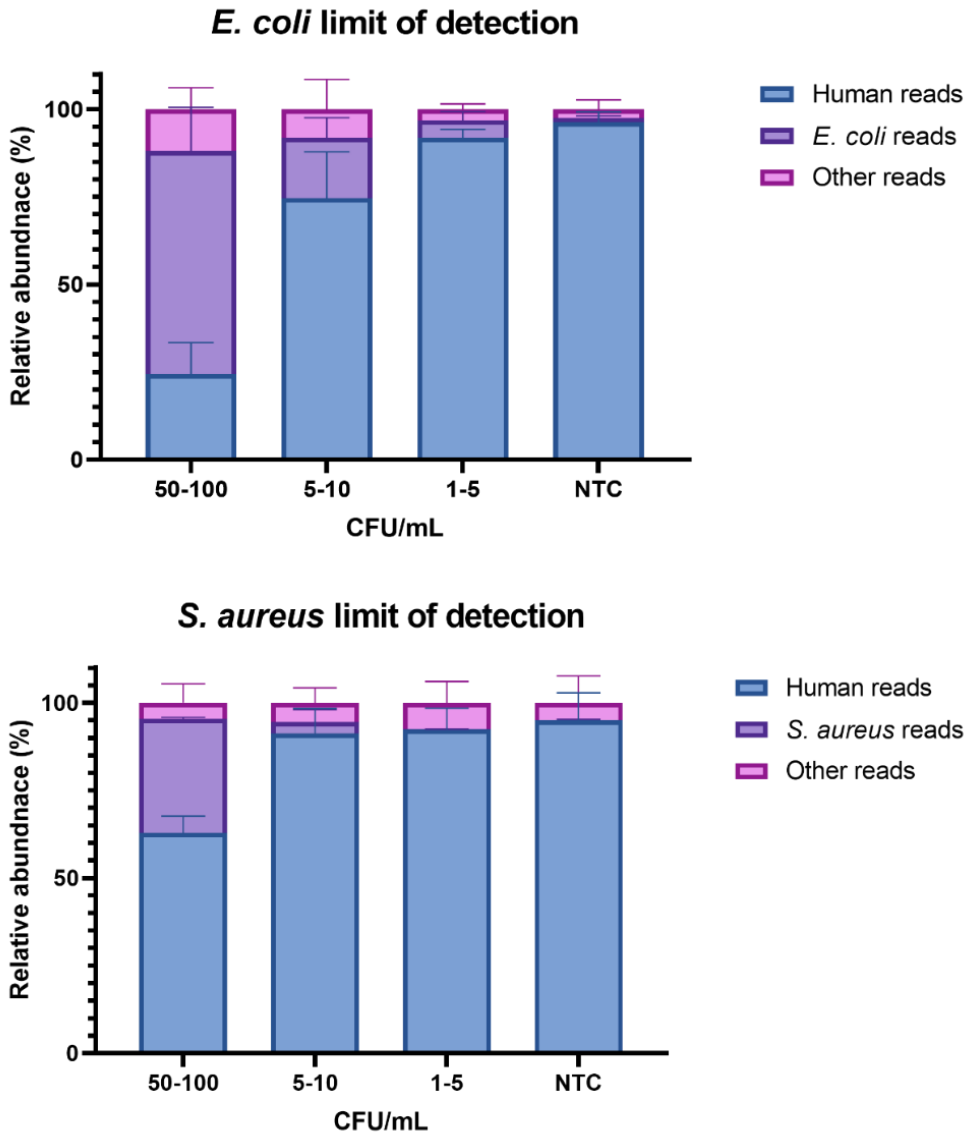


Figure 29. Optimised CMg pipeline (including scaled-up depletion protocol) LoD experiment using samples spiked with *E. coli*. 1mL blood samples were spiked with 50-100, 10-50, 5-10, 5-10 CFU of *E. coli* or *S. aureus* and non-spiked (NTC). Then samples were subjected to the optimsed scaled-up 3% saponin-based host depletion method and the DNA extraction and WGA optimised protocols. Samples were then subjected to library preparation and sequencing. Taxonomically classified reads were classified on 'human', *'E. coli / S.aureus'* and any bacterial read that were not the spiked species ('other read's). Relative abundance of each of the three categories were obtained and samples were classified as positive for bacterial infection if any of the species detected relative abundance reads was ≤1%. Data are means of biological replicates (n=3), ± SD.

For both *E. coli* and *S. aureus*, mapped reads matching their respective references were found in all samples. These were also found in their respective negative controls, but at lower average total numbers when compared to the 1-5 CFU/mL spiked samples (Table 3). When compared to *E. coli*, the average total number of reads classified *S. aureus* were always noticeable lower, which resulted on lower abundance results and consequently a lower sensitivity for this species. As seen on previous pipeline iterations, genome coverage values were relatively low for both species. Nonetheless, in both cases, genome coverage values directly correlated to the spiked inoculum.

E. coli

Sample (CFU/mL)	Average total number of classified reads	Average microbe reads	Average total number *E. coli* reads	Average *E. coli* reads abundance (%)	Average genome coverage (X)
50-100	23310	16564.3	12895.5	63.7	1.31
5-10	18373.2	5893.8	2835	17.3	0.46
1-5	3265.6	1334.0	150	5.0	0.1
NTC	11695.8	573.8	121.5	0.87	-

S. aureus

Sample (CFU/mL)	Average total number of classified reads	Average microbe reads	Average total number *S. aureus* reads	Average *S. aureus* reads abundance (%)	Average genome coverage (X)
50-100	6718	3489	2189	32.6	0.42
5-10	3458.3	495.3	75.7	3.2	000.8
1-5	2288.1	171.2	16.7	0.1	0.004
NTC	965	42.7	5.6	0.1	-

Table 3. Sequencing metrics results obtained on optimised CMg pipeline using samples spiked with *E. coli* or *S. aureus*. 1mL blood samples were spiked with 50-100, 10-50, 5-10, 5-10 CFU of *E. coli* or *S. aureus* and non-spiked (NTC). Then samples were subjected to the optimsed scaled-up 3% saponin-based host depletion method and the DNA extraction and WGA optimised protocols. Table of reads metric obtained at the different spiked samples: microorganism abundances results and average coverage against the *E. coli* or *S.aureus* reference. Data are means of biological replicates (n=3).

2.4 – Discussion.

A direct metagenomics approach for the detection of BSI must overcome the challenge of identifying a low number of microorganisms in blood samples that contain a large number of host cells, whose abundance of DNA mask that of the species of interest. To tackle this challenge, previous efforts have primarily relied on host depletion techniques [270,271] and additional amplification steps to enhance bacterial DNA detection [272–274]. In this study, the initial focus on method development and refinement were centred on creating an effective DNA extraction method, whole genome amplification, and debranching techniques for their integration into the CMg pipeline. Subsequently, our efforts were extended to generate an effective host DNA depletion method. The effectiveness of these optimized methods was evaluated by conducting LoD experiments using blood samples spiked with clinically relevant concentrations of both *E. coli* and *S. aureus*.

In this chapter, it has been described how all the methods were optimised to be used in a CMg pipeline for the detection of BSI. Key findings include:

- Saponin-based host depletion method was not effective when applied on commercial blood cultures as they contain SPS, a potent inhibitor of the reaction. This led us to decide to process blood samples collected in EDTA-coated tubes only.
- The WGA multiple displacement method efficiently amplified low bacterial gDNA concentrations into substantial quantities within hours.
- DNA sub-products and tertiary structures generated during WGA can obstruct nanopore flow-cell pores, reducing sequencing yields.
- Current saponin-based depletion protocols are inadequate for BSI detection, as host DNA still predominates over microbial DNA in blood samples.

- Following saponin-based depletion, the remaining host DNA is predominantly human mitochondrial DNA. Increasing the saponin concentration proved to be the most effective method for reducing mitochondrial presence.
- Host mitochondrial DNA reduction significantly enhanced the sensitivity of the pipeline.
- Combining all optimized methods enables the detection of *E. coli* and *S. aureus* in 1mL blood samples spiked at clinically relevant concentrations.

2.4.1 Saponin-based host depletion cannot be successfully applied to commercial blood cultures.

Given the bacterial enrichment process occurring in positive blood cultures, it was hypothesized that our pipeline would be ideal for the detection of microorganisms in blood cultures and provide extensive gene identification coverage. This metagenomics approach to positive blood cultures has been successfully described by other authors as a sensitive tool for quick ID identification [217,218]. The potential application of our pipeline to positive blood culture samples was evaluated by spiking bacteria into three widely used commercially available blood culture systems designed for detecting aerobic species: BacT/ALERT® and BD BACTECT™ [275,276]. Surprisingly, after the application of saponin-based host DNA depletion, the majority of samples experienced no significant reduction of host DNA content. It was speculated that this phenomenon might be attributed to the presence of SPS, which serves as the primary anticoagulant in most commercial blood culture systems [277]. SPS is widely employed for its ability to inhibit the humoral immune system and host cellular immunity [85,277]. However, it has also been reported as a potent PCR inhibitor, as it tends to co-purify with DNA due to its chemical structure similarity [85]. Consequently, it was hypothesized that SPS could have a similar negative effect to the DNase enzyme added during the saponin-based host depletion protocol, preventing the efficient

depletion of host DNA. Subsequent experiments confirmed our theory: even at minimal concentrations (as low as 0.0005% in the reaction), the presence of SPS prevented effective host DNA depletion. As a result of these findings, a decision was made to not utilize blood samples on blood culture media. Instead, it was determined that the metagenomics pipeline would be exclusively employed for whole blood samples collected in EDTA-coated tubes.

2.4.2 Optimised DNA extraction protocol caused no loss of bacterial DNA.

Initially, the focus of establishing a sensitive clinical metagenomics pipeline was on refining the DNA extraction process. Effective recovery of low bacterial yields from blood through DNA extraction is pivotal for the development of a sensitive sequencing pipeline [106,278]. To ensure future-proofing and user-friendliness, the aim was to automate the DNA extraction method using a robotic extraction system.

To ensure the compatibility of the method across diverse species with varying morphologies and membrane properties, a preliminary bead beating step was incorporated prior to extraction to enable the lysis of membranes that are specially resistant to conventional DNA extraction chemical methods [240], using a mechanical disruption technique. This was tested with the major groups of Gram-negative and Gram-positive causing BSI, *E. coli* and *S. aureus* [36]. The use of a bead beating step for the efficient DNA extraction of tissue samples has been extensively described in the literature [139,279,280]. Nonetheless, to our knowledge, no work on determining the most appropriate extraction buffers for the bead beating step has been published. In this project, it was demonstrated how the selection of this buffer could have significant effects on the extraction efficient and PBS was, among the tested buffers, the most optimal one to introduce the samples into the lysing matrix tubes. Post-bead beating, the samples underwent DNA extraction using PureFood Pathogen Kit Cartridges. The utilization of this kit has been described in other sequencing methods to detect bacteria in food samples [281] and faecal samples [282]. Nonetheless, to date, the application of this kit for the extraction of DNA in blood samples has

never been reported. During this project, the application of this method resulted in the extraction of DNA from both species was extracted without loss, even at the lowest concentrations.

2.4.3 Optimised whole genome amplification resulted in a sensitive method that can amplify large quantities of DNA in a short time.

Following the extraction of microorganisms' gDNA from blood samples the exploration of a whole genome amplification (WGA) step was initiated. The goal of this approach was to potentially magnify the scarce quantities of microorganism gDNA into larger amounts within a few hours of incubation, generating sufficient yields for subsequent nanopore library preparation to analyse samples more extensively. WGA protocols have emerged as valuable tools for investigating microbial communities with low-level presence [144,145]. While acknowledging the inherent species bias often associated with WGA methods and their tendency to amplify certain microorganisms at a higher rate [144], this phenomenon is counteracted when applied to BSI samples due to the predominance of monomicrobial infections, as other authors have reported as a rare scenario BSI infections caused by more than one species [265,283]. The application of a WGA protocol transformed *E. coli* DNA detection from being nearly imperceptible to a large presence of the bacterial species reads, thanks to a notable 10^3-fold increase in number of reads (section 2.3.3.1).

Upon recognizing the importance of integrating a WGA step into the pipeline, diverse commercial protocols were evaluated to determine which could achieve the most substantial amplification of microbial gDNA in the shortest timeframe. Through kit-comparison experiments, the REPLI-g Single Cell protocol emerged as the most effective at amplifying *E. coli* gDNA, particularly at low gDNA concentrations. Notably, the kit could detect and amplify as little as 500fg of gDNA (equivalent to ~100 bacterial cells), demonstrating a very high sensitivity and effectivity. None of the other commercial WGA protocols successfully amplified the template gDNA at this

concentration (section 2.3.3.2). This was the expected outcome as the manufacturer of all the tested kit describes the Single cell kit as the most sensitive MDA method that they commercialise [284]. The REPLI-g Single Cell kit then underwent optimization to enhance both its sensitivity and speed by changing some of the reaction volumes and adding DTT, which is a well-known for its amplification-enhancer properties [285,286]. The WGA kit manufacturer recommends the use of DTT on samples that contain either the whole cells or impure DNA as this helps with the denaturation of the DNA. As such, it was decided to test the effects of DTT onto our pure DNA extractions from blood samples. The newly optimized protocol exhibited a highly significant increase in sensitivity. It was capable of amplifying large quantities of DNA from both species (when this was added at gDNA equivalences of 10 CFU or higher) in 1 hour and 30 minutes. Strikingly, despite its inability to amplify gDNA when this was added at 1 CFU equivalence, the latest LoD experiments (which included the scaled-up saponin-based protocol, section 2.3.8.2) successfully detected bacteria that had been spiked at concentrations of 1-5 CFU/mL. This finding also suggested that the WGA step effectively amplified microbial DNA even when bacteria were present at ~1 CFU. Our hypothesis was that this improved performance was a consequence of the molecular crowding effect during multiple-displacement amplification [287]. In this case, the high presence of host DNA would cause the polymerase used in the WGA step to exhibit more activity than when tested with the spiked bacterial gDNA only , allowing for the amplification of low bacterial DNA concentrations in the presence of the large quantities of host DNA and other molecules present in blood samples.

2.4.4 Whole genome amplification products decrease nanopore flow-cells pore activity over time.

The utilization of the phi29 polymerase in WGA causes the generation of concatemers of branched DNA [288]. This branched DNA often forms artificial structures that can impede the activity of nanopore flow-cell pores [152]. In initial sequencing experiments on WGA-treated

samples, pore activity exhibited a rapid decline. Within 1 hour of sequencing, half of the pores became inactive, with fewer than 10% of them actively sequencing. This pronounced reduction in pore activity significantly affected data generation and potentially compromised the overall pipeline sensitivity. Strategies involving nucleases targeting the generated branches have been previously proposed to counter this issue [289,290]. Regrettably, none of the tested debranching and enzymatic digestion approaches tested substantially reduced the decline in active pores over time.

In our hands, despite their potential debranching capacity, neither S1 Nuclease or T7 Endonuclease I, whether used individually or combined, succeeded in preventing the decline of active pores. Even if the use of these enzymes has previously been described as suitable to reduce DNA branched structures, other authors have also reported the ineffectiveness of this strategy on WGA multiple-displacement amplicons [291]. Furthermore, it is known that nanopore protein pores inherently degrade over time [292], although this doesn't fully account for the abnormal decline in pore activity relative to standard nanopore runs conducted by colleagues. During this project, and considering all the above information, the reduction in active pore percentage was acknowledged as a secondary concern; thus, this was not investigated further. However, even if the reduction in active pores persisted and robust data was limited (n=1), treatment with T7 Endonuclease I was incorporated into the pipeline because of its comparatively lower decrease in pore activity over time (section 2.3.4.1) and its relatively minor impact on overall pipeline speed.

2.4.5 Mitochondrial DNA remains in blood samples after saponin-based depletion.

The utilization of a CMg pipeline for whole blood samples need to overcome a significant constraint in the form of an elevated amount of host genetic material compared to that of pathogens within the sample. This overabundance of host DNA substantially affects the overall

sensitivity of the method and its capacity to detect pathogens [128,206,207]. Initially, the saponin-based host depletion method was chosen to address this challenge, as it had been successfully used in CMg pipelines for the detection of pathogens in sputum samples by Charalampous *et al.* (2019) [136]. Since then, other groups further refined this protocol by using a more efficient "one-pot method" that requires a single incubation mix, as demonstrated by Charalampous *et al.* (2021) [269]. Notably, in 2021, this streamlined protocol was proven effective for blood samples by the group of Justin O'Grady [216], and this data served as the basis for the method initially evaluated in this project.

In initial experiments, the saponin-based host depletion method was applied to blood samples and was highly effective in selectively removing human chromosomal DNA without impacting the presence of spiked bacteria. However, sequencing results on depleted samples still revealed a notable presence of human mitochondrial DNA. Surprisingly, the detected amounts were sufficient to mask bacterial DNA when pathogens were present at low concentrations. The lytic action of saponin on cell membranes is caused by an interaction between the aglycon groups in saponin and sterol molecules, primarily cholesterol, within the membranes [226]. Saponin's effectiveness in disrupting human cells is attributed to the abundance of cholesterol in the membranes of leukocytes [228]. However, saponin's capacity to disrupt the inner and outer membranes of human mitochondria may be limited. Both mitochondrial membranes contain cholesterols, yet in considerably lower quantities (approximately 1-5% of total lipid content) [293,294], rendering mitochondria relatively more resistant to saponin-based host depletion.

In healthy individuals' blood, mitochondria are predominantly found on platelets and leukocytes [295]. These cell types are typically present at around $4.5x10^9$ and $2x10^9$ per millilitre of blood, respectively [131,296], and both contain a large quantity of mitochondria. Approximately 200-600 mitochondria are found in each leukocyte [297], and around 5 in platelets [298]. Factoring in these specific cell types alone yields an estimated presence of approximately 10^{10} mitochondria

organelles per millilitre of blood. Given that each mitochondrion carries a 16.5Kb genome [299], this would equate to about 16,500 Gb of human mitochondrial DNA per mL of blood. On the other hand, the amount microorganisms that can cause BSI (i.e. 1-10 CFU/mL [103]) translates to approximately 5-50Mb of bacteria DNA, considering an average bacterial genome size of 5Mb [133]. Consequently, the amount of mitochondrial DNA in whole blood samples surpasses microorganisms' gDNA by approximately 10^6 times. Initial experiments utilising standard saponin-based host depletion (1% saponin) (section 2.3.1.1) resulted in an approximate 10X reduction in mitochondrial DNA compared to non-depleted blood samples. However, this was not enough to unmask the presence of bacterial DNA, and LoD experiments revealed that detection of bacteria at clinically relevant concentrations was not possible using this depletion method.

To address the challenge caused by the elevated presence of mitochondria, multiple strategies were explored, including Dynabeads®, Calcium, Pneumolysisn, Milteny beads, PLDS, and differential centrifugation (section 2.3.6.2). All these approaches have been documented for their potential membrane-lysis properties, which could either permeate or lyse mitochondrial membranes, eliminate leukocytes prior to cell lysis, or offer other mechanisms to deplete mitochondria from blood samples [300–309]. However, none of these strategies, whether applied individually or in conjunction with the saponin-based protocol, substantially reduced the presence of mitochondrial DNA compared to the saponin method alone. It was hypothesized that these methods, along with their established protocols, were not primarily designed to achieve the extensive host depletion required by our pipeline. Hence, their impact remained marginal in comparison to the saponin-based depletion method. Conversely, elevating the saponin concentration led to a 10-100X increase in mitochondria depletion (section 2.3.6.2.7) compared to the standard approach. This augmented depletion, coupled with the upscaled depletion protocol (1mL of blood), had a significant influence on pipeline sensitivity. To our knowledge, this is the first time that a saponin-based protocol at a saponin 3% final concentration applied in blood samples is described.

2.4.6 Improvements on individual clinical metagenomics pipeline steps resulted in a final sensitivity of 1-5 CFU/mL for *E. coli* and 10-50 CFU/mL for *S. aureus.*

All the optimised methods and their successive iterations underwent validation via LoD experiments. In these experiments, blood samples were artificially infected with predetermined concentrations of microorganisms and subjected to the complete CMg pipeline process. The resulting sequencing data were analysed to detect the concentrations at which the spiked bacterial species' DNA could be detected, as indicated by their taxonomically classified read abundance.

At this project phase, LoD experiments were exclusively conducted by spiking blood samples with *E. coli* and *S. aureus*. A detailed schematic map of all steps undertaken on LoD experiments is shown in Figure 30.

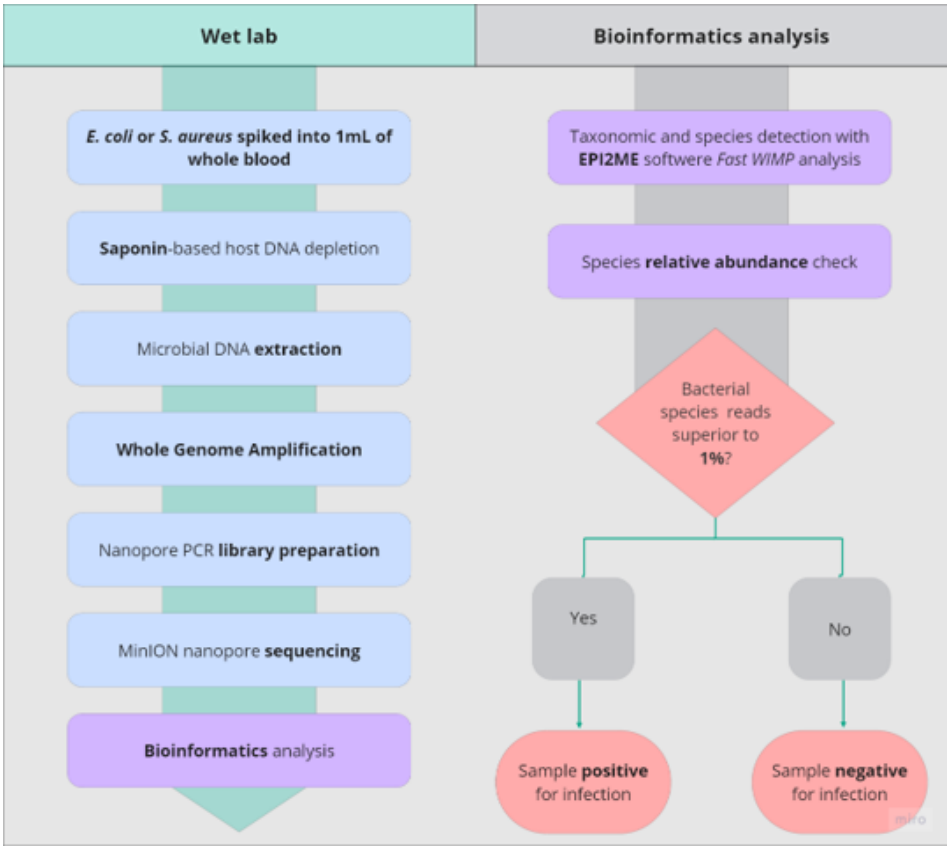


Figure 30. Schematic diagram of clinical metagenomics pipeline. *E. coli or S. aureus* at different clinicaly relevant concentrations would be spiked into whole blood samples and subjected to all the optimised wet lab methods to finally be sequenced using nanopore sequencing. Samples would be considered positive for infection according to their abundance of taxnomically classified reads.

Before optimizing the saponin-depletion method, the detection of *E. coli* was positive when samples were spiked with 100-1000, 50-100, and 10-50 CFU, indicating a limit of detection of 50-250 CFU/mL. This result signalled the need for further pipeline enhancements, as the obtained sensitivity fell short of detecting BSI. The saponin-based host depletion method, although highly effective in respiratory samples [136], demonstrated inadequacy for detecting bacteria at clinically relevant concentrations in blood. This discrepancy can be attributed to the significant difference in microbial concentrations between the respiratory tract (approximately 10^4-10^5 CFU/mL [233,234]) and BSI infections (<10 CFU/mL). Despite the method's insufficient sensitivity, *E. coli* reads were consistently detected across all samples, with read quantities corresponding to the spiked concentrations. These findings, coupled with the absence of *E. coli* reads in the No Template Control (NTC) samples, reinforced our confidence in the results and the potential of this method as a rapid species-identification tool.

The next iteration of the CMg pipeline introduced an improved host depletion method using a higher concentration of saponin. This pipeline iteration was tested by spiking different clinically relevant concentrations of *E. coli*. The increase in saponin concentration to 3% led to an approximate 10X increase in sensitivity as indicated by LoD experiments, allowing for the detection of *E. coli* at 25-50 CFU/mL. This reduction in host DNA presence directly influenced an increase in the number of bacterial reads detected, resulting in significantly improved genomic coverage compared to previous LoD iterations. Furthermore, the saponin method was combined with an upscaled protocol that could handle five times the sample volume, thereby accommodating a higher number of microorganisms. This combination was tested in LoD experiments using 1mL blood samples spiked with *E. coli* or *S. aureus*. The results indicated successful detection at concentrations of 1-5 CFU/mL and 10-50 CFU/mL, respectively. The application of this iteration of the pipeline into proof-of-concept LoD experiments resulted in the detection of the two major microorganisms responsible for BSI at clinically relevant concentrations. Additionally, the application of this CMg pipeline yielded the highest genome coverage values achieved thus far.

Despite the successful increase in sensitivity, it is worth noting that these values might not attain the necessary genomic coverage depth [310] for specific gene detection purposes. This could potentially represent a limitation of this method when trying to use this pipeline for specific antimicrobial resistance (AMR) gene detection.

2.4.7 Conclusion.

While a direct metagenomics approach has the potential to replace the current service standard diagnosis for BSI, none of the previously published CMg clinical protocols have fully addressed the challenges associated with it. Based on recent published results and the group's expertise in saponin-based depletion, it was theorised that the saponin-based depletion method could play a key role on addressing some of these challenges. After using this approach on blood samples, a significant presence of mitochondrial DNA remained, often outnumbering the microorganisms. Increasing the saponin concentration resulted in a noteworthy reduction in human mitochondrial DNA. Furthermore, a DNA extraction technique was optimised to preserve bacterial gDNA in blood samples without loss and optimised a whole genome amplification protocol capable of efficiently amplifying small quantities of extracted bacterial gDNA for nanopore sequencing. The culmination of these efforts allowed us to successfully detect microorganisms when spiked at clinically relevant concentrations. This progress marked a significant step towards the effective detection of microorganisms causing BSI by utilizing a refined combination of all the steps that constitute the CMg pipeline.

In conclusion, our study highlights the potential of a direct metagenomics approach to revolutionize the diagnosis of BSI. This study underscores the potential of a well-crafted combination of these methods to address the challenges of BSI diagnosis, bringing us closer to make possible the use of direct metagenomics as the future service standard in BSI detection. This approach has the potential to improve patient outcomes and contribute significantly to global efforts to combat infectious diseases.

CHAPTER THREE

TESTING THE CMG PIPELINE ON CLINICAL SAMPLES FROM NNUH BSI PATIENTS.

3.1 – Introduction.

Chapter 2 described how the optimised metagenomics pipeline could identify microbial species when spiked at clinically relevant concentrations in whole blood. For the next part of the project, the aim was to test this method on clinical blood samples as a viable alternative to blood culture, while demonstrating the practical application of the pipeline in conjunction with routine methodologies used by hospitals. The pipeline was evaluated on blood samples collected from patients with confirmed bacteraemia at the Norwich and Norfolk University Hospital (NNUH).

This collaboration was a great opportunity to test the pipeline on samples coming from a clinical setting. Clinical samples often have characteristics and are subjected to conditions that differ from those of mock samples prepared in the laboratory. Therefore, it is key to ensure that the pipeline works on clinical samples. Some key issues predicted to impact the efficiency of the test were: the variability in blood sample characteristics (e.g. different levels of host cells present in blood), the potential presence of antibiotics in blood samples and their probable detrimental effect on microbial detection, and the extended periods of time between sample receival and analysis. Thus, multiple pre- and post-analytical considerations needed to be taken for the implementation of CMg pipeline into the clinical setting.

3.1.1 – Proof-of-concept implementation of a clinical metagenomics pipeline into a clinical setting.

Scientific innovations, such as a new diagnostic tool for infectious diseases, are often difficult to implement in routine clinical care, Balas and Boren in 2000 suggested that there was a clinical gap of approximately 17 years for innovative products to become part of clinical practice [311]. This slow adoption of new techniques has massive impacts for the healthcare systems, patients, funding bodies and researchers that have dedicated significant times and resources to advance the scientific medical knowledge [312]. There are many barriers to overcome to bring a

product from development to implementation into a clinical setting. Most of these challenges are policy and regulation related, but others are also related to their robustness and adaptability to the clinical laboratory [313]. When implementing new diagnostics for the detection of BSI, multiple pre-analytical and post-analytical aspects need to be considered to optimise specimen collection (e.g. ensuring the right specimen is collected at the right time) and developing an effective and robust approach for reporting metagenomic findings to clinicians [314].

Pre-analytical considerations studied in this project:

- **Collection:** the method and timing of collection must provide sufficient volumes for the analytical pipeline and be taken considering other factors that can contribute to the specimen degradation. Many patients are treated with antibiotics that will affect the method sensitivity [212,315]. Therefore, collection of samples at a time when patients have no antibiotics circulating in blood is key to avoid false negative results. Also, accurate BSI diagnostics would ideally process volumes of around 10mL of blood due to the low presence of microorganisms in circulation (1-10 CFU/mL) [316]. Nonetheless, this quantity (10mL) might not be feasible in all situations (e.g. neonates, the elderly) [317]. The selection of a suitable vial for sample collection is vital for the proper preservation of its integrity [212]. Sample collection tubes are often coated with anticoagulant substances (such as EDTA) to prevent clot formation which aids in the maintenance of nucleic acid integrity, allowing long-term storage of the samples. However, these anticoagulants can also can impact further downstream applications due to their chemical properties [214].
- **Anonymisation:** to respect patient privacy, in research studies it is important to remove any patient identifiers from the sample while adding new internal (to the research team) identifiers. These identifiers will be used during the validation of the method and will allow backwards traceability to the clinicians [318].

- **Storage:** subjecting blood samples to temperatures below 0˚C is not recommended as it can have negative impacts on their integrity [212]. Nonetheless, when stored at 4°C or at room temperature, any antibiotics present in the samples contribute to the elimination of microorganisms [315] and there is an increased risk of haemolysis and nucleic acid degradation [319]. However, the storage of certain blood samples at room temperature (e.g. for lymphocyte count assays) is a common practice at NHS hospitals [320], which requires collection and processing of blood samples as quickly as possible.
- **Transport:** the minimization of transport time after sample collection is key to diminish sample degradation and exposure to the incorrect conditions (e.g. temperature, light or humidity) [212], and analysis should be performed upon arrival. Transport time can be a limiting factor for the scope of developing clinical metagenomics, as expert facilities may not be on site therefore requiring extended transport times which may have negative impacts for this method [321].
- **Selection:** in most hospital departments, over 90% of tested blood cultures do not yield any microbial growth [322–324]. Therefore, selection of suitable samples for validation requires a balance between directly processing a high number of samples in the laboratory to yield the 10% that contain aetiological agents, or waiting until blood culture testing results are ready and applying the method on samples from confirmed BSI positive patients only.

Post-analytical considerations tested:

- **Analysis of internal results and quality control:** criteria for the analysis of results need to be pre-established based on proof-of-concept experiment results. These are normally based on thresholds that will help the diagnostic decisions process [325]. Importantly, quality control needs to be applied throughout the experimental process (by the addition of appropriate control samples) and be considered for the analysis of the data [325].

- **Comparison of results against the gold standard and cross-validation:** the analysis of results being masked from the outcome obtained through the gold standard test ensures objectivity. To validate the sensitivity and specificity of the new test, its results can then be compared to the gold standard method, in this case, blood cultures [326]. Importantly, the low sensitivity displayed by blood cultures [101] can lead to contradictory results between both methods. In these cases, it might be relevant to cross-compare the results with other methodologies such as qPCR which are more sensitive than blood culture to help interpret potential false-positive results [327].
- **Reporting of results:** results of a diagnostic test are released to trained medical staff as part of a package of information used to make a final diagnosis and determine the best treatment. It is important to select and adapt the information given to the clinicians to aid their interpretation of the results For example, coagulase-negative *Staphylococci* is a very commonly reported finding within blood culture results rather than providing more information such as a specific species determination.

In order to explore how the method under development was impacted by the above considerations, a collaboration was established with the Microbiology and Haematology department at the NUHH. This allowed a proof-of-concept study to assess the diagnostic potential of the CMg pipeline developed in this study in a clinical setting. The direct costs of this work were supported in part by the award of a BBSRC Impact Acceleration Account grant (UKRI). This funding allowed me to collect and process up to 50 clinical blood samples from NNUH patients over a period of three months.

When establishing the collaboration, two major factors were taken into consideration as key differences between mock samples tested at the laboratory and clinical blood samples:

1. In a lot of cases, at the time of sample collection patients had been treated with antibiotics [328]. The circulation of antibiotics in blood can affect the sensitivity of blood culture-based methods [315]. Their potential negative effect on the CMg pipeline was unknown and needed to be explored.
2. Clinical blood samples can contain a higher lymphocyte count ratio compared to healthy individuals [329]. The additional presence of host cells and, consequently host DNA, could have a potential negative effect on the microorganisms detection as their relative abundance in the sample would be even lower than in healthy patients in which this proportion is already largely imbalanced [89,130].

3.1.2 – Clinical metagenomics pipeline, alternative approaches and experimental design.

The successful application of a direct CMg pipeline on clinical blood samples has only been reported before by Parize *et al*. [201]. However, their study presented significant drawbacks. For instance, the identification of infection-causing agents using their approach was only achievable through a 2-3 day long protocol (due to extended sequencing times) and its application was limited to immunocompromised patients (who present low numbers of host cells in blood) [201,330]. Current efforts in the field are focused on developing a CMg tool that is capable of providing results in a shorter time and that can perform specific gene detection [331,332]. A 16S-based amplification has been previously presented as a potential alternative tool for BSI detection. Examples include El Gawhary *et al.* (2016) [333] or Moore *et al.* (2016) [334]. The 16S ribosomal RNA gene (rDNA) is a highly conserved genetic region across prokaryotes which is then followed by several hypervariable regions. The amplification of these variable regions and its subsequent sequencing analysis allows the classification of bacterial organisms present to its species level [335]. The 16S amplification technique has the advantage of providing rapid results (<24 hours) and high sensitivity (70-100%) at the species identification level [333,334,336]. Notably, the 16S-based approach eliminates the need for a depletion phase, as it relies on the amplification of the specific 16S ribosomal RNA gene

[335], leading to an enrichment of microbial DNA over host DNA. Moreover, Charalampous *et al.* (2019) [141] noted that the saponin-based depletion method could result in the loss of bacterial DNA for certain species (e.g. *S. pneumoniae*), highlighting the advantage of the use of 16s in some instances. A 16S approach holds the potential to be more sensitive and could serve as a valuable control to mitigate the issues of using a saponin-based method. However, this technology does not allow the detection of AMR genes, which is one of the best assets of the application of the CMg approach [110]. For all these reasons, the 16S sequencing approach was introduced in this study as a backup method to run in parallel to the CMg pipeline to cross-compare the metagenomics results.

A detailed schematic representation of all samples and methods used in this project is displayed on Figure 31.

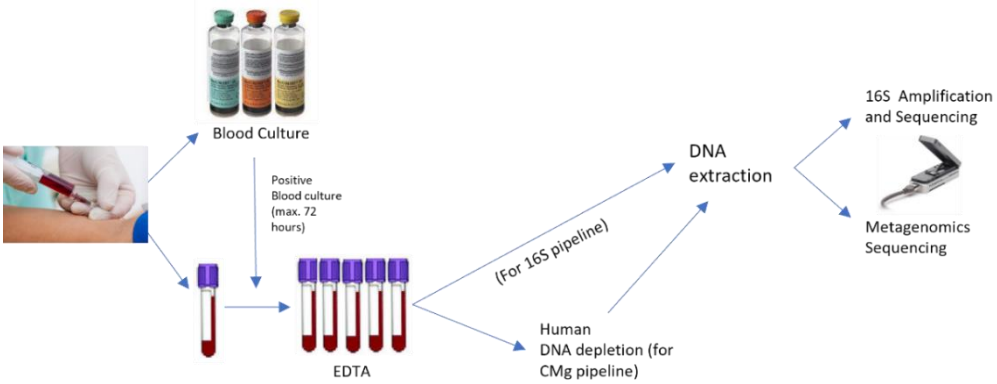


Figure 31. Schematic representation of the QIB-NNUH collaboration for metagenomics detection of BSI. After NNUH detected sepsis positive patients by blood culture, blood collected on EDTA-coated tubes from the same patients would be collected by QIB staff. Samples would be then saponin-based host depleted and subjected to DNA extraction. Then the DNA would be subjected to WGA, debranching and library preparation steps to finally be sequenced using nanopore sequencing. Moreover, the same blood samples would be subjected in parallel to the 16S PCR pipeline by extracting the DNA and amplifying it with 16S primers to finally sequence the amplified product using nanopore sequencing.

3.1.3 Aims and hypothesis.

The CMg pipeline detailed in the previous chapter had outstanding sensitivity results, detecting as few as 1-5 CFU/mL. It was hypothesized that this optimised CMg pipeline had the potential to be effective at detecting BSI in clinical blood samples and detect BSI at a similar rate

compared to standard microbiology testing. To test this hypothesis, the following aims and considerations were designed:

- Clinical blood samples were obtained in EDTA-coated tubes from blood culture positive patients, prioritising the selection of samples with a high chance of containing microorganisms to validate the CMg pipeline. This type of sample offered an efficient use of blood already collected for other purposes, such as white blood cell counting [337], and allowed the execution of this work under existing ethical approval work.
- To process positive-only samples clinical blood samples with a metagenomics and 16S-based sequencing approach.
- To analyse sequencing results blindly from the hospital microbiology results and assess the sensitivity of these methods compared to service standard blood culture methods.

3.2 Methods.

3.2.1 Criteria for the selection and collection of clinical blood samples.

Collaborators at the NNUH provided samples over a three-month period. A comprehensive process map giving an overview of this collaboration can be found in Figure 32. During their weekly review of critical blood culture results, microbiology laboratory consultants identified patients whose samples had been positive in blood culture testing, which were selected to be used in the study. Additional blood samples from these same patients had been concurrently collected in EDTA-coated tubes for other routine clinical tests. After the target patients were identified, technicians in the haematology laboratory at NNUH retrieved and anonymized the surplus sample remaining in these EDTA-coated tubes, which were prepared for transportation to QIB to be used in the study. Importantly, blood samples in EDTA-coated tubes were incubated at room temperature until the outcome of blood culture tests were known. Only samples that had been at room temperature for less than 3 days were selected for this study, although in some instances another half day had passed until they could be collected and processed. Finally, to also assess the specificity of our method, the Microbiology team at NNUH deliberately selected some samples that were negative for infection to comprise approximately 10% of the cases.

Once samples were collected, they were immediately processed using the metagenomics and 16S pipelines. Every time samples were received and processed, a process negative control was included in the batch to detect any potential reagent, manipulation, or environmental contaminants. For the process control, 1mL of PBS was subjected to the same methods as the clinical blood samples. For the process control, 1mL of PBS was subjected to the same methods as the clinical blood samples. At the end of the processing, results were analysed and a QIB results report was produced before knowing the NNUH microbiology results. The report was then shared with the NNUH team and the outcomes were crossed-compared to those obtained with service standard microbiology testing.

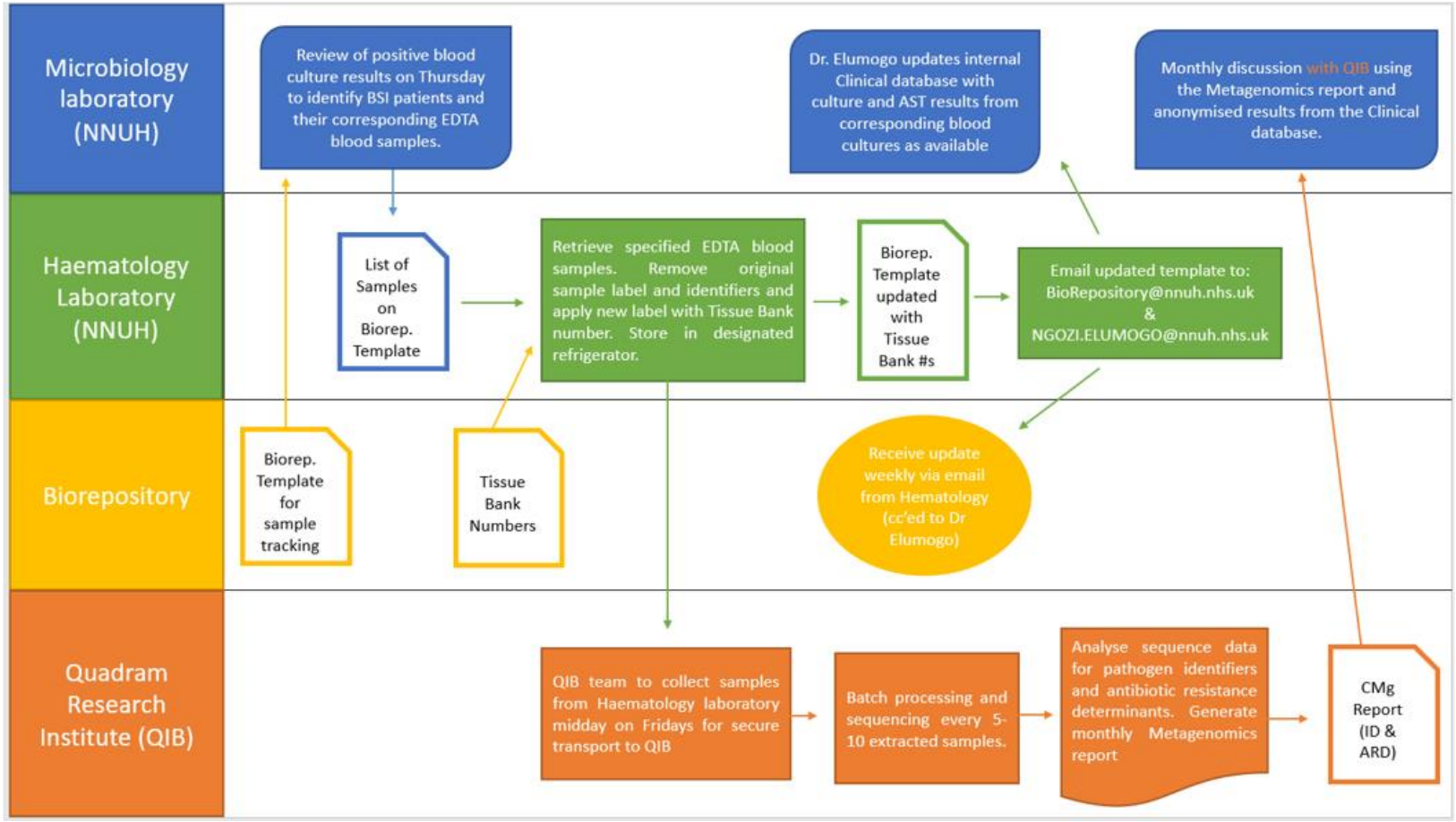


Figure 32. Process map of QIB-NNUH collaboration for the obtention of clinical blood samples. Positive samples were identified by the microbiology laboratory and retrieved by the haematology team. Samples were then anonymised by the Biorepository team and finally collected by the QIB team. Samples were then processed using the metagenomics and 16S pipelines (latter one by the ONT team) and results were reported back to the Microbiology team at NNUH.

3.3.2 – Metagenomics pipeline.

3.3.2.1 Saponin-based host depletion and DNA extraction.

Collected clinical blood samples were subjected to the CMg pipeline. 1mL of each sample was subjected to the saponin-based depletion by adding saponin at a final concentration of 3%, 1mL of HLSAN Buffer and 10µL of HLSAN DNase on a 5.0mL Eppendorf tube. Samples were incubated at 37˚C for 30 minutes at 1000rpm on a ThermoMixer. After that, samples were centrifuged at 8,000g for 6 minutes, before the resulting supernatant was discarded and DNA was extracted from the pellet as described on section 2.2.4. To assess the efficiency of the host depletion step on clinical blood samples, the same saponin-based protocol was ran on blood samples from healthy individuals (section 2.2.1) and DNA was extracted using the same protocol. Then, DNA extractions from clinical blood samples were compared to DNA extractions from healthy control donors by running them on a qPCR for the detection of human chromosomal and mitochondrial DNA as described in section 3.2.2.2.

3.3.2.2 qPCR assay and statistical analysis.

qPCR assays were performed on 8 of the clinical blood samples containing the extracted DNA as described in section 2.2.5 to assess the depletion efficiency on clinical blood samples. These were also compared with samples from healthy patients described in section 2.2.1 and to compare against hospital microbiology results. qPCR assays were used to detect bacterial DNA from *E. coli*, *S. aureus* or K. *pneumoniae* or for the host detection of chromosomal or mitochondrial DNA (Supplementary Table 1). The qPCR results were statistically analysed as described in section 2.2.6 and fold change values were obtained by using depleted blood samples with no SPS added as a control.

3.3.2.3 Whole genome amplification and debranching.

After extracting the DNA from clinical blood saponin-based depleted samples, the extractions were cleaned with AMPure beads as described in section 2.2.11 to finally elute the DNA in 15µL of water. Then, all samples were subjected to the optimised WGA protocol described on 2.2.9. Finally, amplified samples were subjected to a debranching step using T7 Endonuclease I as described on section 2.2.10.

3.3.2.4 Nanopore library preparation and sequencing.

The debranched products were subjected to the Rapid PCR barcoding kit SQK-RPB004, following the manufacturer's guidelines with the modifications described on section 2.2.12. Once libraries were prepared, samples were quantified, pooled in equal concentrations, and cleaned with a 0.6X AMPure as detailed in 2.2.11. Approximately 100fmoles of pooled PCR library was then loaded onto the MinION flow cell (R9.4.1, Oxoford Nanopore Technologies). For all clinical blood samples sequencing run time was 3 hours, after that fastq files were obtained.

3.3.2.5 Bioinformatics analysis.

Pathogens were identified and abundance results were obtained from the sequencing data using the EPI2ME software and the *'What's In My Pot pipeline'* (Oxford Nanopore Technologies). To determine the presence of bacteria and determine the sample as positive, bacterial species had to be present at ≥1% over the total number of classified reads. EPI2ME results were also used to calculate the total number of reads. When analysing the possible presence/absence of microbial species based on total number of classified reads, the reads identified on negative process control samples (non-spiked, underwent all the CMg pipeline) were taken into account. If these were found at similar abundance or total number, the identification of these species was considered negative on clinical blood samples sequencing results.

3.3.3 – 16S amplification pipeline.

For each patient blood sample, 1mL of blood was subjected to a 16S sequencing pipeline. This pipeline was performed by ONT staff and only the following methodological and analytical details were shared to us: to each 1mL of blood they performed a differential centrifugation step by mixing the samples with the same volume of PBS and then centrifuging at low speed for 5 minutes. Then, the plasma layer was taken and the rest was discarded. Plasma was subjected to centrifugation at 12,000g for 5 minutes to pellet the bacteria present in the plasma and then the pellet was subjected to DNA extraction using a column DNA manual extraction kit. The extractions were then amplified using proprietary 16S primers to exponentially increase the presence of the 16S ribosomal RNA gene. Finally, the amplified DNA was subjected to a nanopore library preparation protocol before being sequenced on a MinION.

16S nanopore sequencing results were analysed by considering the number of classified reads obtained for a particular bacterial species and the total or relative absence of these species reads from their respective NTC. The final identification results of each patient sample were only disclosed after their performed the appropriate bioinformatics analysis.

3.3.4 – Ethical approval.

Clinical blood samples collection and their immediate processing was done under the FMH ethics project number UEA FMH reference 2018/19 - 153 "*Study Title: Next generation sequencing based diagnosis for bacterial/fungal/viral infections directly from clinical samples (blood, urine, stool, respiratory, tissue)"* and amended on the version 4 submitted on the 10th May 2022.

3.3 – Results.

In total, 33 whole blood samples in EDTA-coated tubes (from here on referred clinical blood samples) from BSI culture-positive patients were received and processed at QIB using the metagenomics and 16S pipelines.

3.3.1 Host DNA depletion efficiency in clinical blood samples.

First, the efficiency of the saponin-based host depletion method was checked on clinical blood samples. This was done by comparing the remaining quantities of chromosomal and mitochondrial DNA in clinical blood samples treated with the saponin-based depletion method. 8 randomly selected clinical blood samples were used in this experiment. Results were compared to the same clinical samples without treatment and also to commercial non-infected blood samples (both saponin-based depletion treated and non-treated).

The application of the host depletion step on patient samples successfully depleted the host chromosomal DNA at a similar fold change ratio than in non-infected control samples (Figure 33A left). Conversely, the mitochondrial DNA depletion efficiency was approximately $>10^1$-fold less efficient on clinical blood samples than on blood from healthy donors. However, this difference was not statistically significant and therefore it was considered successful at depleting mitochondrial DNA, too (Figure 33A right).

To rule out the possibility of the efficiency being slightly lower (although not significantly) due to clinical blood samples having an increased presence of host DNA (due to higher blood cell counts), the Ct values of chromosomal and mitochondrial before saponin-based depletion were evaluated. The Ct values of both chromosomal and mitochondrial DNA were not significantly different between patient and non-infected blood samples (Figure 33B).

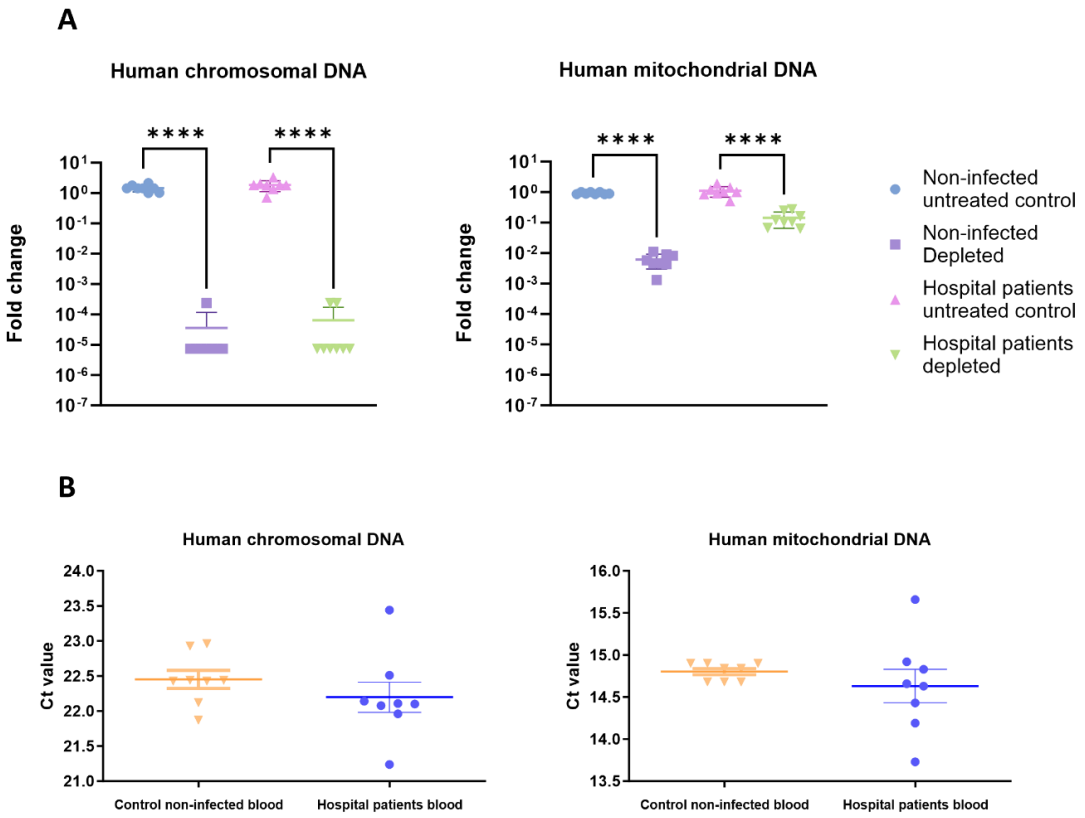


Figure 33. Host depletion efficiency on clinical blood samples compared to non-infected and non-treated controls. (A) Fold change values on both human chromosomal and mitochondrial DNA obtained from healthy donors and patient samples subjected to saponin-based depletion methods compared to their untreated controls. **(B)** Ct values obtained by qPCR of host chromosomal and mitochondrial DNA of non-depleted healthy donor and patient samples. Data are means of biological replicates (n=8), ± SD. ns > 0.05 (not shown), ****$p \leq 0.0001$.

3.3.2 Sequencing results from clinical blood samples.

Despite observing that the host depletion step was potentially less efficient (non-significant) in blood samples from the hospital, the reduction in host DNA compared to untreated samples was highly significant.

3.3.2.1 Identification of causative agents based on relative abundance.

The 33 clinical blood samples received from the hospital were processed using the whole pipeline and the relative abundance sequencing results were divided into two categories: abundance of reads identified as human, and abundance of reads identified as from microbial species. At the beginning of the collaboration, the first 3 batches (samples 1-14) received, produced no reads that were taxonomically classified as microbial (Supplementary Table 2). Instead, all the

detected reads on those samples were classified as human (Figure 34). It was hypothesized that the samples that had been received contained no intact bacteria at the time of host depletion and DNA extraction treatments and that microorganisms could be dying due to the action of immune cells or the presence of antibiotics. Because of the lack of intact microorganisms by the time of processing, logistic changes were introduced to collect samples a day earlier to minimise the chances of this happening. After introducing this change, from sample 15 onwards, reads classified as from bacterial species were observed. Nonetheless, out of the 33 clinical blood samples, the proportion of microbial reads was ≥1% in only 6 samples (sample number 20, 21, 22, 23, 24 and 32). Even in those samples, the human reads represented the large majority of the sequenced reads and microorganism reads ranged from 26% to 2%: S20 (3.8%), S21 (9.0%), S22 (13.5%), S23 (2.3%), S24 (26.5%) and S32 (3.3%) (Figure 34).

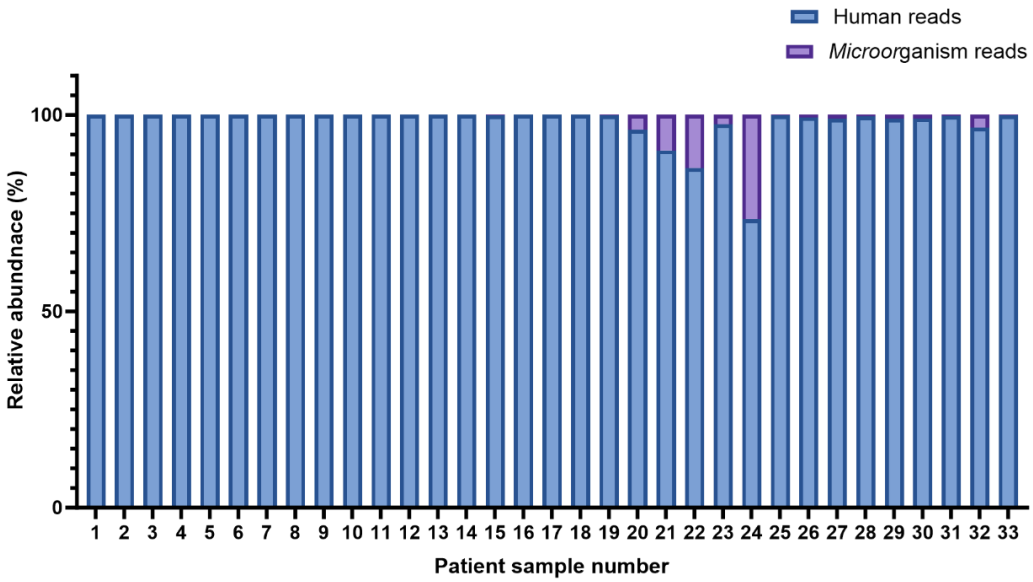


Figure 34. Human and microorganism reads abundance on clinical blood samples. All reads were taxonomically classified and grouped as *'Human'* or *'Microorganism'*. Relative abundance of each of the two groups was calculated and samples were classified as positive for bacterial infection if any of the microbial species relative abundance reads was ≥1%.

In order to identify the causative agent of infection using the metagenomics data, the sequencing data from the 6 samples with microorganisms' abundance ≥1% were taxonomically classified (Figure 35). From the 6 samples, a causative agent of infection was identified in three of them. The species identified were *Clostridium difficile* (sample 22), *Paracoccus yeei* (sample 24)

and *Escherichia coli* (sample 35). In the remaining 3 samples, the agent of infection could not be identified because the species detected in these samples were also detected in the negative control at similar relative abundances.

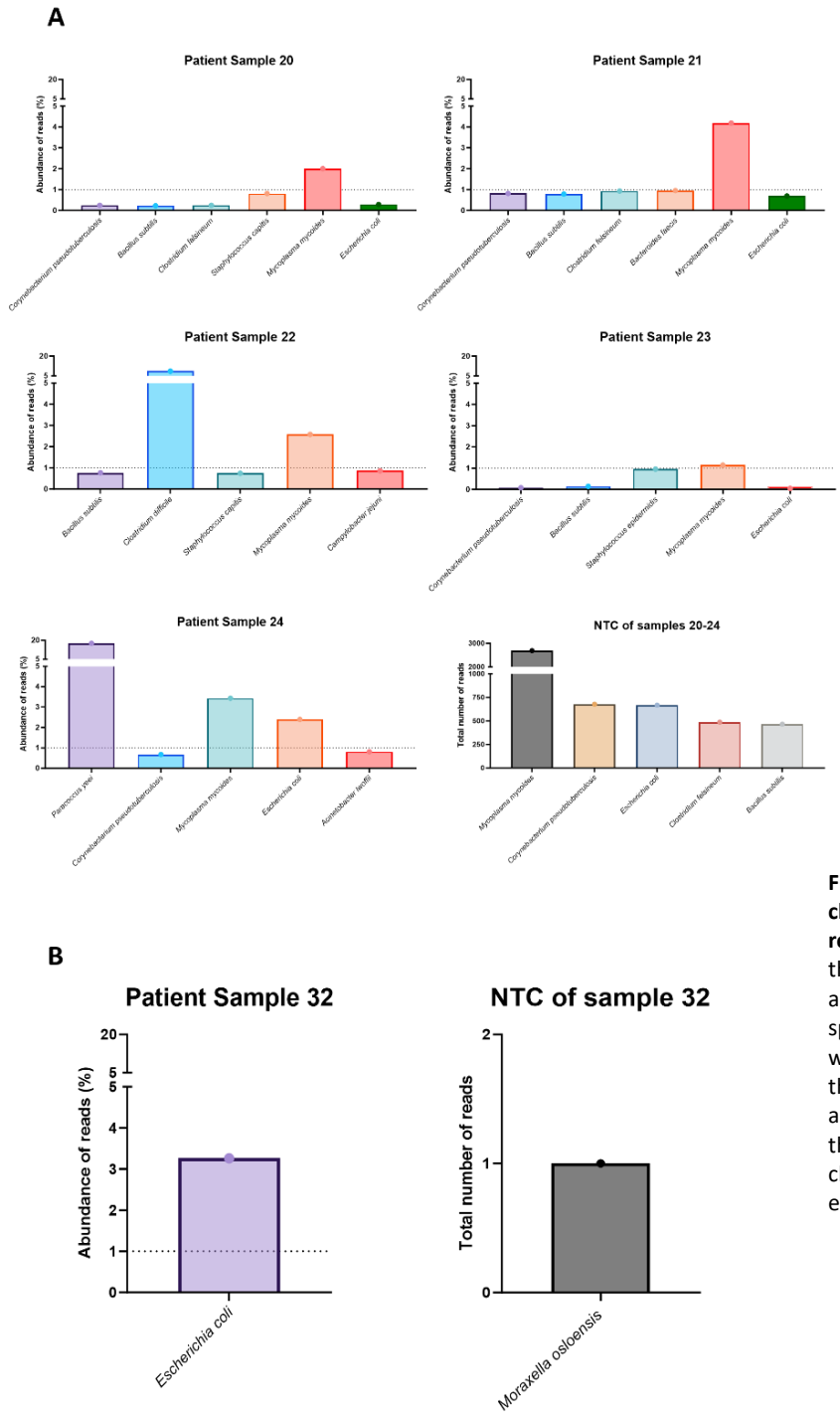


Figure 35. Taxonomic classification of samples and the respective negative controls. All the reads taxonomically classified as different microorganisms species for samples 20-24 and 32 were analysed and compared to their respective NTC. Their relative abundance was calculated over the total number of taxonomically classified reads obtained from each sample.

Overall, after the application of the described criteria (based on abundance), the metagenomics pipeline was only capable of identifying possible causative agents on three of the

33 samples received from the hospital. This represented a 9% positivity rate (16% if only results after changing collection time are considered), 10X less than the expected considering that 90% of the samples received were expected to be blood culture positive (Table 4).

Patient sample number	Metagenomics results
Sample 1	Non-reportable hits
Sample 2	Non-reportable hits
Sample 3	Non-reportable hits
Sample 4	Non-reportable hits
Sample 5	Non-reportable hits
Sample 6	Non-reportable hits
Sample 7	Non-reportable hits
Sample 8	Non-reportable hits
Sample 9	Non-reportable hits
Sample 10	Non-reportable hits
Sample 11	Non-reportable hits
Sample 12	Non-reportable hits
Sample 13	Non-reportable hits
Sample 14	Non-reportable hits
Sample 15	Non-reportable hits
Sample 16	Non-reportable hits
Sample 17	Non-reportable hits
Sample 18	Non-reportable hits
Sample 19	Non-reportable hits
Sample 20	Non-reportable hits
Sample 21	*Clostridium difficile*
Sample 22	Non-reportable hits
Sample 23	Non-reportable hits
Sample 24	*Paracoccus yeei*
Sample 25	Non-reportable hits
Sample 26	Non-reportable hits
Sample 27	Non-reportable hits
Sample 28	Non-reportable hits
Sample 29	Non-reportable hits
Sample 30	Non-reportable hits
Sample 31	Non-reportable hits
Sample 32	*Escherichia coli*
Sample 33	Non-reportable hits

Table 4. **Clinical blood samples sequencing results based on abundance.** List of patient blood and their metagenomics results obtained using the CMg pipeline applied directly on clinical blood samples. Only in 3 samples (highlighted in yellow), the abundance of bacterial reads species was ≥1% and not present on their respective NTC.

3.3.2.2 Identification of causative agents based on total number of reads.

To attempt to identify the causative agent of infection using the metagenomics results, given that most of the received samples were from blood culture positive infected patients, it was decided to perform a detailed case-by-case analysis of the sequencing results. For the remaining samples that did not initially meet the abundance threshold to be deemed positive for infection (>1% abundance over total number of reads), taxonomic composition analysis of all microbial reads was conducted. The taxonomic results on each sample were also compared to the results obtained on their respective NTC samples.

For the majority of samples, taxonomically classified bacterial reads were found at relatively low numbers (<4 reads). Moreover, many of the identified species were also found in the respective negative controls with similar numbers of reads. For only 5 clinical blood samples, a read count of ≥4 was found: sample 15 (7 reads classified as *Escherichia coli*), sample 16 (4 reads classified as *Staphylococcus aureus*), sample 19 (20 reads classified as *Streptococcus dysgalactiae*), sample 27 (4 reads classified as *Cutibacterium acnes*) and sample 30 (4 reads classified as *Moraxella osloensis*). Because *Moraxella osloensis* and *Cutibacterium acnes* were present in their respective negative controls, samples 27 and 30 were considered negative for infection. On the other hand, samples 15, 16 and 19 were considered potentially positive for infection as the species were not present on their respective NTC (Figure 36). In these samples, when analysing the fraction of microorganism classified reads, the relative abundance of the identified species over the total number of identified microbial species was relatively high: *E. coli* reads represented 90.9% over the total number of microbial reads for sample 15, *S. aureus* reads constituted 85.7% for sample 16 and *S. dysgalactiae* 100% for sample 19.

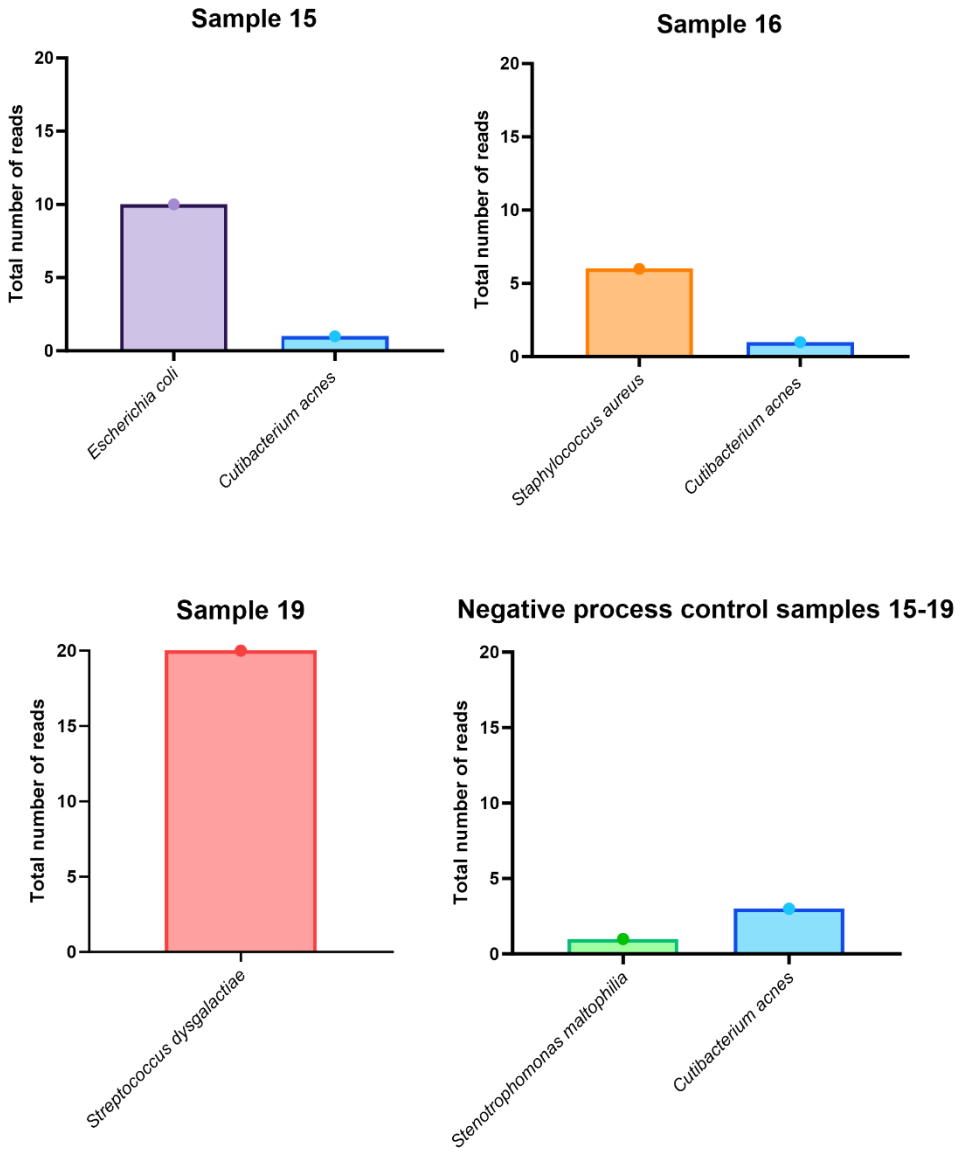


Figure 36. Samples 15, 16 and 19 and their respective NTC taxonomic classification. Total number of reads obtained from samples 15, 16 and 19 and their respective NTC. Samples were considered positive for infection if those species reads were not detected at similar rate on their NTC.

After the detailed taxonomic composition analysis of the microbial reads in each sample, 3 more samples were flagged as positive for infection. These were added to the 3 previous samples that were labelled as positive due to their microorganism reads abundance being ≥1%. With this, a total of 6 samples were identified as positive, a 18% positivity-rate, still a considerably lower positivity than the expected 90% when compared to hospital culture results (Table 5).

Patient sample number	Metagenomics results
Sample 1	Non-reportable hits
Sample 2	Non-reportable hits
Sample 3	Non-reportable hits
Sample 4	Non-reportable hits
Sample 5	Non-reportable hits
Sample 6	Non-reportable hits
Sample 7	Non-reportable hits
Sample 8	Non-reportable hits
Sample 9	Non-reportable hits
Sample 10	Non-reportable hits
Sample 11	Non-reportable hits
Sample 12	Non-reportable hits
Sample 13	Non-reportable hits
Sample 14	Non-reportable hits
Sample 15	*Escherichia coli*
Sample 16	*Staphylococcus aureus*
Sample 17	Non-reportable hits
Sample 18	Non-reportable hits
Sample 19	*Streptococcus dysgalactiae*
Sample 20	Non-reportable hits
Sample 21	*Clostridium difficile*
Sample 22	Non-reportable hits
Sample 23	Non-reportable hits
Sample 24	*Paracoccus yeei*
Sample 25	Non-reportable hits
Sample 26	Non-reportable hits
Sample 27	Non-reportable hits
Sample 28	Non-reportable hits
Sample 29	Non-reportable hits
Sample 30	Non-reportable hits
Sample 31	Non-reportable hits
Sample 32	*Escherichia coli*
Sample 33	Non-reportable hits

Table 5. Metagenomics sequencing results on clinical blood samples considering total number of classified reads. List of clnical blood samples and their metagenomics results obtained using the CMg pipeline applied directly on clinical blood samples. Only in 6 samples (highlighted in yellow), the abundance of reads of abacterial species was ≥1% or their total number of reads was considerable higher than other bacterial species and not present on their respective NTC.

3.3.2.3 16S sequencing results.

In parallel to running the samples on the CMg pipeline, 1mL of blood of each patient sample was subjected to a 16S pipeline.

Of the total 33 clinical blood samples, the 16S pipeline identified 10 samples as positive, almost twice the number of positives identified by the clinical metagenomics pipeline (Table 6). Nonetheless, the positivity rate was still approximately 3X lower than the expected 90% from the hospital microbiology results. Additionally, on only three occasions, both pipelines flagged the same sample as positive for infection. In only one sample (sample 19), both 16S and metagenomics pipelines identified the same species as the causative agent, *Streptococcus dysgalactiae*. On all other instances, the results obtained using metagenomics and 16S pipelines identified different species that were not even from the same order or family: *Clostridium difficile vs Staphylococcus capitis* in Sample 21 and *Paracoccus yeei vs Streptococcus constellatus* in Sample 24 (Table 6). Similarly to the metagenomics results, the introduction of a faster sample collection time had a notable impact on the 16S results positivity-rate. With the initial collection time (samples 1-14) 16S could not detect any causative agent of infection in any of the samples.

Overall, with the combination of the metagenomics and 16S pipelines, 13 samples were detected positive for infection. This represented a positivity rate of 39.4%, roughly half of that which was expected according to microbiology results (Table 6).

Patient sample number	Metagenomics results	16S Results
Sample 1	Non-reportable hits	Nonreportable hits
Sample 2	Non-reportable hits	Nonreportable hits
Sample 3	Non-reportable hits	Nonreportable hits
Sample 4	Non-reportable hits	Nonreportable hits
Sample 5	Non-reportable hits	Nonreportable hits
Sample 6	Non-reportable hits	Nonreportable hits
Sample 7	Non-reportable hits	Nonreportable hits
Sample 8	Non-reportable hits	Nonreportable hits
Sample 9	Non-reportable hits	Nonreportable hits
Sample 10	Non-reportable hits	Nonreportable hits
Sample 11	Non-reportable hits	Nonreportable hits
Sample 12	Non-reportable hits	Nonreportable hits
Sample 13	Non-reportable hits	Nonreportable hits
Sample 14	Non-reportable hits	Nonreportable hits
Sample 15	*Escherichia coli*	Nonreportable hits
Sample 16	*Staphylococcus aureus*	Nonreportable hits
Sample 17	Non-reportable hits	Nonreportable hits
Sample 18	Non-reportable hits	Nonreportable hits
Sample 19	*Streptococcus dysgalactiae*	*Streptococcus agalactiae* and/or *Streptococcus dysgalactiae*
Sample 20	Non-reportable hits	Nonreportable hits
Sample 21	*Clostridium difficile*	*Staphylococcus capitis* and/or *S. caprae*
Sample 22	Non-reportable hits	*Staphylococcus epidermidis*
Sample 23	Non-reportable hits	*Escherichia coli*
Sample 24	*Paracoccus yeei*	*Streptococcus constellatus* and/or *Streptococcus intermedius*
Sample 25	Non-reportable hits	Non-reportable hits
Sample 26	Non-reportable hits	*Staphylococcus epidermidis*
Sample 27	Non-reportable hits	*Staphylococcus capitis*
Sample 28	Non-reportable hits	*Escherichia coli*
Sample 29	Non-reportable hits	Non-reportable hits
Sample 30	Non-reportable hits	*Staphylococcus epidermis*
Sample 31	Non-reportable hits	*Escherichia coli*
Sample 32	*Escherichia coli*	Non-reportable hits
Sample 33	Non-reportable hits	Non-reportable hits

Table 6. Metagenomics and 16S sequencing results on clinical blood samples. List of patient clinical samples and the results obtained using the metagneomics and 16S pipelines. Both methods, combined pipelines identified causative agents of infection in 13 samples (highlighted in yellow). Only for 1 sample (highlighted in green), metagenomics results were confirmed by 16S results.

3.3.2.4 Microbiology *vs* Metagenomics *vs* 16S.

Once the positive samples and its causative agents for both metagenomics and 16S pipelines were established, these were compared to the microbiology results reported by the hospital (Table 7). Generally, it took from 2-5 days for the microbiology team to obtain blood culture and species identification results. After that, within a maximum of three days (2 after changes were introduced), clinical blood samples (on EDTA-coated tubes) were collected and processed on the same day.

Of the 33 samples, microbiology results detected infection in 31 samples, a 94% positivity rate (Table 7). Out of the 6 positive samples detected by the metagenomics pipeline, only 4 (66.7%) had the same species identified by microbiology results. In the case of samples 21 and 24, a high relative abundance of microorganisms reads was identified (8.5% of Clostridium difficile and 17.7% Paracoccus yeei, respectively). However, microbiology results identified other causative agents of infection. Overall, when compared to the service standard culture-based methodology, the metagenomics pipeline had a 18.2% sensitivity. In contrast, when 16S results were compared to the microbiology data, these was species concordance in 7 out of the 10 16S-positive samples (70%). When taking into account the three different diagnostic methodologies, only for 3 samples was the same result identified: samples 4 and 11, which were culture negative, and Sample 19 which was reported as *Streptococcus spp.* (Group G) by microbiology testing, *Streptococcus dysgalactiae* by metagenomics and *Streptococcus agalactiae* and/or *Streptococcus dysgalactiae* for 16S pipeline. Overall, when both (metagenomics and 16S) sequencing results were combined, the same results as the hospital microbiology department (one of the two methods matching) were obtained in 10 samples, a sensitivity of 32.2% when compared to the blood culture service standard method (Table 7).

Patient sample number	NNUH Culture result	Metagenomics results	16S Results
Sample 1	*Pseudomonas aeruginosa*	Non-reportable hits	Nonreportable hits
Sample 2	*Staphylococcus warneri*	Non-reportable hits	Nonreportable hits
Sample 3	*Escherichia coli*	Non-reportable hits	Nonreportable hits
Sample 4	NEGATIVE CULTURE	Non-reportable hits	Nonreportable hits
Sample 5	*Streptococcus sp* (Grp G)	Non-reportable hits	Nonreportable hits
Sample 6	*Streptococcus sanguinis* and *Coagulase negative Staphylococci*	Non-reportable hits	Nonreportable hits
Sample 7	*Staphylococcus epidermidis* and *Streptococcus sanguinis*	Non-reportable hits	Nonreportable hits
Sample 8	*E. coli* (ESBL)	Non-reportable hits	Nonreportable hits
Sample 9	*Klebsiella pneumoniae* (ESBL)	Non-reportable hits	Nonreportable hits
Sample 10	*S. epidermidis* and *Enterobacter cloacae* complex	Non-reportable hits	Nonreportable hits
Sample 11	NEGATIVE CULTURE	Non-reportable hits	Nonreportable hits
Sample 12	*Staphylococcus capitis*	Non-reportable hits	Nonreportable hits
Sample 13	*Acinetobacter lwoffi*	Non-reportable hits	Nonreportable hits
Sample 14	*Staphylococcus aureus*	Non-reportable hits	Nonreportable hits
Sample 15	*Escherichia coli*	*Escherichia coli*	Nonreportable hits
Sample 16	*Staphylococcus aureus*	*Staphylococcus aureus*	Nonreportable hits
Sample 17	*Staphylococcus hominis*	Non-reportable hits	Nonreportable hits
Sample 18	*Staphylococcus epidermidis*	Non-reportable hits	Nonreportable hits
Sample 19	*Streptococcus sp* (Grp G)	*Streptococcus dysgalactiae*	*Streptococcus agalactiae* and/or *Streptococcus dysgalactiae*
Sample 20	*Streptococcus pneumoniae*	Non-reportable hits	Nonreportable hits
Sample 21	*Staphylococcus capitis*	*Clostridium difficile*	*Staphylococcus capitis* and/or *S. caprae*
Sample 22	*Escherichia coli*	Non-reportable hits	*Staphylococcus epidermidis*
Sample 23	*Escherichia coli*	Non-reportable hits	*Escherichia coli*
Sample 24	*Escherichia coli* and/or *Serratia marcescens*	*Paracoccus yeei*	*Streptococcus constellatus* and/or *Streptococcus intermedius*
Sample 25	*Staphylococcus hominis*	Non-reportable hits	Non-reportable hits
Sample 26	*Staphylococcus epidermidis* and/or *Klebsiella oxytoca*	Non-reportable hits	*Staphylococcus epidermidis*
Sample 27	*Proteus mirabilis* and/or *Enterococcus faecalis*	Non-reportable hits	*Staphylococcus capitis*
Sample 28	*Escherichia coli*	Non-reportable hits	*Escherichia coli*
Sample 29	*Escherichia coli*	Non-reportable hits	Non-reportable hits
Sample 30	Mixed coag neg *Staphylococcus*	Non-reportable hits	*Staphylococcus epidermis*
Sample 31	*Escherichia coli*	Non-reportable hits	*Escherichia coli*
Sample 32	*Escherichia coli*	*Escherichia coli*	Non-reportable hits
Sample 33	*Citrobacter koseri*	Non-reportable hits	Non-reportable hits

Table 7. Microbiology blood culture compared to metagenomics and 16S sequencing results on clinical blood samples. List of clinical blood samples and the results obtained using both sequencing pipelines (metagenomics and 16S) and the microbiology results reported by the hospital. Out of the total number positive samples for sequencing pipelines, only in 4 occasions the identified species were not the same as reported by the hospital (highlighted in yellow). For the rest of the positive samples, they all matched the microbiology results (highlighted in green). Only for 1 sample, the three detection methods identified the same species (highlighted in orange).

3.3.3 qPCR detection compared to microbiology results.

Due to the low positivity rate obtained across all the sequencing methodologies applied, selected samples underwent detection by qPCR as an alternative approach to metagenomics analysis to detect the presence of aetiological microorganisms. The samples that were positive for either *E. coli*, *S. aureus* or *K. pnuemoniae* (according to the microbiology data) were analysed by qPCR with specific primers for the detection of DNA of these microorganisms. According to hospital microbiology results, 10 out of 33 samples were positive for *E. coli* infection (Table 5). However, none of these could be detected by qPCR as their fold change values were not significantly different compared to the negative control (Figure 37). The same case applied to the single patient positive for *K. pneumoniae* infection according to microbiology testing results. *K. pneumoniae* DNA could not be detected in sample 9 by qPCR, as fold change ratio values were not significantly different from the respective negative control. Finally, 2 samples were blood culture positive for *S. aureus* infection. Fold change values were significantly higher in one of the patient samples but not in the other. Overall, the hospital microbiology results could only be confirmed by qPCR in 1 out of the 13 samples analysed. These samples were only tested in duplicate and as a result, these results needs be taken with caution.

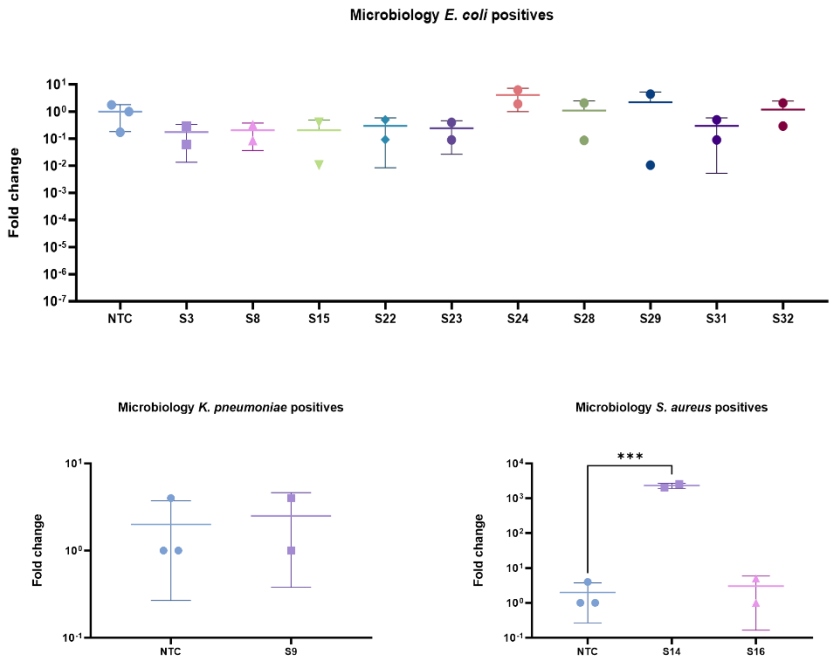


Figure 37. qPCR causative species identification on clinical blood samples. Samples that were microbiology-positive for either *E. coli*, *S. aureus* or *K. pneumoniae* or were analysed by qPCR using specific species detection primers. Ct values were normalised to those obtained from negative control samples and fold change values were obtained. Data are means of qPCR technical replicates (n=2), ± SD. ns > 0.05 (not shown), ***p ≤ 0.001.

3.4 – Discussion.

The results detailed in chapter 2 encouraged us to establish a collaboration with the NNUH to test the CMg pipeline on clinical blood samples from patients with confirmed BSI. Since this condition is the cause of a high number of hospitalization cases across the NHS hospitals [338], samples were readily available and thus this collaboration was an excellent opportunity to test a newly developed diagnostic tool as an alternative to the service standard methods. With this purpose, clinical blood samples were processed by applying both the CMg pipeline and a 16S sequencing pipeline in parallel. The key findings were:

- Saponin-based depletion was effective at removing host DNA from hospital patient samples when compared to blood samples from healthy donors.
- The metagenomics pipeline abundance results detected infective agents only in 3 samples. This number was increased to a total of 6 when all classified reads were analysed regardless of their relative abundance.
- 16S pipeline results also had a low positivity rate and only confirmed the same species as metagenomics on one occasion.
- qPCR results could only confirm standard microbiology results on one occasion.
- The extended period of time (approximately 72h) in which blood samples were left at room temperature before being processed led to almost no bacteria being left in the samples upon receival. This was confirmed by the low positivity rate obtained on the metagenomics and 16S sequencing pipelines and qPCR, which was recovered after advancing the collection time approximately 24 hours.

3.4.1 Host depletion efficiency decreased when tested on clinical blood samples.

The first step to test the CMg pipeline on the clinical blood samples was to assess the effectivity of the saponin-based host depletion. Initially, it was speculated that patients with BSI

might exhibit a higher leukocyte count in their blood compared to the amounts found in the healthy donor samples (used previously to optimise the saponin-based host depletion method). The total white blood cell count (WBC) has long served as a key laboratory test to promptly alert clinicians of potential BSI, with both elevated (>12·10^9/L) and reduced (<4·10^9/L) WBC counts serving as primary criteria for screening [339]. More recently, Surviving Sepsis Campaign guidelines in the UK have also highlighted an elevated lymphocyte count ratio as a quick prognostic indicator for BSI diagnosis [329]. However, it has also been reported by several studies that the increase in WBC among septic patients is minimal and should not be relied upon as a sole infection indicator, as these methods exhibit moderate to poor diagnostic ability [340–342].

The presence of host DNA (both chromosomal and mitochondrial) in samples from hospital patients was examined and host DNA was not significantly higher than in blood samples from healthy donors. However, it is worth noting that these samples had been stored at room temperature for approximately three days, which could have influenced the number of WBCs in the sample as described by other authors [343]. The application of the saponin-based method was able to significantly reduce the presence of host DNA in comparison with untreated samples.

3.4.2 Metagenomic abundance and 16S pipelines had poor positivity rate and sensitivity.

Despite the potential underperformance of the saponin-based depletion method observed by qPCR, host DNA presence was highly significantly reduced compared to untreated controls and therefore all patient samples were subjected to the whole metagenomics pipeline.

Initially, the taxonomic classification results were examined by assessing the relative abundance of each species' reads in comparison to the total number of classified reads obtained. After the processing of the first 14 samples, no microbial reads were obtained, with 100% of the reads classified as human. The 16S pipeline was also unable to identify any microbial species in any of these samples. It was theorized that this could be attributed to the significant time gap between

sample collection from patients and the subsequent metagenomic processing. The EDTA-coated tubes that were used for sample collection alongside blood culture has been stored at room temperature until blood cultures were positive, which took between 1 and 3 days. Then, an additional day passed until QIB staff collected the samples, resulting in an average total storage time of 2-4 days. EDTA is a commonly employed blood anticoagulant known for maintaining blood in a fluid state, crucial for proper WBC testing [344]. However, EDTA is also recognized for its antimicrobial and antibiofilm properties, often used in the treatment of skin and wound infections due its chelating properties [345]. Its antimicrobial properties have been demonstrated at concentrations as low as 1mmol/L [346], 4 times less than the usual EDTA concentration found on Vacutainers (blood vials) [347]. Furthermore, Surviving Sepsis Campaign guidelines mark the collection of blood samples before initiating antibiotic treatment as one of the main points to follow for the correct diagnosis of BSI using service standard methods [348,349]. However, real-world scenarios often conflict with this: a study performed at an ICU suggested that between 50% and 82% of suspected BSI patients were already under antibiotic treatment before their blood samples were taken [350], and another study on samples collected across the hospital suggested that this was the case for approximately 28% to 63% of patients [351,352].

It was hypothesized that the combination of EDTA and, in many cases, the presence of antibiotics in the collected samples might have led to the absence of intact microorganisms at the time of sample processing. The employed saponin-based host depletion method would digest any cell-free DNA or DNA from microorganisms with compromised membranes, effectively eliminating all DNA. Thus, after analysing the first 14 samples, the time between sample collection and processing was reduced by one day, with samples collected immediately after haematology reporting blood culture results had been obtained, rather than the following morning. Although the samples would still be stored at room temperature for an average of 2-3 days, this single-day reduction had a significant positive impact on sequencing results. After the 14th sample, both metagenomics and 16S sequencing results were able to detect microbial reads in all subsequent

samples. However, except for 6 samples, microbial read abundance remained below 1%, while host read abundance remained above 99%. Further scrutiny of these six samples with higher microbial read abundance revealed that three of them contained a significant proportion of *Mycoplasma mycoides* reads. This bacterial species is widely recognized as a common contaminant in biological products [353] and was also present in significant quantities in the respective sample's negative process controls (no blood). Consequently, these three samples were considered negative for infection. In summary, the metagenomics abundance results identified infection in only 3 out of the 33 samples.

The application of a ≥1% abundance threshold was established during the initial stages of our metagenomics pipeline (described in chapter 2). However, when this threshold was applied to clinical blood samples, it proved to be unrealistic as the majority of bacteria species reads were found at much lower abundance rates. Consequently, it was decided to explore whether a more comprehensive analysis of all microbial reads obtained on each sample could lead to the detection of additional causative agents. The individual analysis of taxonomic populations revealed that in 3 samples, there was a notably high presence of reads taxonomically classified as species that were not found in their corresponding NTC samples. Given that all of these species are well-known causes of bloodstream infections (*S. dysgalactiae* [354], *E. coli* [355] and *S. aureus* [356]), it was concluded that these three samples should be considered positive for infection. This approach resulted in a positivity rate of 18% when using the metagenomics pipeline.

The 16S pipeline complementary approach was chosen for its potential to offer high sensitivity and quicker results in diagnosing BSI [335]. Mishra *et al.* in 2020 demonstrated how a 16S-based sequencing approach applied on plasma from BSI patients led to a better sensitivity when compared to standard blood culture results and even, in some instances, the detection of species that were not detected by culture [357]. However, in this study, 16S pipeline sequencing results also showed a low positivity rate and potential sensitivity concerns. Among the 33 samples,

the 16S results managed to identify potential causative agents of infection in 10 samples. Upon comparing the outcomes of 16S and metagenomics analyses, only three instances methods identified the same samples as positive for infection, and in just one case they detected the same species. Given the consistently low positivity rates from both pipelines, sample collection was paused before reaching the initially intended target of processing 50 samples. At this point in the project and considering both metagenomic and 16S results, it was hypothesized that the observed low sensitivity rate on clinical samples compared to spiked blood samples was due to them containing negligible levels of intact bacterial causative agents.

Following the completion of sequencing analyses for both the metagenomics and 16S pipelines, sequencing results were compared to those obtained by the NNUH microbiology laboratory using service standard methodology. Among the 31 samples that tested positive for standard microbiology methods, our sequencing pipelines identified the same species in 10 samples, resulting in a sensitivity rate of 32.2%. Additionally, both pipelines accurately confirmed the two negative patient samples as truly negative, yielding a theoretical specificity of 100%. When metagenomics pipeline detected bacterial species, these were normally confirmed by service standard methods. However, there were two noteworthy instances where metagenomics results diverged from microbiology cultures: (1) In the first case, metagenomics identified *Paraccocus yeeii* as the causative agent due to its high abundance (17.7% of total reads), while the microbiology laboratory reported *E. coli* infection. *P. yeeii* is a recently recognized bloodborne pathogen that can pose challenges in identification using conventional biochemical kits as it normally goes undetected [358,359]. It is possible the microbiology laboratory could have misidentified it as another Gram-negative species that can grow in aerobic conditions such as *E. coli*. (2) The second case involved *Clostridium difficile*, which metagenomics identified as the causative agent based on a significant large abundance of reads (8.5% of total reads). However, the microbiology laboratory reported infection caused by *S. capitis*. *C. difficile* is another uncommon species capable of causing bloodstream infections, typically escaping detection by service standard microbiology testing,

which often reports other Gram-positive species [360,361] or shows no growth on media [362]. As reported by other authors [103,363], a metagenomics approach has the potential to detect species that normally go undetected or are misclassified as others when using culture-based standard methodologies. The findings on these two samples could also be suggesting that the presented CMg pipeline had the ability to do detected these species that go normally undetected. However, further experimentation, such as confirming with qPCR analysis or a more detailed sequencing analysis, would be necessary to confirm the metagenomics results in these two cases.

3.4.3 qPCR validation could only confirm microbiology results on one patient blood sample.

The notably low sensitivity and positivity rates observed with both sequencing pipelines raised the possibility that the patient samples processed may have contained little to no detectable levels of intact bacteria. In pursuit of a more sensitive and specific alternative, samples that had tested positive for *E. coli*, *S. aureus*, or *K. pneumoniae* infection according to NNUH microbiology results were subjected to qPCR analysis. This approach was chosen because qPCR has been recognized as one of the most sensitive methods for detecting bloodstream infections [76]. Although this approach being limited to only detected species that the primers are present for, on samples with known bacterial species it is capable of detecting the majority of bacterial species when present at low levels [111,364]. However, it is important to note that microbiology results could only be confirmed in one out of the 13 samples analysed by qPCR, which contributed even further to the idea that the majority of blood samples contained no intact bacteria at the time of receival.

3.4.4 Conclusion.

When establishing the main pre-analytical considerations of this collaborative project, the sample selection criteria (blood samples only coming from patients that were blood culture

positive) was prioritised over the sample collection time. However, this resulted in poor sensitivity observed in the metagenomics pipeline, 16S pipeline and qPCR testing. The most likely explanation for the low sensitivity observed on all methods was the extended incubation period at room temperature, spanning approximately 4 days from collection to processing. The influence of extended incubation times before processing was confirmed by decreasing this to one day, which had a significant positive impact on the detection of microorganisms through sequencing. This would have to be taken into consideration when establishing new collaborations for the processing of clinical blood samples. The immediate processing of samples after their extraction from the patient combined with the application of the method on patients that have not received antibiotic treatment will be key to clinically demonstrate the efficacy of this method when compared to service standard methodologies.

CHAPTER FOUR

FURTHER DEVELOPMENT OF THE HOST DNA DEPLETION METHODOLOGY AND REFINEMENT OF THE BIOINFORMATICS PIPELINE.

4.1 – Introduction.

The adaptation and implementation of new methodologies in a clinical setup is one of the most challenging steps in the development of novel diagnostic tools. In this study, the efficacy of the developed CMg pipeline in detecting BSI causative species in blood samples has been demonstrated in mock samples. However, when applying the pipeline on clinical blood samples, the observed sensitivity was lower than that of the service standard culture methods. Because of this, potential opportunities for methodological refinement were defined in multiple areas. Thus, the final part of this study focused on improving the performance of the pipeline by changing specific methodological considerations (such as using blood collection tubes with an anticoagulant alternative to EDTA and introducing further methods to enhance the host depletion step) and modifying analytical parameters to refine the bioinformatics pipeline. The ultimate goal was to produce a CMg pipeline with improved sensitivity that will have better prospects for clinical use than blood culture-based diagnostic methods.

4.1.1 Transitioning to a new blood anticoagulant collection tube.

Ethylenediamine Tetraacetic Acid or EDTA is one of the most common anticoagulants present in blood collection tubes used in hospital haematology laboratories to collect samples for diagnostics purposes. EDTA can maintain the physical and chemical properties of blood in its original state for extended periods of time and is especially useful for its ability to preserve the components and morphology of human cells [365]. Due to its common use to support clinical analysis of blood, samples collected in EDTA-coated tubes were used to optimise the CMg pipeline and during the hospital collaboration detailed in previous chapters. The negative effect of an extended incubation time of samples in EDTA-coated tubes may have had an impact on the microbial viability in experiments performed on clinical blood samples (chapter 3). For this reason, it was necessary to find an alternative anticoagulant that would not compromise the microbial viability when samples were stored for extended times. In 2022, Rivolta *et al.* [366] demonstrated,

in an anticoagulant comparison experiment, how EDTA highly inhibited the growth of species such as *Rhodococcus equi,* while Citrate-phosphate-dextrose (CPD) did not have an effect on bacterial growth. CPD is an anticoagulant which is also commonly used to collect blood samples for diagnosis (generally for performing coagulation studies [367]) and it is especially useful because it confers blood an extended shelf life of 21 days [368]. Moreover, to date, the negative effect of CPD on microorganisms has not been described in the literature, which made this anticoagulant an excellent candidate for its use in the collection of blood samples that needed long incubation periods without harming the microorganisms and, overall, to be introduced in the CMg pipeline.

4.1.2 Sepsi*PURE* capture and enrichment method to increase the CMg pipeline sensitivity.

In order to further improve the sensitivity of the CMg pipeline, the implementation of the Sepsi*PURE* capture and enrichment method was tested in collaboration with the company Momentum Bioscience Ltd. Sepsi*PURE* is part of their overall pipeline for the detection of BSI Sepsi*STAT®*, a diagnostic method which ultimately detects BSI causative agents and antimicrobial genes using an amplification based-approach [369]. The Sepsi*PURE* method is capable of capturing microorganisms in 5mL blood samples using a magnetic bead-based system to extract the microorganisms from the rest of the blood components. Then, the captured microorganisms go through a growth phase (microorganism enrichment phase) in Momentum Bioscience's proprietary medium named TRM [370]. This way, microorganisms are isolated and further amplified, while host cells remain in the discarded fraction. Because of this, it was hypothesized that the combination of this method with the saponin-based host depletion protocol would synergistically improve host DNA depletion and increase the quantity of microbial DNA yield, overall resulting in an increase in the sensitivity of the CMg pipeline.

4.1.3 Filtering of reads and taxonomic classification.

In the CMg field, the correct interpretation of the sequencing results is key to perform the most accurate interpretation of the microorganism composition of a sample and decipher the contamination from microbes [170]. To analyse nanopore sequencing data, EPI2ME is a widely used software that was also used to perform the taxonomic classification and abundance check by other group members to diagnose respiratory infections [136]. Despite presenting clear advantages, such as being quick and easy-to-use, using EPI2ME can also have some limitations when applied to metagenomics analysis of blood samples due to its *k*-mer-based approach nature. When performing metagenomics informatic taxonomic classification, there is usually detection of false-positive species (species that are misidentified as others non-present in the samples). This is mainly due to the combination of (1) reads mapping to regions that are highly conserved among similar species and (2) the high presence of background contaminant reads coming from multiple sources [371]. In a benchmarking comparative analysis of metagenomics classification tools, Simon *et. al.* [372] demonstrated in 2019 that the application of the Centrifuge software led to the highest quantity of false positive identifications (a large number of microorganisms species that were not present in the sample were detected). The high false positive quantification was theorized to be mainly due to the use of a *k*-mer approach combined with the use of the wrong reference databases. Conversely, using taxonomic classification pipelines which use an alignment-based approach of the whole read (non-*k*-mer) can potentially eliminate some of those contaminant reads or reduce the number of misclassified species.

When comparing *k*-mer and alignment-based tools to study microbial communities, Chan [373] demonstrated that minimap2 was the one that resulted in the highest specificity at accurately taxonomically classifying reads. Some authors have pointed that the application of a pairwise alignment tool to analyse the reads needs to be combined with a read filtering step based on alignment scores. For that, each reads gets assigned a score based on its identity and coverage

values, while the reads with lowest scores are removed [374]. Finally, it has been reported that metagenomics results obtained from nanopore sequencing can benefit from an extra filtering step based on the removal of short fragments [375]. The taxonomic classification step is key for the sensitivity of the CMg pipeline. As such, efforts were made to implement other bioinformatics tools combined with extra filtering steps and asses if the sensitivity could be improved.

4.1.4 K-mer-based antimicrobial resistance (AMR) genes detection.

The use of CMg approaches for BSI diagnostics brings an opportunity to further exploit sequencing results by assessing the presence of AMR determinants, which will allow more targeted treatments. After implementing the improvements previously described, it was hypothesised that CMg pipeline would provide enough sensitivity and coverage for the successful detection of AMR genes. Adding this key component to the pipeline would greatly contribute to supporting the clinical microbiology aim of improving antibiotic treatment.

In 2016, Clausen *et. al.* [191] made the KmerResistance (KMA) tool public. KMA uses the *k*-mer read-based alignment technology to quickly detect microbial resistance genes by mapping fragments of individual reads (*k*-mer) against the RefSeq database of resistant genes. The authors demonstrated that a read mapping approach offers a superior resistance prediction over assembly-based methods and that the KMA method was more sensitive and accurate when compared to other read mapping tools (ResFinder or SRST2). Despite the inability this read-based approach to assemble full genomes, its use is more suitable for samples with low amounts of DNA (which leads to low coverage) for an approach based on the full assembly of the genome [189].

4.1.5 Aims and hypothesis.

After the use of our CMg pipeline and on clinical blood samples and the subsequent analysis of results, the need to further optimise the method was observed. The hypothesis for the last part of the study were:

- That the coating of EDTA-coated tubes induced microorganism degradation in blood samples and that the use of a different anticoagulant coating would improve the viability of microorganisms and the overall performance of the pipeline.
- That the use of Sepsi*PURE* technology would positively contribute to the depletion of host DNA.
- That the use of a bioinformatics read-based approach and the introduction of new thresholds to analyse all of the obtained bacterial reads would be beneficial for increasing the pipeline sensitivity.
- That the resulting improved CMg pipeline would also be able to bioinformatically detect AMR determinants.

These hypotheses were tested through the following aims:

- To test the saponin-based depletion method on blood samples collected in CPD-coated tubes.
- To increase the method's sensitivity by introducing the Sepsi*PURE* technology into the pipeline in collaboration with Momentum Bioscience Ltd.
- To reassess the bioinformatic analysis by using minimap2 as a classification tool, introducing length-based filtering of reads and a filtering based on the assignation of an alignment score to each microbial DNA read.
- To use the bioinformatics KMA pipeline to investigate the presence of AMR determinants in the CMg output data.

4.2 – Methods.

4.2.1 Blood sample obtention.

Whole human blood samples were commercially acquired from Cambridge Bioscience. These blood samples had been collected from healthy donors aged between 18-60 and were supplied in Ethylenediaminetetraacetic acid (EDTA)- or Citrate-Phosphate-Dextrose-coated (CPD)-coated anticoagulant tubes and stored at 4˚C.

4.2.2 Generation of mock specimens by spiking of pathogens into blood samples and colony count.

Different volumes of blood corresponding to variants of the core protocol (1mL or 5mL) were spiked using serial dilutions (in PBS) of *Escherichia coli* (ATCC 10536 or NCTC 13441), *Staphylococcus aureus* (ATCC 25923), *Klebsiella pneumoniae* (ATCC 13882) and *Enterococcus faecalis* (NCTC 13779) from overnight cultures in Luria-Bertani (LB) broth or TRM medium (Momentum Bioscience Ltd). The spiking and the colony count was performed as detailed in section 2.2.2.

4.2.3 Saponin-based host depletion and anticoagulants tubes comparison.

To compare the application of the saponin-based host depletion effect on CPD-coated compared to blood coming from EDTA-coated tubes, 1mL of blood (EDTA or CPD) underwent the saponin-based depletion (3% final concentration) method described at the end of section 2.2.13. After that, the DNA was extracted exactly as described in section 2.2.4. To assess the efficiency of the host depletion step, extractions were run on a qPCR for the detection of chromosomal DNA as described in section 2.2.5.

4.2.4 Sepsi*PURE* enrichment method.

To test the Sepsi*PURE* method efficiency, 5mL blood samples from CPD-coated tubes were subjected to Sepsi*PURE* microbial isolation and enrichment protocol (Momentum Bioscience Ltd) by simulating a 1-hour transportation phase (mimicking time between phlebotomy and the beginning of sample processing) in 5mL TRM medium (20˚C, static) and then combining the blood with 715µL of binding buffer containing universal microbial capture magnetic beads and incubated for 30 minutes for microbial extraction. Magnetic beads with microbes attached were retained at the side of the tube by magnetisation. The rest of the blood sample was discarded and beads were subjected to a wash using a proprietary washing buffer (Momentum Bioscience Ltd). The retained microorganisms were eventually subjected to a 4-hour growth phase at 37˚C in 150µL of LB or TRM medium (Momentum Bioscience Ltd). Beads were kept in the medium or removed by magnetisation (depending on the experiment) and the medium containing the enriched microorganisms was mixed with PBS for a total volume of 1mL and subjected to saponin-based host DNA depletion (3% final concentration) method described at the end of section 2.2.13 and DNA extracted as in section 2.2.4. The effect of individually applying the Sepsi*PURE* protocol to blood samples as well as in combination with the saponin-based method was assessed by qPCR by measuring the effects of these methods on both host DNA (chromosomal and mitochondrial DNA) and *E. coli* DNA as described in section 2.2.5.

4.2.5 Statistical analysis.

Data on fold change values as well as total number of reads are expressed as means and ± SD and were analysed using GraphPad Prism 10 (version 10.0.2) software. Fold change multiple comparison analysis were performed by obtaining p-values performing an ordinary one-way ANOVA test with a post hoc Tukey's test. Values were considered significant and represented in the graphics when *p ≤ 0.05, **p ≤ 0.01. ***p ≤ 0.001, ****p ≤ 0.0001. Importantly, non-significant differences (ns > 0.05) were not shown.

4.2.6 Whole genome amplification and debranching.

After subjecting 5mL the samples to the SepsiPURE method in combination with the saponin-based depletion (section 4.2.4) or the 1mL samples only to the saponin-based host depletion method (section 4.2.3), samples were DNA extracted and then bead cleaned with AMPure beads as described in section 2.2.10 to finally elute the DNA in 15µL of water. Then, all the samples were subjected to the optimised WGA protocol described on 2.2.8. Finally, amplified samples were subjected to a debranching step using T7 Endonuclease I as described on section 2.2.9.

4.2.7 Nanopore library preparation and sequencing.

The debranched product was subjected to the Rapid PCR barcoding kit SQK-RPB004, following the manufacturer's guidelines with the modifications described on section 2.2.12. Once libraries were prepared, samples were quantified, pooled in equal concentrations, and cleaned with a 0.6X AMPure as detailed in 2.2.12. Approximately 100fmoles of pooled PCR library was loaded onto the MinION flow cell (R9.4.1, Oxoford Nanopore Technologies). All samples sequencing run time was 3 hours or 24 hours (only for AMR experiments), after that fastq files were obtained.

4.2.8 Bioinformatics pipeline.

A custom pipeline for taxonomic profiling and screening antimicrobial resistance determinants was developed (https://github.com/quadram-institute-bioscience/nf-ont-taxamap). The fast5 files acquired were base-called with the Super Accuracy (SUP) model using guppy_basecaller and subsequently demultiplexed using guppy_barcoder of the Guppy software v6.2.x [376]. For the SQK-RBK004 kit, the minimum barcoding score was set at 65 and mid-read barcode filtering enabled, while for the SQK-PBK004, the barcoding score was kept at 60, and set enabled for the options: barcode both ends and mid-read barcode filtering.

The pipeline involved the following steps for taxonomic profiling and screening AMR genes: human reads were removed from each barcode fastq file by mapping each barcode's reads against the *Homo sapiens* reference genome CHM13v2 [377] using minimap2 (version 2.17) [175]. Un-mapped reads to the CHM13v12 were included for mapping against the RefSeq database curated as of the 6th of March 2022 using minimap2 (v2.17). The mapped reads were filtered based on a taxa_score defined as a harmonic mean of identity and coverage of the mapped reads (2*[identity+coverage]/[identity*coverage]), only reads with *taxa score* ≥85 (in a range from 0 to 100) were included for taxonomic profiling, which identified the taxa and their relative abundance. The non-human reads were also used for screening AMR genes using KMA (version 1.4.9) [378] tool with the chosen database amrfinderplus (v3.10). Average read coverage was calculated taking into account both upstream and downstream flanking regions and barcodes introduced during nanopore library preparation. These, combined, added approximately 100-150bp on each read which would represent approximately between 6% and 12% of the bases on each read (averaging between 1000 and 2000 bases) so maximum average read coverage values matching the reference that could be obtained were approximately 88-95%. Final results were aggregated into an Excel format, including taxa relative abundance, AMR genes, and species-based coverage estimated from the total mapped bases against multiple references of the species and then divided by the species genome size. From the relative abundance of the bacteria within each barcode/sample, the sample is concluded to be positive with the taxon if the abundance of reads for the species, over the total population of classified reads, was ≥90% and ≥5 reads matched that species. AMR genes were considered to be present if their template identity and coverage and their depth of coverage reported by KMA were superior to ≥98% and ≥1X, respectively.

4.3 – Results.

4.3.1 – Saponin-based depletion using EDTA or CPD.

In order to assess the suitability of CPD as an anticoagulant for the samples, the saponin-based host DNA depletion method was tested on blood samples collected in CPD-coated tubes. To do that, the 3% scaled-up saponin-based protocol (optimised on chapter 2) was ran in parallel on blood samples collected on both EDTA- and CPD-coated tubes. Host depletion efficiency was compared in both chromosomal and mitochondrial DNA (fold change reduction) after the depletion protocol.

When comparing the depleted samples to their respective non-depleted controls, the saponin-based method applied on blood collected on EDTA- or CPD-coated tubes depleted chromosomal and mitochondrial DNA to comparable levels (Figure 38). The application of this protocol to blood on CPD-coated tubes, resulted in an approximate 10^4-fold reduction of chromosomal DNA and a 10^2-fold reduction of mitochondrial DNA. From this point onward, the experiments were conducted using blood collected in CPD-coated tubes.

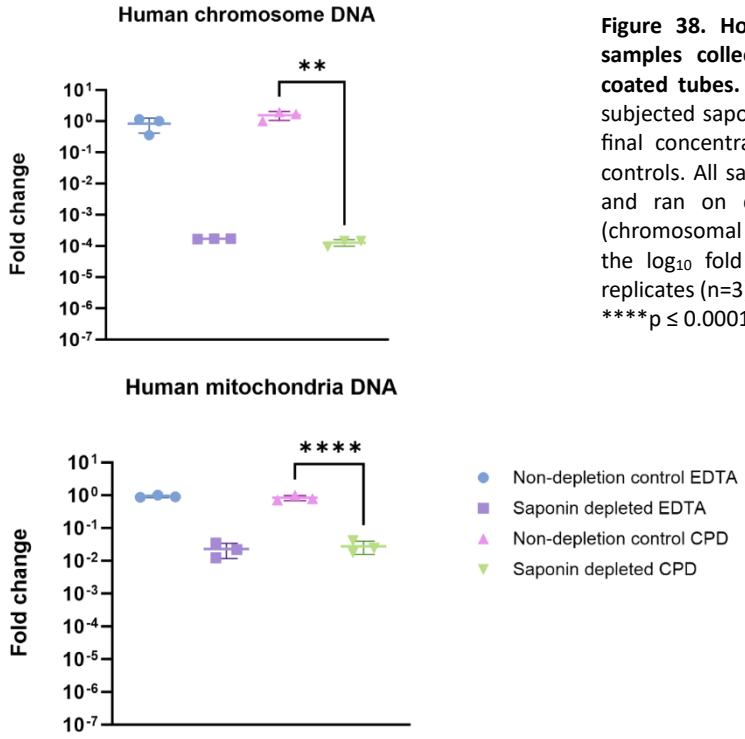


Figure 38. Host DNA depletion efficiency on blood samples collected on EDTA-coated tubes and CPD-coated tubes. 1mL EDTA- or CPD-blood samples were subjected saponin-based porotocol with saponin at 3% final concentration. Untreated samples were used as controls. All samples were subjected to DNA extraction and ran on qPCR for the detection of host DNA (chromosomal and mitochondrial). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01., ****p ≤ 0.0001.

4.3.2 – Sepsi*PURE* enrichment method optimisation.

The implementation and optimisation of the enrichment Sepsi*PURE* method developed by Momentum Bioscience Ltd was a key part of the collaboration with the industrial partner that partly funded this project. This was an excellent opportunity to work on the application of Sepsi*PURE*, a method that would potentially allow us to work with bigger volumes of sample, potentially further eliminate host DNA and increase the abundance of bacteria for their better detection.

4.3.2.1. First combination of Sepsi*PURE* with host DNA depletion method.

The initial experiments with Sepsi*PURE* aimed at checking whether the introduction of this method could result in an increase on the amount of host DNA being depleted. This was assessed by testing the method on its own as well as in combination with the saponin-based method. To combine both methods, 5mL blood samples were spiked with 50-100 CFU/mL of *E. coli* and subjected to the Sepsi*PURE* protocol. The beads with the microorganism attached were immersed in TRM medium and subjected to the saponin-based host depletion protocol.

The application of the Sepsi*PURE* method on its own caused a significant reduction in both chromosome and mitochondrial host DNA, an average of 10^3 and 10^2-fold decrease, respectively (Figure 39). The Sepsi*PURE* method caused a similar reduction of host DNA to the saponin-based method applied on its own. However, the combination of both methods did not result in a significant increase in host DNA depletion when compared to fold change ratio results obtained on samples with the methods applied individually.

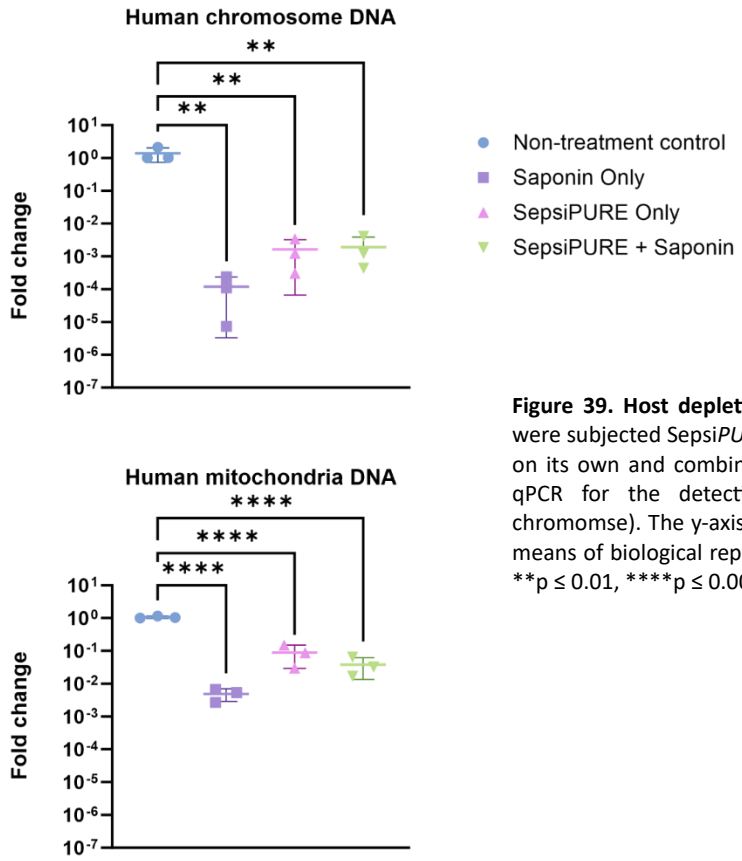


Figure 39. Host depletion efficiency using Sepsi*PURE*. Samples were subjected Sepsi*PURE* and saponin-based depletion methods on its own and combined. DNA extracted samples were ran on qPCR for the detection of host DNA (mitochondrial and chromomse). The y-axis represents the $\log_{10}$fold change.Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01, ****p ≤ 0.0001.

4.3.2.2 Effects of changing Sepsi*PURE* incubation medium.

Initial experiments using the SepsiP*URE* protocol demonstrated that this method was highly efficient at removing host DNA but, unfortunately, its addition to the saponin-based protocol did not synergistically improve the reduction of host depletion efficiency. It was hypothesized that there could be a molecular or chemical element in the Sepsi*PURE* reagents that had an inhibitory effect on the saponin-based reaction. To investigate this inhibitory effect, all the reagents of the Sepsi*PURE* reaction were individually mixed with blood samples which were then subjected to the saponin-based host depletion protocol (results not shown as reagents are considered commercially sensitive information). After detailed analysis, the TRM medium, which is used in both transportation step and to incubate the beads, was the only product used for the Sepsi*PURE* protocol that caused a reduction in depletion efficiency when mixed with the saponin-based host depletion method. A closer inspection of the components present in the TRM medium revealed

that it contained sodium polyanethol sulphonate, or SPS (at concentrations lower than 0.03%). It was previously demonstrated how the presence of this chemical was highly detrimental to the saponin-based host depletion even at concentrations as low as 0.0005% (section 2.3.1.3). For this reason, the next experiment consisted in combining both methods again, but replacing the TRM medium that was added for the incubation step with Luria Broth (LB) medium.

Similarly to what was observed when using Sepsi*PURE* method with TRM medium (section 4.3.2.1), the individual application of the method (using LB medium) resulted in a significant reduction on both chromosomal and mitochondrial host DNA. Additionally, the combination of Sepsi*PURE* using LB medium with the saponin-based method increased the depletion of chromosome host DNA by approximately 10^2-fold when compared to samples only subjected to the saponin method, although this difference was not statistically significant. The depletion of mitochondrial DNA remained the same as with the application of each method individually (Figure 40).

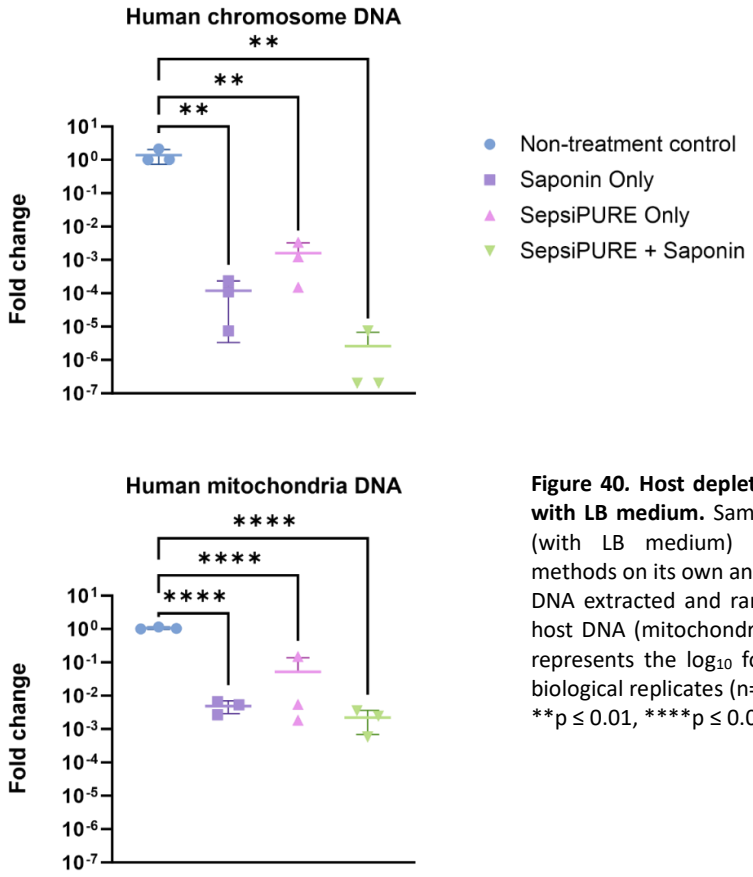


Figure 40. Host depletion efficiency using Sepsi*PURE* with LB medium. Samples were subjected Sepsi*PURE* (with LB medium) and saponin-based depletion methods on its own and combined. Samples were then DNA extracted and ran on qPCR for the detection of host DNA (mitochondrial and chromomse). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), **p ≤ 0.01, ****p ≤ 0.0001.

4.3.2.3 Comparing Sepsi*PURE* medium and beads fractions.

The substitution of the TRM medium for an SPS-free medium (LB) caused a moderate increase in depletion efficiency. However, more improvements were required to take the most advantage of combining both methods. The application of the Sepsi*PURE* method on blood samples included the selective binding of bacteria to the magnetic beads, which allowed for the blood components to be washed away. After that, the beads were incubated in LB medium (for 4 hours) and then both beads and medium were subjected to the saponin-based depletion methods. We hypothesized that one of the two fractions (medium or beads) could potentially still contain a large presence of host cells which was still impacting the depletion efficacy. To test this hypothesis, the beads and the medium fraction were analysed separately by qPCR for the presence of host (chromosomal and mitochondrial) and *E. coli* DNA.

Upon analysis of the qPCR results, a large presence of host DNA was observed in the beads fraction: chromosomal DNA fold change values on beads were very similar to those obtained from untreated samples. Only fold change ratios obtained on mitochondrial DNA were significantly different to the untreated controls (depletion efficiency was still lower when compared to saponin only samples; Figure 41). In contrast, a significant reduction on host DNA (both chromosomal and mitochondrial) was observed in the medium fraction. Fold change values obtained in this fraction were similar to the ones obtained from samples subjected to saponin-based host depletion, which confirmed that beads not only had affinity for bacteria, but they also captured host cells. When analysing the *E. coli* fold change, a large increase in the bacteria presence was observed in samples subjected to the Sepsi*PURE* method (both beads and medium fraction), with approximately $>10^2$ and $>10^3$-fold increase (respectively) when compared to non-treated or saponin-only controls (Figure 41).

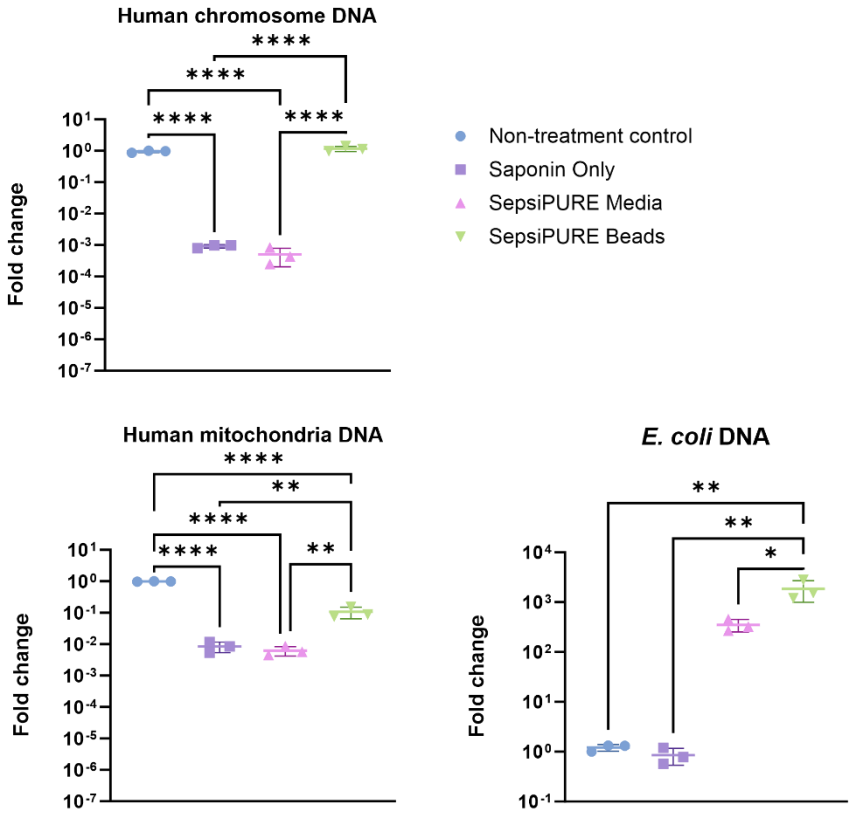


Figure 41. Beads and media fractions comparison after applying Sepsi*PURE* method. Spiked blood samples were untreated or subjected to the saponin-based host depletion or Sepsi*PURE* method. At the end of the Sepsi*PURE* method, beads and media fractions were analysed separatly. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse DNA). The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *$p \leq 0.05$, **$p \leq 0.01$, ****$p \leq 0.0001$.

4.3.2.4 Testing the optimised Sepsi*PURE* method.

After analysing the media and beads fractions separately, it was concluded that the depletion method could highly benefit from only working with the medium fraction and discarding the beads. The beads fraction had a large proportion of host cells bound to them compared to the medium fraction (10^2-fold more chromosomal and mitochondrial DNA). At the same time, the beads fraction only enriched the bacterial presence 10^1-fold more than the medium fraction. Thus, assessment of whether the removal of the beads would positively contribute to the host DNA depletion was performed by subjecting blood samples to Sepsi*PURE* and then applying the saponin-based depletion protocol to the media fraction only.

The addition of Sepsi*PURE* to the saponin based method culminated in a significant 10^1-fold decrease in both chromosomal and mitochondrial host DNA (Figure 42) compared to saponin-only samples. For the first time, the methods synergistically benefit from each other. When observing the levels of bacterial DNA, the Sepsi*PURE* protocol resulted in a significant increase in *E. coli* DNA detection both when used alone and in combination with the saponin-based depletion by an approximate 10^3-fold increase when compared to non-treated controls.

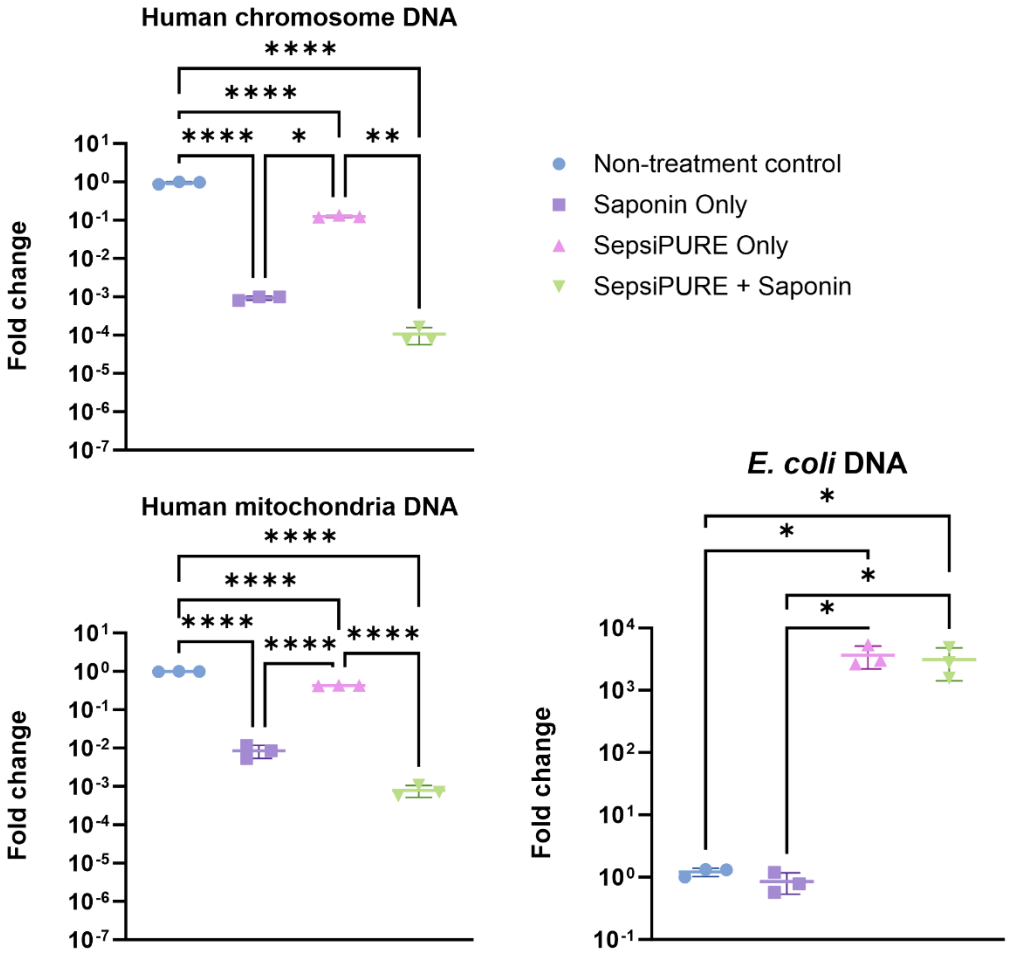


Figure 42. Assessment of host DNA depletion and *E. coli* DNA enrichment when combining Sepsi*PURE* and saponin-based depletion methods. Spiked blood samples were untreated or subjected to different combinations of treatments. For samples subjected to Sepsi*PURE* method, only the medium fraction was saponin-based depleted. Samples were then DNA extracted and ran on qPCR for the detection of *E. coli* and host (mitochondrial and chromomse) DNA. The y-axis represents the $\log_{10}$ fold change. Data are means of biological replicates (n=3), ± SD. ns > 0.05 (not shown), *$p \leq 0.05$, **$p \leq 0.01$, ****$p \leq 0.0001$.

4.3.3 –PCR-free library preparation approach – testing the pore activity.

As a result of microorganism enrichment, the extracted DNA yields were highly increased (10-100X) in comparison to the ones obtained from blood samples which were only subjected to the saponin-based host depletion method. In previous CMg pipeline iterations, a PCR-based library preparation kit (SQK-RPB004) was used due to the need to further amplify the limited amount of input material. However, the obtention of superior yields in this new version of the pipeline allowed the use of a PCR-free library preparation kit such as the Rapid barcoding kit (SQK-RBK004).

In initial sequencing experiments in which flow cells were loaded with libraries prepared with the PCR-based kit (SQK-ROB004), the pore activity highly decreased within the first sequencing hour (section 2.2.3). However, it was important to assess if switching to a PCR-free based approach would result in a difference in nanopore pore activity.

After comparative experiments of samples prepared with one of the two library preparation kits, the use of the new RBK kit (PCR-free) significantly reduced the speed at which pores were decreasing their activity. While with RBP (PCR-based) samples experienced an average decrease in pore activity of approximately 30% and 40% of their pores after 1 and 2 hours, respectively (Figure 43). On RBK samples, most pores that were active on time 0, were still active after 1 and 2 hours of sequencing. After 3 hours of sequencing, the activity pore loss was more noticeable, but the decrease in pore activity was still significantly lower than that of the RPB (PCR-based library) samples.

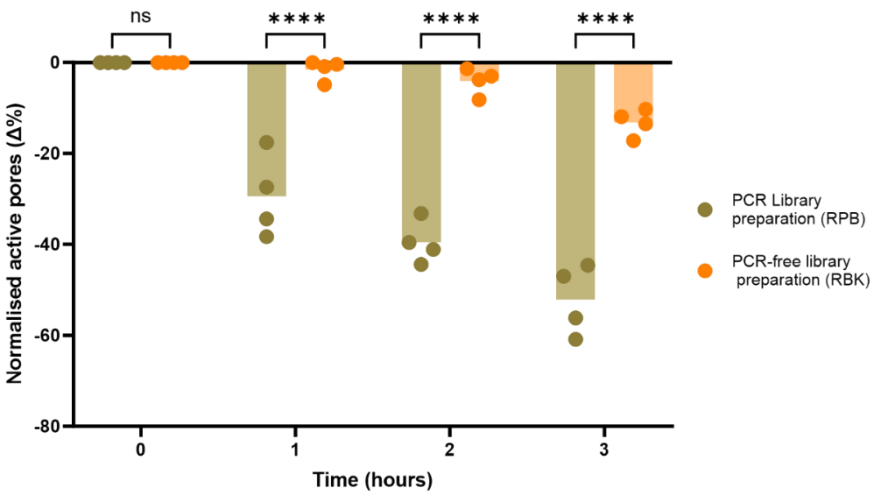


Figure 43. Comparison of pore activity on samples subjected to PCR-based or PCR-free library preparation methods. Measurement of Δ% activity evey hour by substracting pore actvity evey hour to the pore actvity % obtained at 0 hour. Samples underwent the Sepsi*PURE* and saponin-based host depletion protocols combined, WGA, debranching and different library prepeartion (RPB or RBK kits) and Δ% were compared. Data are means of biological replicates (n=3). ns > 0.05, ****p ≤ 0.0001.

4.3.4 – Limit of detection.

Once the Sepsi*PURE* method was optimized, the effect of its implementation on the overall pipeline sensitivity was tested by performing limit of detection (LoD) experiments. LoD experiments were the ultimate test that combined all the previously described optimized steps to measure the lowest analytical concentration that the CMg pipeline was able to detect.

4.3.4.1 Limit of detection experiments with *E. coli* and *S. aureus* - analysis based on relative abundance.

In brief, the new CMg pipeline consisted of: 5mL blood underwent the Sepsi*PURE* and saponin-based methods. Then, DNA was extracted and subjected to WGA and debranching. Finally, debranched DNA products were subjected to RBK (PCR-free) library preparation followed by nanopore sequencing. This pipeline was termed the *'5mL enrichment protocol'* and was tested by spiking clinically relevant concentrations of *E. coli* or *S. aureus*. Sequencing results were analysed by taxonomically classifying all reads as human, *E. coli*/*S. aureus* or other bacterial species.

Finally, the *'5mL enrichment protocol'* abundance results were compared to the ones obtained with the 1mL protocol (described in chapter 2 section 2.3.8.2 and named *'1mL standard protocol'*) to assess the effect of the addition of the optimised Sepsi*PURE* protocol to the pipeline. Note that for the comparative experiments, these two protocols contained different volumes of processed blood (1mL vs 5mL) but bacteria were spiked at the same concentration. A saponin-based depletion method directly applied (without Sepsi*PURE*) to 5mL of blood was not used as a control because it led to poor depletion efficiency. Instead, the *'5mL enrichment protocol'* was compared to the *'1mL standard protocol'*, which was the most effective protocol that did not require the Sepsi*PURE* enrichment step.

When using the *'5mL enrichment protocol'* on samples spiked with *E. coli*, the abundance of *E. coli* reads was ≥1% for all spiking concentrations: 50-100 (72.3%), 10-50 (23.8%) and 1-5 CFU/mL (21.8%), which revealed a limit of detection of 1-5CFU/mL (Figure 44). These values were evidently higher than the ones obtained for the 1mL protocol (data from section 2.3.2.8.2), in which *E. coli* reads had an average abundance of 63.67% for samples spiked with 50-100 CFU/mL, 17.32% for 5-10 CFU/mL and 5.01% for 1-5 CFU/mL samples. NTC samples contained no reads that were classified as *E. coli* or any other bacterial species; all reads were classified as human.

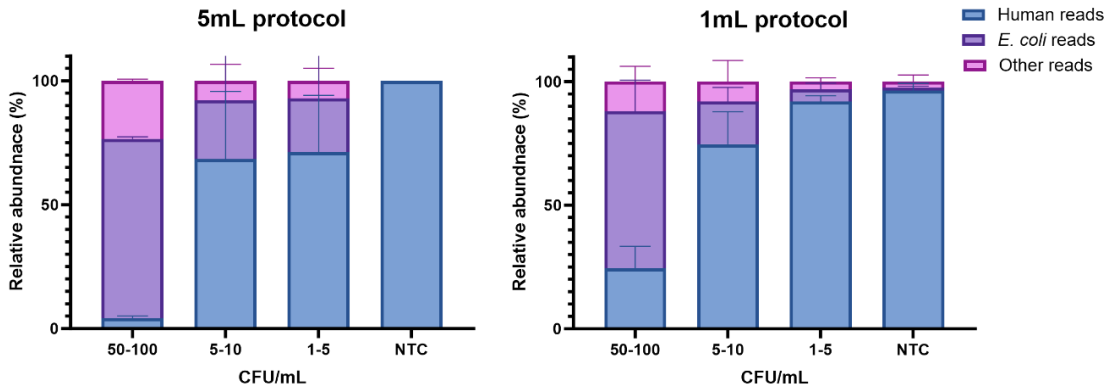


Figure 44. LoD abundance results on *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines spiked with *E. coli*. 5mL or 1mL blood samples were spiked with 50-100, 10-50, 5-10, 5-10 CFU/mL of *E. coli.* Non-spiked samples were used as controls (NTC). 5mL samples were subjected to Sepsi*PURE* and then both 1mL and 5mL samples underwent saponin-based depletion. Taxonomically classified reads were classified as '*human*', '*E. coli*' and any bacterial read that were not the spiked species (*other reads*). Relative abundance of each of the three categories were obtained and samples were classified as positive for bacterial infection if the relative abundance of any of the species detected was ≥1%. Data are means of biological replicates (*n*=3), ± SD.

When using the *'5mL enrichment protocol'* on samples spiked with *S. aureus,* the abundance of reads pertaining to this bacterial species was only ≥1% in the samples spiked with 50-100 CFU/mL (10.5%) (Figure 45). In the rest of the samples, taxonomical abundance results were highly dominated by the presence of human reads. Conversely, the LoD results on samples subjected to the *'1mL standard protocol'* had a superior sensitivity for the detection of *S. aureus* (data from section 2.3.2.8.2). Using the '1ml standard protocol', *S. aureus* abundance reads were 32.62% and 3.20% on 50-100 and 10-50 CFU/mL samples, respectively.

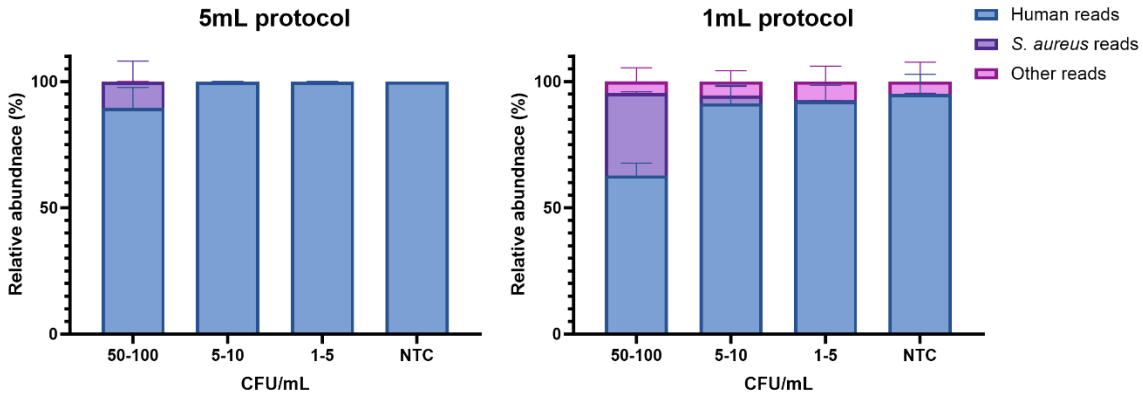


Figure 45. **LoD abundance results on *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *S. aureus*.** 5mL or 1mL blood samples were spiked with 50-100, 10-50, 5-10, 5-10 CFU of *S. aureus* and non-spiked (NTC). 5mL samples were subjected to SepsiPURE and then both 1mL and 5mL samples underwent the saponin-based depletion. Taxonomically classified reads were classified as '*human*', '*S. aureus*' and any bacterial read that were not the spiked species ('*other read's*). Relative abundance of each of the three categories were obtained and samples were classified as positive for bacterial infection if any of the species detected relative abundance reads was ≥1%. Data are means of biological replicates (*n*=3), ± SD.

4.3.4.2 *S. aureus* limit of detection experiments, checking for total number of reads and bioinformatics thresholds optimisation.

Despite demonstrating how Sepsi*PURE* method successfully increased the depletion of host DNA, LoD experiments revealed a lower detection of *S. aureus* when using the '*5ml enrichment method*' versus when applying the *'1mL standard protocol'* (no Sepsi*PURE* approach). It was decided to further investigate the sequencing results obtained with the *'5mL enrichment protocol'* and assess if the detection could be improved by refining the bioinformatics analysis. During the clinical collaboration with NNUH (described in chapter 3) the bioinformatics detection of bacteria benefited, in some scenarios, from an analysis based on the total number of microorganisms

classified reads (compared to analysing their relative abundance (section 3.3.2.2). Therefore, a similar approach was applied to the sequencing results of the samples spiked with *S. aureus* and subjected to the *'5mL enrichment protocol'*. This time, all the reads that were classified as human were bioinformatically eliminated as they were no longer needed for the analysis. Then, the bacterial reads were filtered according to their size (minimum 250 bp) and their *taxa score*. Finally, the remaining classified bacterial reads were grouped as either *S. aureus* or any other classified species reads (other reads).

By using the *'5mL enrichment protocol'* and applying this detailed bioinformatics filtering, *S. aureus* reads were detected at all spiked concentrations and their average total number was always superior to that of any other bacterial read (Figure 46; Table 8).

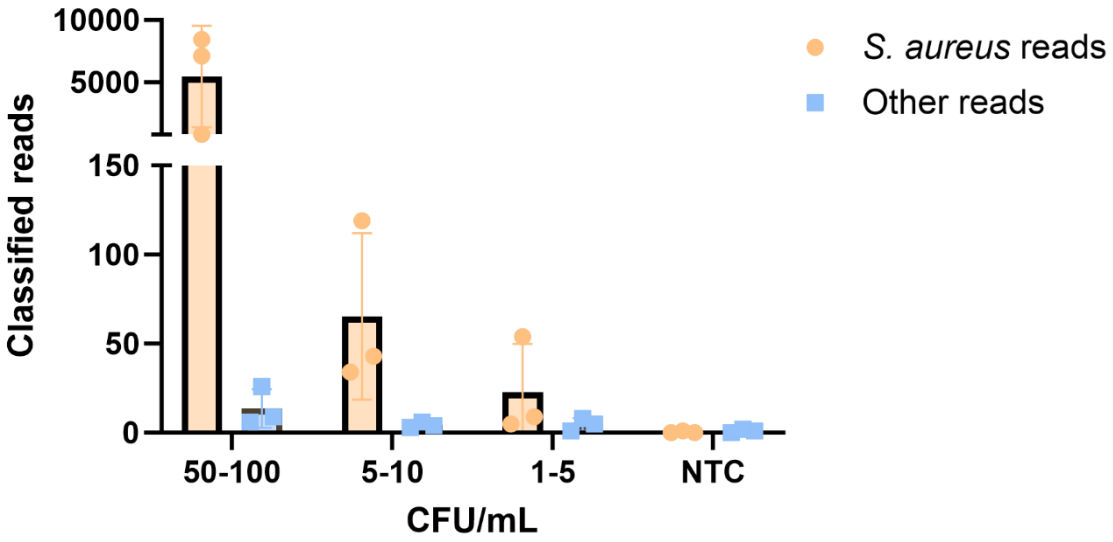


Figure 46. Total number of taxonomically classified *S. aureus* reads using the *'5mL enrichment protocol'*. Number of classified reads obtained as *S. aureus* when spiked at different concentrations (50-100, 5-10 or 1-5 CFU/mL). The different concentrations were spiked into 5mL of whole blood and were subjected to enrichment pipeline. Sequencing results were grouped into two different categories: *S. aureus* reads = taxonomically classified reads that matched *S. aureus*; Other reads = reads that were classified as any other microbial species. Data are means ± SD. n= 3 are biological replicates.

Importantly, there were almost no bacterial reads detected in NTC (non-spiked samples) samples and their *S. aureus* reads presence was minimal, with only 1 read detected in one of the

replicates. With the elimination of the human reads and the appropriate filtering, the relative abundance of *S. aureus* reads (over to total number of bacterial reads) was recalculated. Consequently, the abundance of *S. aureus* was the highest in all samples, compared to that of other classified species (Table 8). However, in samples spiked with lower concentrations (1-5 CFU/mL) there was still detection of a considerable quantity of reads for other identified species (Supplementary Table 3), which led to a reduction in the relative abundance of *S. aureus* over other classified microorganisms reads.

The application of the bioinformatic filtering of host reads followed up by the analysis of the total classified reads on *S. aureus* spiked samples resulted in an overall more sensitive bioinformatics approach to detect microorganisms in blood samples.

	Average total number of *S. aureus* reads	Average total number of other microorganism reads	*S. aureus* average relative abundance (%)
50-100 CFU/mL	5465.4 ± 2353.9	13.6 ± 6.2	99.5 ± 0.2
5-10 CFU/mL	65.3 ± 26.9	4.3 ± 0.8	92.7 ± 1.7
1-5 CFU/mL	22.7 ± 15.7	4.6 ± 2.0	67.1 ± 15.6
NTC	0.3 ± 0.3	1 ± 0.5	33.3 ± 33.3

Table 8. Sequencing metrics on *S. aureus* reads using the *'5mL enrichment protocol'*. Sequencing results metrics obtained at experiments spiking with 50-100, 10-50 and 1-5 CFU/mL *S. aureus* into 5mL of blood and subjecting them to the *'5mL enrichment protocol'*. Total number of *S. aureus* reads = number of reads matching *S. aureus* reference. Total number of other microorganisms reads = reads matching other bacterial species (non-*S. aureus*). *S. aureus* relative abundance (%) = abundance of *S. aureus* reads over the total number of taxonomically classified reads obtained. Data are means ± SD. n= 3 are biological replicates.

After the application of the new filtering, it was still observed a large background of non-spiked microorganisms reads. Hence, next efforts were put on eliminating these background reads by further optimize the bioinformatics filtering. The obtained *S. aureus* sequencing results were reanalysed with the application of a higher *taxa score* filtering. In the previous experiment (Figure 46), metagenomics reads were obtained by applying a *taxa score* of ≥65. This time, these reads were compared to those obtained with the application of *taxa scores* filtering of ≥75, ≥85 and ≥90

to the same results. The higher the score, the more filtering is performed, as reads do not meet the minimum alignment requirements.

The application of more stringent filtering directly impacted the number of bacterial reads obtained. At *taxa score* ≥65, there was a high presence of reads from *S. aureus* and from other microorganisms. Increasing the filter to ≥75 or ≥85 led to a near total absence of non-*S. aureus* reads, while still allowing the detection of *S. aureus* reads at all spiked concentrations. Finally, a *taxa score* of ≥90 caused a high reduction of all reads (including *S. aureus*), to the point that almost no reads were detected in the 5-10 and 1-5 CFU/mL conditions (Figure 47; Table 9).

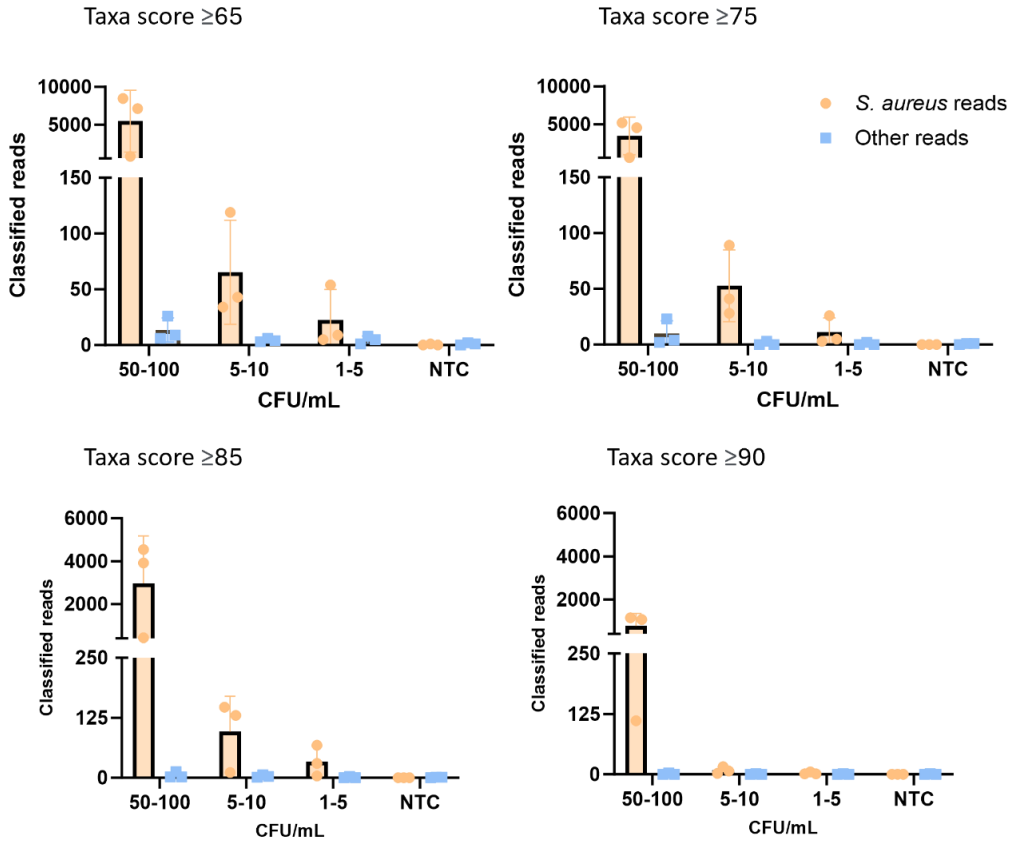


Figure 47. Total number of taxonomically classified reads as *S. aureus* when different *taxa score* were applied. Number of classified reads obtained as *S. aureus* when spiked at different concentrations (50-100, 5-10 or 1-5 CFU/mL). The same metagenomics results were run through the same bioinformatics pipeline, only changing the taxa score filtering applied. Sequencing results were grouped into two different categories: *S. aureus* reads = taxonomically classified reads that matched *S. aureus*; Other reads = reads that were classified as any other microbial species. Data are means ± SD. n= 3 are biological replicates.

The increase of the *taxa score* filtering caused an increment in the relative abundance of *S. aureus* classified reads (over total bacterial population). This trend was especially visible in samples spiked at 1-5 CFU/mL concentrations. As the filtering values became more stringent, the proportion of reads that were classified as non-*S. aureus* species was reduced (Supplementary Tables 4-6) compared to reads classified as *S. aureus,* which also decreased although at a lower level. Due to the decrease of non-*S. aureus* reads and the relative conservation of the total number of *S. aureus* reads, the taxa score filtering of ≥85 led to the highest abundance detection of *S. aureus* at a relative abundance of ≥90% at all spiked concentrations (Table 9). As such, a filtering of ≥85 was selected as the default filtering to apply from this moment.

	Average total number of *S. aureus* reads	Average total number of other microorganism reads	*S. aureus* average relative abundance (%)
Taxa score ≥65			
50-100 CFU/mL	5465.4 ± 2353.9	13.6 ± 6.2	99.5 ± 0.2
5-10 CFU/mL	65.3 ± 26.9	4.3 ± 0.8	92.7 ± 1.7
1-5 CFU/mL	22.7 ± 15.7	4.6 ± 2.0	67.1 ± 15.6
NTC	0.3 ± 0.3	1 ± 0.5	33.3 ± 33.3
Taxa score ≥75			
50-100 CFU/mL	3444 ± 1434.4	9.6 ± 6.7	99.7 ± 0.3
5-10 CFU/mL	52.6 ± 18.5	1 ± 1	98.9 ± 1.1
1-5 CFU/mL	11.3 ± 7.4	0.6 ± 0.6	86.7 ± 13.3
NTC	0 ± 0	0.6 ± 0.3	0 ± 0
Taxa score ≥85			
50-100 CFU/mL	2963 ± 1279.6	5.6 ± 3.6	99.7 ± 0.1
5-10 CFU/mL	37.6 ± 15.8	1 ± 1	98.6 ± 1.4
1-5 CFU/mL	6 ± 3.1	0.3 ± 0.3	97.5 ± 2.5
NTC	0 ± 0	0.6 ± 0.3	0 ± 0
Taxa score ≥90			
50-100 CFU/mL	778 ± 334.4	1 ± 1	99.9 ± 0.08
5-10 CFU/mL	8.3 ± 4.1	0.3 ± 0.3	98.0 ± 1.9
1-5 CFU/mL	2.3 ± 1.3	0.3 ± 0.3	83.4 ± 16.7
NTC	0 ± 0	0 ± 0	0 ± 0

Table 9. Sequencing metrics on *S. aureus* reads using the *'5mL enrichment protocol'* and processed with different *taxa score* filtering. Sequencing results metrics obtained from 5mL blood samples spiked with *S. aureus* at 50-100, 10-50 or 1-5 CFU/mL and subjected to the *'5mL enrichment protocol'*. Metagenomic run was subjected to different taxa score filtering levels. Total number of *S. aureus* reads = number of reads matching *S. aureus* reference. Total number of other microorganisms reads = reads matching other bacterial species (non-*S. aureus*). *S. aureus* relative abundance (%) = abundance of *S. aureus* reads over the total number of taxonomically classified reads obtained. Data are means ± SD. n = 3 are biological replicates.

4.3.4.3 Limit of detection on '5mL enrichment protocol' and '1mL standard protocol' CMg pipeline.

Based on the findings described in previous section, the following modifications were made: (1) species detection was changed from an approach based on overall abundance of reads to a more detailed study of the total number of classified reads, (2) reads matching human genome were eliminated for a detailed analysis based only on the bacterial population present on blood (relative abundances from only microorganism population were recalculated), (3) after removing human reads, the reads were filtered according to their fragment length (≤250bp) and the *taxa score* (≥85), (4) samples would be deemed positive if the total bacterial species reads was ≥5 and their relative abundance ≥90%.

After implementing the Sepsi*PURE* method and establishing the new bioinformatics analysis approach, all the individual steps were combined under the same CMg pipeline (which constitutes the *'5mL enrichment protocol'*) to test again their analytical limit of detection and assess the effects of all the new described improvements. Then the *'5mL enrichment protocol'* was compared to the 1mL protocol optimised and described in chapter 2 – termed the '*1mL standard protocol'*. The comparison was done by performing LoD experiments on blood samples spiked with *E. coli, S. aureus. K. pneumoniae and E. faecalis*.

4.3.4.3.1 *Escherichia coli* Limit of detection experiments.

The first LoD experiments were performed by spiking samples with relevant concentrations of *E. coli* and run through the '*1mL standard protocol'* or the *'5mL enrichment protocol'* pipelines. Taxonomically classified reads matching the reference were detected at all concentrations in both protocols. Using either method, the amount of obtained reads directly correlated to the number of microorganisms spiked in the blood samples (Figure 48).

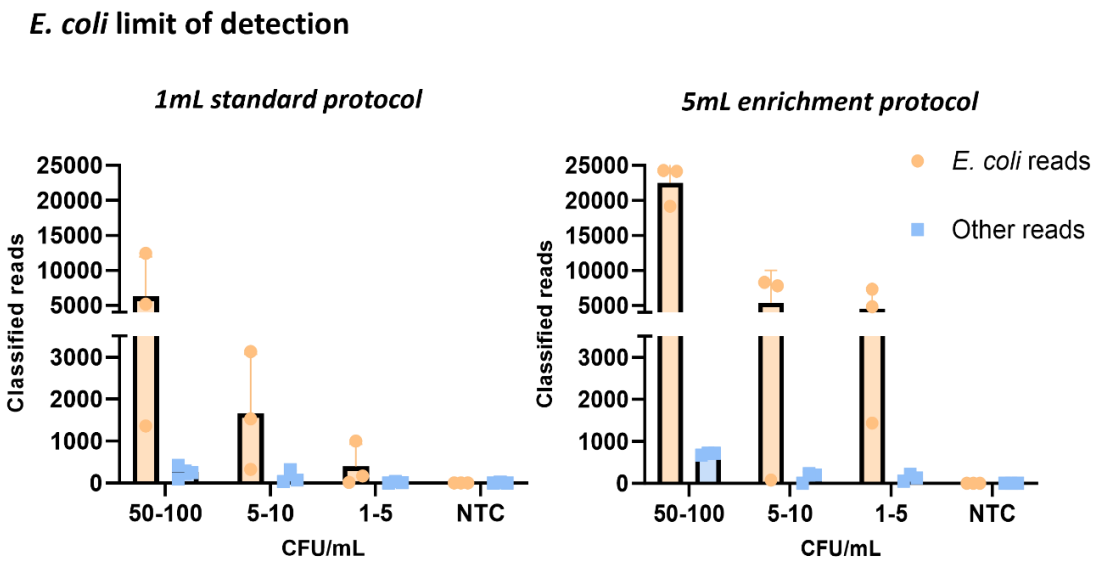


Figure 48. **Taxonomic classification of reads obtained using *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines spiked with *E. coli*.** Number of classified *E. coli* reads obtained at different spiking concentrations (50-100, 5-10 or 1-5 CFU/mL) using the '1mL standard protocol' or the *'5mL enrichment protocol'*. Sequencing results were grouped into two different categories: *E. coli* reads = taxonomically classified reads that matched *E. coli* reference; Other reads = reads classified as any other microbial species. Data are means ± SD. n = 3 biological replicates.

In both protocols and at any concentration, the *E. coli* average abundance was ≥90%, which translates into a 1-5 CFU/mL sensitivity (Table 10). Moreover, none of the NTC samples were positive (which is defined by abundance ≥90% and total number of classified reads ≥5) for any bacterial species (Supplementary Table 8, Samples 1-3). The *'5mL enrichment protocol'* resulted in approximately 5X more *E. coli* taxonomically classified reads than the '*1mL standard protocol'* across all spiked concentrations. This increase in identified reads had a direct impact on the average coverage values, as the highest coverage was continuously achieved when using the 5mL protocol (approximately 2X superior to the results obtained on the 1mL protocol). Nonetheless, the overall genome coverage results were low (<10X) for both protocols. As a result of subjecting the pipelines to different library preparation methods, average bp amplicon size values matching the reference were different, approximately 1.8-2.5Kbp and 1.1Kbp for the '*1mL standard protocol'* and *'5mL enrichment protocol'* pipelines, respectively. For both protocols, out of the total bp per read, between 88% and 92% matched their respective *E. coli* genomic reference (Table 10).

***E. coli* limit of detection**

	Total number of reads matching reference		Bacteria population relative abundance (%)		Average bp length of mapped reads to reference (average read coverage(%))		Average coverage to reference (X)	
	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol
50-100 CFU/mL	6306.7 ± 3238.1	22516.3 ± 1677.5	95.2 ± 0.9	96.9 ± 0.2	1878.8.3 ± 129.4 (91.0 ± 0.6)	1132.3 ± 54.5 (92.1 ± 4.0)	2.4 ± 1.25	4.7 ± 0.5
5-10 CFU/mL	1662.3 ± 811.9	5384.7 ± 2655.6	90.5 ± 4.4	95.9 ± 1.4	2221.4 ± 108.9 (92.2 ± 0.4)	1197.1 ± 26.6 (88.2 ± 0.2)	0.7 ± 0.4	1.1 ± 0.6
1-5 CFU/mL	392.7 ± 306.8	4521.3 ± 1701.9	94.9 ± 1.2	96.9 ± 0.2	2519.2 ± 71.8 (89.1 ± 3.6)	1148.9 ± 12.2 (87.8 ± 0.1)	0.2 ± 0.1	0.9 ± 0.3

Table 10. Sequencing metrics on samples spiked with *E. coli* and subjected to the '5mL enrichment protocol' and '1mL standard protocol' CMg pipelines. Sequencing results obtained in experiments on 1mL and 5mL of blood spiked with 50-100, 10-50 and 1-5 CFU/mL of *E. coli*. Total number of reads matching reference = number of reads matching *E. coli* reference. Bacteria population relative abundance (%) = abundance of *E. coli* reads over the total number of taxonomically classified reads. Average bp length of mapped reads to reference (average read coverage(%)) = average length of reads that matched the reference and percentage of identical nucleotides per read between template and consensus. Average coverage to reference (X) = estimated depth of coverage taking into account the number of reads matching *E. coli* reference and their average length. Data are means ± SD. n = 3 biological replicates.

4.3.4.3.2 *Staphylococcus aureus* limit of detection experiments.

The same analysis was performed on blood samples spiked with *S. aureus*. When using either pipeline, *S. aureus* reads could be detected at all different spiking concentrations and the average number of reads obtained directly correlated to quantity of CFU spiked into blood samples. Moreover, the total number of reads classified as *S. aureus* obtained by either protocol were very similar. However, the use of the '*1mL standard protocol'* led to a substantially higher detection of reads from other bacterial species. No *S. aureus* reads were detected in the NTC samples. (Figure 49).

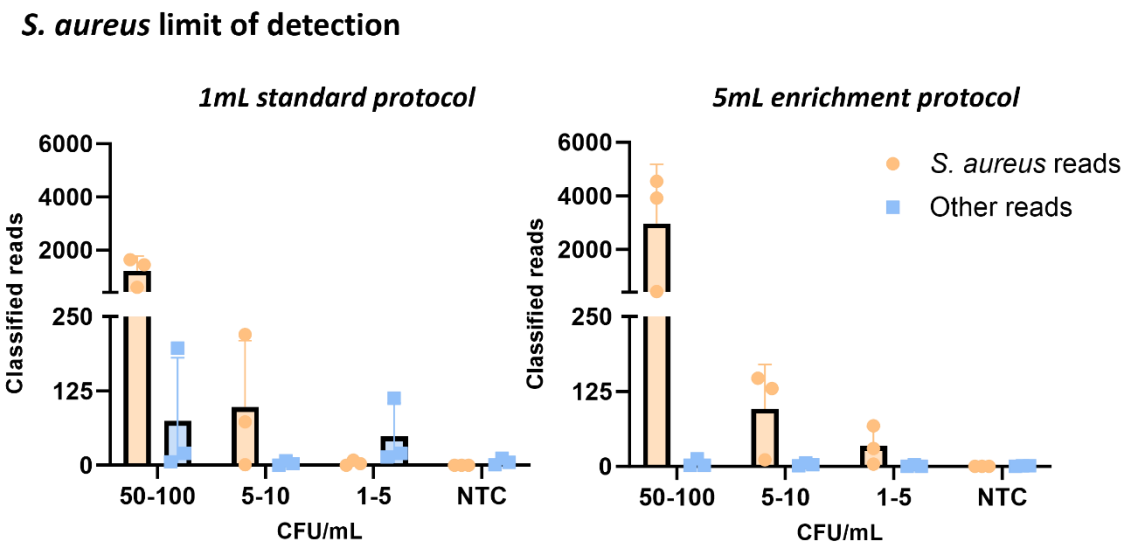


Figure 49. Taxonomic classification of reads obtained using *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *S. aureus*. Number of classified S. aureus reads obtained at different spiking concentrations (50-100, 5-10 or 1-5 CFU/mL) using the *'1mL standard protocol'* or the *'5mL enrichment protocol'*. Sequencing results were grouped into two different categories: *S. aureus* reads = taxonomically classified reads that matched *S. aureus* reference; Other reads = reads classified as any other microbial species. Data are means ± SD. n = 3 biological replicates.

When analysing the results obtained using the '*1mL standard protocol*', the relative abundance of *S. aureus* reads was ≥90% only for samples spiked with 50-100 and 5-10 CFU/mL. In contrast, when using the *'5mL enrichment protocol'* the relative abundance was ≥90% at all spiked concentrations, which indicated a sensitivity of 5-10 CFU/mL for the *'1mL standard protocols'* and of 1-5 CFU/mL for the *' 5mL enriched protocol'* (Table 11). For both CMg pipelines, none of the NTC samples were positive for any bacterial species (Supplementary Table 8, Samples 4-6 (1mL) and 1-3 (5mL)). Overall genome coverage results were relatively low (<1X). Since the number of classified reads was very similar between protocols, their genome coverage was almost identical. Average amplicon sizes varied depending on the protocol (which was higher when using the *'1mL standard protocol'*) and the average read coverage was similar between both pipelines (85%-95%) (Table 11).

***S. aureus* limit of detection**

	Total number of reads matching reference		Bacteria population relative abundance (%)		Average bp length of mapped reads to reference (average read coverage(%))		Average coverage to reference (X)	
	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol
50-100 CFU/mL	1224.7 ± 320.6	2963.0 ± 1279.6	98.9 ± 3.6	99.7 ± 0.1	1947.7 ± 117.3 (90.6 ± 0.4)	1336.5 ± 12.6 (89.0 ± 0.2)	0.8 ± 0.21	0.9±0.4
5-10 CFU/mL	98.0 ± 64.4	96.0 ± 42.8	96.2 ± 3.1	95.1 ± 1.8	2496.2 ± 390.1 (85.7 ± 6.0)	687.2 ± 21.2 (93.6 ± 2.3)	0.06 ± 0.04	0.02±0.01
1-5 CFU/mL	4.0 ± 2.6	34.0 ± 18.6	17.4 ± 3.2	98.6 ± 1.4	1892.7 ± 17.9 (91.6 ± 0.7)	797.1 ± 83.8 (94.7 ± 1.7)	0.002 ± 0.001	0.005±0.002

Table 11. Sequencing metrics on samples spiked with *S. aureus* and subjected to the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. Sequencing results obtained in experiments on 1mL and 5mL of blood spiked with 50-100, 10-50 and 1-5 CFU/mL of *S. aureus*. Total number of reads matching reference = number of reads matching S. aureus reference. Bacteria population relative abundance (%) = abundance of *S. aureus* reads over the total number of taxonomically classified reads. Average bp length of mapped reads to reference (average read coverage (%)) = average length of the reads that matched the reference and percentage of identical nucleotides per read between template and consensus. Average coverage to reference (X) = estimated depth of coverage taking into account the number of reads matching *S. aureus* reference and their average length. Data are means ± SD. n= 3 are biological replicates.

4.3.4.3.3 *Klebsiella pneumoniae* limit of detection experiments.

LoD experiments in 5mL blood samples spiked with *E. coli* or *S. aureus* demonstrated how the application of an extra depletion and enrichment method (Sepsi*PURE*) could result in an exponential increase in sensitivity. In order to further assess the pipeline, this was also tested by spiking other common causative agents of BSI in blood samples and subjecting them to the '*1mL standard protocol'* and *'5mL enrichment protocol'* in parallel. *K. pneumoniae* is one of the top Gram-negative causative species of BSI worldwide and thus it was the next one to be tested.

The application of the *'1mL standard protocol'* on samples spiked with *K. pneumoniae* led to the detection of *K. pneumoniae* reads at all spiking concentrations. However, in samples spiked with 10-50 and 1-5 CFU/mL, the number of reads classified as other bacterial species was highly superior to the number of reads classified as *K. pneumoniae*. Notably, the number of *K. pneumoniae* reads obtained using the *'5mL enrichment protocol'* was substantially higher (>10X) than those classified as "other reads" (Figure 50).

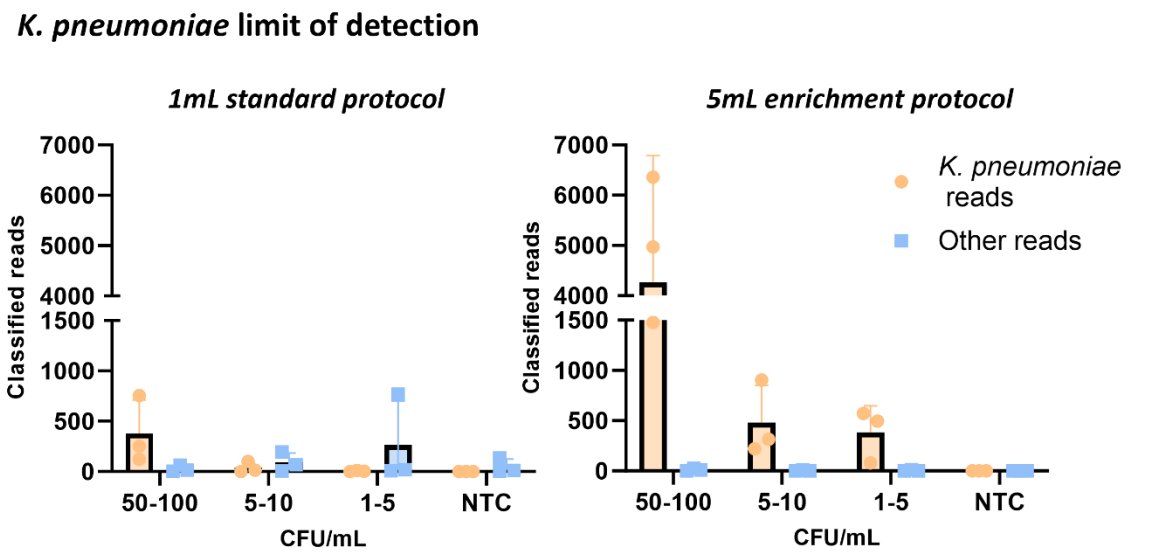


Figure 50. Taxonomic classification of reads obtained using *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *K. pneumoniae*. Number of classified *K. pneumoniae* reads obtained at different spiking concentrations (50-100, 5-10 or 1-5 CFU/mL) using the *'1mL standard protocol'* or the *'5mL enrichment protocol'*. Sequencing results were grouped into two different categories: *K. pneumoniae* reads = taxonomically classified reads that matched *K. pneumoniae* reference; Other reads = reads classified as any other microbial species. Data are means ± SD. n = 3 biological replicates.

For the *'1mL standard protocol'* samples, the abundance of *K. pneumoniae* reads was ≥90% only at the highest spiking concentration. At 10-50 and 1-5 CFU/mL spiking concentrations, *K. pneumoniae* reads were outnumbered by those of other bacterial species (Table 12). For the samples spiked at the lowest concentration (1-5 CFU/mL), the abundance of other bacterial reads was ≥90% (Supplementary Table 7). The abundance of each of these other species was checked individually to detect potential false positives, but none of the species gave a positive result after the application of the taxonomic thresholds (≥90% abundance, ≥5 total number of identified reads). In contrast, the application of the *'5mL enrichment protocol'* led to the detection (with ≥90% abundance) of *K. pneumoniae* at all concentrations, which translates into an analytical sensitivity of 1-5 CFU/mL. In both pipelines, none of the NTC samples were positive for any bacterial species (Supplementary Table 8, Samples 7-9 (1mL) and 4-6 (5mL)). As observed in previous experiments with spiking of other bacterial species, the average genome coverage results were generally low for both CMg pipelines, being especially low for the 1mL samples (0.06-0.0007X). In comparison with the *'1ml standard protocol'*, the *'5mL enrichment protocol'* provided a genome coverage

approximately 100X higher. As seen in the two previous LoD experiments sections, amplicon sizes varied depending on the library protocol that was used and the average read coverage was similar between both pipelines (88%-94%) (Table 12).

***K. pneumoniae* limit of detection**

	Total number of reads matching reference		Bacteria population relative abundance (%)		Average bp length of mapped reads to reference (average read coverage(%))		Average coverage to reference (X)	
	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol
50-100 CFU/mL	377.7 ± 192.4	4268.7 ± 1452.8	94.6 ± 1.6	99.3 ± 0.6	2481.8 ± 587.2 (92.0 ± 1.3)	1379.9 ± 55.6 (89.5 ± 0.4)	0.06 ± 0.02	2.7 ± 1.5
5-10 CFU/mL	39.3 ± 31.1	479.3 ± 213.7	21.4 ± 6.7	98.9 ± 1.1	2219.5 ± 263.7 (92.5 ± 1.6)	1262.6 ± 50.3 (88.7±0.2)	0.003 ± 0.002	0.1 ± 0.05
1-5 CFU/mL	4.7 ± 3.3	382.3 ± 152.7	5.5 ± 4.1	95.3 ± 4.6	2345.9 ± 31.8 (93.9 ± 0.1)	1216.5 ± 49.6 (88.5 ± 0.4)	0.0007 ± 0.0	0.07 ± 0.03

Table 12. Sequencing metrics on samples spiked with *K. pneumoniae* and subjected to the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. Sequencing results obtained in experiments on 1mL and 5mL of blood spiked with 50-100, 10-50 and 1-5 CFU/mL of *K. pneumoniae*. Total number of reads matching reference = number of reads matching *K. pneumoniae* reference. Bacteria population relative abundance (%) = abundance of *K. pneumoniae* reads over the total number of taxonomically classified reads. Average bp length of mapped reads to reference (average read coverage (%)) = average length of the reads that matched the reference and percentage of identical nucleotides per read between template and consensus. Average coverage to reference (X) = estimated depth of coverage taking into account the number of reads matching *K. pneumoniae* reference and their average length. Data are means ± SD. n= 3 are biological replicates.

4.3.4.3.4 *Enterococcus faecalis* limit of detection experiments.

Finally, the same LoD experiments were performed spiking blood samples with clinically relevant concentrations of *E. faecalis* and subsequently running the 1mL and 5mL protocols. The application of the *'1mL standard protocol'* on resulted in the identification of *E. faecalis* reads only in samples with the highest concentration of spiked bacteria (50-100 and 10-50 CFU/mL). Conversely, other bacterial species reads were identified in samples of all groups subjected to this protocol (except for NTCs), with especially higher presence in the 5-10 and 1-5 CFU/mL spiked samples. Conversely, the application of the *'5mL enrichment protocol'* resulted in the detection of

E. faecalis reads at all concentrations and these reads found at a significant higher number than other bacterial species reads (Figure 51).

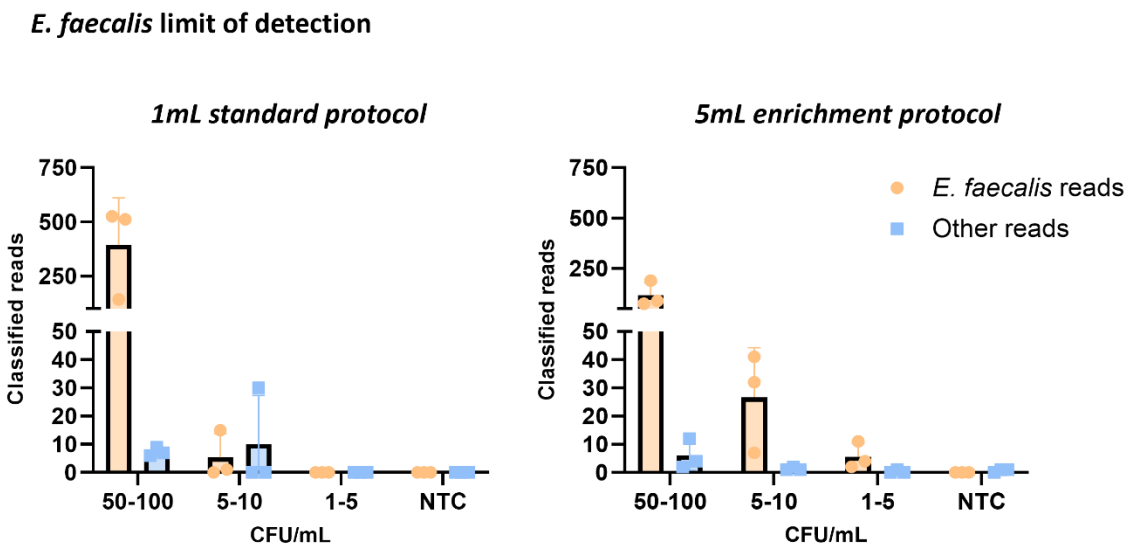


Figure 51. Taxonomic classification of reads obtained using *'5mL enrichment protocol' and '1mL standard protocol'* CMg pipelines spiked with *E. faecalis*. Number of classified *E. faecalis* reads obtained at different spiking concentrations (50-100, 5-10 or 1-5 CFU/mL) using the '1mL standard protocol' or the *'5mL enrichment protocol'*. Sequencing results were grouped into two different categories: *E. faecalis* reads = taxonomically classified reads that matched *E. faecalis* reference; Other reads = reads classified as any other microbial species. Data are means ± SD. n = 3 biological replicates.

On the metagenomics results obtained on samples spiked with *E. faecalis* and subjected to the '*1mL standard protocol',* the abundance of this bacterial species was only ≥90% at the highest spiked concentration, which gives a limit of detection of 50-100 CFU/mL. For the samples that were spiked at the lowest concentration (1-5 CFU/mL), the abundance of the aggregate of other bacterial reads was ≥90%. Thus, the abundance of each of the classified species individually was checked to detect potential false positive results. After the application of the taxonomic thresholds, none of the samples were considered positive for any bacterial species (Supplementary table 7). Instead, when using the *'5mL enrichment protocol',* the average proportion of bacterial reads was ≥90% at all spiked concentrations, indicating a limit of detection of 1-5 CFU/mL when using this pipeline (Table 13). As observed in all previous LoD experiments, in neither of the two pipelines the NTC samples were positive for any bacterial species (Supplementary Table 8, Samples 10-12 (1mL) and 4-6 (5mL)). The average genomic coverage results, which was similar between the two pipelines,

was the lowest obtained from all the different LoD experiments spiked with different species, ranging from 0.2-0.002X. The amplicon average size varied depending on the library protocol used and the average read coverage was similar between both pipelines (90%-97%) (Table 13).

***E. faecalis* limit of detection**

	Total number of reads matching reference		Bacteria population relative abundance (%)		Average bp length of mapped reads to reference (average read coverage(%))		Average coverage to reference (X)	
	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol	1mL protocol	5mL protocol
50-100 CFU/mL	393.7 ± 125.4	268.0 ± 29.3	97.2 ± 0.8	97.9 ± 1.0	2350.3 ± 138.9 (91.3 ± 0.6)	1161.3 ± 66.7 (92.1 ± 4.0)	0.26 ± 0.07	0.04 ± 0.02
5-10 CFU/mL	5.4 ± 4.8	78.0 ± 38.0	44.5 ± 29.4	98.2 ± 1.4	2474.4 ± 181.7 (90.6 ± 0.6)	1091.0 ± 109.9 (97.6 ± 0.3)	0.006 ± 0.005	0.008 ± 0.004
1-5 CFU/mL	0.0 ± 0	7 ± 1.7	0.0 ± 0	100 ± 0	-	1107.7 ± 377.4 (96.3 ± 1.1)	-	0.002 ± 0.001

Table 13. Sequencing metrics on samples spiked with *E. faecalis* and subjected to the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. Sequencing results obtained in experiments on 1mL and 5mL of blood spiked with 50-100, 10-50 and 1-5 CFU/mL of *E. faecalis.* Total number of reads matching reference = number of reads matching *E. faecalis* reference. Bacteria population relative abundance (%) = abundance of *E. faecalis* reads over the total number of taxonomically classified reads. Average bp length of mapped reads to reference (average read coverage (%)) = average length of the reads that matched the reference and percentage of identical nucleotides per read between template and consensus. Average coverage to reference (X) = estimated depth of coverage taking into account the number of reads matching *E. faecalis* reference and their average length. Data are means ± SD. n= 3 are biological replicates.

4.3.4.4 Antimicrobial determinants detection: 1mL vs 5mL pipelines.

After demonstrating the successful detection of the 4 bacterial species spiked at clinically relevant concentrations, the suitability of both pipelines for the detection of AMR determinants was assessed. The AMR detection capacity of the pipelines was tested by spiking different concentrations of β-lactamase resistant *E. coli* CXTM-15 (containing the extended-resistance Extended-Spectrum β-Lactamase (ESBLs) pEK499 plasmid) on whole blood samples, which were then subsequently run through their respective pipelines and subjected to sequencing for 3 hours.

When using the *'1mL standard protocol'* pipeline, 6 out of the 10 AMR genes that are present in the pEK499 plasmid could be detected in samples spiked at the highest concentration (100 CFU/mL). 2 out of 10 were detected at the 10 CFU/mL concentration, and none at the lowest

spiked concentrations. When these AMR genes were detected, the average coverage was low (<10X) but the high template identity (≥98%,) and high coverage values (between 100-102%) confirmed the presence of all of them (Table 14). When assessing the *'5mL enrichment protocol'* samples, all resistance genes could be detected at any input concentration, except for *tet(A)*, *catB4* and *blaTEM-1* at the lowest spiking concentration. Most of the detected genes were classified as genes present in the pEK499 plasmid. However, on two occasions, genes were misidentified: *aac(6')-Ib-W104R, aac(6')-Ib-D181Y* and *aac(6')-Ib-AKT* were falsely detected instead of the expected *aac6'-Ib-cr* and *catB3* instead of the expected *catB4*. The obtained average coverage varied among the different genes and spiked concentrations. Average coverage obtained from samples spiked with low amounts of bacteria (1-10 CFU/mL) was generally low (<10X). However, the high average template identity (≥99%) and coverage (between 99-101%) values confirmed the presence of all these genes when using the *'5mL enrichment protocol'.*

Because specific gene detection for some of the genes was challenging due to the low coverage results (<10X), it was hypothesized that AMR detection would be enhanced by increasing the sequencing time. However, this was only possible on the *'5mL enrichment protocol'* since, as previously seen in section 4.2.2, the use of a PCR-based library preparation method caused a significant reduction of active pores after a few hours of sequencing. As the pore blocking issue was minimal when using a PCR-free library, it was decided to test this on 5mL blood samples using the *'5mL enrichment protocol'.*

pEK499 expected genes	Average Coverage (X)						Average template identity (%) / Average Template Coverage (%)					
	1mL protocol			5mL protocol			1mL protocol			5mL protocol		
	50-100 CFU/mL	5-10 CFU/mL	1-5 CFU/mL	50-100 CFU/mL	5-10 CFU/mL	1-5 CFU/mL	50-100 CFU/mL	5-10 CFU/mL	1-5 CFU/mL	50-100 CFU/mL	5-10 CFU/mL	1-5 CFU/mL
blaCTX-M-15	4.1 ± 0.7	1.0 ± 1.0	0	48.9 ± 10.1	11.3 ± 1.8	17.3 ± 7.5	100.0±0 / 100.0±0	100.0±0 / 100.3 ± 0	-	99.0±0.0 / 100.0±0	100.0±0 / 100.0±0	99.0±0.1 / 100.8±0.8
blaOXA-1	0	0	0	6.4 ± 1.4	2.0 ± 1.0	5.5 ± 2.8	-	-	-	100.0±0 / 100.0±0	99.9±0.1 / 100.0±0	100.0±0 / 100.8±0.8
blaTEM-1	1.0 ± 0.1	0	0	2.7 ± 0.1	0.7 ± 0.7	0	98.8±0.3 / 100.5±0.1	-	-	99.7±0 / 101.7±1.2	99.4±0.0 / 101.4±1.4	-
aac6'-lb-cr	0	0	0	3.9 ± 2.9	1.7 ± 0.9	1.9 ± 1.1	-	-	-	99.8±0.0 / 100.3±0.3	98.9±1.1 / 100.3±0.3	99.7±0.1 / 100.5±0.5
mph(A)	5.6 ± 2.6	1.0± 0.6	0	39.8 ± 6.5	6.9 ± 1.5	13.0 ± 7.7	99.4±0.6 / 100.0±0	100.0±0 / 100.3±0	-	100.0±0 / 100.0±0	100.0±0 / 100.0±0	100.0±0 / 100.0±0
catB4	0		0	7.3 ± 1.7	1.4 ± 0.2	0.7 ± 0.7	-	-	-	92.5±1.2 / 102.1±0.7	86.9±1.0 / 101.7±1.0	94.5±0 / 99.8±0
tet(A)	2.3 ± 2.3	0	0	0.3 ± 0.3	0	0.3 ± 0.3	92.5±0 / 101.7±0	-	-	98.3±0.0 / 100.2±0	-	73.2±0 / 77.4±0
dfrA7	4.5± 1.5	0	0	16.5 ± 2.6	5.7 ± 0.3	3.5 ± 1.9	99.1±0.8 / 100.1±0.1	-	-	99.8±0.0 / 100.0±0	99.8±0.0 / 100.0±0	99.3±0.5 / 100.1±0.1
sulI	3.6 ± 2.0	0	0	18.7 ± 5.8	3.6 ± 1.6	3.6 ± 2.0	99.8±0.2 / 100.1±0.1	-	-	100.0±0 / 100.1±0	98.9±1.2 / 100.0±0	99.9±0.1 / 100.1±0.1
aadA5	4.9± 0.1	0	0	19.6 ± 3.8	9.1 ± 0.5	4.4 ± 2.6	99.0±0.1 / 100.0±0	-	-	99.9±0.1 / 100.0±0	100.±0.0 / 100.0±0	100.±0 / 100.0±0

Table 14. AMR gene detection metrics after 3 hours of sequencing using the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. 1mL or 5mL of blood were spiked with *E. coli* CTX-M-15 (peK499 plasmid) at different concentrations (50-100, 10-50 and 1-5 CFU/mL) and subjected to *'1mL standard protocol'* and *'5mL enrichment protocol'*. AMR determinants were detected and samples were compared by assessing their coverage and template identity to the reference. Average Coverage (X) = average depth of coverage of the template. Average template identity (%) = percentage of identical nucleotides between template and consensus. Average template coverage (%) = percentage of bases in the template that is covered by consensus sequence. Data are means ± SD. n= 3 are biological replicates.

Different clinically relevant concentrations of the *E. coli* CXTM-15 strain were spiked into 5mL blood samples, which were then subjected to the *'5mL enrichment protocol'* and sequenced for 3 or 24 hours. The obtained results were compared to check the effects of extended sequencing times. The 24-hour sequencing noticeably increased gene coverage across all targets (Table 15). The increment on coverage enabled successful detection of the *blaTEM-1* gene across all spiked concentrations, which had remained elusive during the previous 3-hour sequencing. Conversely, as per the 3-hour sequencing results, the *catB3* and *tet(A)* genes remained undetected. The extended sequencing time led to a decrease in misidentification of genes from the pEK499 plasmid: only

aac(6')-Ib-W104R was identified in one replicate instead of the expected *aac6'-Ib-cr*. Similarly, *catB4* was consistently detected across all replicates, as opposed to the expected *catB3* gene.

pEK499 expected genes	Average Coverage (X)			Average template identity (%) / Average Template Coverage (%)		
	5mL protocol					
	50-100 CFU/mL	5-10 CFU/mL	1-5 CFU/mL	50-100 CFU/mL	5-10 CFU/mL	1-5 CFU/mL
blaCTX-M-15	172 ± 45.3	36.5 ± 3.1	46.8 ± 21.7	99.9 ± 0 / 100.0 ± 0	99.9 ± 0 / 100.0 ± 0	99.9 ± 0 / 100.0 ± 0
blaOXA-1	25.0 ± 6.7	7.6 ± 3.2	11.4 ± 6.4	100.0 ± 0 / 100.0 ± 0	100.0 ± 0 / 100.0 ± 0	100.0 ± 0 / 100.0 ± 0
blaTEM-1	11.1 ± 1.3	4.4 ± 1.4	1.4 ± 0.7	99.9 ± 0 / 100.0 ± 0	99.8 ± 0.2 / 101.4 ± 1.4	96.5 ± 3.4 / 101.5 ± 1.2
aac6'-Ib-cr	13.0 ± 1.7	3.7 ± 0.6	4.1 ± 2.5	99.8 ± 0 / 100.0 ± 0	99.8 ± 0.1 / 100.6 ± 0.5	99.7 ± 0.1 / 100.5 ± 0.5
mph(A)	121.6 ± 28.6	23.9 ± 2.8	37.2 ± 15.9	100.0 ± 0 / 100.0 ± 0	99.9 ± 0.1 / 100.0 ± 0	100.0 ± 0 / 100.0 ± 0
catB4	15.4 ± 4.5	2.7 ± 0.5	3.7 ± 2.1	91.9 ± 0.9 / 102.4 ± 0.7	91.4 ± 1.3 / 101.7 ± 1.0	94.5 ± 0 / 99.8 ± 0
tet(A)	3.2 ± 0.9	0	1.0 ± 0.6	92.8 ± 7.0 / 96.8 ± 3.2	-	73.2 ± 0 / 77.4 ± 0
dfrA7	16.5 ± 2.6	5.7 ± 0.3	3.5 ± 1.9	99.8 ± 0 / 100.0 ± 0	99.8 ± 0 / 100.0 ± 0	99.3 ± 0.5 / 100.1 ± 0.1
sulI	53.3 ± 11.2	11.6 ± 3.7	10.5 ± 5.2	100.0 ± 0 / 100.0 ± 0	100.0 ± 0 / 100.0 ± 0	99.9 ± 0.1 / 100.1 ± 0.1
aadA5	66.4 ± 17.9	19.3 ± 0.8	14.4 ± 6.3	99.9 ± 0 / 100.0 ± 0	100.0 ± 0 / 100.0 ± 0	100.0 ± 0 / 100.0 ± 0

Table 15. AMR gene detection metrics after 24 hours of sequencing using the *'5mL enrichment protocol'* and *'1mL standard protocol'* CMg pipelines. 5mL of blood were spiked with *E. coli* CTX-M-15 (peK499 plasmid) at different concentrations (50-100, 10-50 and 1-5 CFU/mL) and subjected *'5mL enrichment protocol'* with an extended sequencing time (24 hours). Average Coverage (X) = average depth of coverage of the template. Average template identity (%) = percentage of identical nucleotides between template and consensus. Average template coverage (%) = percentage of bases in the template that is covered by consensus sequence. Data are means ± SD. n = 3 are biological replicates.

4.4 – Discussion.

During the initial stages of this project, our efforts to develop a BSI CMg diagnostics tool with high sensitivity led to generation of a pipeline that effectively identified microorganisms at clinically relevant concentrations (chapter 2). However, when applied to clinical blood samples which had been confirmed positive for infection, the method exhibited a lower sensitivity compared to the standard BSI detection methods (chapter 3). Consequently, it was determined the need for further method optimization to enhance its readiness for future clinical application. The goal was to improve both the sensitivity and the specificity of the CMg pipeline, focusing on the refinement of the host depletion and bioinformatics analysis steps. In this chapter, it has been described how the CMg pipeline was further optimised by adding additional steps to the host depletion protocol and improving the bioinformatics pipeline. The key findings were:

- Sepsi*PURE* method can be used in combination with the saponin-based method to both enrich bacterial presence and deplete host DNA. This combination allowed the processing of higher volumes of blood (5mL) compared to the previous saponin-only method (1mL).
- Changing to a PCR-free library preparation protocol caused a great improvement in pore activity sustainment on Nanopore flow cells.
- A detection threshold based on abundance over the total number of reads based resulted in poor sensitivity in experiments using spiking with *S. aureus*. Switching to an abundance over bacterial population drastically increased the sensitivity.
- Two CMg pipelnies were developed and compared: one that can process 1mL of blood (*'1mL standard protocol'*) and another one with an extra enrichment step that can process 5mL of blood (*'5mL enrichment protocol'*).
- Four of the top bacterial BSI causative species were detected when spiked in 5mL of blood and using the optimised CMg pipeline.

- AMR determinants could be detected when using both pipelines, and their detection benefited from longer incubation times.

4.4.1 Bacterial enrichment significantly improved host depletion efficiency.

Host DNA depletion method had already been optimised in early stages of the project by increasing the saponin concentration to 3%. This increase led to a decrease in both chromosomal and mitochondrial DNA in the samples without affecting the bacterial yield. However, this proved insufficient when applied to clinical samples and therefore, to further optimize the depletion of mitochondrial DNA, a collaboration with Momentum Ltd (in Oxford) was established to implement their Sepsi*PURE* [370] method within our CMg pipeline. To date, this was the first time that a group implemented this method onto a metagenomics pipeline.

Initial experimentation demonstrated how, on its own, the Sepsi*PURE* method effectively depleted both host chromosomal and mitochondrial DNA. It was hypothesized that the beads selectively bound to microorganisms, allowing for their capture through magnetization. This process resulted in the removal of a substantial portion of host blood cells from the sample. However, when both methodologies were combined, a non-significant reduction in host DNA compared to when either method was used individually was obtained. A closer examination of the TRM growth medium composition revealed the presence of SPS. As observed previously (described in section 2.3.1.3), SPS is an inhibitor of enzymatic reactions [277] and it notably inhibits the saponin-based host depletion reaction. Consequently, a decision was made to replace the TRM medium with LB, a generic and rich general-purpose medium capable of supporting the growth of a wide variety of bacterial species [379]. The change in medium resulted in a moderate decrease in both host chromosomal and mitochondrial DNA when both methodologies were combined. The decrease on host DNA was likely due to the medium change not interfering with the saponin-based method, as the SPS-containing medium did.

Because of the moderate success of the combination of both methods, the host cells' interaction with the beads during the Sepsi*PURE* reaction was investigated to attempt further optimisation. It was hypothesised that a proportion of these host cells could also be binding to the magnetic beads, alongside the bacteria, and subsequently be carried downstream in the process. Following the application of the Sepsi*PURE* method, a significant amount of host DNA was found to be associated with the bead fraction, while notably lower quantities of host DNA were detected in the medium fraction. The working hypothesis was that the same physical and chemical forces responsible for bacteria binding to the beads were also leading to a substantial portion of host cells attaching. At the same time, both the beads and medium fractions contained *E. coli*. It was theorized that during the 4-hour incubation step, the microorganisms attached to the beads would gradually detach and transfer to the medium fraction, hence their high concentration in the medium. Due to the finding that the magnetic beads also retained host cells and that the medium fraction was sufficient to enrich the bacterial yield, both depletion methods were combined by initially applying the Sepsi*PURE* method and then subjecting the incubated medium fraction to the saponin-based host depletion method. This approach resulted in an approximate 10^1-fold improvement in the removal of mitochondrial DNA compared to applying the saponin-based method on its own. Simultaneously, this procedure substantially increased the presence of spiked bacteria by 10^2-fold. In summary, the integration of the Sepsi*PURE* method into our pipeline allowed us to process larger blood volumes (up to 5mL) while enhancing the depletion of host DNA. Consequently, this led to a noteworthy increase in the CMg pipeline's sensitivity, culminating in the development of a new Clinical Metagenomics protocol termed the *'5mL enrichment protocol'*.

4.4.2 Using a PCR-free library preparation method helped maintaining flow cell pore activity constant.

The implementation of the Sepsi*PURE* method in the CMg pipeline resulted in a remarkable approximate 100X increase in the abundance of spiked *E. coli*. Consequently, after DNA extraction

and WGA, significantly higher DNA yields were achieved. These elevated DNA yields allowed us to adopt the Rapid Barcoding (SQK-RBK004) library preparation method [380] (compared to the PCR-based method used until this point). This library preparation method was preferred due to being known as one of the quickest and simplest hands-on protocols. Additionally, this PCR-free library preparation method offered the advantage of an amplification-free approach, potentially eliminating additional sources of bias [381]. In the context of our CMg pipeline, the introduction of this library preparation kit not only made the library preparation step faster but also had a key impact on the activity of nanopore sequencing flow cells.

Sequencing runs involving previous samples that had been prepared with the PCR-based library method displayed a significant decrease in pore activity within a few hours. Conversely, the application of a PCR-free library preparation protocol resulted in minimal reductions in pore activity. A similar reduction in pore activity has been described by other authors such as Kovaka *et al.* (2021) [382] or Laszlo *et al.* (2016) [292] suggesting that nanopore pores temporarily become unavailable while sequencing, which, in turn, permitted longer sequencing incubation periods. In chapter 2, subjecting the samples to WGA potentially generated a substantial number of DNA secondary and tertiary structures that might interfere with the sequencing activity of nanopore pores. In both the PCR and non-PCR library preparation methods employed in this project, samples underwent MuA-mediated transpose-based fragmentation steps [380,383]. During these transposition steps, the enzyme randomly cleaved the template DNA and then ligated DNA fragments, which served as barcodes for the PCR-free approach or as adapters used as primers in the PCR library approach [384]. Theoretically, during the transpose-based step, a significant proportion of the structures formed during WGA [152] would be fragmented into smaller amplicons, digesting the structures formed during WGA. Consequently, these samples should not lead to a reduction in pore activity. The digestion step could potentially explain why, after subjecting the samples to a Rapid Barcoding kit (which includes a transpose-based step), there was minimal loss of pore activity. It was hypothesized that the subsequent PCR-based library step following

fragmentation might be able to reassemble some of these structures, which could then obstruct the nanopore sequencing pores, as observed in this project. To date, no definitive explanation for this issue has been identified, and this challenge has not been reported by colleagues in the field.

4.4.3 Changing to a total number of reads-over-abundance based criteria increased the sensitivity and specificity of the pipeline.

The impact of using the Sepsi*PURE* method along with saponin-based depletion, as well as transitioning from a PCR-based to a PCR-free library preparation approach, was ultimately assessed by LoD experiments. These experiments were key to test the sensitivity of the described CMg pipeline. When of *E. coli* was spiked and processed using the *'5mL enrichment protocol'*, *E. coli* was detected at all concentrations. Remarkably, the same level of sensitivity when using the '*1mL standard protocol'* was achieved. However, with the *'5mL enrichment protocol'*, a significantly higher abundance of *E. coli* reads was observed. Conversely, when samples were spiked with *S. aureus*, the *'5mL enrichment protocol'* was only able to detect the spiked bacteria for samples with 50-100 CFU/mL, representing a 10X reduction in sensitivity compared to the '*1mL standard protocol'.* It was hypothesized that this happened as a result of increasing the volume of processed blood, which could be leading to a substantial increase in host DNA cells, resulting in a lower relative abundance of bacterial reads. Furthermore, the high abundance of *E. coli* reads obtained could be attributed to the faster growth rate of *E. coli* in LB medium [385], which is superior to the growth rate of *S. aureus* on this culture [386]. Given the strong evidence obtained that the '*5mL enrichment protocol'* effectively removed host DNA, it was decided to adjust the bioinformatic approach. The new approach shifted from applying thresholds based on relative abundance over the total number of reads to an analysis based on the abundance of reads of each species within the total number of obtained bacterial reads. To focus on the bacterial population, it was decided to introduce a step at the beginning of the bioinformatics pipeline to remove the remaining host DNA reads from the sample's data. This method has also been employed by other researchers, particularly when host

read abundance surpasses that of bacterial reads, as seen in the work published by Grumaz *et. al.* in 2016 [158]. To biomimetically remove host DNA, all reads were mapped against a human reference genome, classifying any reads displaying ≥80% identity with the reference as human and subsequently eliminating them. To do this, minimap2 was used, a precise and rapid alignment tool that has been commonly used for long-read analysis by other authors [175,387].

The elimination of host DNA reads (which usually accounted for ≥90% of the sequenced reads, even after host DNA depletion) resulted in processed datasets containing only bacterial reads. Subsequently, bacterial reads were selected or discarded based on their identification and coverage values in relation to the reference. As described by Kunin *et al.* (2018) in its *Bioinformatician's guide to metagenomics* [388], given that metagenomics samples often contain numerous sequences from sources other than the samples themselves, it is crucial to apply some level of filtering to remove such reads before any further data processing. Similar to the approaches used by other research groups such as Schmider *et al.* (2011) [389] and Ivy *et al.* (2018) [390], contamination read filtering was successfully implemented. This involved assigning a score to each read based on their alignment metrics, encompassing query coverage and identity, which we named *taxa score*. Reads that did not achieve a high alignment score were deemed unsuitable for subsequent taxonomic classification. This score was tested in LoD experiments on blood samples spiked with *S. aureus* and subjected to the *'5mL enrichment protocol'*. Initially, low filtering score values (≥65) had led to the identification of a large presence of non-*S. aureus* reads. This filtering significantly affected the abundance results for samples with low spiking concentrations (≤70% on 1-5 CFU/mL). Increasing the *taxa score* filtering to (≥85) resulted in a substantial reduction in non-spiked/contaminant reads, while the impact on *S. aureus* reads was minimal. Consequently, higher relative abundance results were achieved for *S. aureus* reads. Upon analysing the abundance and total number of reads obtained using various taxa score thresholds, it was determined that: (1) a *taxa* score of ≥85 would be consistently applied, (2) samples would be considered positive if their relative abundance (over total bacterial population identified was ≥90% and (3) the total number

of reads taxonomically identified reads of one species was at least ≥5. Conceptually, the applied *taxa score* (≥85) was equivalent to only accepting reads that had a 98-100% match with the references. This was because average read coverage was calculated considering both upstream and downstream flanking regions and barcodes introduced during nanopore library preparation. These flanking regions and barcodes added approximately 100-150bp to each read, accounting for roughly 6% to 12% of the bases in each read (which averaged between 1000 and 2000 bases) [391]. Therefore, the maximum average read coverage values matching the reference ranged from 88% to 94%. By implementing these stringent thresholds, the risk of obtaining false positive results, as described by other authors [389,390], was minimized

4.4.4 All spiked species were detected when present at clinically relevant concentrations.

All the wet-lab and bioinformatic methods described in chapter 2 and in this chapter were integrated into the same CMg pipeline to evaluate the analytical sensitivity and specificity. 5mL blood samples were spiked with different bacterial species and concentrations and underwent the *'5mL enrichment protocol'* which included an initial enrichment step by Sepsi*PURE*. However, an enrichment step could compromise the ability of the pipeline to detect BSI causative species that are fastidious or result in the overrepresentation of rapidly growing species [392]. This would defeat the purpose of using metagenomics as an unbiased approach to the detection of organisms present in blood. As such, the CMg pipeline was also compared with the *'1mL standard protocol'* detailed in chapter 2, a pipeline that can directly work on 1mL blood samples and does not require an enrichment phase, avoiding the mentioned overrepresentation of rapid-growing species and the under-detection of fastidious species. The two pipelines were tested on mock blood samples spiked with 4 of the main causative species of BSI *E. coli*, *S. aureus*, *K. pneumoniae* and *E. faecalis* [36]. This process took 9 hours for the 1ml pipeline and 12 hours for the 5mL pipeline, from sample collection to the obtention of results (Figure 52).

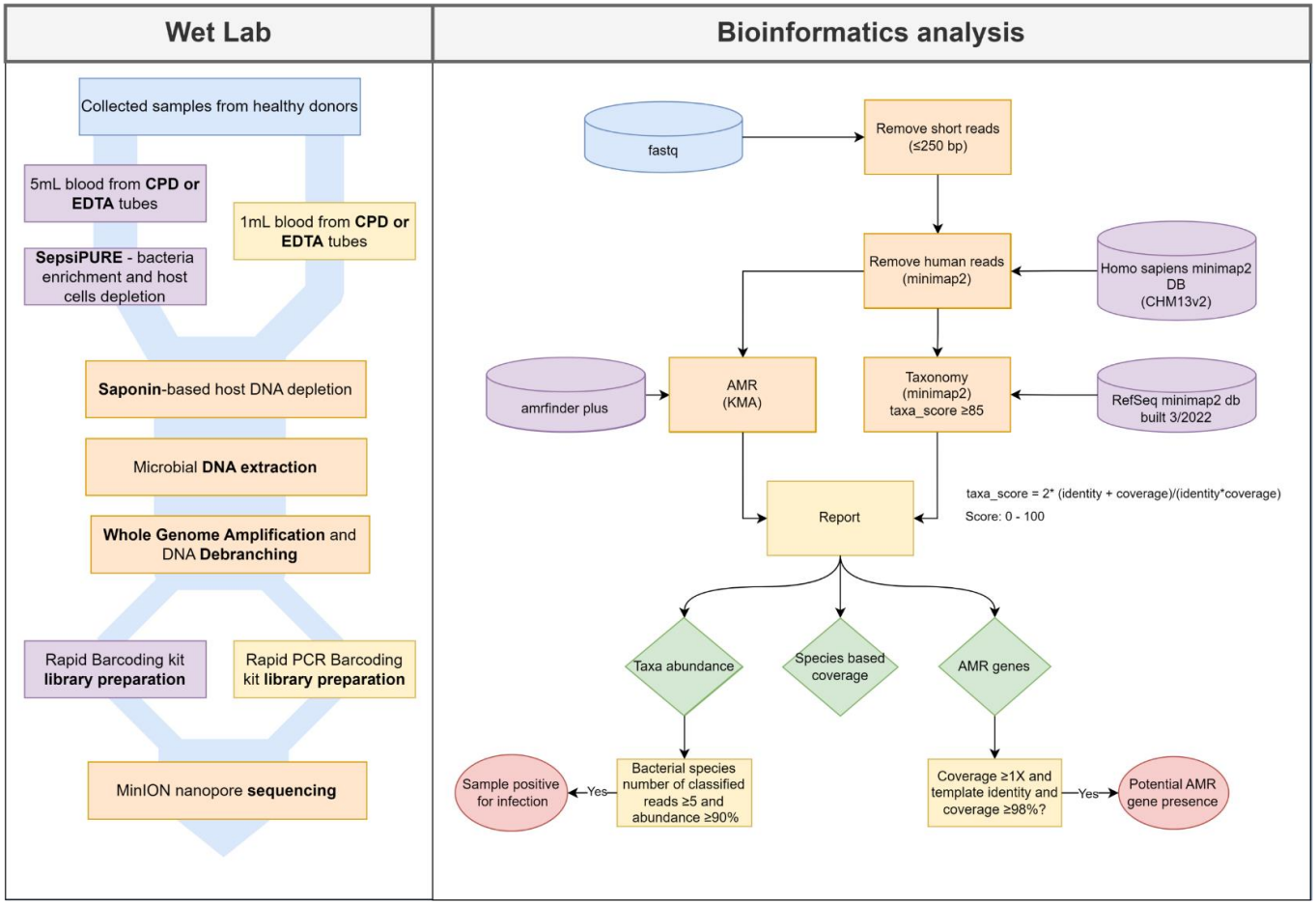


Figure 52. Schematic representation of steps involved in 1mL and 5mL CMg pipelines. Left panel: In triplicates, 50-100 CFU/mL, 5-10 CFU/mL, 1-5 CFU/mL (and NTC) of *E. coli* (CTX-M-15), *S. aureus*, *K. pneumoniae* or *E. faecalis* cultures were spiked into 1mL (Standard protocol) or 5mL (Enrichment protoco*l*) of whole blood samples in EDTA or CPD-coated tubes. 5mL samples underwent Sepsi*PURE* extraction and both 5mL and 1mL samples went through saponin-based host DNA depletion and subsequent DNA extraction. DNA extracts were whole genome amplified and debranched/digested. Samples were subjected to either Rapid PCR Barcoding (SQK-RPB004) or Rapid barcoding kit (SQK-RBK004) library preparation and nanopore sequenced (3 hr) with Oxford Nanopore MinION. Right panel: Short (<250bp) fastq were removed and then mapped against the human chromosome to eliminate remaining host DNA reads. Remaining bacterial reads were taxonomically classified using minimap2 and filtered according to their *taxa score* and AMR determinants identified using KMA. Reports for taxonomic abundance, species coverage mapped metrics and AMR determinants detection data were produced. Samples were deemed positive for infection if bacterial species number of reads was ≤5 and relative abundance over the total number of reads ≤90%. For AMR gene detection, gene specific coverage had to be ≤5 and template identity and coverage ≤99%.

Importantly, the proof-of-concept presented method detected as little as 1-5 CFU/mL for *E. coli*, 5-10 CFU/mL for S*. aureus* and 50-100 CFU/mL for *K. pneumoniae* and *E. faecalis* with the *'1mL standard protocol*' pipeline. When using the *'5mL enrichment protocol',* the sensitivity was of 1-5 CFU/mL for all species. The extra depletion and enrichment step in this protocol resulted in a significant increase in the method's sensitivity, detecting all the spiked species at the lowest concentration. For the *'1mL standard protocol'*, the application of the saponin-based method as the only host depletion protocol was not sufficient to deplete all the necessary host DNA (mainly mitochondrial) to detect spiked microorganisms reads for certain bacteria species when these were present at low concentrations. It was hypothesised that the LoD values obtained for different spiked species when using the *'1mL standard protocol'* could be due to the species bias introduced at several steps: (1) the saponin-based method, which has been reported to potentially lyse certain bacterial species leading to a loss of their DNA [141]; (2) the DNA extraction, which has shown differences in effectivity between bacterial species, as some are known for being hard-to-lyse [106,393]; (3) the WGA method or the Rapid PCR barcoding library kit, which use polymerases (phi29 or LongAmp^{TM}) that are known for having more affinity for amplifying certain bacterial species [394,395]. This could be a potential limitation of this method. Despite these problems would apply to both 1mL and 5mL protocols, the superior sensitivity of the *'5mL enrichment protocol'* made these problems negligible by detecting species at all concentrations. As commonly described in other metagenomic studies, this pipeline could be susceptible to contamination and false positive results [396]. However, after the application of the appropriate thresholds, all spiked samples were negative for any non-targeted species. Similarly, all the non-template control samples were negative for any bacterial species which meant that CMg pipeline was a highly sensitive method that could also be used to rule out infections.

The introduction of an enrichment phase as part of the *'5mL enrichment protocol'* resulted in increased yields of microbial DNA. Consequently, adding higher yields to the WGA reaction resulted in increased yields post-WGA when compared to non-enriched samples. The higher DNA

quantities allowed the use of a rapid barcoding library instead of the PCR-based approach used in the *'1mL standard protocol'*, which is one of the fastest and simplest manual library preparation protocols [380,381,383]. Nonetheless, the application of a PCR-free library preparation approach caused a slightly shorter average amplicon size and consequently, the augment on reads matching the reference did not result in a proportional increase in genome coverage.

Genomic coverage varied among the different spiked species and, because read length values were similar across samples, it was directly impacted by the number of reads obtained (which also varied among species). Despite having a limit of detection of 1-5 CFU/mL for all species when using the *'5mL enrichment protocol',* some species like *E. coli*, had a significantly higher number of reads compared to other species and consequently a higher average genome coverage. It was hypothesized that the short enrichment phase could be introducing species-differences due to both the quantity of bacteria bound to the beads and their multiplication rate during the medium incubation. This, combined with all the previously listed steps of the pipeline that can potentially introduce species-bias, had a direct impact on the final number of reads and the genomic coverage and could potentially be a limitation of the protocol. In the future, it would be necessary to test a bigger panel of species to evaluate the real impact of all the listed steps introducing species bias.

4.4.5 Antimicrobial determinants were identified at all spiked concentrations.

Because of the excellent analytical sensitivity of the CMg pipelines, it was decided to test the pipelines by spiking them with clinically relevant concentrations of a species containing AMR resistant genes. The detection of antimicrobial genes could be challenging due to the low coverage numbers obtained using either CMg protocol (at least 10X coverage is necessary for reliable gene detection [310]), making it difficult to establish phenotypic correlations [187,397]. However, when the pipeline was used to study samples spiked with different concentrations of *E. coli* CXT-M-15, most of the pEK499 plasmid resistance genes [398] were detected. Compared to obtained genomic coverage values on LoD experiments, all resistance genes had a higher depth of coverage (>1X).

This, together with the high template identity and obtained template coverage, allowed the detection of almost all resistance genes at the lowest spiking concentrations. As pEK499 is a low-copy number plasmid [399,400], its yield would not be superior to the chromosomic bacterial DNA. The increased coverage on plasmid AMR genes could be explained by the plasmid amplification bias introduced during WGA, which has also been described by other authors such as Pinard *et. al.* (2006) [394] or Dean *et. al.* (2001) [401] in which the Phi29 polymerase used on WGA had a preference for plasmid amplification over genomic. This bias could be problematic when trying to detect bacterial genomic AMR genes. Nonetheless, the most prevalent BSI causative antimicrobial resistance strains (MRSA, Vancomycin-Resistant *Enterococcus*, Multidrug-Resistant *Enterobacteriaceae*, ESBLs Gram-negative species or Carbapenem-Resistant *Enterobacterales* [36]), generally have plasmid-mediated resistant mechanisms [402–406]. Because of the low coverage obtained from two of the genes present in the plasmid (*aac6'-Ib-cr* and *catB4*), these were misidentified as closely related genes: *aac(6')-Ib-W104R, aac(6')-Ib-D181Y, aac(6')-Ib-AKT and catB3*. However, all these variants were mutations of the expected genes and are part of the same resistance mechanisms [407–409]. To overcome this, our method could be adapted to have longer sequencing incubation times, resulting in higher coverage. As nanopore sequencing allows real-time visualisation of results, it was proposed that within the first hours of sequencing the causative microorganisms could be identified and extended incubation times (3-36 hours) would allow, thanks to the higher coverage, an improved detection of specific AMR determinants.

4.4.6 Conclusion.

In conclusion, the different methodological steps that were optimised in chapter 2 were further refined in this chapter ultimately generating a pipeline that is suited for detecting BSI. To the best of our knowledge, this was the first time that a CMg pipeline with a sensitivity of 1-5 CFU/mL across several species and under 12 hours could offer reliable species detection and AMR

determinants information. Our method, compared to blood cultures, has the potential to be used as a more rapid and sensitive diagnostic tool to improve the management of patients with BSI.

When compared to similarly published CMg approaches, the presented method outperformed them in terms of speed and sensitivity. The CMg pipeline described by Parize *et al.* (2017) [201] was significantly slower than the described method (12 hours compared to 2-3 days) while reaching similar sensitivity. Other approaches that have focused on sequencing cell-free DNA, such as Blauwkamp *et al.* (2019) [124] or Grumaz *et al.* (2019) [197], were either slower (>24 hours) or less sensitive (<65%) than the described CMg pipeline. Additionally, to date, none of the published CMg approaches have reported successful AMR determinants identification without the need of long periods of sequencing time (2-3 days). Proof-of-concept results demonstrated a method that is significantly faster and more sensitive than those published in literature. However, the presented methodology would still have to be tested with clinical blood samples.

CHAPTER FIVE

DISCUSSION

Addressing the diagnosis of BSI is crucial for reducing its incidence and mortality rates. Current blood culture methods are insufficient in terms of both speed and sensitivity, hindering the ability of clinicians to provide precise treatment and initiate early antibiotic de-escalation. Theoretically, a comprehensive CMg pipeline for BSI diagnosis appears highly beneficial, offering genetic-level insights on the infection cause in under 24 hours. However, in practice, implementing a CMg pipeline for BSI detection faces unresolved challenges, such as the substantial imbalance between host DNA and bacterial DNA presence in samples and the difficulty to distinguish between contaminants or causative species of infection. Previous efforts to apply a CMg pipeline for BSI detection have been limited in scope. Some focused solely on sequencing the cell-free DNA on plasma (e.g., Grumaz *et al.* in 2016 [158]), which makes the detection of AMR genes difficult, while others developed CMg pipelines that required extended sequencing times (e.g., Parize *et al.* in 2016 [207]). Despite these limitations, the scientific community agrees on two key points: these challenges are surmountable and, in consequence, metagenomics holds the future of diagnostics for infectious diseases [76,120,125,128,145]. The development of this CMg diagnostics tool is the goal of many laboratories in both the public and the private sector. This is illustrated by the fact that a company such as DayZero Diagnostics in California (USA) has attracted approximately $50 million in investment since its founding in 2017. This company is working on an automated CMg pipeline for BSI detection, claiming to be on the verge of market release by using a very similar approach to the one used in the project (host DNA depletion step, followed by WGA and nanopore sequencing) [410]. Their poster at the recent European Congress of Clinical Microbiology and Infectious Diseases (2023 ECCMID) earned the “Best poster of the day” award, emphasising the importance given by the medical microbiology community to finding alternative diagnostic approaches for BSI. In this project, the overall aim has been to develop a CMg pipeline that overcomes the issues associated with the high levels of host DNA present in samples and that allows the rapid detection of low concentrations of BSI-causing microorganisms in blood.

5.1 Assets of the study and implementation of new technologies in the medical field.

From the conception of the project, the primary objective was to develop a CMg pipeline that is faster and more sensitive than the current service standard diagnostic techniques and has potential for clinical implementation. To achieve this, a non-linear model for the development and implementation of new medical technologies similar to the one described by Clark *et al.* (2018) [314]. was followed. This model consists of two distinct phases: one in which a new medical technology is conceptualised, created and tested (on mock samples) and another phase in which the methodology is implemented and tested in the clinical facility. Importantly, this approach acknowledges that the generation of new clinical methodologies does not usually end after the first implementation phase. Instead, the various stages influence each other, allowing for interactive feedback that will help to improve the method. The implementation phase often provides essential insights for researchers to revisit their invention phase to then attempt the implementation of an enhanced version of the technology. A generic schematic representation of this model can be seen in Figure 53.

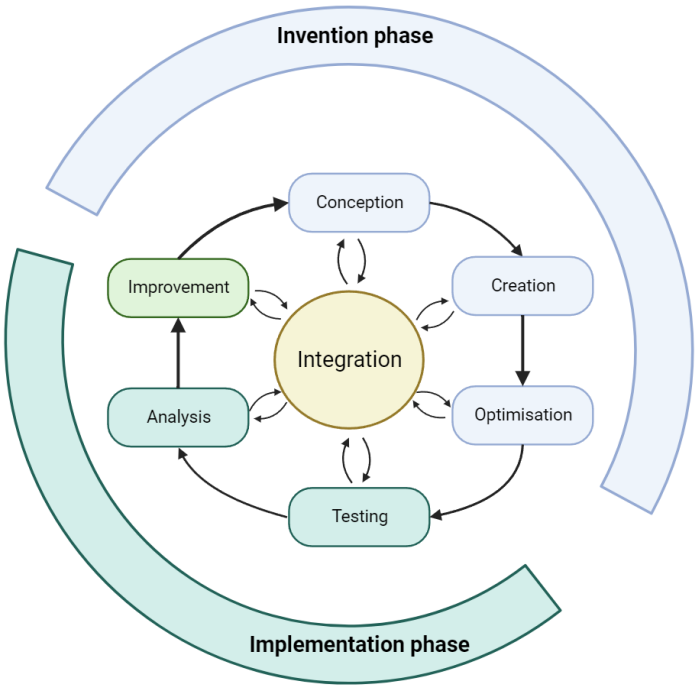


Figure 53. Schematic representation of a medical technology innovation cycle. Figure adapted from Clark *et al.* 2020.

During the initial conceptual stage of the project, the need for an alternative BSI diagnostic method was identified. The design of the new CMg pipeline considered the limitations encountered by previous CMg methods for the diagnosis of BSI, often limited by the abundant presence of host DNA and the low quantity of bacterial DNA found in septic clinical blood samples. Two key methodologies were pinpointed and hypothesised to have potential to address these issues: the application of a saponin-based host depletion method to eliminate the large presence of host DNA in blood samples and the subsequent WGA of the microbial DNA. To our knowledge, the combination of these two methods had not been previously explored in the context of BSI detection and presented a unique opportunity to significantly advance the field of BSI diagnosis.

During the generation and optimization of the host DNA depletion and WGA steps, which were described in chapter 2, several challenges were encountered. Despite the high efficiency of the saponin-based host DNA depletion method, a significant amount of host mitochondrial DNA persisted in the samples after treatment. The resistance of human mitochondria to the saponin-based method reduced the sensitivity of the pipeline. Only after optimising the saponin-based method (by increasing the saponin concentration and scaling up the reaction volume) and the WGA protocol (by changing reaction buffers and increasing the sample volume), the detection of *E. coli* and *S. aureus* spiked at clinically relevant concentrations was possible. To our knowledge, this marked the first successful application of a CMg approach for the identification of bacteria at concentrations of 1-5 CFU/mL in whole blood samples in under 9 hours.

After establishing the potential analytical sensitivity of the developed CMg pipeline, the project progressed to the implementation phase. Here, the method was tested using clinical blood samples from confirmed NNUH BSI patients and compared against service standard methods. For this collaboration with the NNUH, various pre- and post-analytical factors were considered: the clinical samples would be collected in EDTA-coated tubes, samples would only come from patients with blood culture confirmed BSI, samples would also be analysed with other non-CMg methods

(16S sequencing and potentially qPCR), and results from all methodologies would be compared to the service standard microbial identification results. As contemplated by the medical technology innovation cycle model previously discussed (Figure 53), the proof-of-concept implementation phase highlighted key findings that suggested limitations in both the CMg method and the pre- and post-analytical considerations taken when establishing the collaboration. Thus, the last part of the project consisted in addressing some of these limitations.

First, the pipeline sensitivity was increased by improving the host DNA depletion step to detect smaller quantities of microorganisms. This was done to account for the partial loss of microbial viability that takes place in the collection tube from the moment of collection to the time of processing. In addition, this microbial DNA loss was further addressed by changing the type of tube in order to improve the microbial viability. By the end of the project, a CMg pipeline (referred to as the *'5mL enrichment protocol'*) had been developed. This protocol was capable to, for the first time, reliably detect several microorganism's species when they were present at concentrations of 1-5 CFU/mL of blood in as little as 12 hours, compared to the 2-5 days needed to obtained results from blood culture-based tests. The potential implementation of this tool for BSI diagnosis would mean that patients would be able to receive targeted antibiotic treatment after the first generic dose that is usually administered. This is highly relevant as studies have shown that mortality rates highly increase every hour that the appropriate antibiotic is not given, and it would therefore have a tremendous effect on patient prognosis and survival [69,71,72]. In the long term, the use of faster and more accurate BSI diagnostics tools, like the one developed in this project, will potentially contribute to the improvement of the currently limited epidemiological data and, most importantly, to reduce the increasing burden of antibiotic resistance.

5.2 Limitations of the study and future directions.

While the CMg pipeline developed in this study exhibited high reliability and sensitivity, its current version may prove too complex for its practical application in clinical laboratories. Several

steps within the method demand expertise, both in the wet lab and *in silico* steps. Authors such as Cherkaoui and Schrenzel [411] argue that the complete automation of new diagnostic technologies is essential for the complete substitution of the current culture-based service standard techniques. CMg methods like the one presented in this thesis need further simplification, with the ultimate goal of achieving automated high-throughput sample processing, which would ultimately surpass current service standard methods.

Furthermore, the developed CMg pipeline heavily depends on an initial saponin-based host depletion method which, because it involves the addition of DNAse in the digestion step, requires that the membranes of microorganisms to remain intact. This may be difficult since many patients might be under antibiotic treatment at the time of blood sample collection (as observed in 20% of cases in a study by Hirosawa *et al.* in 2023 [412]). Thus, applying the host depletion method to blood samples containing antibiotics could result in the loss of bacterial DNA. As such, sample selection and timing will be key for testing this method on clinical blood samples. Furthermore, the saponin-based method may also adversely affect microorganisms that have membrane properties that are similar to those of eukaryotic cells, leading to their digestion during the depletion process. This could be the case in infections caused by *Candida spp*., as fungal cell walls contain sterol groups [232] which saponin would interact with, leading to their lysis. Consequently, the proposed pipeline would have to be tested on a larger array of species with different membrane characteristics to assess how these would be affected by the saponin-based depletion method.

Lastly, the presented CMg pipeline encompasses multiple steps where species bias could be introduced, as in previous chapters: the enrichment phase of the Sepsi*PURE* method, the saponin-based depletion step, the WGA step or the PCR-based library amplification phase. This species bias can lead to the display of variable efficiencies at detecting different potential BSI causative microbial species. This problem, also associated with service standard blood culture

methodologies, would have to be further investigated by testing the method on mock samples spiked with a large panel of different BSI causative species.

In terms of pre-analytical considerations needed for the integration of this method into clinical practice, it would be worthwhile to do more experimentation to determine the most optimal blood collection tube for this approach. Despite having substantial indirect evidence hinting at potential microbial degradation when using EDTA-coated tubes for blood collection, no experiments have been reported exploring their effect on microbial viability over time. Indeed, conducting comparative studies to assess the impact of different anticoagulants on microbial membrane integrity and viability before the application of the saponin-based method would be highly valuable. Despite the inhibitory effect of SPS on the saponin-based reaction that has been shown in this study, SPS is recognized as one of the primary anticoagulants that can preserve microbial viability in blood DayZero Diagnostics, a leading company in the development of a CMg kit for BSI detection, filed a patent in 2021 proposing the use of SPS as the anticoagulant in blood collection tubes, which is then removed by the addition of albumin [413] before the subsequent downstream processing. Albumin interacts with SPS and precipitates after a brief centrifugation step. This approach shows promise for enhancing the viability of microorganisms in clinical blood samples that need to be processed using CMg pipelines. Future work could further investigate how to best utilise SPS for the protection of the microorganisms in blood and how to eliminate SPS from the sample before saponin-based host DNA depletion.

It would also be worthwhile to further study host DNA depletion methods. Despite the extensive reduction in host DNA achieved with the developed pipeline (over 99.999% chromosomal DNA depletion), sequencing results still revealed significantly higher quantities of host DNA in comparison to microbial DNA in spiked and clinical blood samples. The high presence of host DNA not only makes pathogen identification challenging, but it makes detection of AMR genes problematic. The low number of sequenced microbial reads resulted in poor coverage values that

do not meet the minimum 10X coverage required to be confident about the identification of specific AMR genes and SNPs. Enhancing the depletion efficiency would be essential to further improve the sensitivity of the CMg pipeline, enabling the detection of more microbial reads and the obtention of deeper sequencing coverage of their genomes. The higher sequencing coverage would, in turn, also improve the detection of AMR genes. One approach to achieve higher host DNA depletion would be combining saponin with other detergents that have proven effective in generating a differential lysis approach, such as Tween 20 or Triton-X [138], in the host DNA depletion step. This could involve creating a "cocktail" of detergents for the improved lysis of host cells and mitochondria membranes.

The bioinformatics pipeline used for the identification of microbial reads could also be further improved. Throughout this project, efforts were made to refine the bioinformatic filtering of contaminant and host DNA reads, as well as to enhance the accuracy of the taxonomic species classification. This was especially challenging since specific guidelines for the generation of an optimal bioinformatics approach for CMg-based BSI diagnosis have not yet been established. Furthermore, given the rapid development of bioinformatic technology in general and bioinformatic pipelines in particular, novel advances in the field will have to be integrated into the CMg pipeline as they are being published. Constant revision of the pipelines used for the CMg-based BSI diagnosis will help to establish more accurate thresholds for the positive detection of species and read filtering.

Finally, as the next step of the circular model for the implementation of new technologies, the updated CMg pipeline described in chapter 4 would need to be tested on clinical blood samples to establish its suitability for BSI diagnosis. One of the main pre-analytical considerations to be taken in account would be to prioritize the selection of samples from patients who have not been administered antibiotics as well as to process clinical blood samples immediately after their extraction. To facilitate this new implementation step, a new collaboration has been recently

granted to our group by the Barts Health NHS Trust (Barts Health) to process 250 clinical blood samples from endocarditis patients from Queen Mary's Hospital in London. For this collaborative project, the mentioned considerations are being implemented and samples are sent to the laboratory immediately after collection and processed in less than 24 hours by QIB staff. A high positivity rate has been observed in preliminary results from the first samples, but sequencing results have not been compared to microbiology results yet.

5.3 Conclusion

The generation of improved BSI diagnostics methods is key for the reduction of the mortality of this condition, helping its targeted treatment and antibiotic resistance de-escalation. In this study, a CMg pipeline has been developed for the analysis of whole blood samples through the identification of microbial species and genes associated with antimicrobial resistance. At the beginning of the project, the fact that current saponin-based host depletion methods lack the necessary efficiency to detect low concentrations of bacterial species in blood due to a relatively high proportion of retained human mitochondrial DNA was demonstrated. Even after initial optimisation, the application of this method on clinical blood samples resulted in poor sensitivity outcomes. Additional reduction of persistent mitochondrial DNA in the samples was achieved with the addition of a bead-based microbial enrichment protocol followed by the application of a higher concentration of saponin during host depletion steps, which led to the successful detection of *E. coli*, *S. aureus*, *K. pneumoniae* and *E. faecalis* in samples spiked at clinically relevant concentrations.

Classification of the microbial sequences obtained with the CMg pipeline was completed using a customised bioinformatics approach with a threshold for the determination of reportable aetiological agents. In LoD experiments with mock samples, only true positive results were observed in spiked samples and no false positives were observed in control samples. AMR genes were also detectable in sequencing data if there was sufficient genome coverage. Overall, four of the main causative species of BSI could be detected when spiked into 5mL of whole blood at

clinically relevant concentrations (1-5 CFU/mL blood) in under 12 hours, from sample collection to results. Further reduction of human mitochondrial DNA are anticipated to be achievable by further optimising the host depletion step and to potentially have an incremental effect on the sensitivity of CMg pipelines for BSI.

In conclusion, this study presents several methodological developments and optimization steps for the generation of a CMg pipeline that has the potential to be used for the diagnosis of BSI. Compared to blood culture, this methodology could be used as a more rapid and sensitive diagnosis tool, positively contributing to the management of patients with BSI.

Publications, conferences and collaborations.

Publications

- Moragues-Solanas L., Le-Viet T., Halford C., et. al. Development of a proof-of-concept rapid clinical metagenomics diagnostic tool for the detection of bloodstream infections and sepsis management directly from whole blood. Manuscript in review, submitted at BMC Medical Genomics in October 2023.
- Maria Diaz, Heather Aird, Thanh Le Viet,…*et al*…, Lluis Moragues-Solanas, Nicolle Som, Jim McLauchlin, Frieda Jørgensen, Matthew Gilmour. Microbial composition and dynamics in environmental samples from a ready-to-eat food production facility with a long-term colonisation of Listeria monocytogenes. Manuscript in preparation, to be submitted in December 2023.
- Moragues-Solanas L., Scotti R., & O'Grady J. (2021). Rapid metagenomics for diagnosis of bloodstream and respiratory tract nosocomial infections: current status and future prospects. Expert Review of Molecular Diagnostics, 21(4), 371-380.

Conference attendance

- Poster Presentation at European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) 2023. Poster Title: Metagenomics diagnosis for bloodstream infections using a highly efficient host depletion and bacterial genome amplification protocol from whole blood.
- Poster Presentation at International Conference on Clinical Metagenomics (ICCMg) 2022 – Development of a metagenomics diagnostic methodology for bloodstream infections.
- Poster Presentation at European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) 2020. Poster Title: Rapid, sensitive diagnosis of sepsis using clinical metagenomics.

Collaborations and Institutional responsibility

- Created and wrote the laboratory protocols that helped establishing the AMetIP (*Accuracy of Metagenomic Blood Sampling to Identify Pathogen in Infective Endocarditis Patients*) project with NHS Barts Health Trust in 2023 for processing endocarditis patients blood samples.
- Created the protocols and worked on the ethical approval for establishing the NNUH collaboration in 2022. Obtained an Impact Acceleration Account grant for this project.

Bibliography

1. Viscoli C. Bloodstream Infections: The peak of the iceberg. Virulence [Internet]. 2016;7(3):248–51. Available from: http://dx.doi.org/10.1080/21505594.2016.1152440

2. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet [Internet]. 2016;388(10063):3027–35. Available from: http://dx.doi.org/10.1016/S0140-6736(16)31593-8

3. Bearman GML, Wenzel RP. Bacteremias: A leading cause of death. Arch Med Res. 2005;36(6):646–59.

4. Diekema DJ, Beekmann SE, Chapin KC, Morel KA, Munson E, Doern G V. Epidemiology and outcome of nosocomial and community-onset bloodstream infection. J Clin Microbiol. 2003;41(8):3655–60.

5. Laupland KB. Incidence of bloodstream infection: A review of population-based studies. Clin Microbiol Infect. 2013;19(6):492–500.

6. Angus DC, Van Der Poll T. Severe sepsis and septic shock. N Engl J Med. 2013;369(9):840–51.

7. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Jama. 2016;315(8):801–10.

8. Bone, R. C., Balk RA, Cerra FB, Dellinger RP, Fein AM, Knaus WA, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Chest. 1992;101(6):1644–55.

9. Al Mohajer M, Darouiche RO. Sepsis syndrome, bloodstream infections, and device-related infections. Med Clin North Am [Internet]. 2012;96(6):1203–23. Available from: http://dx.doi.org/10.1016/j.mcna.2012.08.008

10. Shankar-Hari M, Harrison DA, Rubenfeld GD, Rowan K. Epidemiology of sepsis and septic shock in critical care units: Comparison between sepsis-2 and sepsis-3 populations using a national critical care database. Br J Anaesth. 2017;119(4):626–36.

11. Surveillance AS. Adult Sepsis Surveillance. 2018;(May).

12. Rhee C, Rhee C, Wang R, Zhang Z, Fram D, Kadri SS, et al. Epidemiology of hospital-onset versus community-onset sepsis in U.S. Hospitals and association with mortality: A retrospective analysis using electronic clinical data. Crit Care Med. 2019;47(9):1169–76.

13. Bizzarro MJ, Dembry LM, Baltimore RS, Gallagher PG. Changing patterns in neonatal Escherichia coli sepsis and ampicillin resistance in the era of intrapartum antibiotic prophylaxis. Pediatrics. 2008;121(4):689–96.

14. Batimore R. Neonatal Sepsis. Pediatr drugs. 2003;5(11):719–32.

15. Alberts B, Johnson A, Lewis J, Raff, M. R, K., Walter P. Molecular Biology of the Cell. NY: Garlan. New York; 2002.

16. Castillo DJ, Rifkin RF, Cowan DA, Potgieter M. The healthy human blood microbiome: Fact or fiction? Front Cell Infect Microbiol. 2019;9(MAY):1–12.

17. Smith, D. A., & Nehring SM. Bacteremia. 2017;

18. Morris JA, Harrison LM, Biswas J, Telford DR. Transient bacteraemia: A possible cause of sudden life threatening events. Med Hypotheses. 2007;69(5):1032–9.

19. Lewis T. Observations relating to subacute infective endocarditis. Heart. 1923;10:21–9.

20. Del Giudice C, Vaia E, Liccardo D, Marzano F, Valletta A, Spagnuolo G, et al. Infective endocarditis: A focus on oral microbiota. Microorganisms. 2021;9(6):1–18.

21. Orsini J, Mainardi C, Muzylo E, Karki N, Cohen N, Sakoulas G. Microbiological profile of organisms causing bloodstream infection in critically ill patients. J Clin Med Res. 2012;4(6):371.

22. Chancharoenthana W, Kamolratanakul S, Schultz MJ, Leelahavanichkul A. The leaky gut and the gut microbiome in sepsis – targets in research and treatment. Clin Sci. 2023;137(8):645–62.

23. Minasyan H. Sepsis: mechanisms of bacterial injury to the patient. Scand J Trauma Resusc Emerg Med. 2019;27(1):1–22.

24. Vanuytsel T, Tack J, Farre R. The Role of Intestinal Permeability in Gastrointestinal Disorders and Current Methods of Evaluation. Front Nutr. 2021;8(August).

25. Deitch EA. Gut-origin sepsis: Evolution of a concept. Surgeon. 2012;10(6):350–6.

26. Whittle E, Leonard MO, Harrison R, Gant TW, Tonge DP. Multi-method characterization of the human circulating microbiome. Front Microbiol. 2019;10(JAN):1–12.

27. Cogen AL, Nizet V, Gallo RL. Skin microbiota: A source of disease or defence? Br J Dermatol. 2008;158(3):442–55.

28. Franco-Paredes C. Chapter 2 - Bloodstream Infections. Core Concepts Clin Infect Dis. 2016;(Cccid):9–15.

29. Waters D, Jawad I, Ahmad A, Lukšić I, Nair H, Zgaga L, et al. Aetiology of community-acquired neonatal sepsis in low- and middle-income countries. J Glob Health. 2011;1(2):154–70.

30. Stearns-Kurosawa DJ, Osuchowski MF, Valentine C, Kurosawa S, Remick DG. The pathogenesis of sepsis. Annu Rev Pathol Mech Dis. 2011;6:19–48.

31. Huang M, Cai S, Su J. The pathogenesis of sepsis and potential therapeutic targets. Int J Mol Sci. 2019;20(21).

32. Wiersinga WJ, van der Poll T. Immunopathophysiology of human sepsis. eBioMedicine [Internet]. 2022;86:104363. Available from: https://doi.org/10.1016/j.ebiom.2022.104363

33. Wiersinga WJ, Seymour CW. Handbook of sepsis. Handb Sepsis. 2018;1–253.

34. Jarczak D, Kluge S, Nierhaus A. Sepsis—Pathophysiology and Therapeutic Concepts. Front Med. 2021;8(May):1–22.

35. Van der Poll T, Wiersinga WJ. Chpater 73: sepsis and septic shock. In: Principles and practice of infectious diseases. 9th ed. Saunders; 2019.

36. Diekema DJ, Hsueh PR, Mendes RE, Pfaller MA, Rolston K V., Sader HS, et al. The microbiology of bloodstream infection: 20-year trends from the SENTRY antimicrobial surveillance program. Antimicrob Agents Chemother. 2019;63(7):e00355-19.

37. Public Health England. Laboratory surveillance of polymicrobial bacteraemia and fungaemia in England, Wales and Northern Ireland: 2017. Heal Prot Rep [Internet]. 2018;12(10). Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/691268/hpr1018_polymcrbls.pdf

38. McDanel J, Schweizer M, Crabb V, Nelson R, Samore M, Khader K, et al. Incidence of extended-spectrum β-Lactamase (ESBL)-producing Escherichia coli and klebsiella infections in the United States: A systematic literature review. Infect Control Hosp Epidemiol. 2017;38(10):1209–15.

39. Johnson AP, Davies J, Guy R, Abernethy J, Sheridan E, Pearson A, et al. Mandatory surveillance of methicillin-resistant Staphylococcus aureus (MRSA) bacteraemia in England: The first 10 years. J Antimicrob Chemother. 2012;67(4):802–9.

40. Turnidge J. Impact of antibiotic resistance on the treatment of sepsis. Scand J Infect Dis. 2003;35(9):677–82.

41. Delaloye J, Calandra T. Invasive candidiasis as a cause of sepsis in the critically ill patient. Virulence. 2014;5(1):154–62.

42. Guinea J. Global trends in the distribution of Candida species causing candidemia. Clin Microbiol Infect. 2014;20(6):5–10.

43. Pfaller MA, Diekema DJ, Isolates C, Ic F. Epidemiology of Invasive Candidiasis : a Persistent Public Health Problem. Clin Microbiol Rev. 2007;20(1):133–63.

44. Montini F, Grenouillet F, Capellier G, Piton G. Strongyloidiasis : an unusual cause of septic shock with pneumonia and enteropathy in western countries. 2015;1–3.

45. Lucet, J. C., Bailly MP, Bedos JP, Wolff M, Gachot B, Vachon F. Septic shock due to toxoplasmosis in patients infected with the human immunodeficiency virus. Chest. 1993;104(4):1054–8.

46. Njim T, Dondorp A, Mukaka M, Ohuma EO. Identifying risk factors for the development of sepsis during adult severe malaria. Malar J [Internet]. 2018;1–10. Available from: https://doi.org/10.1186/s12936-018-2430-2

47. Srinivas R, Agarwal R, Gupta D. Severe sepsis due to severe falciparum malaria and leptospirosis co-infection treated with activated protein C. 2007;5:1–5.

48. Boissier F, Vermersch C, Spagnolo S, Bruneel F, Prost N De. Septic shock complicating Plasmodium falciparum malaria in a pregnant patient with low parasitemia. 2014;44:539–40.

49. The G, Pattern S, Incoming THE, Inoculum V. Pathogenesis of Virus Infections. 2017;

50. Kimura H, Ito Y, Futamura M, Ando Y, Yabuta Y, Hoshino Y, et al. Quantitation of Viral Load in Neonatal Herpes Simplex Virus Infection and Comparison Between Type 1 and Type 2. 2002;353(December 2001):349–53.

51. Fleischmann C, Scherag A, Adhikari NKJ, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of global incidence and mortality of hospital-treated sepsis current estimates and limitations. Am J Respir Crit Care Med. 2016;193(3):259–72.

52. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. Lancet. 2020;395(10219):200–11.

53. Metrics GH. Global , regional , and national incidence , prevalence , and years lived with disability for 354 diseases and injuries for 195 countries and territories , 1990 – 2017 : a systematic analysis for the Global Burden of Disease Study 2017. 2018;1990–2017.

54. Fleischmann C, Reichert F, Cassini A, Horner R, Harder T, Markwart R, et al. Global incidence and mortality of neonatal sepsis : a systematic review and meta- - analysis. 2021;745–52.

55. Haque M, Sartelli M, Mckimm J, Abu Bakar M. Health care-associated infections – an overview. Infect Drug Resist [Internet]. 2018;11(1):2321–33. Available from: http://dx.doi.org/10.2147/IDR.S177247

56. Bassler D, Stoll BJ, Schmidt B, Asztalos E V., Roberts RS, Robertson CMT, et al. Using a count of neonatal morbidities to predict poor outcome in extremely low birth weight infants: Added role of neonatal infection. Pediatrics. 2009;123(1):313–8.

57. Schlapbach LJ, Aebischer M, Adams M, Natalucci G, Bonhoeffer J, Latzin P, et al. Impact of sepsis on neurodevelopmental outcome in a swiss national cohort of extremely premature infants. Pediatrics. 2011;128(2).

58. Page, D. B., Donnelly, J. P., & Wang HE. Community-, healthcare-and hospital-acquired severe sepsis hospitalizations in the University HealthSystem Consortium. Crit Care Med. 2015;43(9):1945.

59. Agodi A, Barchitta M, Auxilia F, Brusaferro S, D'Errico MM, Montagna MT, et al. Epidemiology of intensive care unit-acquired sepsis in Italy: Results of the SPIN-UTI network. Ann di Ig Med Prev e di Comunita. 2018;30(5):15–21.

60. Markwart R, Saito H, Harder T, Tomczyk S, Cassini A, Fleischmann-Struzek C, et al. Epidemiology and burden of sepsis acquired in hospitals and intensive care units: a systematic review and meta-analysis. Intensive Care Med [Internet]. 2020;46(8):1536–51. Available from: https://doi.org/10.1007/s00134-020-06106-2

61. Singer M, Inada-Kim M, Shankar-Hari M. Sepsis hysteria: excess hype and unrealistic expectations. Lancet [Internet]. 2019;394(10208):1513–4. Available from: http://dx.doi.org/10.1016/S0140-6736(19)32483-3

62. Statistics NDHE. NHS Digital. 2019 [cited 2023 Sep 27]. Discharges for sepsis broken down by age, 2015-16 to 2017-18. Available from: https://digital.nhs.uk/data-and-information/find-data-and-publications/supplementary-information/2019-supplementary-information-files/discharges-for-sepsis-broken-down-by-age-2015-16-to-2017-18

63. Statistics O for N. https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/deathsfromsepsisintheuk2006to202/deathsfromsepsisintheuk2006to2020. Deaths from sepsis in the UK 2006 to 2020.

64. Hex N, Retzler J, Bartlett C, Arber M. The Cost of Sepsis Care in the UK. Yhec [Internet]. 2017;(February):1–46. Available from: http://sepsistrust.org/wp-content/uploads/2017/02/YHEC-Sepsis-Report-17.02.17-FINAL.pdf

65. Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. Crit Care Med. 2004;32(3):858–73.

66. Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. Intensive Care Med [Internet]. 2021;47(11):1181–247. Available from: https://doi.org/10.1007/s00134-021-06506-y

67. Daniels R, Nutbeam T. THE SEPSIS MANUAL 4th edition 2017 - 2018 [Internet]. 2017. 120 p. Available from: https://sepsistrust.org/wp-content/uploads/2018/06/Sepsis_Manual_2017_web_download.pdf

68. Christopher W. Seymour, M.D. FG, M.D. HCP, M.D. ME, Friedrich MD, Iwashyna TJ, M.D. PD, et al. Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. N Engl J Med. 2017;376(23):2235–44.

69. Kumar A, Haery C, Paladugu B, Kumar A, Symeoneides S, Taiberg L, et al. The Duration of Hypotension before the Initiation of Antibiotic Treatment Is a Critical Determinant of Survival in a Murine Model of Escherichia coli Septic Shock: Association with Serum Lactate and Inflammatory Cytokine Levels. J Infect Dis. 2006;193(2):251–8.

70. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med. 2006;34(6):1589–96.

71. Gaieski, D. F., Mikkelsen ME, Band RA, Pines JM, Massone R, Furia FF, et al. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. Crit Care Med. 2010;38(4):1045–105.

72. Weiss SL, Fitzgerald JC, Balamuth F, Alpern ER, Lavelle J, M. C, et al. Delayed antimicrobial therapy increases mortality and organ dysfunction duration in pediatric sepsis. Crit Care Med. 2014;42(11):2409.

73. National Institute for Health and Care Excellence. Sepsis: recognition, diagnosis and early

management: NICE (2017) Sepsis: recognition, diagnosis and early management. BJU Int. 2018;121(4):497–514.

74. Hunt A. Sepsis: An overview of the signs, symptoms, diagnosis, treatment and pathophysiology. Emerg Nurse. 2019;27(5):32–41.

75. Faridi, M. A., Ramachandraiah, H., Banerjee, I., Ardabili S, Zelenin S, Russom A. Elasto-inertial microfluidics for bacteria separation from whole blood for sepsis diagnostics. J Nanobiotechnology. 2017;15:1–9.

76. Costa SP, Carvalho CM. Burden of bacterial bloodstream infections and recent advances for diagnosis. Pathog Dis. 2022;80(1):1–13.

77. Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. Intensive Care Med [Internet]. 2018;44(6):925–8. Available from: https://doi.org/10.1007/s00134-018-5085-0

78. Public Health England. Microbiology Services. UK Standards for Microbiology Investigations. Stand Unit, Natl Infect Serv Bacteriol. B 37(8.2):1–55.

79. Opota O, Croxatto A, Prod'Hom G, Greub G. Blood culture-based diagnosis of bacteraemia: state of the art. Clin Microbiol Infect. 2015;21(4):313–22.

80. Diaz E, Valle J. Risks for multidrug-resistant pathogens in the ICU. 2014;20(5):516–24.

81. Arulkumaran N, Routledge M, Schlebusch S, Lipman J, Conway Morris A. Antimicrobial-associated harm in critical care: a narrative review. Intensive Care Med [Internet]. 2020;46(2):225–35. Available from: https://doi.org/10.1007/s00134-020-05929-3

82. Guo Y, Gao W, Yang H, Sui S. De-escalation of empiric antibiotics in patients with severe sepsis or septic shock: a meta-analysis. Hear Lung. 2016;45(5):454-459.

83. Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. Intensive Care Med. 2018;44(6):925–8.

84. Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M. The era of molecular and other non-culture-based methods in diagnosis of sepsis. Clin Microbiol Rev. 2010;23(1):235–51.

85. Palarasah Y, Skjoedt M ole, Vitved L, Andersen TE, Skjoedt K, Koch C. Sodium Polyanethole Sulfonate as an Inhibitor of Activation of Complement Function in Blood Culture Systems . 2010;48(3):908–14.

86. Cohen J, Vincent JL, Adhikari NKJ, Machado FR, Angus DC, Calandra T, et al. Sepsis: A roadmap for future research. Lancet Infect Dis. 2015;15(5):581–614.

87. Lamy B, Dargère S, Arendrup MC, Parienti JJ, Tattevin P. How to Optimize the Use of Blood Cultures for the Diagnosis of Bloodstream Infections ? A. Front Microbiol. 2016;7:697.

88. Skvarc M, Stubljar D, Rogina P, Kaasch AJ. Non-culture-based methods to diagnose bloodstream infection: does it work? Eur J Microbiol Immunol. 2013;3(2):97–104.

89. Opota O, Jaton K, Greub G. Microbial diagnosis of bloodstream infection: towards molecular diagnosis directly from blood. Clin Microbiol Infect. 2015;21(4):323–31.

90. Kirn TJ, Weinstein MP. Update on blood cultures: how to obtain, process, report, and interpret. Clin Microbiol Infect. 2013;19(6):513–20.

91. Azrad M, Keness Y, Nitzan O, Pastukh N, Tkhawkho L, Freidus V, et al. Cheap and rapid in-house method for direct identification of positive blood cultures by MALDI-TOF MS technology. 2019;1–7.

92. Beuving J, Van Der Donk CF, Linssen CF, Wolffs PF, Verbon A. Evaluation of direct inoculation of the BD PHOENIX system from positive BACTEC blood cultures for both Gram-positive cocci and Gram-negative rods. BMC Microbiol [Internet]. 2011;11(1):156. Available from: http://www.biomedcentral.com/1471-2180/11/156

93. Idelevich EA, Schüle I, Grünastel B, Wüllenweber J, Peters G, Becker K. Acceleration of antimicrobial susceptibility testing of positive blood cultures by inoculation of Vitek 2 cards with briefly incubated solid medium cultures. J Clin Microbiol. 2014;52(11):4058–62.

94. Kristóf K, Pongrácz J. Interpretation of blood microbiology results–function of the clinical microbiologist. Ejifcc. 2016;27(2):147.

95. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012;36(2):380–407.

96. Altun O, Botero-Kleiven S, Carlsson S, Ullberg M, Özenci V. Rapid identification of bacteria from positive blood culture bottles by MALDI-TOF MS following short-term incubation on solid media. J Med Microbiol. 2015;64(11):1346–52.

97. Lin HH, Tseng KH, Tien N, Lin YT, Yu J, Hsueh PR, et al. Evaluation of the Rapid Sepsityper protocol and specific MBT-Sepsityper module for the identification of bacteremia and fungemia using Bruker Biotyper MALDI-TOF MS. J Microbiol Immunol Infect [Internet]. 2022;55(6):1330–3. Available from: https://doi.org/10.1016/j.jmii.2022.07.005

98. Lo-Ten-Foe JR, De Smet AMGA, Diederen BMW, Kluytmans JAJW, Van Keulen PHJ. Comparative evaluation of the VITEK 2, disk diffusion, etest, broth microdilution, and agar dilution susceptibility testing methods for colistin in clinical isolates, including heteroresistant Enterobacter cloacae and Acinetobacter baumannii strains. Antimicrob Agents Chemother. 2007;51(10):3726–30.

99. Chen P, Li S, Li W, Ren J, Sun F, Liu R, et al. Rapid diagnosis and comprehensive bacteria profiling of sepsis based on cell-free DNA. J Transl Med. 2020;18(1):1–10.

100. Bochud PY, Bonten M, Marchetti O, Calandra T. Antimicrobial therapy for patients with severe sepsis and septic shock: an evidence-based review. Crit Care Med. 2004;32(11):S495–512.

101. Gupta S, Sakhuja A, Kumar G, McGrath E, Nanchal RS, Kashani KB. Culture-Negative Severe Sepsis: Nationwide Trends and Outcomes. Chest. 2016;150(6):1251–9.

102. Cardoso JR, Pereira LM, Iversen MD, Ramos AL. What is gold standard and what is ground truth? Dental Press J Orthod. 2014;19(5):27–30.

103. Poole S, Kidd SP, Saeed K. A review of novel technologies and techniques associated with identification of bloodstream infection etiologies and rapid antimicrobial genotypic and quantitative phenotypic determination. Expert Rev Mol Diagn. 2018;18(6):543–55.

104. Yagupsky P, Nolte FS. Quantitative aspects of septicemia. Clin Microbiol Rev. 1990;3(3):269–79.

105. Kellogg JA, Manzella JP, Bankert DA. Frequency of low-level bacteremia in children from birth to fifteen years of age. J Clin Microbiol. 2000;38(6):2181–5.

106. Dalla-Costa LM, Morello LG, Conte D, Pereira LA, Palmeiro JK, Ambrosio A, et al. Comparison of DNA extraction methods used to detect bacterial and yeast DNA from spiked whole blood by real-time PCR. J Microbiol Methods. 2017;140:61–6.

107. Lee CW, Chang HY, Wu JK, Tseng FG. Ultra-sensitive electrochemical detection of bacteremia enabled by redox-active gold nanoparticles (raGNPs) in a nano-sieving microfluidic system (NS-MFS). Biosens Bioelectron. 2019;133:215–22.

108. Kumar A, Ellis P, Arabi Y, Roberts D, Light B, Parrillo JE, et al. Initiation of inappropriate antimicrobial therapy results in a fivefold reduction of survival in human septic shock. Chest. 2009;136(5):1237–48.

109. Paolucci M, Landini MP, Sambri V. Conventional and molecular techniques for the early diagnosis of bacteraemia. Int J Antimicrob Agents. 2010;36:S6–16.

110. Peker N, Couto N, Sinha B, Rossen JW. Diagnosis of bloodstream infections from positive blood cultures and directly from blood samples: recent developments in molecular approaches. Clin Microbiol Infect. 2018;24(9):944–55.

111. Florio W, Morici P, Ghelardi E, Barnini S, Lupetti A. Recent advances in the microbiological diagnosis of bloodstream infections. Crit Rev Microbiol. 2018;44(3):351–70.

112. Deck MK, Anderson ES, Buckner RJ, Colasante G, Coull, J. M., Crystal B, Fiandaca MJ. Multicenter evaluation of the Staphylococcus Quick FISH method for simultaneous identification of Staphylococcus aureus and coagulase-negative staphylococci directly from blood culture bottles in less than 30 minutes. J Clin Microbiol. 2012;50(6):1994–8.

113. Oliveira K, Procop GW, Wilson D, Coull J, Stender H. Rapid identification of Staphylococcus aureus directly from blood cultures by fluorescence in situ hybridization with peptide nucleic acid probes. J Clin Microbiol. 2002;40(1):247–51.

114. Chang SS, Hsieh WH, Liu TS, Lee SH, Wang CH, Chou HC, et al. Multiplex PCR System for Rapid Detection of Pathogens in Patients with Presumed Sepsis - A Systemic Review and Meta-Analysis. PLoS One. 2013;8(5).

115. Wellinghausen, N., Kochem AJ, Disqué C, Mühl H, Gebert, S., Winter, J. Sakka SG. Diagnosis of bacteremia in whole-blood samples by use of a commercial universal 16S rRNA gene-based PCR and sequence analysis. J Clin Microbiol. 2009;47(9):2759–65.

116. Grif, K., Heller I, Prodinger WM, Lechleitner K, Lass-Flörl C, Orth D. Improvement of detection of bacterial pathogens in normally sterile body sites with a focus on orthopedic samples by use of a commercial 16S rRNA broad-range PCR and sequence analysis. J Clin Microbiol. 2012;50(7):2250–4.

117. Warhurst, G., Dunn G, Chadwick P, Blackwood B, McAuley D, Perkins GD, et al. Rapid detection of health-care-associated bloodstream infection in critical care using multipathogen real-time polymerase chain reaction technology: a diagnostic accuracy study and systematic review. Health Technol Assess (Rockv). 2015;19(35):1–142.

118. Straub J, Paula H, Mayr M, Kasper D, Assadian O, Berger A, et al. Diagnostic accuracy of the ROCHE Septifast PCR system for the rapid detection of blood pathogens in neonatal sepsis—A prospective clinical trial. PLoS One. 2017;12(11):e0187688.

119. Claudio U. Köser, Ellington MJ, Cartwright EJP, Gillespie SH, Ko CU, Brown NM, et al. Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. PLoS Pathog. 2012;8:e1002824:e1002824.

120. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20(6):341–55.

121. Ruppé E, Lazarevic V, Girard M, Mouton W, Ferry T, Laurent F, et al. Clinical metagenomics of bone and joint infections: a proof of concept study. Sci Rep. 2017;7(1):1–12.

122. Yan Q, Cui S, Chen C, Li S, Sha S, Wan X, et al. Metagenomic analysis of sputum microbiome as a tool toward culture-independent pathogen detection of patients with ventilator-associated pneumonia. Am J Respir Crit Care Med. 2016;194(5):636–9.

123. 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 Res. 2019;29(5):831–42.

124. Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol. 2019;4.4: 663-6.

125. Moragues-Solanas L, Scotti R, O'Grady J. Rapid metagenomics for diagnosis of bloodstream and respiratory tract nosocomial infections: current status and future prospects. Expert Rev Mol Diagn [Internet]. 2021;21(4):371–80. Available from: https://doi.org/10.1080/14737159.2021.1906652

126. Decuypere S, Meehan CJ, Van Puyvelde S, De Block T, Maltha J, Palpouguini, L., Deborggraeve S. Diagnosis of bacterial bloodstream infections: a 16S metagenomics approach. PLoS Negl Trop Dis. 2016;10(2):e0004470.

127. Sabat, A. J., van Zanten, E., Akkerboom V, Wisselink G, van Slochteren K, de Boer RF, Kooistra-Smid AM. Targeted next-generation sequencing of the 16S-23S rRNA region for culture-independent bacterial identification-increased discrimination of closely related species. Sci Rep. 2017;7(1):3434.

128. Peker N, Couto N, Sinha B, Rossen JW. Diagnosis of bloodstream infections from positive blood cultures and directly from blood samples : recent developments in molecular approaches. Clin Microbiol Infect. 2018;24(9):944–55.

129. Gyarmati P, Kjellander C, Aust C, Song Y, Öhrmalm L, Giske CG. Metagenomic analysis of bloodstream infections in patients with acute leukemia and therapy-induced neutropenia. Sci Rep. 2016;6(1):23532.

130. Forbes JD, Knox NC, Ronholm J, Pagotto F, Reimer A. Metagenomics: The next culture-independent game changer. Front Microbiol. 2017;8:1069.

131. Dacey LJ, DeSimone J, Braxton JH, Leavitt BJ, Lahey SJ, Klemperer JD, et al. Preoperative white blood cell count and mortality and morbidity after coronary artery bypass grafting. Ann Thorac Surg. 2003;76(3):760–4.

132. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, et al. A gene map of the human genome. Science (80-). 1996;274(5287):540–6.

133. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics. 2015;15(2):141–61.

134. Vijayvargiya P, Jeraldo PR, Thoendel, M. J., Greenwood-Quaintance KE, Esquer Garrigos Z, Sohail MR, Patel R. Application of metagenomic shotgun sequencing to detect vector-borne pathogens in clinical blood samples. PLoS One. 2019;14(10):e0222915.

135. Oechslin CP, Lenz N, Liechti N, Ryter S, Agyeman P, Bruggmann R, et al. Limited correlation of shotgun metagenomics following host depletion and routine diagnostics for viruses and bacteria in low concentrated surrogate and clinical samples. Front Cell Infect Microbiol. 2018;8:375.

136. 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. Nat Biotechnol. 2019;37(7):783–92.

137. Shi Y, Wang G, Lau HCH, Yu J. Metagenomic sequencing for microbial DNA in human samples: emerging technological advances. Int J Mol Sci. 2022;23(4):2181.

138. Hasan MR, Rawat A, Tang P, Jithesh P V., Thomas, E., Tan R, Tilley P. Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. J Clin Microbiol. 2016;54(4):919–27.

139. 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):1–22.

140. Wen Y, Xiao F, Wang C, Wang Z. The impact of different methods of DNA extraction on microbial community measures of BALF samples based on metagenomic data. Am J Transl Res. 2016;8(3):1412.

141. 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. Nat Biotechnol. 2019;37(7):783–92.

142. Israeli O, Makdasi E, Cohen-Gihon I, Zvi A, Lazar S, Shifman O, et al. A rapid high-throughput sequencing-based approach for the identification of unknown bacterial pathogens in whole blood. Futur Sci OA. 2020;6(6).

143. Heravi FS, Zakrzewski M, Vickery K, Hu H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. J Microbiol Methods. 2020;170(January):105856.

144. Ahsanuddin S, Afshinnekoo E, Gandara J, Hakyemezoğlu M, Bezdan D, Minot S, et al. Assessment of

REPLI-g multiple displacement whole genome amplification (WGA) techniques for metagenomic applications. J Biomol Tech. 2017;28(1):46–55.

145. Goldberg B, Sichtig H, Geyer C, Ledeboer N, Weinstock GM. Making the leap from research laboratory to clinic: Challenges and opportunities for next-generation sequencing in infectious disease diagnostics. MBio. 2015;6(6):e01888-15.

146. 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.

147. Zhang, L., Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N. Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci. 1992;89(13):5847–51.

148. Spits C, Le Caignec C, De Rycke M, Van Haute L, Van Steirteghem, A., Liebaers I, Sermon K. Whole-genome multiple displacement amplification from single cells. Nat Protoc. 2006;1(4):1965–70.

149. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell. 2012;148(5):886–95.

150. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science (80-). 2012;338(6114):1622–6.

151. Hammond M, Homa F, Andersson-Svahn, H., Ettema TJ, Joensson HN. Picodroplet partitioned whole genome amplification of low biomass samples preserves genomic diversity for metagenomic analysis. Microbiome. 2016;4(1):1–8.

152. Arakawa K. Ultralow-Input Genome Library Preparation for Nanopore Sequencing with Droplet MDA. 2023;2632:91–100.

153. Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., & Sermon K. Optimization and Evaluation of Single-Cell Whole- Genome Multiple Displacement Amplification. Hum Mutat. 2006;0(5):496–503.

154. Voelkerding K V., Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. Clin Chem. 2009;55(4):641–58.

155. Jeon SA, Park JL, Kim JH, Kim JH, Kim YS, Kim JC, et al. Comparison of the MGISEQ-2000 and Illumina HiSeq 4000 sequencing platforms for RNA sequencing. Genomics Inform. 2019;17(3).

156. Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21(1):1–16.

157. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Hum Immunol. 2021;82(11):801–11.

158. Grumaz S, Stevens P, Grumaz C, Decker SO, Weigand MA, Hofer S, et al. Next-generation sequencing diagnostics of bacteremia in septic patients. Genome Med [Internet]. 2016;8(1):1–13. Available from: http://dx.doi.org/10.1186/s13073-016-0326-8

159. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother. 2017;72(1):104–14.

160. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015;1–13.

161. de Lorenzi-Tognon M, Ruppé E, Schrenzel J. Messages from the seventh international conference on clinical metagenomics (ICCMg). Microbes Infect. 2023;25(4).

162. Chiang AD, Dekker JP. From the pipeline to the bedside: Advances and challenges in clinical metagenomics. J Infect Dis. 2020;221(3):S331–40.

163. Quainoo S, Coolen JPM, Sacha A. F. T. van Hijum C, Martijn A. Huynen c WJGM, Willem van Schaik E, Wertheim HFL. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial. Clin Microbiol Rev. 2017;30(4):1015–64.

164. Gu W, Deng X, Lee M, Sucu YD, Arevalo S, Stryke D, et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. Nat Med. 2020;27(1):115–24.

165. Fan, Jeremy, Steven Huang and SDC. BugSeq: a highly accurate cloud platform for long-read metagenomic analyses. bioRxiv. 2020;

166. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. Gigascience. 2020;9(10):1–14.

167. Leggett RM, Heavens D, Caccamo M, Clark MD, Davey RP. NanoOK: Multi-reference alignment analysis of nanopore sequencing data, quality and error profiles. Bioinformatics. 2016;32(1):142–4.

168. Taxt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Sci Rep. 2020;10(1):1–11.

169. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: Rapid and sensitive classification of metagenomic sequences. Genome Res. 2016;26(12):1721–9.

170. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7(1):11257.

171. Nasko DJ, Koren S, Phillippy AM, Treangen TJ. RefSeq database growth influences the accuracy of k-mer-based lowest common ancestor species identification. Appl Radiol. 2018;19(1):1–10.

172. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3).

173. Tran Q, Phan V. Assembling reads improves taxonomic classification of species. Genes (Basel). 2020;11(8):946.

174. Ayling M, Clark MD, Leggett RM. New approaches for metagenome assembly with short reads. Brief Bioinform. 2020;21(2):584–94.

175. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.

176. Stefan CP, Koehler JW, Minogue TD. Targeted next-generation sequencing for the detection of ciprofloxacin resistance markers using molecular inversion probes. Nat Publ Gr [Internet]. 2016;(May):1–12. Available from: http://dx.doi.org/10.1038/srep25904

177. Rossen JWA, Friedrich AW, Moran-Gilad J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. Clin Microbiol Infect [Internet]. 2018;24(4):355–60. Available from: https://doi.org/10.1016/j.cmi.2017.11.001

178. Bossé JT, Li Y, Rogers J, Crespo RF, Li Y, Chaudhuri RR, et al. Whole genome sequencing for surveillance of antimicrobial resistance in Actinobacillus pleuropneumoniae. Front Microbiol. 2017;8(MAR):6–11.

179. Maiden MCJ, Harrison OB. Population and functional genomics of neisseria revealed with gene-by-gene approaches. J Clin Microbiol. 2016;54(8):1949–55.

180. Oniciuc EA, Likotrafiti E, Alvarez-Molina A, Prieto M, Santos JA, Alvarez-Ordóñez A. The present and future of whole genome sequencing (WGS) and whole metagenome sequencing (WMS) for surveillance of antimicrobial resistant microorganisms and antimicrobial resistance genes across the food chain. Genes (Basel). 2018;9(5):1–28.

181. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST

Subcommittee. Clin Microbiol Infect. 2017;23(1):2–22.

182. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J Antimicrob Chemother. 2013;68(4):771–7.

183. Neuert S, Nair S, Day MR, Doumith M, Ashton PM, Mellor KC, et al. Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal Salmonella enterica. Front Microbiol. 2018;9(MAR):1–11.

184. Eyre DW, Silva D De, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. J Antimicrob Chemother. 2017;72(7):1937–47.

185. Nguyen M, Wesley Long S, McDermott PF, Olsen RJ, Olson R, Stevens RL, et al. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. J Clin Microbiol. 2019;57(2).

186. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75(12):3491–500.

187. 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. Antimicrob Agents Chemother. 2019;63(11):10–1128.

188. Lakin SM, Dean C, Noyes NR, Dettenwanger A, Ross AS, Doster E, et al. MEGARes: An antimicrobial resistance database for high throughput sequencing. Nucleic Acids Res. 2017;45(D1):D574–80.

189. Boolchandani M, D'Souza AW, Dantas G. Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet [Internet]. 2019;20(6):356–70. Available from: http://dx.doi.org/10.1038/s41576-019-0108-4

190. Ko KKK, Chng KR, Nagarajan N. Metagenomics-enabled microbial surveillance. Nat Microbiol. 2022;7(4):486–96.

191. Clausen PTLC, Zankari E, Aarestrup FM, Lund O. Benchmarking of methods for identification of antimicrobial resistance genes in bacterial whole genome data. J Antimicrob Chemother. 2016;71(9):2484–8.

192. Gu W, Miller S, Chiu CY. Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. Annu Rev Pathol Mech Dis. 2019;14:319–38.

193. Strong MJ, Xu G, Morici L, Splinter Bon-Durant S, Baddoo M, Lin Z, et al. Microbial Contamination in Next Generation Sequencing: Implications for Sequence-Based Analysis of Clinical Samples. PLoS Pathog. 2014;10(11):1–6.

194. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. Arch Pathol Lab Med. 2017;141(6):776–86.

195. Leggett RM, Alcon-Giner C, Heavens D, Caim S, C. BT, M. K, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nat Microbiol. 2020;5(3):430–42.

196. Sinha M, Jupe J, Mack H, Coleman TP, Lawrence SM, Fraley I. Emerging Technologies for Molecular Diagnosis of Sepsis. Clin Microbiol Rev. 2018;31(2):10–1128.

197. 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. Crit Care Med. 2019;47(5):e394–402.

198. Grumaz C, Hoffmann A, Vainshtein Y, Kopp M, Grumaz S, Stevens P, et al. Rapid Next-Generation Sequencing Based Diagnostics of Bacteremia in Septic Patients. J Mol Diagnostics [Internet]. 2020;22(3):405–18. Available from: https://doi.org/10.1016/j.jmoldx.2019.12.006

199. Thair, Simone et al. The SEP-SEQ Trial-Clinical Validation of the Karius Plasma Next-Generation Sequencing Test for Pathogen Detection in Sepsis. Am J Respir Crit Care Med. 2018;197(A2789).

200. Wanda, Lisa et al. Direct Detection and Quantification of Bacterial Cell-free DNA in Patients with Bloodstream Infection (BSI) Using the Karius Plasma Next Generation Sequencing (NGS) Test. Open Forum Infect Dis Oxford Univ Press. 2017;4(suppl_1):S12–S12.

201. Parize P, Muth E, Richaud C, Gratigny M, Pilmis B, Lamamy A, et al. Untargeted next-generation sequencing-based fi rst-line diagnosis of infection in immunocompromised adults : a multicentre , blinded , prospective study *. Clin Microbiol Infect. 2017;23(8):574.e1-574.e6.

202. Hong DK, Blauwkamp TA, Kertesz M, Bercovici S, Truong C, Banaei N. Liquid biopsy for infectious diseases : sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. Diagnostic Microbiol Infect Dis. 2018;92(3):210–3.

203. Dinakaran V, Rathinavel A, Pushpanathan M, Sivakumar R, Gunasekaran P, Rajendhran J. Elevated levels of circulating DNA in cardiovascular disease patients: Metagenomic profiling of microbiome in the circulation. PLoS One. 2014;9(8):e105221.

204. Grumaz C, Hoffmann A, Vainshtein Y, Kopp M, Grumaz S, Stevens P, et al. Rapid Next-Generation Sequencing Based Diagnostics of Bacteremia in Septic Patients. J Mol Diagnostics. 2020;22(3):405–18.

205. O'Grady J. A powerful, non-invasive test to rule out infection. Nat Microbiol. 2019;4(4):554–5.

206. Lecuit M, Eloit M. The potential of whole genome NGS for infectious disease diagnosis. Expert Rev Mol Diagn. 2015;15(12):1517–9.

207. Parize P, Pilmis B, Lanternier F, Lortholary O, Lecuit M, Muth E, et al. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. Clin Microbiol Infect. 2017;23(8):574.e1-574.e6.

208. Couto N, Schuele L, Raangs EC, Machado MP, Mendes CI, Jesus TF, et al. Critical steps in clinical shotgun metagenomics for the concomitant detection and typing of microbial pathogens. 2018;(September):1–13.

209. Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, et al. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J Clin Microbiol. 2014;52(1):139–46.

210. Willmann M, El-Hadidi M, Huson DH, Schütz M, Weidenmaier C, Autenrieth IB, et al. Antibiotic selection pressure determination through sequence-based metagenomics. Antimicrob Agents Chemother. 2015;59(12):7335–45.

211. Hasan MR, Rawat A, Tang P, Jithesh P V., 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. J Clin Microbiol. 2016;54(4):919–27.

212. Conrad S, Kanegusuku AG, Conklin SE. Taking a step back from testing: Preanalytical considerations in molecular infectious disease diagnostics. Clin Biochem. 2023;115:22–32.

213. Putra AAMS, Hernaningsih Y. Comparison of complete blood count parameters using EDTA, sodium citrate, and heparin anticoagulants. Age (Omaha). 2022;41(66):1.

214. Baskin L, Dias V, Chin A, Abdullah A, Naugler C. Effect of patient preparation, specimen collection, anticoagulants, and preservatives on laboratory test results. In: Accurate Results in the Clinical Laboratory. 2013. p. 19–34.

215. Kotikalapudi R, Patel RK. Comparative study of the influence of EDTA and sodium heparin on long term storage of cattle DNA. Cell J. 2015;17(1):181.

216. O'Grady J, Kay GL, Charalampous T, Aydin A, Scotti R. Method for digesting nucleic acid in a sample. Patent WO2021/105659A1. United Kingdom; 2021. p. 1–35.

217. Horiba K, Kawada JI, Okuno Y, Tetsuka N, Suzuki T, Ando S, et al. Comprehensive detection of pathogens in immunocompromised children with bloodstream infections by next-generation sequencing. Sci Rep. 2018;8(1):1–9.

218. Ashikawa S, Tarumoto N, Imai K, Sakai J, Kodana M, Kawamura T, et al. Rapid identification of pathogens from positive blood culture bottles with the MinION nanopore sequencer. J Med Microbiol. 2018;67(11):1589–95.

219. Mulcahy-O'Grady H, Workentine ML. The challenge and potential of metagenomics in the clinic. Front Immunol. 2016;7(FEB):1–8.

220. Sanabria A, Hjerde E, Johannessen M, Sollid JE, Simonsen GS, Hanssen AM. Shotgun-Metagenomics on Positive Blood Culture Bottles Inoculated With Prosthetic Joint Tissue: A Proof of Concept Study. Front Microbiol. 2020;11(July):1–15.

221. Kamau E, Yang S. Metagenomic Sequencing of Positive Blood Culture Fluid for Accurate Bacterial and Fungal Species Identification: A Pilot Study. Microorganisms. 2023;11(5).

222. Kok NA, Peker N, Schuele L, de Beer JL, Rossen JWA, Sinha B, et al. Host DNA depletion can increase the sensitivity of Mycobacterium spp. detection through shotgun metagenomics in sputum. Front Microbiol. 2022;13(October):1–11.

223. Wu N, Ranjan P, Tao C, Liu C, Yang E, He B, et al. Rapid identification of pathogens associated with ventilator-associated pneumonia by Nanopore sequencing. Respir Res. 2021;22(1):1–14.

224. 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. J Clin Microbiol. 2020;58(3).

225. Simner PJ, Miller HB, Breitwieser FP, Monsalve GP, Pardo CA, Salzberg SL, et al. Development and optimization of metagenomic next-generation sequencing methods for cerebrospinal fluid diagnostics. J Clin Microbiol. 2018;56(9).

226. Francis G, Kerem Z, Makkar HPS, Becker K. The biological action of saponins in animal systems : a review. Br J Nutr. 2002;88(6):587–605.

227. Böttger S, Melzig MF. The influence of saponins on cell membrane cholesterol. Bioorganic Med Chem. 2013;21(22):7118–24.

228. Cooper GM. The Cell: A Molecular Approach. Structure of the Plasma Membrane. Eighth edi. Oxford University Press; 2019. 768 p.

229. Seeman P. Transient holes in the erythrocyte membrane during hypotonic hemolysis and stable holes in the membrane after lysis by saponin and lysolecithin. J Cell Biol. 1967;32(1):55–70.

230. Jacob MC, Favre M, Bensa J -C. Membrane cell permeabilisation with saponin and multiparametric analysis by flow cytometry. Cytometry. 1991;12(6):550–8.

231. Brender JR, Mchenry AJ, Ramamoorthy A. Does cholesterol play a role in the bacterial selectivity of antimicrobial peptides? Front Immunol. 2012;3:1–4.

232. Perczyk P, Wójcik A, Broniatowski M. The role of phospholipid composition and ergosterol presence in the adaptation of fungal membranes to harsh environmental conditions – membrane modeling study. BBA - Biomembr. 2020;1862(2):183136.

233. Tang J, Chen J, He T, Jiang Z, Zhou J, Hu B, et al. Diversity of upper respiratory tract infections and prevalence of Streptococcus pneumoniae colonization among patients with fever and flu-like symptoms 11 Medical and Health Sciences 1108 Medical Microbiology 06 Biological Sciences 0605 Microbiology. BMC Infect Dis. 2019;19(1):1–10.

234. Singh, S., Sharma, A., & Nag VL. Bacterial pathogens from lower respiratory tract infections: A study from Western Rajasthan. J Fam Med Prim Care [Internet]. 2020;9(3):1407. Available from: http://www.jfmpc.com/article.asp?issn=2249-

4863;year=2017;volume=6;issue=1;spage=169;epage=170;aulast=Faizi

235. Ghaheri M, Kahrizi D, Yari K, Babaie A, Suthar RS, Kazemi E. A comparative evaluation of four DNA extraction protocols from whole blood sample. Cell Mol Biol. 2016;62(3):120–4.

236. Mardan-Nik M, Saffar Soflaei S, Biabangard-Zak A, Asghari M, Saljoughian S, Tajbakhsh A, et al. A method for improving the efficiency of DNA extraction from clotted blood samples. J Clin Lab Anal. 2019;33(6):1–7.

237. White PL, Perry MD, Loeffler J, Melchers W, Klingspor L, Bretagne S, et al. Critical stages of extracting DNA from Aspergillus fumigatus in whole-blood specimens. J Clin Microbiol. 2010;48(10):3753–5.

238. 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. Int J Med Microbiol. 2020;310(6).

239. Tan SC, Yiap BC. DNA, RNA, and protein extraction: The past and the present. J Biomed Biotechnol. 2009;2009.

240. Fujimoto S, Nakagami Y, Kojima F. Optimal Bacterial DNA Isolation Method Using Bead-Beating Technique. Med Sch. 2004;3(March):33–8.

241. Hwang KY, Kwon SH, Jung SO, Lim HK, Jung WJ, Park CS, et al. Miniaturized bead-beating device to automate full DNA sample preparation processes for Gram-positive bacteria. Lab Chip. 2011;11(21):3649–55.

242. Müller FMC, Werner KE, Kasai M, Francesconi A, Chanock SJ, Walsh TJ. Rapid extraction of genomic DNA from medically important yeasts and filamentous fungi by high-speed cell disruption. J Clin Microbiol. 1998;36(6):1625–9.

243. Smit E, Leeflang P, Glandorf B, Van Elsas JD, Wernars K. Analysis of fungal diversity in the wheat rhizosphere by sequencing of cloned PCR-amplified genes encoding 18S rRNA and temperature gradient gel electrophoresis. Appl Environ Microbiol. 1999;65(6):2614–21.

244. Claudel L, Valeix N, Basmaciyan L, Pereira B, Costa D, Vincent A, et al. Comparative study of eleven mechanical pretreatment protocols for Cryptosporidium parvum DNA extraction from stool samples. Microorganisms. 2021;9(2):1–11.

245. Lehle S, Emons J, Hack CC, Heindl F, Hein A, Preuß C, et al. Evaluation of automated techniques for extraction of circulating cell-free DNA for implementation in standardized high-throughput workflows. Sci Rep [Internet]. 2023;13(1):1–11. Available from: https://doi.org/10.1038/s41598-022-27216-5

246. Oldham AL, Drilling HS, Stamps BW, Steven BS, Duncan KE. Automated DNA extraction platforms offer solutions to challenges of assessing microbial biofouling in oil production facilities. AMB Express. 2012;2(1):1–11.

247. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, et al. Comprehensive human genome amplification using multiple displacement amplification. Proc Natl Acad Sci U S A. 2002;99(8):5261–6.

248. Blanco L, Salas M. Characterization and purification of a phage Ø29-encoded DNA polymerase required for the initiation of replication. Proc Natl Acad Sci U S A. 1984;81(17 I):5325–9.

249. Luthra R, Medeiros LJ. Isothermal multiple displacement amplification: A highly reliable approach for generating unlimited high molecular weight genomic DNA from clinical specimens. J Mol Diagnostics [Internet]. 2004;6(3):236–42. Available from: http://dx.doi.org/10.1016/S1525-1578(10)60516-8

250. Hughes S, Jones JL. The use of multiple displacement amplified DNA as a control for mMethylation specific PCR, pyrosequencing, bisulfite sequencing and methylation-sensitive restriction enzyme PCR. BMC Mol Biol. 2007;8:1–7.

251. Sobol MS, Kaster AK. Back to Basics: A Simplified Improvement to Multiple Displacement Amplification for Microbial Single-Cell Genomics. Int J Mol Sci. 2023;24(5).

252. Mai M, Hoyer JD, McClure RF. Use of multiple displacement amplification to amplify genomic DNA before sequencing of the α and β haemoglobin genes. J Clin Pathol. 2004;57(6):637–40.

253. Sheka D, Alabi N, Gordon PMK. Oxford nanopore sequencing in clinical microbiology and infection diagnostics. Brief Bioinform. 2021;22(5):1–17.

254. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med [Internet]. 2015;1–13. Available from: http://dx.doi.org/10.1186/s13073-015-0220-9

255. Pendleton KM, Erb-Downward JR, Bao Y, Branton WR, Falkowski NR, Newton DW, et al. Rapid pathogen identification in bacterial pneumonia using real-time metagenomics. Am J Respir Crit Care Med. 2017;196(12):1610–2.

256. Oxford Nanopore Technologies. Rapid sequencing DNA - PCR Barcoding (SQK-RPB004). 2019.

257. Christopher Noone J, Helmersen K, Leegaard TM, Skråmm I, Aamot HV. Article rapid diagnostics of orthopaedic-implant-associated infections using nanopore shotgun metagenomic sequencing on tissue biopsies. Microorganisms. 2021;9(1):1–11.

258. Marquet M, Zöllkau J, Pastuschek J, Viehweger A, Schleußner E, Makarewicz O, et al. Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. Sci Rep [Internet]. 2022;12(1):1–10. Available from: https://doi.org/10.1038/s41598-022-08003-8

259. Oxford Nanopore Technologies. Rapid sequencing gDNA - barcoding (SQK-RBK004). 2019.

260. Lu H, Giordano F, Ning Z. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics, Proteomics Bioinforma [Internet]. 2016;14(5):265–79. Available from: http://dx.doi.org/10.1016/j.gpb.2016.05.004

261. Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 2018;19(1):1–11.

262. Kerkhof LJ. Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? FEMS Microbiol Ecol. 2021;97(3):1–10.

263. Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and accurate classificaton of metagenomic sequences. Genome Res [Internet]. 2016;26(12):1721–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27852649%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5131823%0Ahttp://www.biorxiv.org/content/early/2016/05/25/054965.abstract

264. Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, Yao JZ, Chia N, Hanssen AD, et al. Identification of prosthetic joint infection pathogens using a shotgun metagenomics approach. Clin Infect Dis. 2018;67(9):1333–8.

265. Zheng C, Zhang S, Chen Q, Zhong L, Huang T, Zhang X, et al. Clinical characteristics and risk factors of polymicrobial Staphylococcus aureus bloodstream infections. Antimicrob Resist Infect Control. 2020;9(1):1–11.

266. Greninger AL, Naccache SN. Metagenomics to Assist in the Diagnosis of Bloodstream Infection. J Appl Lab Med. 2019;3(4):643–53.

267. Ebinger A, Fischer S, Höper D. A theoretical and generalized approach for the assessment of the sample-specific limit of detection for clinical metagenomics. Comput Struct Biotechnol J. 2021;19:732–42.

268. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. BMC Res Notes. 2019;12(1):1–7.

269. Charalampous T, Alcolea-Medina A, Snell LB, Williams TGS, Batra R, Alder C, et al. Evaluating the potential for respiratory metagenomics to improve treatment of secondary infection and detection of nosocomial transmission on expanded COVID-19 intensive care units. Genome Med. 2021;13(1):1–16.

270. Feehery GR, Yigit E, Oyola SO, Langhorst BW, Schmidt VT, Stewart FJ, et al. A Method for Selectively Enriching Microbial DNA from Contaminating Vertebrate Host DNA. PLoS One. 2013;8(10):e76096.

271. Horz HP, Scheer S, Huenger F, Vianna ME, Conrads G. Selective isolation of bacterial DNA from human clinical specimens. J Microbiol Methods. 2008;72(1):98–102.

272. Anscombe C, Gharbia S, England PH, Kingdom U, Chi H, City M, et al. Whole genome amplification and sequencing of low cell numbers directly from a bacteria spiked blood model. 2018;

273. Seth-Smith HMB, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, et al. Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture. Genome Res. 2013;23(5):855–66.

274. Lasken RS. Genomic DNA amplification by the multiple displacement amplification (MDA) method. Biochem Soc Trans. 2009;37(2):450–3.

275. Coorevits L, Van den Abeele AM. Evaluation of the BD BACTEC FX blood volume monitoring system as a continuous quality improvement measure. Eur J Clin Microbiol Infect Dis. 2015;34(7):1459–66.

276. McDonald CP, Rogers A, Cox M, Smith R, Roy A, Robbins S, et al. Evaluation of the 3D BacT/ALERT automated culture system for the detection of microbial contamination of platelet concentrates. Transfus Med. 2002;12(5):303–9.

277. Fredricks DN. Improved Amplification of Microbial DNA from Blood Cultures by Removal of the PCR Inhibitor Sodium Polyanetholesulfonate. 1998;36(10):2810–6.

278. Smith K, Diggle MA, Clarke SC. Comparison of commercial DNA extraction kits for extraction of bacterial genomic DNA from whole-blood samples. J Clin Microbiol. 2003;41(6):2440–3.

279. 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. J Microbiol Methods. 80(2):209-211.

280. Lim MY, Song EJ, Kim SH, Lee J, Nam Y Do. Comparison of DNA extraction methods for human gut microbial community profiling. Syst Appl Microbiol [Internet]. 2018;41(2):151–7. Available from: https://doi.org/10.1016/j.syapm.2017.11.008

281. Bloomfield SJ, Zomer AL, O'Grady J, Kay GL, Wain J, Janecko N, et al. Determination and quantification of microbial communities and antimicrobial resistance on food through host DNA-depleted metagenomics. Food Microbiol. 2023;110:104162.

282. Mcgaughey KD, Yilmaz-swenson T, Elsayed NM, Cruz DA, Rodriguez RR, Kritzer MD, et al. Comparative evaluation of a new magnetic bead-based DNA extraction method from fecal samples for downstream next- generation 16S rRNA gene sequencing. 2018;1–16.

283. Bartlett JG. Nosocomial bloodstream infections in US hospitals: Analysis of 24,179 cases from a prospective nationwide surveillance study. Infect Dis Clin Pract. 2004;12(6):376.

284. QIAGEN. REPLI-g Advanced DNA Single Cell Kit [Internet]. 2023. Available from: https://www.qiagen.com/us/products/discovery-and-translational-research/next-generation-sequencing/single-cell-low-input/repli-g/repli-g-advanced-dna-single-cell-kit

285. Nagai, M., Yoshida, A., & Sato N. Additive effects of bovine serum albumin, dithiothreitol and glycerolon PCR. IUBMB Life. 1998;44(1):157–63.

286. Ralser M, Querfurth R, Warnatz HJ, Lehrach H, Yaspo ML, Krobitsch S. An efficient and economic enhancer mix for PCR. Biochem Biophys Res Commun. 2006;347(3):747–51.

287. Ballantyne KN, van Oorschot RAH, John Mitchell R, Koukoulas I. Molecular crowding increases the

amplification success of multiple displacement amplification and short tandem repeat genotyping. Anal Biochem. 2006;355(2):298–303.

288. Huang L, Ma F, Chapman A, Lu S, Xie XS. Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. Annu Rev Genomics Hum Genet. 2015;16:79–102.

289. Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW, et al. Sequencing genomes from single cells by polymerase cloning. Nat Biotechnol. 2006;24(6):680–6.

290. Heavens D, Chooneea D, Giolai M, Cuber P, Aanstad P, Martin S, et al. How low can you go? Driving down the DNA input requirements for nanopore sequencing. 2021;1–20.

291. Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H, et al. Assembling the marine metagenome, one cell at a time. PLoS One. 2009;4(4).

292. Laszlo AH, Derrington IM, Gundlach JH. MspA nanopore as a single-molecule tool: From sequencing to SPRNT. Methods. 2016;105(2016):75–89.

293. Horvath, S. E., & Daum G. Lipids of mitochondria. Prog Lipid Res. 2013;52(4):590–614.

294. Casares, D., Escribá, P. V., & Rosselló CA. Membrane lipid composition: effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues. Int J Mol Sci. 2019;20(9):2167.

295. Picard M. Blood mitochondrial DNA copy number: What are we counting? Mitochondrion. 2021;60:1–11.

296. Giles C, Trent S on. The Platelet Count. 1981;C(1971):31–7.

297. Rausser S, Trumpff C, McGill MA, Junker A, Wang W, Ho SH, et al. Mitochondrial phenotypes in purified human immune cell subtypes and cell mixtures. Elife. 2021;10:1–35.

298. Melchinger H, Jain K, Tyagi T, Hwa J. Role of Platelet Mitochondria: Life in a Nucleus-Free Zone. Front Cardiovasc Med. 2019;6(October):1–11.

299. Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P, et al. MITOMAP: A human mitochondrial genome database - 2004 update. Nucleic Acids Res. 2005;33(DATABASE ISS.):2004–6.

300. Vutukuru MR, Sharma DK, Ragavendar MS, Schmolke S, Huang Y, Gumbrecht W, et al. A rapid, highly sensitive and culture-free detection of pathogens from blood by positive enrichment. J Microbiol Methods. 2016;131:105–9.

301. Rashighi M, Harris JE. Mitochondrial Ca2+ and Regulation of the Permeability Transition Pore. Vol. 176, Physiology & behavior. 2017. 139–148 p.

302. Strubbe-Rivera JO, Schrad JR, Pavlov E V., Conway JF, Parent KN, Bazil JN. The mitochondrial permeability transition phenomenon elucidated by cryo-EM reveals the genuine impact of calcium overload on mitochondrial structure and function. Sci Rep [Internet]. 2021;11(1):1–15. Available from: https://doi.org/10.1038/s41598-020-80398-8

303. Lemasters JJ, Theruvath TP, Zhong Z, Nieminen AL. Mitochondrial calcium and the permeability transition in cell death. Biochim Biophys Acta - Bioenerg [Internet]. 2009;1787(11):1395–401. Available from: http://dx.doi.org/10.1016/j.bbabio.2009.06.009

304. Rai P, He F, Kwang J, Engelward BP, Chow VTK. Pneumococcal Pneumolysin Induces DNA Damage and Cell Cycle Arrest. Nat Publ Gr. 2016;(March):1–12.

305. Nerlich A, Mieth M, Letsiou E, Fatykhova D, Zscheppang K, Imai-Matsushima A, et al. Pneumolysin induced mitochondrial dysfunction leads to release of mitochondrial DNA. Sci Rep. 2018;8(1):1–13.

306. Franko A, Baris OR, Bergschneider E, Von Toerne C, Hauck SM, Aichler M, et al. Efficient isolation of pure and functional mitochondria from mouse tissues using automated tissue disruption and

enrichment with anti-TOM22 magnetic beads. PLoS One. 2013;8(12).

307. Selvy PE, Lavieri RR, Lindsley CW, Brown HA. Phospholipase D: Enzymology, functionality, and chemical modulation. Chem Rev. 2011;111(10):6064–119.

308. Petersen EN, Gudheti M, Pavel MA, Murphy KR, Ja WW, Jorgensen EM, et al. Phospholipase D Transduces Force to TREK-1 Channels in a Biological Membrane. bioRxiv. 2019;

309. Vuk T, Strauss Patko M, Gulan-Harcet J, Očić T, Šarlija D, Jukić I. Quality control of leucocyte-depleted platelet concentrates obtained by buffy-coat method. Transfus Med. 2013;23(5):338–43.

310. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: Key considerations in genomic analyses. Nat Rev Genet. 2014;15(2):121–32.

311. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. Yearb Med Inform. 2000;9(1):65–70.

312. Kirchner JAE, Smith JL, Powell BJ, Waltz TJ, Proctor EK. Getting a clinical innovation into practice: An introduction to implementation strategies. Psychiatry Res. 2020;283(June 2019).

313. Palamountain KM, Baker J, Cowan EP, Essajee S, Mazzola LT, Metzler M, et al. Perspectives on introduction and implementation of new point-of-care diagnostic tests. J Infect Dis. 2012;205(SUPPL. 2):181–90.

314. Clark D, Dean G, Bolton S, Beeson B. Bench to bedside: the technology adoption pathway in healthcare. Health Technol (Berl). 2020;10:537–45.

315. Katsouli A, Massad MG. Current issues in the diagnosis and management of blood culture-negative infective and non-infective endocarditis. Ann Thorac Surg. 2013;95(4):1467–74.

316. Afshari A, Schrenzel J, Ieven M, Harbarth S. Bench-to-bedside review: Rapid molecular diagnostics for bloodstream infection-a new frontier? Crit Care. 2012;16(3):1–12.

317. Klouche M, Schröder U. Rapid methods for diagnosis of bloodstream infections. Clin Chem Lab Med. 2008;46(7):888–908.

318. Li XB, Qin J. Anonymizing and sharing medical text records. Inf Syst Res. 2017;28(2):332–52.

319. Huang LH, Lin PH, Tsai KW, Wang LJ, Huang YH, Kuo HC, et al. The effects of storage temperature and duration of blood samples on DNA and RNA qualities. PLoS One. 2017;12(9):e0184692.

320. Sheffield Teaching Hospitals NHS Foundation Trust. Storage of Clinical Specimens in Primary / Secondary Care. 2020;(January).

321. Zhao Y, Li Y, Chen P, Li S, Luo J, Xia H. Performance comparison of blood collection tubes as liquid biopsy storage system for minimizing cfDNA contamination from genomic DNA. J Clin Lab Anal. 2019;33(2):e22670.

322. Fabre V, Carroll KC, Cosgrove SE. Blood Culture Utilization in the Hospital Setting: a Call for Diagnostic Stewardship. J Clin Microbiol. 2022;60(3).

323. Fabre V, Sharara SL, Salinas AB, Carroll KC, Desai S, Cosgrove SE. Does this patient need blood cultures? A scoping review of indications for blood cultures in adult nonneutropenic inpatients. Clin Infect Dis. 2020;71(5):1339–47.

324. Linsenmeyer K, Gupta K, Strymish JM, Dhanani M, Brecher SM, Breu AC. Culture if spikes? Indications and yield of blood cultures in hospitalized medical patients. J Hosp Med. 2016;11(5):336–40.

325. Krleza JL, Honovic L, Tanaskovic JV, Podolar S, Rimac V, Jokic A. Post-analytical laboratory work: national recommendations from the Working Group for Post-analytics on behalf of the Croatian Society of Medical Biochemistry and Laboratory Medicine. 2019;6(2):1–34.

326. Sikaris K. Performance criteria of the post-analytical phase. Clin Chem Lab Med. 2015;53(6):949–58.

327. O'Grady J. A powerful, non-invasive test to rule out infection. Nat Microbiol [Internet]. 2019;4(4):554–5. Available from: http://dx.doi.org/10.1038/s41564-019-0424-7

328. da Fonseca Pestana Ribeiro JM, Park M. Less empiric broad-spectrum antibiotics is more in the ICU. Intensive Care Med. 2020;46(4):783–6.

329. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Vol. 43, Intensive Care Medicine. Springer Berlin Heidelberg; 2017. 304–377 p.

330. Clinical Applications | PathoQuest [Internet]. [cited 2020 Nov 17]. Available from: https://www.pathoquest.com/?page_id=730

331. Caliendo AM, Gilbert DN, Ginocchio CC, Hanson KE, May L. Better Tests , Better Care : Improved Diagnostics for Infectious Diseases. 2013;57(Suppl 3):139–70.

332. Sinha M, Jupe J, Mack H, Coleman TP, Lawrence SM, Fraley I. Emerging Technologies for Molecular Diagnosis of Sepsis. Clin Microbiol Rev. 2018;Volume 31(2 e00089-17).

333. El Gawhary S, El-Anany M, Hassan R, Ali D, El Gameel EQ. The role of 16S rRNA gene sequencing in confirmation of suspected neonatal sepsis. J Trop Pediatr. 2016;62(1):75–80.

334. Moore MS, McCarroll MG, McCann CD, May L, Younes N, Jordan JA. Direct screening of blood by PCR and pyrosequencing for a 16S rRNA gene target from emergency department and intensive care unit patients being evaluated for bloodstream infection. J Clin Microbiol. 2016;54(1):99–105.

335. Rutanga JP, Van Puyvelde S, Heroes AS, Muvunyi CM, Jacobs J, Deborggraeve S. 16S metagenomics for diagnosis of bloodstream infections: opportunities and pitfalls. Expert Rev Mol Diagn. 2018;18(8):749–59.

336. Wellinghausen N, Kochem AJ, 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. J Clin Microbiol. 2009;47(9):2759–65.

337. Van Rood JJ, Leeuwen V. Leukocyte grouping. A method and its application. 1963;42(9):1382–90.

338. Gerver SM, Mihalkova M, Bion JF, Wilson APR, Chudasama D, Johnson AP, et al. Surveillance of bloodstream infections in intensive care units in England, May 2016–April 2017: epidemiology and ecology. J Hosp Infect. 2020;106(1):1–9.

339. Bone RC. Toward an epidemiology and natural history of SIRS (systemic inflammatory response syndrome). JAMA - J Am Med Assoc. 1992;268(24):3452–5.

340. Magrini L, Gagliano G, Travaglino F, Vetrone F, Marino R, Cardelli P, et al. Comparison between white blood cell count, procalcitonin and C reactive protein as diagnostic and prognostic biomarkers of infection or sepsis in patients presenting to the emergency department. Clin Chem Lab Med. 2014;52(10):1465–72.

341. Seigel TA, Cocchi MN, Salciccioli J, Shapiro NI, Howell M, Tang A, et al. Inadequacy of temperature and white blood cell count in predicting bacteremia in patients with suspected infection. J Emerg Med. 2012;42(3):254–9.

342. Marik PE, Stephenson E. The ability of Procalcitonin, lactate, white blood cell count and neutrophil-lymphocyte count ratio to predict blood stream infection. Analysis of a large database. J Crit Care. 2020;60:135–9.

343. Hill VL, Simpson VZ, Higgins JM, Hu Z, Stevens RA, Metcalf JA, et al. Evaluation of the performance of the sysmex XT-2000i hematology analyzer with whole blood specimens stored at room temperature. Lab Med. 2009;40(12):709–18.

344. Banfi G, Salvagno GL, Lippi G. The role of ethylenediamine tetraacetic acid (EDTA) as in vitro anticoagulant for diagnostic purposes. Clin Chem Lab Med. 2007;45(5):565–76.

345. Finnegan S, Percival SL. EDTA: An Antimicrobial and Antibiofilm Agent for Use in Wound Care. Adv Wound Care. 2015;4(7):415–21.

346. Deng Z, Liu F, Li C. Therapeutic effect of ethylenediaminetetraacetic acid irrigation solution against wound infection with drug-resistant bacteria in a rat model: an animal. Bone Joint Res. 2019;8(5):189–98.

347. Kemppainen RJ, Behrend EN, Carter SF, Cole JE. Excess EDTA interferes with cortisol measurement using a solid-phase, chemiluminescent enzyme immunoassay. Diagnostic Investig. 2018;30(3):438–41.

348. Levy MM, Evans LE, Rhodes A. The surviving sepsis campaign bundle: 2018 update. Crit Care Med. 2018;46(6):997–1000.

349. Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R, et al. Surviving sepsis campaign: Association between performance metrics and outcomes in a 7.5-year study. Crit Care Med. 2015;43(1):3–12.

350. Zadroga R, Williams DN, Gottschall R, Hanson K, Nordberg V, Deike M, et al. Comparison of 2 blood culture media shows significant differences in bacterial recovery for patients on antimicrobial therapy. Clin Infect Dis. 2013;56(6):790–7.

351. Flayhart D, Borek AP, Wakefield T, Dick J, Carroll KC. Comparison of BACTEC PLUS blood culture media to BacT/Alert FA blood culture media for detection of bacterial pathogens in samples containing therapeutic levels of antibiotics. J Clin Microbiol. 2007;45(3):816–21.

352. Weinstein MP, Towns ML, Quartey SM, Mirrett S, Reimer LG, Parmigiani G, et al. The Clinical Significance of Positive Blood Cultures in the 1990s: A Prospective Comprehensive Evaluation of the Microbiology, Epidemiology, and Outcome of Bacteremia and Fungemia in Adults. Vol. 24, Clinical Infectious Diseases. 1997. p. 584–602.

353. Tabatabaei-Qomi R, Sheykh-Hasan M, Fazaely H, Kalhor N, Ghiasi M. Development of a PCR assay to detect mycoplasma contamination in cord blood hematopoietic stem cells. Iran J Microbiol. 2014;6(4):281–4.

354. Oppegaard, O., Glambek, M., Skutlaberg, D. H., Skrede, S., Sivertsen, A., & Kittang BR. Streptococcus dysgalactiae Bloodstream Infections, Norway, 1999–2021. Emerg Infect Dis. 2023;29(2):260.

355. Laupland KB, Gregson DB, Church DL, Ross T, Pitout JDD. Incidence, risk factors and outcomes of Escherichia coli bloodstream infections in a large Canadian region. Clin Microbiol Infect. 2008;14(11):1041–7.

356. Thomer L, Schneewind O, Missiakas D. Pathogenesis of Staphylococcus aureus Bloodstream Infections. Annu Rev Pathol Mech Dis. 2016;11:343–64.

357. Mishra D, Satpathy G, Wig N, Fazal F, Ahmed NH, Panda SK (2020). Evaluation of 16S rRNA broad range PCR assay for microbial detection in serum specimens in sepsis patients. J Infect Public Health. 2020;13(7):998–1002.

358. Daneshvar MI, Hollis DG, Weyant RS, Steigerwalt AG, Whitney AM, Douglas MP, et al. Paracoccus yeeii sp. nov. (Formerly CDC group EO-2), a novel bacterial species associated with human infection. J Clin Microbiol. 2003;41(3):1289–94.

359. Wallet F, Blondiaux N, Foy CL de Sainte, Loïez C, Armand S, Pagniez D, et al. Paracoccus yeei: a new unusual opportunistic bacterium in ambulatory peritoneal dialysis. Int J Infect Dis. 2010;14(2):173–4.

360. Libby DB, Bearman G. Bacteremia due to Clostridium difficile –— review of the literature. 2009;

361. Falcone M, Russo A, Iraci F, Carfagna P, Goldoni P, Vullo V. Risk Factors and Outcomes for Bloodstream Infections Secondary to Clostridium difficile Infection. 2016;60(1):252–7.

362. Abid H, Bischof E. An unusual presentation of severe sepsis due to Clostridium difficile enteritis. Cureus. 2019;11(3).

363. Klaerner H, Eschenbach U, Kamereck K, Lehn N, Wagner H. Failure of an Automated Blood Culture System To Detect Nonfermentative Gram-Negative Bacteria. 2000;38(3):1036–41.

364. Gosiewski T, Flis A, Sroka A, K˛dzierska A, Pietrzyk A, K˛dzierska J, et al. Comparison of nested, multiplex, qPCR; FISH; SeptiFast and blood culture methods in detection and identification of bacteria and fungi in blood of patients with sepsis. BMC Microbiol. 2014;14(1):1–8.

365. Banfi G, Salvagno GL. The role of ethylenediamine tetraacetic acid (EDTA) as in vitro anticoagulant for diagnostic purposes. 2007;45(5):565–76.

366. Rivolta AA, Pittman DC, Kappes AJ, Stancil RK, Kogan C, Sanz MG. The type of anticoagulant used for plasma collection affects in vitro Rhodococcus equi assays. BMC Res Notes [Internet]. 2022;15(1):1–7. Available from: https://doi.org/10.1186/s13104-022-05933-4

367. Bayot ML, Tadi P. Laboratory Tube Collection. [Internet]. 2020 [cited 2023 Oct 10]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK555991/#:~:text=The %22citrate tube%22 is for,recommended to form the plasma.

368. Meledeo MA, Peltier GC, McIntosh CS, Bynum JA, Cap AP. Optimizing whole blood storage: hemostatic function of 35-day stored product in CPD, CP2D, and CPDA-1 anticoagulants. Transfusion. 2019;59(S2):1549–59.

369. D. Lockhart, Turner J, Bennett H, Kadoom A, Faui S La, Kadoom Y, et al. A novel diagnostic system for BSI that provides rapid detection and microbial isolation direct from blood. ECCMID 2022 Abstr. 2022;(3581).

370. Lockhart D, Turner J, Kadoom A, Fauci S La, McSorley E, Jay P, et al. A novel diagnostic system for BSI that provides rapid detection and microbial isolation direct from blood. European Congress of Clinical Microbiology and Infectious Diseases; 2022.

371. R. Marcelino V, Holmes EC, Sorrell TC. The use of taxon-specific reference databases compromises metagenomic classification. BMC Genomics. 2020;21(1):1–5.

372. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell [Internet]. 2019;178(4):779–94. Available from: https://doi.org/10.1016/j.cell.2019.07.010

373. Chan KC wei. Phylogenetic classification of long read sequences. (2019).

374. Daber R, Sukhadia S, Morrissette JJD. Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets. Cancer Genet. 2013;206(12):441–8.

375. Ulrich JU, Lutfi A, Rutzen K, Renard BY. ReadBouncer: Precise and scalable adaptive sampling for nanopore sequencing. Bioinformatics. 2022;38:I153–60.

376. Oxford Nanopore Technologies. 2020. Guppy: Accurate base calling for Oxford Nanopore sequencing. Available from: https://nanoporetech.com/products/guppy

377. Bachtrog D, Charlesworth B. Towards a complete sequence of the human Y chromosome. Genome Biol. 2001;2(5):1–47.

378. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics. 2018;19(1):1–8.

379. Macwilliams MP, Liao M ken. Luria Broth (LB) and Luria Agar (LA) Media and Their Uses Protocol Resource Type : Publication Date : Authors. Am Soc Microbiol. 2019;(October 2006):7–9.

380. Oxford Nanopore Technologies. Rapid sequencing gDNA - barcoding (SQK-RBK004) [Internet]. 2019 [cited 2023 May 8]. Available from:

https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-barcoding-sequencing-sqk-rbk004/v/rbk_9054_v2_revae_14aug2019

381. Freed NE, Vlková M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. Biol Methods Protoc. 2021;5(1):1–7.

382. Kovaka S, Fan Y, Ni B, Timp W, Schatz MC. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. Nat Biotechnol [Internet]. 2021;39(4):431–41. Available from: http://dx.doi.org/10.1038/s41587-020-0731-9

383. Oxford Nanopore Technologies. Rapid sequencing DNA - PCR Barcoding (SQK-RPB004) [Internet]. 2019 [cited 2023 May 8]. Available from: https://community.nanoporetech.com/docs/prepare/library_prep_protocols/rapid-pcr-barcoding/v/rpb_9059_v1_revp_14aug2019

384. Radukic MT, Brandt D, Haak M, Muller KM, Kalinowski J. Nanopore sequencing of native adeno-Associated virus (AAV) single-stranded DNA using a transposase-based rapid protocol. NAR Genomics Bioinforma. 2020;2(4):1–16.

385. Sezonov G, Joseleau-Petit D, D'Ari R. Escherichia coli physiology in Luria-Bertani broth. J Bacteriol. 2007;189(23):8746–9.

386. Missiakas DM, Schneewind O. Growth and Laboratory Maintenance of Staphylococcus aureus. Curr Protoc Microbiol. 2018;1–12.

387. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37(23):4572–4.

388. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P. A Bioinformatician's Guide to Metagenomics. Microbiol Mol Biol Rev. 2008;72(4):557–78.

389. Schmieder R, Edwards R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS One. 2011;6(3).

390. Ivy MI, Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, Hanssen AD, Abdel MP, et al. Direct detection and identification of prosthetic joint infection pathogens in synovial fluid by metagenomic shotgun sequencing. J Clin Microbiol. 2018;56(9).

391. Smith MA, Ersavas T, Ferguson JM, Liu H, Lucas MC, Begik O, et al. Barcoding and demultiplexing Oxford Nanopore native RNA sequencing reads with deep residual learning. bioRxiv [Internet]. 2019;864322. Available from: https://www.biorxiv.org/content/10.1101/864322v1%0Ahttps://www.biorxiv.org/content/10.1101/864322v1.abstract

392. Ecker DJ, Sampath R, Li H, Massire C, Matthews HE, Toleno D, et al. New technology for rapid molecular diagnosis of bloodstream infections. Expert Rev Mol Diagn. 2010;10(4):399–415.

393. Gosiewski T, Szała L, Pietrzyk A, Brzychczy-Włoch M, Heczko PB, Bulanda M. Comparison of methods for isolation of bacterial and fungal DNA from human blood. Curr Microbiol. 2014;68(2):149–55.

394. Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, et al. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics. 2006;7:1–21.

395. Kai S, Matsuo Y, Nakagawa S, Kryukov K, Matsukawa S, Tanaka H, et al. Rapid bacterial identification by direct PCR amplification of 16S rRNA genes using the MinION™ nanopore sequencer. FEBS Open Bio. 2019;9(3):548–57.

396. 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.

397. Zhao S, Tyson GH, Chen Y, Li C, Mukherjee S, Young S, et al. Whole-genome sequencing analysis accurately predicts antimicrobial resistance phenotypes in Campylobacter spp. Appl Environ Microbiol. 2016;82(2):459–66.

398. Woodford N, Carattoli A, Karisik E, Underwood A, Ellington MJ, Livermore DM. Complete nucleotide sequences of plasmids pEK204, pEK499, and pEK516, encoding CTX-M enzymes in three major Escherichia coli lineages from the United Kingdom, all belonging to the international O25:H4-ST131 clone. Antimicrob Agents Chemother. 2009;53(10):4472–82.

399. Gekenidis MT, Rigotti S, Hummerjohann J, Walsh F, Drissner D. Long-term persistence of blactx-m-15 in soil and lettuce after introducing extended-spectrum β-lactamase (Esbl)-producing escherichia coli via manure or water. Microorganisms. 2020;8(11):1–18.

400. Chen S, Larsson M, Robinson RC, Chen SL. Direct and convenient measurement of plasmid stability in lab and clinical isolates of E. coli. Sci Rep. 2017;7(1):1–11.

401. Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Res. 2001;11(6):1095–9.

402. Malachowa N, Deleo FR. Mobile genetic elements of Staphylococcus aureus. Cell Mol Life Sci. 2010;67(18):3057–71.

403. Flannagan SE, Chow JW, Donabedian SM, Brown WJ, Perri MB, Zervos MJ, et al. Plasmid Content of a Vancomycin-Resistant Enterococcus faecalis Isolate from a Patient also Colonized by Staphylococcus aureus with a VanA Phenotype. Antimicrob Agents Chemother. 2003;47(12):3954–9.

404. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clin Microbiol Rev. 2015;28(3):565–91.

405. Marra AR, Wey SB, Castelo A, Gales AC, Cal RGR, do Carmo Filho JR, et al. Nosocomial bloodstream infections caused by Klebsiella pneumoniae: Impact of extended-spectrum β-lactamase (ESBL) production on clinical outcome in a hospital with high ESBL prevalence. BMC Infect Dis. 2006;6:1–8.

406. Rabaan AA, Eljaaly K, Alhumaid S, Albayat H, Al-Adsani W, Sabour AA, et al. An Overview on Phenotypic and Genotypic Characterisation of Carbapenem-Resistant Enterobacterales. Medicina (B Aires). 2022;58(11):1–19.

407. Ramirez MS, Nikolaidis N, Tolmasky ME. Rise and dissemination of aminoglycoside resistance: The aac(6')-Ib paradigm. Front Microbiol. 2013;4:1–13.

408. Schwarz S, Kehrenberg C, Doublet B, Cloeckaert A. Molecular basis of bacterial resistance to chloramphenicol and florfenicol. FEMS Microbiol Rev. 2004;28(5):519–42.

409. Williams CT, Musicha P, Feasey NA, Adams ER, Edwards T. ChloS-HRM, a novel assay to identify chloramphenicol-susceptible Escherichia coli and Klebsiella pneumoniae in Malawi. J Antimicrob Chemother. 2019;74(5):1212–7.

410. Asundi A, Billings N, Rogers ZH, Roche S, Bombard J, Kumcu ME, et al. 153. Pilot study of a novel whole-genome sequencing based rapid bacterial identification assay in patients with bacteremia. Open Forum Infect Dis. 2020;7(1):S206–S206.

411. Cherkaoui A, Schrenzel J. Total laboratory automation for rapid detection and identification of microorganisms and their antimicrobial resistance profiles. Front Cell Infect Microbiol. 2022;12:67.

412. Hirosawa T, Sakamoto T, Hanai S, Harada Y, Shimizu T. Effect of Prior Antibiotic Treatment on Blood Culture in an Outpatient Department of General Internal Medicine: A Retrospective Case–Control Analysis. Int J Gen Med. 2023;2709–17.

413. TSUI C. An improved method of preparing clinical samples for nucleic acid amplification. 2022.

Supplementary material.

Organism	Gene target	Forward primer (5'-3')	Reverse primer (5'-3')	Probe (5'-3')	Reference
Human Chromosome	RNA polymerase A	TGAAGCCGTGC GGAAGG	ACAAGAGAGCC AAGTGTCG	[6FAM]TACCACGT CATCTCCTTTGATG GCTCC TAT[BHQ1]	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. Nat Biotechnol. 2019;37(7):783–92.
Human Mitochondria	*MT-TL1*	CACCCAAGAAC AGGGTTTGT	TGGCCATGGGT ATGTTGTTA	Sybr Green Master experiment	Weerts MJA, Sieuwerts AM, Smid M, Look MP, Foekens JA, Sleijfer S, et al. Mitochondrial DNA content in breast cancer: Impact on in vitro and in vivo phenotype and patient prognosis. Oncotarget. 2016;7(20):29166–76
E. coli	*cyaA*	CGATAATCGCC AGATGGC	CCTAAGTTGCA GGAGATGG	[6FAM]TAGAGCGC CTTCGGTGTCGGT [BHQ1]	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. Nat Biotechnol. 2019;37(7):783–92.
S. aureus	*eap*	ACTGTAACTTT GGCACTGG	CAAAGCTTCAA AAGCAGCCTCT A	[6FAM]CTCAAGTT GGAAACCACGAGT AAGAGTGATGAA [BHQ1]	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. Nat Biotechnol. 2019;37(7):783–92.
K. pneumoniae	*mdh*	CGGGCGTAGC GCGTAA	GATACCCGCAT TCACATTAAACA G	[6FAM]CCCGGCAT GGATCGTTCCGA[B HQ1]	Fukumoto, H., Sato, Y., Hasegawa, H., Saeki, H. & Katano, H. Development of a new real-time PCR system for simultaneous detection of bacteria and fungi in pathological samples. Int. J. Clin. Exp. Pathol. 8, 15479–15488 (2015)

Supplementary Table 1. qPCR gene targets and primers sequences used in this project.

Sample number	Taxonomically classified species (number of reads)	Taxonomically classified on their respective NTC (number of reads)
1	None	*Cutibacterium acnes* (1)
2	None	*Cutibacterium acnes* (1)
3	None	*Cutibacterium acnes* (1)
4	None	*Cutibacterium acnes* (1)
5	None	*Moraxella osloensis* (1) *Escherichia coli* (1)
6	None	*Moraxella osloensis* (1) *Escherichia coli* (1)
7	None	*Moraxella osloensis* (1) *Escherichia coli* (1)
8	None	*Moraxella osloensis* (1) *Escherichia coli* (1)
9	None	*Moraxella osloensis* (1) *Escherichia coli* (1)
10	None	*Cutibacterium acnes* (2)
11	None	*Cutibacterium acnes* (2)
12	None	*Cutibacterium acnes* (2)
13	None	*Cutibacterium acnes* (2)
14	None	*Cutibacterium acnes* (2)
15	*Cutibacterium acnes* (1) *Escherichia coli* (10)	*Stenotrophomonas maltophilia* (1) *Cutibacterium acnes* (4)
16	*Staphylococcus aureus* (6) *Cutibacterium acnes* (1)	*Stenotrophomonas maltophilia* (1) *Cutibacterium acnes* (4)

17	*Cutibacterium acnes* (1) *Rhodococcus sp. H-CA8f* (1)	*Stenotrophomonas maltophilia* (1) *Cutibacterium acnes* (4)
18	*Clostridium felsineum* (1) *Mycoplasma mycoides* (1)	*Stenotrophomonas maltophilia* (1) *Cutibacterium acnes* (4)
19	*Streptococcus dysgalactiae* (20)	*Stenotrophomonas maltophilia* (1) *Cutibacterium acnes* (4)
25	*Escherichia coli* (1)	*Cutibacterium acnes* (1) *Escherichia coli* (1) *Paracoccus marcusii* (1)
26	*Mycoplasma mycoides* (2)	*Cutibacterium acnes* (1) *Escherichia coli* (1) *Paracoccus marcusii* (1)
27	*Cutibacterium acnes* (4)	*Cutibacterium acnes* (1) *Escherichia coli* (1) *Paracoccus marcusii* (1)
28	*Ralstonia pickettii* (1) *Pseudomonas putida* (1)	*Cutibacterium acnes* (1) *Escherichia coli* (1) *Paracoccus marcusii* (1)
29	*Mycoplasma mycoides* (1) *Cutibacterium acnes* (1)	*Moraxella osloensis* (1)
30	*Moraxella osloensis* (4)	*Moraxella osloensis* (1)
31	*Cutibacterium acnes* (2) *Acinetobacter lwoffii* (1) *Moraxella osloensis* (1)	*Moraxella osloensis* (1)
33	Moraxella osloensis (1)	*Moraxella osloensis* (1)

Supplementary Table 2. CMg sequencing taxonomic classification results on clinical blood samples 1 to 33.

Limit of detection experiments sample	Species identified (Number of reads)
50-100 CFU/mL samples	
Replicate 1	*Escherichia coli* (7) *Moraxella osloensis* (2)
Replicate 2	*Corynebacterium amycolatum* (3) *Klebsiella pneumoniae* (1) *Salmonella enterica* (2)
Replicate 3	*Bifidobacterium catenulatum* (4) *Escherichia coli* (9) *Acinetobacter sp. SCLZS86* (3) *Acinetobacter towneri* (10)
5-10 CFU/mL samples	
Replicate 1	*Cutibacterium acnes* (2) *Ralstonia solanacearum* (2) *Escherichia coli* (1) *Klebsiella pneumoniae* (1)
Replicate 2	*Corynebacterium amycolatum* (2) *Bacillus paralicheniformis* (1) Staphylococcus auricularis (1)
Replicate 3	*Bacillus paralicheniformis* (1) *Ralstonia solanacearum* (2)
1-5 CFU/mL samples	
Replicate 1	*Corynebacterium amycolatum* (3) *Escherichia coli* (1) *Klebsiella pneumoniae* (1)
Replicate 2	*Corynebacterium amycolatum* (3) *Ralstonia solanacearum* (2) *Escherichia coli* (1) *Moraxella osloensis* (1) *Neisseria gonorrhoeae* (1)
Replicate 3	*Corynebacterium amycolatum* (1)
NTC samples	
Replicate 1	*Klebsiella pneumoniae* (1)
Replicate 2	*Klebsiella pneumoniae* (2)
Replicate 3	-

Supplementary Table 3. All species detect at taxa score ≤65.

Limit of detection experiments sample	Species identified (Number of reads)
50-100 CFU/mL samples	
Replicate 1	*Corynebacterium amycolatum* (1) *Escherichia coli* (1) *Streptococcus mitis* (1) *Moraxella osloensis* (1)
Replicate 2	*Pseudomonas aeruginosa* (1) *Paracoccus marcusii* (1)
Replicate 3	*Bifidobacterium catenulatum* (3) *Cutibacterium acnes* (2) *Staphylococcus haemolyticus* (4) *Escherichia coli* (4) *Acinetobacter towneri* (10)
5-10 CFU/mL samples	
Replicate 1	-
Replicate 2	*Escherichia coli* (2) *Cutibacterium acnes* (1)
Replicate 3	-
1-5 CFU/mL samples	
Replicate 1	*Paracoccus marcusii* (1) *Moraxella osloensis* (1)
Replicate 2	-
Replicate 3	-
NTC samples	
Replicate 1	*Klebsiella pneumoniae* (1)
Replicate 2	*Klebsiella pneumoniae* (1)
Replicate 3	-

Supplementary Table 4. All species detect at taxa score ≤75.

Limit of detection experiments sample	Species identified (Number of reads)
50-100 CFU/mL samples	
Replicate 1	*Escherichia coli* (1) *Pasteurella multocida* (1)
Replicate 2	*Ralstonia solanacearum* (1) *Pseudomonas aeruginosa* (1)
Replicate 3	*Acinetobacter towneri* (8) *Escherichia coli* (4) *Ralstonia solanacearum* (1)
5-10 CFU/mL samples	
Replicate 1	-
Replicate 2	*Cutibacterium acnes* (2) *Hymenobacter nivis* (1)
Replicate 3	-
1-5 CFU/mL samples	
Replicate 1	*Ralstonia solanacearum* (1)
Replicate 2	-
Replicate 3	-
NTC samples	
Replicate 1	*Klebsiella pneumoniae* (1)
Replicate 2	*Klebsiella pneumoniae* (1)
Replicate 3	-

Supplementary Table 5. All species detect at taxa score ≤85.

Limit of detection experiments sample	Species identified (Number of reads)
50-100 CFU/mL samples	
Replicate 1	
Replicate 2	
Replicate 3	*Bifidobacterium bifidum* (1) *Cutibacterium acnes* (1) *Acinetobacter towneri* (1)
5-10 CFU/mL samples	
Replicate 1	-
Replicate 2	*Cutibacterium acnes* (2) *Hymenobacter nivis* (1)
Replicate 3	-
1-5 CFU/mL samples	
Replicate 1	*Cutibacterium acnes* (1)
Replicate 2	-
Replicate 3	-
NTC samples	
Replicate 1	-
Replicate 2	-
Replicate 3	-

Supplementary Table 6. All species detect at taxa score ≤90.

Limit of detection experiments sample	Species identified (Number of reads/Bacterial population relative abundance)
Klebsiella pneumoniae 1-5 CFU/mL samples	
Replicate 1	*Lactobacillus iners* (3/30%) *Gardnerella vaginalis* (3/30%) *Aerococcus christensenii* (2/20%) *Fannyhessea vaginae* (1/10%) *Gardnerella leopoldii* (1/10%)
Replicate 2	*Serratia marcescens* (388/50.4%) *Cutibacterium acnes* (227/29.5%) *Staphylococcus epidermis* (127/16.5%) *Staphylococcus capitis* (14/1.8%) *Klebsiella pneumoniae* (11/1.4%) *Citrobacter freundii* (3/0.4%)
Replicate 3	*Curibacterium acnes* (11/55%) *Klebsiella pneumoniae* (3/15%) *Ralstonia picketii* (2/10%) *Escherichia coli* (2/10%) *Rhodococcus erythropolis* (1/5%) *Stenotrophomonas maltophilia* (1/5%)
Enterococcus faecalis 1-5 CFU/mL samples	
Replicate 1	*No bacterial reads identified*
Replicate 2	*No bacterial reads identified*
Replicate 3	*No bacterial reads identified*

Supplementary Table 7. All species detect at 1mL pipeline 1-5 CFU/mL of *K. pneumoniae* or *E. faecalis* samples.

1mL standard protocol		5mL quick-enrichment protocol	
Sample	Species identified (Number of reads/Bacterial population relative abundance)	Sample	Species identified (Number of reads/Bacterial population relative abundance)
1	*Cutibacterium acnes* (4/80%) *Corynebacterium kroppenstedtii* (1/20%)	1	*Cutibacterium acnes* (1/100%)
2	*Staphylococcus capitis* (1/100%)	2	*No bacterial reads identified*
3	*Cutibacterium acnes* (19/82.5%) *Neisseria animaloris* (1/3.5%) *Acinetobacter lwoffii* (1/3.5%) *Cutibacterium acnes* (1/3.5%) *Streptococcus mitis* (1/3.5%) *Staphylococcus capitis* (1/3.5%)	3	*No bacterial reads identified*
4	*Cutibacterium acnes* (3/25%) *Cupriavidus metallidurans* (2/17.5%) *Acinetobacter johnsonii* (2/17.5%) *Staphylococcus epidermidis* (1/8%) *Bradyrhizobium sp. CCBAU* (1/8%) *Brucella anthropic* (1/8%) *Sphingobium fuliginis* (1/8%) *Corynebacterium kroppenstedtii* (1/8%)	4	*Klebsiella pneumoniae* (1/100%)
5	*Staphylococcus epidermidis* (3/60%) *Cutibacterium acnes* (2/40%)	5	*No bacterial reads identified*
6	*Cutibacterium acnes* (1/100%)	6	*No bacterial reads identified*
7	*No bacterial reads identified*		
8	*Moraxella osloensis* (2/40%) *Cutibacterium acnes* (1/20%) *Corynebacterium kroppenstedtii* (1/20%) *Malassezia restricta* (1/20%)		
9	Corynebacterium kroppenstedtii (28/20.6%) Paracoccus marcusii (27/19.8%) Pseudomonas aeruginosa (24/17.6%) Streptococcus mitis (23/16.9%) Cutibacterium acnes (18/13.3%) Citrobacter freundii (16/11.8%)		
10	*No bacterial reads identified*		
11	*No bacterial reads identified*		
12	*No bacterial reads identified*		

Supplementary Table 8. All species detect at NTC samples on 1mL and 5mL CMg pipelines.